

ADVANCING PSYCHOLOGICAL METHODS ACROSS BORDERS

EDITED BY: Kai Ruggeri, Gabriela Diana Roman, Agnieszka Walczak,
Sam Norton, Pietro Cipresso, Rocio Del Pino and
Kristina Egumenovska

PUBLISHED IN: *Frontiers in Psychology*





frontiers

Frontiers eBook Copyright Statement

The copyright in the text of individual articles in this eBook is the property of their respective authors or their respective institutions or funders. The copyright in graphics and images within each article may be subject to copyright of other parties. In both cases this is subject to a license granted to Frontiers.

The compilation of articles constituting this eBook is the property of Frontiers.

Each article within this eBook, and the eBook itself, are published under the most recent version of the Creative Commons CC-BY licence.

The version current at the date of publication of this eBook is CC-BY 4.0. If the CC-BY licence is updated, the licence granted by Frontiers is automatically updated to the new version.

When exercising any right under the CC-BY licence, Frontiers must be attributed as the original publisher of the article or eBook, as applicable.

Authors have the responsibility of ensuring that any graphics or other materials which are the property of others may be included in the CC-BY licence, but this should be checked before relying on the CC-BY licence to reproduce those materials. Any copyright notices relating to those materials must be complied with.

Copyright and source acknowledgement notices may not be removed and must be displayed in any copy, derivative work or partial copy which includes the elements in question.

All copyright, and all rights therein, are protected by national and international copyright laws. The above represents a summary only. For further information please read Frontiers' Conditions for Website Use and Copyright Statement, and the applicable CC-BY licence.

ISSN 1664-8714
ISBN 978-2-88945-948-3
DOI 10.3389/978-2-88945-948-3

About Frontiers

Frontiers is more than just an open-access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers Journal Series

The Frontiers Journal Series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the Frontiers Journal Series operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to Quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews. Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the Frontiers Journals Series: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area! Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers Editorial Office: researchtopics@frontiersin.org

ADVANCING PSYCHOLOGICAL METHODS ACROSS BORDERS

Topic Editors:

Kai Ruggeri, Columbia University, United States

Gabriela Diana Roman, University of Cambridge, United Kingdom

Agnieszka Walczak, Cambridge English, United Kingdom

Sam Norton, King's College London, United Kingdom

Pietro Cipresso, Istituto Auxologico Italiano (IRCCS), Italy

Rocio Del Pino, BioCruces Health Research Institute, Spain

Kristina Egumenovska, Scuola Internazionale Superiore di Studi Avanzati (SISSA), Italy

Citation: Ruggeri, K., Roman, G. D., Walczak, A., Norton, S., Cipresso, P., Del Pino, R., Egumenovska, K., eds. (2020). Advancing Psychological Methods Across Borders. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-88945-948-3

Table of Contents

05 Editorial: Advancing Methods for Psychological Assessment Across Borders

Kai Ruggeri, Lana Bojanić, Lindsey van Bokhorst, Hannes Jarke, Silvana Mareva, Olatz Ojinaga-Alfageme, David T. Mellor and Sam Norton

CHAPTER 1

PROTOCOLS FROM THE JUNIOR RESEARCHER PROGRAM (2015-2018)

09 Study Protocol on Intentional Distortion in Personality

Assessment: Relationship With Test Format, Culture, and Cognitive Ability

Eline Van Geert, Altan Orhon, Iulia A. Cioca, Rui Mamede, Slobodan Golušin, Barbora Hubená and Daniel Morillo

17 Examining the Importance of the Teachers' Emotional Support for Students' Social Inclusion Using the One-With-Many Design

Zarina Hoge Kamp, Johanna K. Blomster, Aslı Bursalıoğlu, Mihaela C. Călin, Melis Çetinçelik, Lauge Haastrup and Yvonne H. M. van den Berg

27 "It's Always the Judge's Fault": Attention, Emotion Recognition, and Expertise in Rhythmic Gymnastics Assessment

Lindsey G. van Bokhorst, Lenka Knapová, Kim Majoranc, Zea K. Szebeni, Adam Táborský, Dragana Tomić and Elena Cañadas

34 Study Protocol on Ecological Momentary Assessment of Health-Related Quality of Life Using a Smartphone Application

Silvana Mareva, David Thomson, Pietro Marengo, Víctor Estal Muñoz, Caroline V. Ott, Barbara Schmidt, Tobias Wingen and Angelos P. Kassianos

42 Study Protocol on Cognitive Performance in Bulgaria, Croatia, and the Netherlands: The Normacog Brief Battery

Lea Jakob, Lana Bojanić, Desislava D. Tsvetanova, Eike K. Buabang, Nienke J. de Bles, Alexandra Sarafoglou, Annet Dijkzeul and Rocio Del Pino

51 Formal and Informal Learning and First-Year Psychology Students' Development of Scientific Thinking: A Two-Wave Panel Study

Demet Soyylmaz, Laura M. Griffin, Miguel H. Martín, Šimon Kucharský, Ekaterina D. Peycheva, Nina Vaupotič and Peter A. Edelsbrunner

62 Developing a Taxonomy of Dark Triad Triggers at Work – A Grounded Theory Study Protocol

Annika Nübold, Josef Bader, Nera Bozin, Romil Depala, Helena Eidast, Elisabeth A. Johannessen and Gerhard Prinz

72 Communicating the Neuroscience of Psychopathy and its Influence on Moral Behavior: Protocol of Two Experimental Studies

Robert Blakey, Adrian D. Askelund, Matilde Boccanera, Johanna Immonen, Nejc Plohl, Cassandra Popham, Clarissa Sorger and Julia Stuhlfreyer

89 Exploring Musical Activities and Their Relationship to Emotional Well-Being in Elderly People Across Europe: A Study Protocol

Jennifer Grau-Sánchez, Meabh Foley, Renata Hlavová, Ilkka Muukkonen, Olatz Ojinaga-Alfageme, Andrijana Radukic, Melanie Spindler and Bodil Hundevad

- 98** *'Talkin' 'Bout My Generation': Using a Mixed-Methods Approach to Explore Changes in Adolescent Well-Being Across Several European Countries*
Alina Cosma, Jelisaveta Belić, Ondřej Blecha, Friederike Fenski, Man Y. Lo, Filip Murár, Darija Petrovic and Maria T. Stella
- 107** *A Study Protocol for Testing the Effectiveness of User-Generated Content in Reducing Excessive Consumption*
Atar Herziger, Amel Benzerga, Jana Berkessel, Niken L. Dinartika, Matija Franklin, Kamilla K. Steinnes and Felicia Sundström
- 117** *Eating for Two? Protocol of an Exploratory Survey and Experimental Study on Social Norms and Norm-Based Messages Influencing European Pregnant and Non-pregnant Women's Eating Behavior*
Kirsten E. Bevelander, Katharina Herte, Catherine Kakoulakis, Inés Sanguino, Anna-Lena Tebbe and Markus R. Tünte
- 129** *Parental Decision-Making on Childhood Vaccination*
Kaja Damnjanović, Johanna Graeber, Sandra Ilić, Wing Y. Lam, Žan Lep, Sara Morales, Tero Pulkkinen and Loes Vingerhoets
- 143** *From Face-to-Face to Facebook: Probing the Effects of Passive Consumption on Interpersonal Attraction*
Amy C. Orben, Augustin Mutak, Fabian Dablander, Marlene Hecht, Jakub M. Krawiec, Natália Valkovičová and Daina Kosite
- 153** *Internet Users' Valuation of Enhanced Data Protection on Social Media: Which Aspects of Privacy are Worth the Most?*
Jasmin Mahmoodi, Jitka Čurdová, Christoph Henking, Marvin Kunz, Karla Matic, Peter Mohr and Maja Vovko
- 167** *The Power of Choice: A Study Protocol on How Identity Leadership Fosters Commitment Toward the Organization*
Mafalda F. Mascarenhas, Felix Dübbers, Magdalena Hoszowska, Aylin Köseoğlu, Ralitsa Karakasheva, Ayse B. Topal, David Izydorczyk and Jérémy E. Lemoine
- 177** *The Effect of Moral Congruence of Calls to Action and Salient Social Norms on Online Charitable Donations: A Protocol Study*
Nikola Erceg, Matthias Burghart, Alessia Cottone, Jessica Lorimer, Kiran Manku, Hannah Pütz, Denis Vlašiček and Manou Willems

CHAPTER 2

EXTERNAL CONTRIBUTORS

- 191** *Cross-Validation of the Spanish HP-Version of the Jefferson Scale of Empathy Confirmed With Some Cross-Cultural Differences*
Adelina Alcorta-Garza, Montserrat San-Martín, Roberto Delgado-Bolton, Jorge Soler-González, Helena Roig and Luis Vivanco
- 200** *Depressive Symptomatology Among Norwegian Adolescent Boys and Girls: The Patient Health Questionnaire-9 (PHQ-9) Psychometric Properties and Correlates*
Jasmina Burdzovic Andreas and Geir S. Brunborg



Editorial: Advancing Methods for Psychological Assessment Across Borders

Kai Ruggeri^{1,2*}, Lana Bojanić³, Lindsey van Bokhorst⁴, Hannes Jarke⁵, Silvana Mareva⁶, Olatz Ojinaga-Alfageme⁷, David T. Mellor⁸ and Sam Norton⁹

¹ Department of Health Policy and Management, Columbia University, New York City, NY, United States, ² Centre for Business Research, Judge Business School, University of Cambridge, Cambridge, United Kingdom, ³ Centre for Mental Health and Safety, University of Manchester, Manchester, United Kingdom, ⁴ Faculty of Psychology and Neuroscience, Maastricht University, Maastricht, Netherlands, ⁵ RAND Europe, Cambridge, United Kingdom, ⁶ MRC Cognition and Brain Sciences Unit, University of Cambridge, Cambridge, United Kingdom, ⁷ Department of Psychological Sciences, Birkbeck College, University of London, London, United Kingdom, ⁸ Center for Open Science, Charlottesville, VA, United States, ⁹ Psychology Department, Institute of Psychiatry, King's College London, London, United Kingdom

Keywords: psychological methods, reproducibility, multinational research, replication, early career researchers

Editorial on the Research Topic

Advancing Methods for Psychological Assessment Across Borders

A NEW GENERATION OF PSYCHOLOGICAL AND BEHAVIORAL SCIENTISTS

OPEN ACCESS

Edited and reviewed by:

Pietro Cipresso,
Istituto Auxologico Italiano (IRCCS),
Italy

*Correspondence:

Kai Ruggeri
kai.ruggeri@columbia.edu

Specialty section:

This article was submitted to
Quantitative Psychology and
Measurement,
a section of the journal
Frontiers in Psychology

Received: 18 January 2019

Accepted: 20 February 2019

Published: 19 March 2019

Citation:

Ruggeri K, Bojanić L, van Bokhorst L,
Jarke H, Mareva S,
Ojinaga-Alfageme O, Mellor DT and
Norton S (2019) Editorial: Advancing
Methods for Psychological
Assessment Across Borders.
Front. Psychol. 10:503.
doi: 10.3389/fpsyg.2019.00503

The nascent push for greater transparency and reproducibility in psychological and behavioral sciences has created a clear call for better standards for research methods across borders and languages. Exponential growth in computing power, access to secondary data, and widespread interest in new statistical methods have ensured a new generation of behavioral researchers will have opportunities for discovery and practice on a level without precedent. As of publication, more than 200 institutions in over 100 countries globally have launched behavioral policy units, and there is immeasurable interest in applications from across specialty areas in psychology. With these trends, there is an undeniable and growing demand for improved methods for assessment in scientific study, industry, and policy.

To maintain progress, it is critical that the next generation of researchers have the awareness, training, and practice for conducting high-quality research in psychological assessment, particularly when studying across populations, borders, and languages. This Editorial summarizes key insights for the Research Topic *Advancing methods for psychological assessment across borders*, followed by general guidance for early career researchers working in multiple languages and countries, or when adapting existing methods for new settings and populations.

EDITION INSIGHTS

This Research Topic was launched to support professional development of students and early career behavioral scientists in the Junior Researcher Programme, an initiative that supports six multi-country psychological research projects annually. Senior academics were also invited to contribute manuscripts of their own multinational studies. Mirroring the field generally, there is considerable diversity in the 19 published manuscripts, with protocols and early-stage findings in education, health, development, technology, personality, data privacy, social media, organizational leadership, and financial decision-making.

The first wave of papers covered a variety of techniques for testing existing paradigms with greater cross-cultural appreciation. This included the development of a multilingual app for assessing quality of life, an expanded measurement for standardizing cognitive ability scores across Europe, new approaches to personality measures to link with cognitive ability, measurement of how teachers support students emotionally, psychological constructs and accuracy of subjective scoring in gymnastic judges, and a cross-validation of an empathy scale. Following topics covered new approaches in assessment across mental health and decision-making. Papers included learning methods, the dark triad of personality, moral behavior, and psychopathy from a neuroscience communication lens, the impact of music on the well-being of elderly people, adolescent well-being, validation of PHQ-9 in Norwegian, and assessing videos as an intervention tool on social media. In the final wave, the focus shifted toward more narrow assessment of behaviors, such as the influence of social norms on eating behavior during pregnancy, parental decision to have children vaccinated, Facebook use and preference for privacy options, how identity leadership builds organizational commitment, and the decision to donate to charity. Such diversity in papers highlights the need for early career researchers to have robust training and experience in responsible, replicable scientific methods.

This Research Topic is primarily geared toward Protocols, meaning there is less in the way of new evidence to summarize. Instead, we cogitate on the approaches, challenges, reviewer feedback, and general direction from these manuscripts as a means of guiding the next wave of junior researchers.

GUIDANCE FOR EARLY CAREER RESEARCHERS

Across these Protocols, a number of themes emerge in attempts to conduct psychological research across borders, for instance:

1. Translation into a new language is not the only adaptation existing measurements need to function in new settings—there is no one clear answer on how best to conduct comparisons in all cases (i.e., using the items that work well enough in all settings vs. using the items that best fit each location);
2. When designing studies to be carried out in multiple countries, particularly with limited resources to conduct the study, it may be more useful to focus on common, narrowly-defined groups in each, than attempt to be representative across all populations for all countries tested
3. Hungry for data: while positive in terms of how it reflects the enthusiasm of the next generation of behavioral scientists, the omnipresent wish to add more items, recruit more participants, and do more rounds of collection may serve as a barrier to producing high-powered studies on clear questions;
4. Highly cited papers that appear in textbooks over several generations of students may have a sample size (power) that would not even pass as a pilot study under present research standards;
5. WEIRD (Western, educated, industrialized, rich, and democratic) populations are easier to access; studying non-WEIRD populations may offer substantially more insight than simply improving existing instruments in tested populations;
6. Particularly when using extensive assessments, knowing how to position a study for eventual publication is a major challenge, and can slow the development of new insights, particularly when a key feature is the adaptation into multiple languages. This is particularly relevant for junior researchers (see **Table 1**).

RECOMMENDATIONS

While difficult to provide uniform guidance for all international, multilingual, and other multi-site studies in psychological sciences, there are some common steps that can be taken to ensure a smoother process with greater possibility of meaningful insights, while avoiding common pitfalls. We highlight some of these here:

Focus the Content on What Matters

Empirical papers usually revolve around a primary research question. All text written in the build up to the research question should serve the reader to understand why this question is important. Although researchers often wish to be as thorough and detailed as possible, too much information creates confusion and detracts from a primary message. Be concise and stay on point. Unless absolutely necessary and directly relevant, do not go back to Freud and Jung, and remove tangents or overstated contingencies. Focus on the critical assumptions and make sure no reader has to guess what the question or hypothesis are.

Utilize the Open Science Framework

At the time of writing, the mandate for researchers to ensure transparency and reproducibility in research is young but building. To meet these standards, there are a broad range of resources, tools, guidelines, and examples produced by the Open Science Framework (osf.io). Early career researchers are encouraged to manage project materials, such as questionnaires, instructions, analysis scripts, and datasets in an OSF project (See **Supplement 1**).

Pre-register Studies

Pre-registration has been suggested as an important tool to combat publication bias and questionable research practices and improve the transparency of the research process (Munafò et al., 2017). Pre-registrations are time-stamped documents specifying all plans for methods, data collection, and analysis. These are produced prior to conducting study. Such documents are expected to settle crucial decisions of the research process *a priori*, along with transparency about initial hypotheses (Nosek et al., 2018). There may be instances where simulated or pilot data may be necessary to aid in certain aspects of methodological decision-making, but this should be justified if done. Another alternative is to randomly split datasets in order to conduct exploratory analyses

TABLE 1 | List of common study designs for early career researchers.

	Features	Considerations
Proof of concept	New method or intervention No strong outcome hypothesized Emphasis on practical delivery of study, not insights	An ideal way to simply assess if a given approach is useful from a study perspective, but typically harder to publish unless truly novel and useful
Pilot study	A small-scale trial of method or modality Emphasis on ability to elicit effect rather than on size, strength, or direction of effect Insights reported but not interpreted	Approach to be used when all the necessary features are present for a major scientific study but concerns about practical delivery give reason to withhold a full wave of data collection that has a clear risk of a systematic flaw
Feasibility study	Research to test study practicality, materials, management, outcomes Answers the question "Can this study be done?" Emphasis on the feasibility of the study design, rather than on effects Insights only reported on process	Methods that are used in a feasibility study range from interviews to forecasting. Commonly overlooked aspects include budget planning and attitudes of potential participants toward the topic of research; latter is crucial in multicultural research.
Exploratory study	Clear theoretical framework for the primary question No strong hypothesis or prediction about patterns or specific outcomes/interactions, but broad reporting of insights	Preferable when the literature and research team have strong basis for no clear hypothesis but can create scenario for cherry-picking or overwrought methods/conclusions
Original confirmation	Full study with strong theoretical justification Specific hypotheses and prediction	When there is a clear theoretical argument that benefits from further testing, but a modified approach is preferable to direct replication. Especially useful when seeking convergence on a topic, rather than explicit reproduction.
Replication	A study that aims to replicate the findings of another study on another population or in a different setting.	Precise operationalization is the key. Only when the concept of interest is precisely defined, the success or the failure to replicate may be attributed to the (non)existence of the effect.
Validation of a metric	Measures the ability of a new/adapted instrument to capture a phenomenon. Emphasis on discriminant features between relevant individuals and groups, and possible change.	Validation is likely to be an iterative process, involving collaborative thinking, and a various number of pilot studies. It is wise to always keep face-validity in mind.
Revalidation of metric in new setting or language	Investigation to check if an instrument measures the same construct in different populations	Useful if no measure for a construct exists in a specific setting or if there are reasons to assume a metric does not equally apply to specific circumstances

prior to confirmation with the held off, un-analyzed dataset (Anderson and Magruder, 2017; **Supplement 2**).

Pre-registrations must eventually be made public in order to address underreporting biases, but many may be embargoed for various lengths of time (e.g., up to 4 years on the OSF¹) be private or public. They also may be submitted for peer review as Registered Reports² prior to data collection, such that valuable feedback could be obtained ahead of the project execution when early career researchers would be most likely to benefit from it. Researchers should also adhere to institutional review board (IRB) guidelines and include relevant approvals and ethical guidance when they pre-register.

Replicate Before You Explore

Registered reports reviewed and accepted prior to data collection provide incentives for researchers to conduct valuable replication studies. Ambitious early career researchers may initially be more drawn to the thought of testing a completely new idea of their own. However, in attempting to build new hypotheses based on theories that may not have substantial validation, the researcher may find themselves with a lot of unpublishable material. Instead, consider first replicating a critical finding that produces the assumptions for your own work, and see if it holds. If it does, then finding a new avenue to explore can be possible.

¹<http://help.osf.io/m/registrations/1/524205-register-your-project#Choose-your-privacy-settings>

²<https://cos.io/rr>

If it does not, then you have ample opportunity to discover unexpected moderating variables. Either way, you have now made an important contribution to the field while concurrently allowing yourself both confirmatory and exploratory hypotheses to test. Using the Registered Report model for this first step can be crucial, as "successful" replications are often deemed "too boring" to publish, whereas "unsuccessful" replications may be subject to more intense scrutiny than warranted and face obstacles to publication.

Publish Null Findings

We optimize the possibility of finding an effect through power calculations to inform sample sizes, but sometimes our results turn out to be null. Albeit often less desirable, such findings are equally relevant and should not be overlooked by researchers or publishers. Data syntheses and systematic reviews rely on publications of statistically significant effects as well as null findings to yield an accurate and generalized conclusion. Not publishing null findings therefore results in a skewed representation of the reality. Null findings could also inform future studies by providing context to consider potential confounds and moderators. So no matter if it is reject or fail to reject, start with publication in mind.

Apply for Ethical Approval Early

All researchers should seek guidelines for obtaining necessary ethical or IRB clearance as early as possible. As a minimum,

review should be obtained at the institution of the principal investigator. When testing in multiple settings, approval may need to be sought from additional boards, such as schools (for collecting student data) or organizations (for collecting data on a centralized platform). Ethical reviews will usually include comments on methods and legal documents such as privacy notices, so earlier review would be better for finalizing research plans, particularly pre-registration. No two IRBs are guaranteed to function in the same way, but most will focus more on protecting participants in the study and the institution, and less on getting you a high-impact publication. We advise you consider methodological input from trusted experts before and after ethical review to ensure the highest quality study. Finally, try to consider aspects of data sharing (Meyer, 2018) and overall project transparency in your IRB applications in order to not face hurdles to transparency later on (**Supplement 3**).

Use the Oxford Comma

Please.

CHALLENGES

Put bluntly: the goalposts have changed in research. While concerns about replication and sufficiently-powered studies have raised standards across the behavioral sciences, they have also created unprecedented challenges for early career researchers. For example, classic rules of thumb for testing new surveys are no longer permissible and should be replaced with systematic power calculations preceding data collection. While this will unambiguously improve scientific quality, it also requires both the statistical knowledge to produce those estimates as well as the resources to meet those participation thresholds. Likewise, while we are fully in support of pre-registering studies, journals note the difficulty in finding reviewers willing to engage with these, which can slow down the completion of study on time-limited research projects conducted by students.

Current standards developed over time and were not used by previous generations of researchers, meaning new behavioral scientists are being trained by academics who were not subject to them at the same stage in their careers. Even when these new approaches become standard in university

REFERENCES

- Anderson, M., and Magruder, J. (2017). *Split-Sample Strategies for Avoiding False Discoveries* (No. w23544). Cambridge, MA: National Bureau of Economic Research.
- Meyer, M. N. (2018). Practical tips for ethical data sharing. *Adv. Methods Pract. Psychol. Sci.* 1, 131–144. doi: 10.1177/2515245917747656
- Munafò, M. R., Nosek, B. A., Bishop, D. V., Button, K. S., Chambers, C. D., Du Sert, N. P., et al. (2017). A manifesto for reproducible science. *Nat. Hum. Behav.* 1:0021. doi: 10.1038/s41562-016-0021
- Nosek, B. A., Ebersole, C. R., DeHaven, A. C., and Mellor, D. T. (2018). The preregistration revolution. *Proc. Natl. Acad. Sci. U.S.A.* 115, 2600–2606. doi: 10.1073/pnas.1708274114

Conflict of Interest Statement: DM is an employee of the Center for Open Science, which builds and maintains the free and open source platform, OSF.

lectures, alternative learning resources, hands-on experience with research, and peer-learning are crucial for the development of junior researchers.

CONCLUSIONS

Take every word here as a positive. With new challenges come new opportunities, and the next wave of behavioral researchers will have a tremendous impact on society. The earlier that the new standards in the field can be applied across all studies, the (likely) better this will be for the advancement of the field and public perception of the work. As psychological and behavioral scientists cover all domains of life for individuals and societies, this will surely promote the greatest impact for the well-being of the science and of populations. Your professional ancestors are cheering for you.

AUTHOR CONTRIBUTIONS

KR directed, drafted, framed, edited, proofed, and finalized the entire paper. All other authors contributed equally by proofing, editing, and making direct contributions to various sections of the text.

ACKNOWLEDGMENTS

We are grateful to many contributors to the Junior Researcher Programme between 2015 and 2018: Augustin Mutak, Maja Vovko, Thomas Lind Andersen, Ondrej Kacha, Daphnee Chabal, Irina Camps Ortueta, Eduardo Garcia Garzon, Lea Jakob, Guillermo Varela, Felicia Sundstrom, Irina Gioaba, Charles Jacob, Richard Griffith, Brian Nosek, Gabriela Roman, Agnieszka Walczak, and Kristina Egumenovska. Finally, we sincerely thank Dr. Pietro Cipresso for diligence, selflessness, enthusiasm, and relentlessness at supporting our initiative.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpsyg.2019.00503/full#supplementary-material>

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The handling editor is currently editing co-organizing a Research Topic with the author KR, and confirms the absence of any other collaboration.

Copyright © 2019 Ruggeri, Bojanić, van Bokhorst, Jarke, Mareva, Ojinaga-Alfageme, Mellor and Norton. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Study Protocol on Intentional Distortion in Personality Assessment: Relationship with Test Format, Culture, and Cognitive Ability

*Eline Van Geert*¹, *Altan Orhon*², *Iulia A. Cioca*³, *Rui Mamede*⁴, *Slobodan Golušin*⁵, *Barbora Hubená*⁶ and *Daniel Morillo*^{7*}

¹ Faculty of Psychology and Educational Sciences, KU Leuven (University of Leuven), Leuven, Belgium, ² Department of Psychology, Istanbul Bilgi University, Istanbul, Turkey, ³ ScienceForWork, Milan, Italy, ⁴ Formerly affiliated with Faculty of Psychology and Education Sciences, University of Coimbra, Coimbra, Portugal, ⁵ Faculty of Philosophy, University of Novi Sad, Novi Sad, Serbia, ⁶ Department of Psychology, Faculty of Arts, Masaryk University, Brno, Czech Republic, ⁷ Chair of Psychometric Models and Applications, Department of Social Psychology and Methodology, Faculty of Psychology, Autonomous University of Madrid, Madrid, Spain

OPEN ACCESS

Edited by:

Gabriela Diana Roman,
University of Cambridge, UK

Reviewed by:

Daniel Saverio John Costa,
University of Sydney, Australia
Keith M. Harris,
University of Queensland, Australia

*Correspondence:

Daniel Morillo
daniel.morillo@uam.es

Specialty section:

This article was submitted to
Quantitative Psychology
and Measurement,
a section of the journal
Frontiers in Psychology

Received: 10 March 2016

Accepted: 06 June 2016

Published: 28 June 2016

Citation:

Van Geert E, Orhon A, Cioca IA, Mamede R, Golušin S, Hubená B and Morillo D (2016) Study Protocol on Intentional Distortion in Personality Assessment: Relationship with Test Format, Culture, and Cognitive Ability. *Front. Psychol.* 7:933. doi: 10.3389/fpsyg.2016.00933

Self-report personality questionnaires, traditionally offered in a graded-scale format, are widely used in high-stakes contexts such as job selection. However, job applicants may intentionally distort their answers when filling in these questionnaires, undermining the validity of the test results. Forced-choice questionnaires are allegedly more resistant to intentional distortion compared to graded-scale questionnaires, but they generate ipsative data. Ipsativity violates the assumptions of classical test theory, distorting the reliability and construct validity of the scales, and producing interdependencies among the scores. This limitation is overcome in the current study by using the recently developed Thurstonian item response theory model. As online testing in job selection contexts is increasing, the focus will be on the impact of intentional distortion on personality questionnaire data collected online. The present study intends to examine the effect of three different variables on intentional distortion: (a) test format (graded-scale versus forced-choice); (b) culture, as data will be collected in three countries differing in their attitudes toward intentional distortion (the United Kingdom, Serbia, and Turkey); and (c) cognitive ability, as a possible predictor of the ability to choose the more desirable responses. Furthermore, we aim to integrate the findings using a comprehensive model of intentional distortion. In the Anticipated Results section, three main aspects are considered: (a) the limitations of the manipulation, theoretical approach, and analyses employed; (b) practical implications for job selection and for personality assessment in a broader sense; and (c) suggestions for further research.

Keywords: personality assessment, personnel selection, forced-choice, Thurstonian IRT, faking, ipsativity, cross-cultural comparison

Abbreviations: BFI, big five inventory; ICAR, international cognitive ability resource; IRT, item-response theory; MUPP, multi-unidimensional pairwise preference [model].

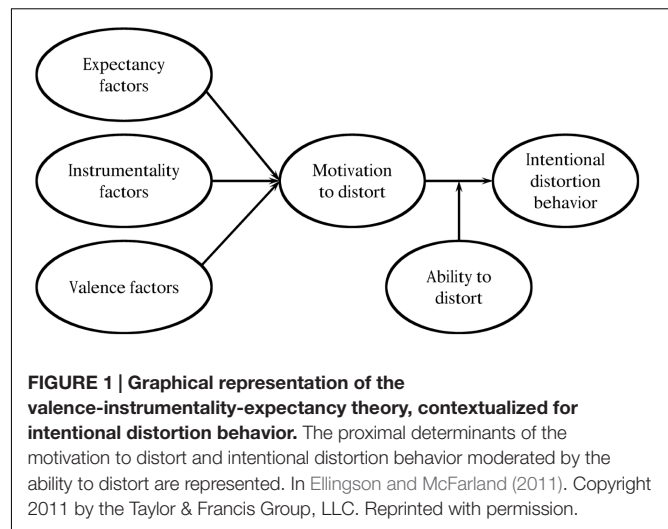
INTRODUCTION

Self-report personality questionnaires are increasingly popular in high-stakes contexts such as personnel selection (Rothstein and Goffin, 2006), college admissions (Sjöberg, 2015), and determining eligibility to stand trial (Archer et al., 2006). However, in these situations, instead of answering honestly, test takers often intentionally distort their answers to increase their chances of being hired (Birkeland et al., 2006). It has been estimated that roughly 30 percent of job applicants intentionally distort their responses (Griffith and Converse, 2011). Intentional distortion is detrimental to the psychometric properties of the assessment instrument, hiring decisions, and the utility of whole-job selection systems (Donovan et al., 2014), although human resources practitioners are largely unaware of the implications (Rothstein and Goffin, 2006). Furthermore, research on intentional distortion suffers from weak theoretical support and over-reliance on empirical and statistical methods (Griffith and Peterson, 2011).

In the literature there is considerable debate on the question whether intentional distortion also decreases the predictive validity of self-report questionnaires. Donovan et al. (2014) conducted a study utilizing a within-subjects design in an actual organizational setting and found not only a negative impact of intentional distortion on the psychometric properties of the non-cognitive self-report measure, but also a negative impact on the quality of the hiring decisions made by the organization. Additionally, people intentionally distorting their answers were found to exhibit lower levels of performance than people answering honestly. This implies that intentional distortion has negative consequences for the predictive validity of the personality test. The opposite argument, however, is based on seeing intentional distortion as a type of intelligence, mostly related to social or emotional intelligence, which consists of the ability to correctly read and interpret cues in social situations. This ability allows test takers to identify correctly the meaning of the test items and the desirable characteristics for the job in question, and later on will also help them perform better at their job, especially if it involves social interactions (Kleinmann et al., 2011). Thus, in this view, the influence of intentional distortion on the personality test leads to an equal or increased predictive validity of the test.

Even when intentional distortion would lead to a better predictive or criterion-related validity of personality tests, it is also important to consider the construct validity of the test. If the test does not measure what it is expected to measure, in this case personality factors, then the construct validity is low. Understanding and reducing the influence of intentional distortion on these measures of personality should lead toward an ideal situation in which a personality test assesses personality (and not intentional distortion), and another test assesses intentional distortion or a related ability, if this variable would have predictive validity for job performance (Kleinmann et al., 2011).

The most comprehensive theoretical model of intentional distortion (see **Figure 1**; Ellingson and McFarland, 2011) is based on the valence-instrumentality-expectancy theory of motivation



(Vroom, 1964). This model states that the predictors of intentional distortion work through three proximal motivational factors: (a) valence, the affective reaction an individual has to a particular outcome of an action; (b) instrumentality, the belief that the action will lead to a particular outcome; and (c) expectancy, the belief that one can perform the action. These three factors determine a person's motivation to engage in intentional distortion; however, the individual's actual ability to enact intentional distortion moderates the effect of the motivation to do so (Ellingson and McFarland, 2011).

Situational characteristics such as test format may offset individuals' capacities for intentional distortion. Forced-choice response formats have been proposed as a way of controlling for intentional distortion in personality assessments (Christiansen et al., 2005). In forced-choice questionnaires, instead of rating items on a graded scale, respondents rank groups of personality statements that seem equivalent in terms of social desirability. Forced-choice questionnaires hinder the identification of advantageous response patterns (Stark et al., 2014), rule out uniform biases such as acquiescence and extreme responding, and are recommended for use in cross-cultural comparisons and high-stakes situations (He et al., 2014). On the other hand, another type of scale format, that of dichotomous answers (yes/no) is rarely used (e.g., Eysenck Personality Questionnaire, Eysenck and Eysenck, 1975), being advantageous in terms of time, it takes to complete the test. However, problems arise with extremely unbalanced response distributions (e.g., everyone answers "yes" to a certain item; Clark and Watson, 1995) which indicates intentional distortion, and the measurement of continuous personality variables through completely polarized items, which minimizes the information obtained for those with real scores in extremities of the distribution (Furr, 2011).

Despite their advantages, forced-choice questionnaires have traditionally been discounted due to problems arising from conventional approaches to scoring them, which produce ipsative scores. These are able to show the relative levels of a trait within an individual (e.g., an individual appears more agreeable than

conscientious), but they cannot be used to compare absolute levels between individuals (Christiansen et al., 2005). An increase on one dimension in an ipsative measurement necessitates a corresponding decrease on other dimensions. This property makes ipsative measures incompatible with methods such as correlation or factor analysis (Cornwell and Dunlap, 1994) and creates issues relating to construct validity, criterion-related validity, and reliability estimates (Brown and Maydeu-Olivares, 2013). Hicks (1970, p. 181) cautioned researchers against using purely ipsative instruments, writing, “[researchers] cannot legitimately manipulate the variables assessed by the test utilizing statistical procedures which assume that independent dimensions are involved.” Methods proposed to address this issue have included increasing the number of dimensions being measured (Hicks, 1970) and compositional data analysis (Aitchison and Egozcue, 2005), yet the relative nature of the inferences remained unresolved (van Eijnatten et al., 2015). However, recent models based on IRT allow the extraction of normative scores from forced-choice responses (Stark et al., 2014; Joubert et al., 2015). Among these, the two state-of-the-art models are the Thurstonian IRT model (Brown and Maydeu-Olivares, 2011) and multi-unidimensional pairwise preference model (MUPP; Stark et al., 2005). These models overcome the problems posed by scoring ipsative measures via classical test theory by explicitly proposing a measurement model, that describes the relationship between items and traits, and a decision model, that describes how the individual selects one item over another (Brown, 2016). This additional level of information is what allows the recovery of normative scores from a forced-choice instrument – on Thurstonian IRT, a structure of correlated latent factors derived from multiple blocks of forced-choice items is fitted to binary outcomes of pairwise comparisons (Brown and Maydeu-Olivares, 2013), whereas MUPP combines multidimensional items with unidimensional pairings and a Bayes modal procedure as means of estimating trait scores (Stark et al., 2005).

The aim of our study is twofold: (a) to present an integrated view of intentional distortion formulated on sound theoretical underpinnings and (b) to reduce the effects of intentional distortion on personality assessment in high-stakes contexts by testing a viable method of scoring forced-choice questionnaires that can overcome earlier difficulties in their use (i.e., the ipsativity problem). Along with this, we will investigate three variables previously found to affect intentional distortion and present the theoretical reasoning behind these hypothesized effects.

First, responses for forced-choice questionnaires should exhibit lower levels of intentional distortion than those for graded-scale questionnaires. Besides the effects of forced-choice format on the ability to distort discussed above (i.e., more difficult identification of advantageous response patterns), having to choose between statements with similar levels of social desirability could induce higher levels of test-taking anxiety in applicants (Converse et al., 2008), lowering cognitive performance and ability to distort. According to Converse et al. (2008), this may come from a perception that in forced-choice format they do not have free choice of answers as well as less opportunity to express their personality qualities related to the

job. Additionally, the forced-choice format could diminish their expectancy beliefs about intentional distortion of their answers (Ellingson and McFarland, 2011).

Second, attitudes toward the appropriateness of a candidate’s use of intentional distortion are associated with several cultural dimensions suggested by the GLOBE study (House et al., 2004), such as uncertainty avoidance (which decreases the appropriateness due to lack of control about the result), or power distance (enhancing the appropriateness due to perceived lack of fairness in societies high in power distance; Fell et al., 2015). These attitudes may act on intentional distortion through (a) valence beliefs, by informing personal standards of behavior, or (b) instrumentality beliefs, by leading to the belief in a more positive outcome of intentional distortion (Ellingson and McFarland, 2011).

Third, because forced-choice questionnaires are more cognitively demanding compared to graded-scale questionnaires (Converse et al., 2008), intentional distortion is expected to relate more strongly to cognitive ability in forced-choice questionnaires than in graded-scale questionnaires. Cognitive ability is on one hand expected to relate positively to the ability of applicants to distort their answers (Christiansen et al., 2005), as it is assumed that more cognitively able applicants will be better able to identify advantageous response patterns in relation to the job requirements. On the other hand, there has also been evidence showing that people with higher cognitive ability distort their answers less often (Austin et al., 2002; Levashina et al., 2009) and do not respond in a more socially desirable manner than other participants (Ones et al., 1996). Reasons for avoiding intentional distortion of their answers include high self-efficacy and believing in one’s own abilities to succeed in assessment without distorting (De Fruyt et al., 2006), short-term outcomes (such as being excluded from the applicants pool for failing social desirability items), or long-term outcomes (such as not being suitable for the role or not fitting into the working team). However, if this would be the case, this relationship would also be evident in the graded-scale questionnaires.

Consequently, our research questions are as follows:

1. Is intentional distortion lower in forced-choice questionnaires than in graded-scale questionnaires?
2. Are there differences in levels of intentional distortion across cultural groups?
3. Do people with higher cognitive ability show more intentional distortion in forced-choice questionnaires than people with lower cognitive ability?

MATERIALS AND EQUIPMENTS

Measures

Big Five Inventory

The BFI is a popular instrument for international studies and it is recommended for use in cross-cultural settings (Schmitt et al., 2007). This 44-item graded-scale-format operationalization (Pervin and John, 1999) of the Big Five theory (John et al., 2008) will be used to assess personality traits. Adaptations of the BFI to

the languages of the target countries, as well as country-specific psychometric properties, are available (Schmitt et al., 2007; Neşe Alkan, “Reliability and Validity of the Turkish Version of the Big Five Inventory,” unpublished manuscript, 2006).

Tailored Forced-Choice Questionnaires

Equivalent forced-choice questionnaires for each country will be constructed by pairing positively keyed items measuring personality traits from the International Personality Item Pool (Goldberg, 1999). Each Big Five trait is represented by 30 items that were selected to reflect the diversity of their respective domains.

In order to ensure that the items being paired to form the blocks in the forced-choice questionnaire are as closely matched in social desirability as possible, we are conducting a study to gauge social desirability levels for each item. Approximately 250 respondents (as in Stark et al., 2005) in each country will be asked to rate the items for their attractiveness. Given that social desirability may be a context-dependent property (Rothstein and Goffin, 2000), the participants will be presented with the job description of the high-stakes condition and prompted to rate the social desirability “as if” applying for that job. Next, the items will be paired using a procedure that (a) generates a list of possible pairs of items on different dimensions of the Big Five, numbering 10 pairs initially; (b) sorts the items from most desirable to least desirable, according to mean ratings; (c) identifies the two items whose means are closest; (d) removes the pair constituted by the two items from the search space; and (e) repeats the process of pairing the closest items until no more pairs remain in the search space, after which the procedure enters the next round of matching. Pairing 150 items in this manner requires eight rounds.

International Cognitive Ability Resource

We will use the 16-item ICAR Sample Test (The International Cognitive Ability Resource Team, 2014) to measure cognitive ability. This instrument, designed for online administration (Condon and Revelle, 2014), is a public-domain measure with four subscales: Letter-Number Series, Matrix Reasoning, 3D Rotation, and Verbal Reasoning. The test has been adapted for use in the native languages of the countries in this study. (Scores will be used for within-culture comparisons only.) Condon and Revelle (2014) report adequate internal consistency for the ICAR Sample Test (Cronbach’s $\alpha = 0.81$, total $\omega = 0.83$) and provide evidence of adequate convergent validity with several widely accepted measures of cognitive ability.

STEPWISE PROCEDURES

Participants

Data will be collected from university students or recent graduates in their early adulthood (aged 18–30) in three countries: the United Kingdom, Serbia, and Turkey. Approximately 250 participants from each country will take part in the study to construct the tailored forced-choice questionnaires and 500 participants from each country will take part in the experimental study. They will be recruited online

(mostly resorting to social media, e.g., Facebook, Twitter), by using university resources (i.e., using online subject pool programs or by administering them to students during or after classes), and by advertising the study in university facilities. In order to maximize participation, the advertisements will be timed to avoid periods that would be associated with decreased participation. The participants of the experimental study will be motivated by the opportunity to enter a raffle for financial prizes and the opportunity to get individual feedback on their personality.

The targeted group of participants are students and graduates in their early adulthood according to Erikson’s (1993) stage of human development. This stage is, besides completing formation adult identity and establishing intimate relationships, typical of finishing one’s education and entering the job market. University students and fresh graduates are likely to be familiar with the situation of applying for jobs, going through job interviews and assessment, including personality assessment. Moreover, the role of assistant manager which is used to introduce the high-stakes condition might be quite realistic and relatively attractive job for wide range of university students and fresh graduates of different specializations with limited work experience (Kleinmann and Klehe, 2011).

Participating countries were chosen based on differences in attitudes toward intentional distortion in job interviews (Fell et al., 2015), which were related to the cultural dimensions (e.g., power distance, in-group collectivism) assessed by the international GLOBE study (House et al., 2004). Our choices are representative of presumed minimum, intermediate, and maximum levels on this attitude index (the United Kingdom, Serbia, and Turkey, respectively), on which a higher score represents a more positive attitude toward intentional distortion. Although Serbia was not included in the GLOBE study, later research provided information on the cultural dimensions in question (Vukonjanski et al., 2012).

Ethics Statement

The study has been given full clearance by the Ethics Committee of Universidad Autónoma de Madrid, which abides law 14/2007 of July 3, 2007 regarding biomedical research, and is fully compliant with the Declaration of Helsinki.

Design and Procedure

Participants will be invited to take a set of online tests in a single session. The tests will be administered via the Qualtrics platform (Qualtrics, Provo, UT, USA). The set includes two self-report questionnaires (graded-scale and forced-choice format), each administered in two conditions (high-stakes and low-stakes), and a test of cognitive ability. In the low-stakes condition, participants will be instructed to respond as sincerely as possible. In the high-stakes condition, they will be instructed to answer as if they were applying for a job—in this case, a management trainee position. Every participant will go through both the high-stakes and the low-stakes condition, with order determined by random assignment (see **Figure 2**). The within-subject design is recommended for studying intentional distortion because it accounts for individual tendencies in the behavior (Viswesvaran

and Ones, 1999). Between the two conditions, respondents will answer a cognitive ability measure, which should have the additional benefit of reducing practice or memory effects for the questionnaires (Grieve and de Groot, 2011). Finally, respondents will answer a single item asking to what extent the described job is attractive for individual participants on a four-point scale (from very unattractive to very attractive). This will allow us to operationalize job attractiveness and possibly control for it. In return for participation, respondents who complete the whole questionnaire will have the possibility to participate in a raffle containing several monetary reward. The participants will also be offered personalized feedback based on the BFI scores in the low-stakes condition which should increase the respondents' motivation to answer the questionnaire in an accurate and honest manner under this instruction.

Proposed Analysis

The Thurstonian IRT model (Brown and Maydeu-Olivares, 2011) has proved to be a flexible, robust model for obtaining normative trait scores from individual responses to forced-choice questionnaires. In contrast to the MUPP (Stark et al., 2005), it does not require precalibration of the item parameters. It can be estimated using the widespread software Mplus (Muthén and Muthén, 2015) and thus does not require any specialized software. Finally, the International Personality Item Pool items that will be used in the forced-choice questionnaires are written in a way that assumes a dominance response model, in which an individual is more likely to answer positively to items assessing traits they are high on; as such, these items are better fit by the Thurstonian IRT than an unfolding model such as the MUPP (Brown and Maydeu-Olivares, 2010). Therefore, Thurstonian IRT is the model of choice to analyze the ipsative data.

This model is based on Thurstone's (1927) Law of Comparative Judgement. It links the utility of each response option to the latent trait it intends to measure, by means of a linear function (Brown and Maydeu-Olivares, 2011). As a result, the probability that a respondent chooses item *i* in a binary comparison between items *i* and *k* in block *l* is expressed by (p. 473),

$$P(Y_l = 1 | \eta_a, \eta_b) = \Phi \left(\frac{-\gamma_l + \lambda_i \eta_a + \lambda_k \eta_b}{\sqrt{\psi_i^2 + \psi_k^2}} \right),$$

where $\Phi_{(x)}$ is the cumulative standard normal distribution function at *x*, γ_l is the threshold for the binary comparison of the two items block *l*, λ_i and λ_k are the factor loadings of items *i* and *k* on their respective factors *a* and *b*, ψ_i^2 and ψ_k^2 the unique variances of items *i* and *k*, and η_a and η_b a respondent's scores in factors *a* and *b*, respectively. By fitting a confirmatory factor-analytic model to the data, item and population parameters can be estimated for the model. Then, normative person parameters can be obtained through a maximum *a posteriori* estimator. Brown and Maydeu-Olivares (2012) provide and document an Excel macro that can be used to generate the necessary input files to Mplus for a given forced-choice questionnaire – the

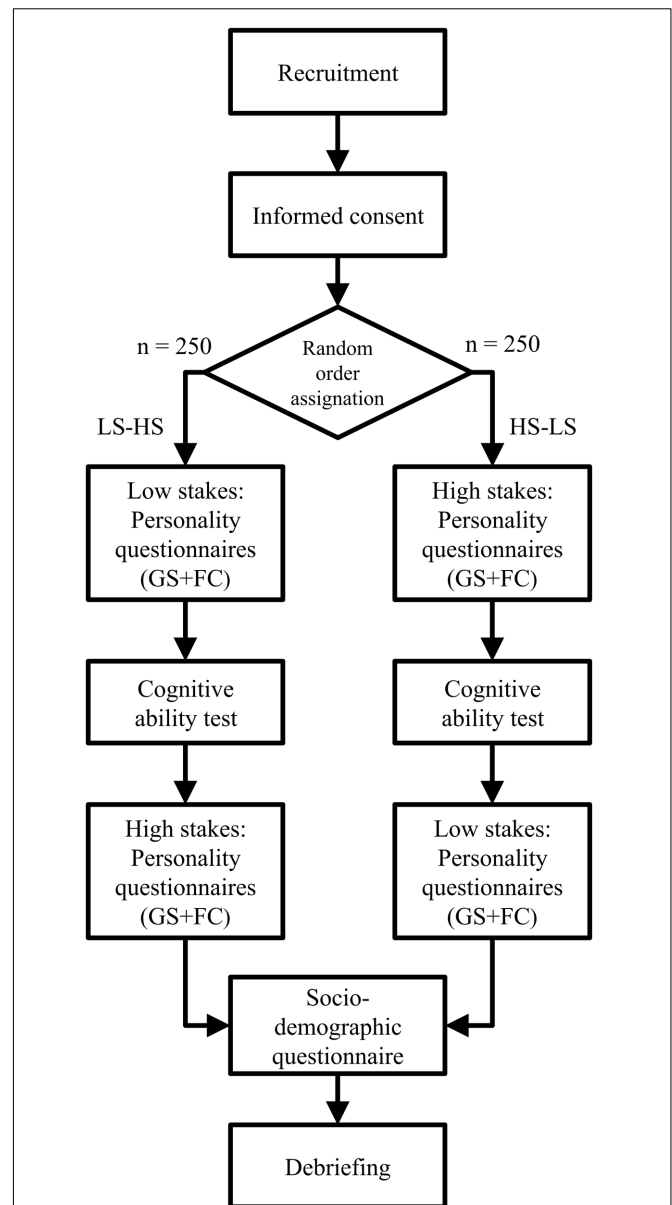


FIGURE 2 | Research design flowchart. LS, low stakes; HS, high stakes; GS, graded scale; FC, forced choice.

output allows estimation and scoring according to a Thurstonian IRT model that fits the data, computing item loadings, item thresholds, and factor scores.

The Thurstonian IRT model will be integrated into a wider structural equation model, where the responses to the forced-choice questionnaire and the graded-scale questionnaire will be jointly modeled. The bivariate information from the low-stakes condition will then be used to fit the structural equation model, and an invariance analysis will be conducted to check for invariance of the two order conditions. Then, a multitrait-multimethod matrix will be assessed for construct, convergent, and discriminant validity. After that, *maximum a posteriori*

scores for the respondents' latent traits in both conditions will be obtained using Mplus (Brown and Maydeu-Olivares, 2012).

Two intentional distortion scores will be obtained for each respondent by subtracting the IRT-based trait scores in the low-stakes (reference score) from those in the high-stakes condition: one concerning each test format (graded-scale versus forced-choice). To answer the first research question, those intentional distortion scores will be tested for significant differences using Rao's *F*-test (Christiansen et al., 2005). To test the second research question, intentional distortion scores of the three cultural samples will be tested for differences across country groups using an analysis of variance test (Converse et al., 2010). Finally, a linear regression analysis will be conducted between intentional distortion scores and cognitive ability scores to answer the third research question.

ANTICIPATED RESULTS

The present study intends to clarify the influence of test format, culture, and cognitive ability on intentional distortion in self-report personality measures. Hypotheses made concerning the influence of test format, culture, and cognitive ability are based on and integrated in the theoretical model of intentional distortion by Ellingson and McFarland (2011). However, the proposed underlying processes are still to be tested in further research.

Firstly, tests that use a forced-choice item format have been proposed to reduce the effects of respondents' intentional distortion on the test results. However, they have proved to be impractical up to now because forced-choice questionnaire items generate ipsative data. By using an IRT-based data analysis, the present study aims to increase the applicability of the forced-choice tests, and provide a valuable alternative for practitioners to reduce the effects of intentional distortion in personality assessment. As forced-choice format makes it more difficult to identify advantageous response patterns (Stark et al., 2014) and might also decrease expectancy beliefs (i.e., belief in ability to successfully distort), it is expected that intentional distortion will be lower in forced-choice questionnaires than in graded-scale questionnaires.

The results of our study regarding the test format will be of practical relevance for the assessment in high-stakes situations, such as personnel selection, where important decisions are made based on candidates' scores on personality tests. Future research could explore the utility of the assessment method for other high-stakes contexts, such as establishing eligibility for trial. In the long term, this will enable a more accurate and fairer assessment of participants in high-stakes contexts.

Secondly, it is expected that cultures differ in the extent of intentional distortion they display. More specifically, it is expected that participants from cultures scoring low, medium, or high on the index of positive attitude toward intentional distortion (the United Kingdom, Serbia, and Turkey, respectively), will show, respectively, low, medium, and high levels of intentional distortion. This influence of culture on intentional distortion may act through valence beliefs (i.e., informing personal attitude toward intentional distortion) and

instrumentality beliefs (i.e., affecting belief that intentional distortion will lead to positive outcomes; Ellingson and McFarland, 2011).

Cross-national work-related mobility is increasing nowadays, and likewise with the reach of multinational enterprises. Practitioners conducting personality assessment in such cross-national contexts need to understand the differences in their respondents' tendencies to complete personality tests in certain ways. By investigating the phenomenon of intentional distortion in three countries that differ in their attitude toward this practice, the present study will have further implications for international assessment.

Thirdly, we will also explore the relationship between a person's general cognitive ability and intentional distortion, both on graded-scale and forced-choice items. In graded-scale questionnaires, no influence of cognitive ability on intentional distortion is expected. In forced-choice questionnaires, a positive relation of cognitive ability and the ability to distort is hypothesized, as it is expected that participants should be more able to identify the advantageous response patterns. Moreover, cognitive ability might also reinforce a person's motivation to distort by raising their expectancy beliefs of how successful they will be at distorting their answers.

Nevertheless, a potential rejection of this hypothesis could indicate support for an alternative explanation. Participant's cognitive ability can be negatively related to their motivation to distort as more cognitively able applicants would be more aware of possible short-term consequences (such as being excluded from the applicants pool for failing social desirability items), or long-term consequences (such as not being suitable for the role or not fitting into working team) of distorting answers in high-stakes contexts. Yet another reason for choosing not to distort in participants with high cognitive skills is associated with higher self-efficacy and belief that they can score high without distorting (Levashina et al., 2009), so their expectancy belief may be that distorting is not worth the effort and risk-taking. However, because of the simulated nature of the high-stakes manipulation, the motivational processes to distort may differ from those in an actual high-stakes situation, for example because the long-term consequences are less taken into consideration, which threatens the ecological validity of the results. Simulating the high-stakes situations is a common practice in this field of research (see, e.g., Christiansen et al., 2005), but future studies with real job applicants would be recommended to validate our findings and their applicability in real-life situations. Additionally, although the nature of the specific instruction set given in the high-stakes context (i.e., "respond as if applying for a job") was chosen to be as ecologically valid as possible in a simulated context, this instruction set does not distinguish between the short- and long-term consequences possibly influencing the motivation to distort, thereby compromising internal and external validity. To disentangle both motivations, further studies could include an additional high-stakes condition focusing on short-term consequences specifically (e.g., "respond so as to maximize your chances of getting hired").

Understanding how cognitive ability and intentional distortion relate in the context of assessment is important to clarify aspects of predictive and construct validity of personality tests. Although a high predictive validity is useful in practice, it is essential to understand what the test actually measures. We have tried to achieve this by anchoring the study design in a solid theoretical framework that not only contributes to explaining the interrelations between concepts but also can guide future research to build a deeper and more comprehensive understanding of intentional distortion.

Limitations of our experimental design include the use of student groups as representative populations, lack of control over the physical testing environment and sample equivalence, and the possibility of a high rate of attrition leading to less diversity in sample. However, we try to mitigate the effect of the first aspect by advertising the study to recent graduates and students in final years, who are confronting (or will soon be confronting) the challenge of obtaining their first job. Regarding control over the physical environment, online assessment is an increasingly common practice, with 81% of the companies that use assessment administering it online (Kantrowitz, 2014), despite its potential disadvantages. Furthermore, it appears that online tests and pen-and-paper versions are roughly equal in their susceptibility to intentional distortion (Grieve and de Groot, 2011); therefore, research on intentional distortion in online assessment is still needed. Weigold et al. (2013) describe two studies comparing results for surveys administered via traditional means (e.g., on paper and in lab settings) and surveys administered either online or in a take-home format. The instruments used in these studies purportedly measured personality and social desirability. The authors reported that paper-and-pencil and

online survey administration were generally equivalent except for some auxiliary aspects such as response rates and completion time. However, Joinson (1999) described an effect whereby participants reported lower social anxiety and social desirability influence in an online survey compared to a paper-based survey, and when they were anonymous compared to being identified. In the case of the present study, it is expected that most participants will provide some personally identifying information in the course of enrolling for the raffle. The present study attempts to reproduce the conditions of high-stakes assessment in a job selection context. Having participants identify themselves matches more closely the conditions of real-life job selection, and a hypothetical increase in susceptibility to social desirability likewise matches what we intend to study. Because of this, our choice of methodology might be more appropriate for drawing conclusions for this type of assessment.

AUTHOR CONTRIBUTIONS

All authors listed, have made substantial, direct and intellectual contribution to the work, and approved it for publication.

ACKNOWLEDGMENTS

This research was made possible by the Junior Researcher Programme (<http://jrp.psycholars.org/>). We would like to thank everyone involved in the organization of the Programme for their assistance.

REFERENCES

- Aitchison, J., and Egozcue, J. J. (2005). Compositional data analysis: where are we and where should we be heading? *Math. Geol.* 37, 829–850. doi: 10.1007/s11004-005-7383-7
- Archer, R. P., Buffington-Vollum, J. K., Stredny, R. V., and Handel, R. W. (2006). A survey of psychological test use patterns among forensic psychologists. *J. Pers. Assess.* 87, 84–94. doi: 10.1207/s15327752jpa8701_07
- Austin, E. J., Deary, I. J., Whiteman, M. C., Fowkes, F. G. R., Pedersen, N. L., Rabbitt, P., et al. (2002). Relationships between ability and personality: does intelligence contribute positively to personal and social adjustment? *Pers. Individ. Dif.* 32, 1391–1411. doi: 10.1016/S0191-8869(01)00129-5
- Birkeland, S. A., Manson, T. M., Kisamore, J. L., Brannick, M. T., and Smith, M. A. (2006). A meta-analytic investigation of job applicant faking on personality measures. *Int. J. Sel. Assess.* 14, 317–335. doi: 10.1111/j.1468-2389.2006.00354.x
- Brown, A. (2016). Item response models for forced-choice questionnaires: a common framework. *Psychometrika* 81, 135–160. doi: 10.1007/s11336-014-9434-9
- Brown, A., and Maydeu-Olivares, A. (2010). Issues that should not be overlooked in the dominance versus ideal point controversy. *Ind. Organ. Psychol.* 3, 489–493. doi: 10.1111/j.1754-9434.2010.01277.x
- Brown, A., and Maydeu-Olivares, A. (2011). Item response modeling of forced-choice questionnaires. *Educ. Psychol. Meas.* 71, 460–502. doi: 10.1177/0013164410375112
- Brown, A., and Maydeu-Olivares, A. (2012). Fitting a Thurstonian IRT model to forced-choice data using Mplus. *Behav. Res. Methods* 44, 1135–1147. doi: 10.3758/s13428-012-0217-x
- Brown, A., and Maydeu-Olivares, A. (2013). How IRT can solve problems of ipsative data in forced-choice questionnaires. *Psychol. Methods* 18, 36–52. doi: 10.1037/a0030641
- Christiansen, N. D., Burns, G. N., and Montgomery, G. E. (2005). Reconsidering forced-choice item formats for applicant personality assessment. *Hum. Perform.* 18, 267–307. doi: 10.1207/s15327043hup1803_4
- Clark, L. A., and Watson, D. B. (1995). Constructing validity: basic issues in scale development. *Psychol. Assess.* 7, 309–319. doi: 10.1037/1040-3590.7.3.309
- Condon, D. M., and Revelle, W. (2014). The international cognitive ability resource: development and initial validation of a public-domain measure. *Intelligence* 43, 52–64. doi: 10.1016/j.intell.2014.01.004
- Converse, P. D., Oswald, F. L., Imus, A., Hedricks, C., Roy, R., and Butera, H. (2008). Comparing personality test formats and warnings: effects on criterion-related validity and test-taker reactions. *Int. J. Sel. Assess.* 16:155. doi: 10.1111/j.1468-2389.2008.00420.x
- Converse, P. D., Pathak, J., Quist, J., Merbedone, M., Gotlib, T., and Kostic, E. (2010). Statement desirability ratings in forced-choice personality measure development: implications for reducing score inflation and providing trait-level information. *Hum. Perform.* 23, 323–342. doi: 10.1080/08959285.2010.501047
- Cornwell, J. M., and Dunlap, W. P. (1994). On the questionable soundness of factoring ipsative data: a response to Saville and Willson (1991). *J. Occup. Organ. Psychol.* 67, 89–100. doi: 10.1111/j.2044-8325.1994.tb00553.x
- De Fruyt, F., Aluja, A., Garcia, L. F., Rolland, J. P., and Jung, S. C. (2006). Positive presentation management and intelligence and the personality differentiation by intelligence hypothesis in job applicants. *Int. J. Sel. Assess.* 14, 101–112. doi: 10.1111/j.1468-2389.2006.00337.x

- Donovan, J. J., Dwight, S. A., and Schneider, D. (2014). The impact of applicant faking on selection measures, hiring decisions, and employee performance. *J. Bus. Psychol.* 29, 479–493. doi: 10.1007/s10869-013-9318-5
- Ellingson, J. E., and McFarland, L. A. (2011). Understanding faking behavior through the lens of motivation: an application of VIE theory. *Hum. Perform.* 24, 322–337. doi: 10.1080/08959285.2011.597477
- Erikson, E. H. (1993). *Childhood and Society*. New York, NY: W. W. Norton and Company.
- Eysenck, H. J., and Eysenck, S. B. G. (1975). *Manual of the Eysenck Personality Questionnaire (Junior and Adult)*. Kent: Hodder and Stoughton.
- Fell, C. B., König, C. J., and Kammerhoff, J. (2015). Cross-cultural differences in the attitude toward applicants' faking in job interviews. *J. Bus. Psychol.* 31, 65–85. doi: 10.1007/s10869-015-9407-8
- Furr, R. M. (2011). *Scale Construction and Psychometrics for Social and Personality Psychology*. London: Sage Publications.
- Goldberg, L. R. (1999). "A broad-bandwidth, public domain, personality inventory measuring the lower-level facets of several five-factor models," in *Personality Psychology in Europe*, eds I. Mervielde, I. Deary, F. de Fruyt, and F. Ostendorf (Tilburg: Tilburg University Press), 7–28.
- Grieve, R., and de Groot, H. T. (2011). Does online psychological test administration facilitate faking? *Comput. Hum. Behav.* 27, 2386–2391. doi: 10.1016/j.chb.2011.08.001
- Griffith, R. L., and Converse, P. D. (2011). "The rules of evidence and the prevalence of applicant faking," in *New perspectives on faking in personality assessments*, eds M. Ziegler, C. McCann, and R. Roberts (Oxford: Oxford University Press).
- Griffith, R. L., and Peterson, M. H. (2011). One piece at a time: the puzzle of applicant faking and a call for theory. *Hum. Perform.* 24, 291–301. doi: 10.1080/08959285.2011.597474
- He, J., Bartram, D., Inceoglu, I., and van de Vijver, F. J. R. (2014). Response styles and personality traits: a multilevel analysis. *J. Cross. Cult. Psychol.* 45, 1028–1045. doi: 10.1177/0022022114534773
- Hicks, L. E. (1970). Some properties of ipsative, normative, and forced-choice normative measures. *Psychol. Bull.* 74, 167–184. doi: 10.1037/h0029780
- House, R. J., Hanges, P. J., Javidan, M., Dorfman, P. W., and Gupta, V. (eds). (2004). *Culture, Leadership, and Organizations: The GLOBE Study of 62 Societies*. Thousand Oaks, CA: Sage Publications.
- John, O. P., Naumann, L. P., and Soto, C. J. (2008). "Paradigm shift to the integrative big-five trait taxonomy: history, measurement, and conceptual issues," in *Handbook of Personality: Theory and Research*, eds O. P. John, R. W. Robins, and L. A. Pervin (New York, NY: Guilford Press), 114–158.
- Joinson, A. (1999). Social desirability, anonymity, and internet-based questionnaires. *Behav. Res. Methods Instrum. Comput.* 31, 433–438. doi: 10.3758/BF03200723
- Joubert, T., Inceoglu, I., Bartram, D., Dowdeswell, K., and Lin, Y. (2015). A comparison of the psychometric properties of the forced choice and Likert scale versions of a personality instrument. *Int. J. Sel. Assess.* 23, 92–97. doi: 10.1111/ijsa.12098
- Kantrowitz, T. M. (2014). *2014 Global Assessment Trends Report*. Arlington, VA: CEB.
- Kleinmann, M., Ingold, P. V., Lievens, F., Jansen, A., Melchers, K. G., and König, C. J. (2011). A different look at why selection procedures work: the role of candidates' ability to identify criteria. *Organ. Psychol. Rev.* 1, 128–146. doi: 10.1177/2041386610387000
- Kleinmann, M., and Klehe, U. C. (2011). Selling oneself: construct and criterion-related validity of impression management in structured interviews. *Hum. Perform.* 24, 29–46. doi: 10.1080/08959285.2010.530634
- Levashina, J., Morgeson, F. P., and Campion, M. A. (2009). They don't do it often, but they do it well: exploring the relationship between applicant mental abilities. *Int. J. Select. Assess.* 17, 271–281. doi: 10.1111/j.1468-2389.2009.00469.x
- Muthén, L. K., and Muthén, B. O. (2015). *Mplus User's Guide*, 7th Edn. Los Angeles, CA: Muthén and Muthén.
- Ones, D. S., Viswesvaran, C., and Reiss, A. D. (1996). Role of social desirability in personality testing for personnel selection: the red herring. *J. Appl. Psychol.* 81, 660–679. doi: 10.1037/0021-9010.81.6.660
- Pervin, L. A., and John, O. P. (eds) (1999). *Handbook of Personality: Theory and Research*, 2nd Edn. New York, NY: Guilford Press.
- Rothstein, M. G., and Goffin, R. D. (2000). "The assessment of personality constructs in industrial-organizational psychology," in *Problems and Solutions in Human Assessment*, eds R. D. Goffin and E. Helmes (New York, NY: Springer), 215–248.
- Rothstein, M. G., and Goffin, R. D. (2006). The use of personality measures in personnel selection: what does current research support? *Hum. Resour. Manag. Rev.* 16, 155–180. doi: 10.1016/j.hrmmr.2006.03.004
- Schmitt, D. P., Allik, J., McCrae, R. R., and Benet-Martinez, V. (2007). The geographic distribution of Big Five personality traits. *J. Cross. Cult. Psychol.* 38, 173–212. doi: 10.1177/0022022106297299
- Sjöberg, L. (2015). Correction for faking in self-report personality tests. *Scand. J. Psychol.* 56, 582–591. doi: 10.1111/sjop.12231
- Stark, S., Chernyshenko, O. S., and Drasgow, F. (2005). An IRT approach to constructing and scoring pairwise preference items involving stimuli on different dimensions: the multi-unidimensional pairwise-preference model. *Appl. Psychol. Meas.* 29, 184–203. doi: 10.1177/0146621604273988
- Stark, S., Chernyshenko, O. S., Drasgow, F., Nye, C. D., White, L. A., Heffner, T., et al. (2014). From ABLE to TAPAS: a new generation of personality tests to support military selection and classification decisions. *Mil. Psychol.* 26, 153–164. doi: 10.1037/mil0000044
- The International Cognitive Ability Resource Team (2014). "International Cognitive Ability Resource" (Measurement Instrument). Available at: <http://icar-project.com/>
- Thurstone, L. L. (1927). A law of comparative judgment. *Psychol. Rev.* 34, 273–286. doi: 10.1037/h0070288
- van Eijnatten, F. M., van der Ark, L. A., and Holloway, S. S. (2015). Ipsative measurement and the analysis of organizational values: an alternative approach for data analysis. *Qual. Quant.* 49, 559–579. doi: 10.1007/s11135-014-0009-8
- Viswesvaran, C., and Ones, D. S. (1999). Meta-analyses of fakability estimates: implications for personality measurement. *Educ. Psychol. Meas.* 59, 197–210. doi: 10.1177/00131649921969802
- Vroom, V. H. (1964). *Work and Motivation*. New York, NY: Wiley.
- Vukonjanski, J., Nikolić, M., Hadžić, O., Terek, E., and Nedeljković, M. (2012). Relationship between GLOBE organizational culture dimensions, job satisfaction, and leader-member exchange in Serbian organizations. *J. East Eur. Manag. Stud.* 17, 333–368. doi: 10.2307/23281716
- Weigold, A., Weigold, I. K., and Russell, E. J. (2013). Examination of the equivalence of self-report survey-based paper-and-pencil and internet data collection methods. *Psychol. Methods* 18, 53–70. doi: 10.1037/a0031607

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2016 Van Geert, Orhon, Cioca, Mamede, Golušin, Hubená and Morillo. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Examining the Importance of the Teachers' Emotional Support for Students' Social Inclusion Using the One-with-Many Design

Zarina Hoge Kamp^{1*}, Johanna K. Blomster², Aslı Bursaloğlu³, Mihaela C. Călin⁴, Melis Çetinçelik³, Lauge Haastrup⁵ and Yvonne H. M. van den Berg⁶

¹ Department of Basic Psychological Research and Research Methods, University of Vienna, Vienna, Austria, ² Department of Psychology, University of Oslo, Oslo, Norway, ³ Department of Psychology, Koç University, Istanbul, Turkey, ⁴ Institute of Health and Society, University of Worcester, Worcester, UK, ⁵ Department of Psychology, University of Southern Denmark, Odense, Denmark, ⁶ Behavioural Science Institute, Radboud University, Nijmegen, Netherlands

OPEN ACCESS

Edited by:

Sam Norton,
King's College London, UK

Reviewed by:

Heather M. Buzick,
Educational Testing Service, USA
Joshua Fredrick Wiley,
Australian Catholic University, Australia

*Correspondence:

Zarina Hoge Kamp
zarina_hogekamp@yahoo.de

Specialty section:

This article was submitted to
Quantitative Psychology and
Measurement,
a section of the journal
Frontiers in Psychology

Received: 09 March 2016

Accepted: 20 June 2016

Published: 04 July 2016

Citation:

Hoge Kamp Z, Blomster JK, Bursaloğlu A, Călin MC, Çetinçelik M, Haastrup L and van den Berg YHM (2016) Examining the Importance of the Teachers' Emotional Support for Students' Social Inclusion Using the One-with-Many Design. *Front. Psychol.* 7:1014. doi: 10.3389/fpsyg.2016.01014

The importance of high quality teacher–student relationships for students' well-being has been long documented. Nonetheless, most studies focus either on teachers' perceptions of provided support or on students' perceptions of support. The degree to which teachers and students agree is often neither measured nor taken into account. In the current study, we will therefore use a dyadic analysis strategy called the one-with-many design. This design takes into account the nestedness of the data and looks at the importance of reciprocity when examining the influence of teacher support for students' academic and social functioning. Two samples of teachers and their students from Grade 4 (age 9–10 years) have been recruited in primary schools, located in Turkey and Romania. By using the one-with-many design we can first measure to what degree teachers' perceptions of support are in line with students' experiences. Second, this level of consensus is taken into account when examining the influence of teacher support for students' social well-being and academic functioning.

Keywords: dyadic analysis, one-with-many design, teacher emotional support, social inclusion, academic functioning

INTRODUCTION

Students spend on average 7751 h with their teachers during their primary and lower secondary education (Organisation for Economic Co-operation Development, 2013). Through the many hours of instruction and interaction, teachers help students acquire academic knowledge and skills. However, teachers also prepare children for later functioning in society by teaching students to successfully navigate in the social world, both in and outside of the school. Previous studies already showed that teacher's emotional support is very important for students' social functioning and academic engagement (Farmer et al., 2011). Unfortunately, these studies did not look at the constant interplay between teachers' intended level of support and a student's experienced support. Therefore, there is much important information left unstudied. Arguably, teacher support may only be of importance for student's well-being when a teacher's intention to be supportive is also experienced as supportive by the student. In the current study, we will therefore use a dyadic analysis strategy called the one-with-many design to gain a better and more detailed insight in the importance of teacher support for students' academic and social functioning.

Students' Social and Academic Adjustment at School

Students not only interact with their teachers at school, but also interact to a large extent with their peers. Therefore, school is not only a place where children learn to read and write, it is also one of the most important contexts in which they acquire social skills (Hughes, 2012). The classroom is where children interact the most with their peers, and through these interactions children develop social competence (Hughes, 2012). Furthermore, the school is a place where children experience feelings of social inclusion for one of the first times. However, the classroom is often also the context where some children experience being socially excluded for the first time. The consequences of being socially excluded are severe both for the individual and for the society as a whole. Excluded people show reduced abilities to self-regulate, which leads to aggression or even crime (Baumeister et al., 2005; United Nations Educational Scientific Cultural Organization, 2010).

Feelings of social inclusion or exclusion are not only important for children's general well-being and social-emotional development. Importantly, feelings of social inclusion also make students benefit more from education (Holz, 2004). For instance, previous research has found that academic engagement of students correlated with feelings of relatedness with teachers and parents (Skinner et al., 2009), and students' school engagement has been found to be an important predictor of their school dropout and academic success in their later education (Croninger and Lee, 2001; Fredricks et al., 2004; Balfanz et al., 2007; Hafen et al., 2012). Importantly, research has shown that social exclusion is likely to promote gradual disengagement as students progress from primary or elementary level to middle school and high school (Skinner et al., 2008; Martin, 2009). Thus, it is very important for children's general well-being and academic success to feel safe and socially included at school.

Affective Quality of Teacher–Student Relationship

Numerous studies have shown that the affective quality of teacher–student relationships is predictive of students' academic functioning and performance (for a review, see Hamre and Pianta, 2006). In addition, students who experience high levels of positive and supportive interactions with their teachers are better liked and more accepted by their peers (Hughes and Kwok, 2006; Hughes, 2012). Providing emotional support is one factor through which teachers can impact students well-being (Buyse et al., 2009), and academic engagement (Skinner et al., 2009). Emotional supportive teacher–student relationships involve teachers being emotionally positive toward students, and setting clear social rules while still allowing students to develop their own social norms (Farmer et al., 2011; Hughes, 2012).

Previous studies have mainly looked at one-sided perceptions of teacher–student relationships, namely teachers' own perceptions on the level of support they provide to students. Student experiences of teachers' emotional support is often not examined, nor have studies looked at the correspondence between students' experienced support from their teachers and teachers' intended level of support. This means that a

great amount of information is left unexplored: whilst teachers might aim to provide emotional support, whether this support is in fact perceived and thus experienced by students is of utmost importance. Perceptions of emotional support can validate its receipt and can ensure pupils benefit from the aforementioned positive outcomes. In a study examining the perceived therapeutic alliance by both therapists and their clients, Marcus et al. (2009) found that therapists' with general tendency to form strong therapeutic alliance—as reported by their clients—had clients with better outcomes. However, therapists' own perception of their alliances were not associated with better therapeutic outcomes. In this particular case the clients' opinions of their therapist were associated with better outcomes, whereas his own opinion was not. This information can be used to inform therapists of their work-efficacy and inform interventions to enhance clients' view of their therapeutic alliance with their therapists. These results underline the importance of studying reciprocity in order to detect elements that influence outcomes for any side of a dyad.

Therefore, the first step is to explore the level of correspondence and reciprocity between teacher's own perceptions of emotional support and students' experienced support from their teacher. Next, we will examine whether students' academic and social functioning can be explained by teacher's intended level of support, students' experienced support, or by the reciprocity in teachers' intended and students' experienced level of support.

With this in mind, the present study will answer the following questions:

1. Does emotional support toward the student as perceived by the teacher correspond with students' experienced emotional support?
 - a. On a general level, are teachers who report to provide high degrees of emotional support also perceived as giving high levels of such support?
 - b. On an individual level, if a student experiences a lot of emotional support from their teacher, relative to the level of his/hers classmates, does the teacher then also report to give more emotional support to said student, relative to other students?
2. How is teacher's emotional support as perceived by both teachers and students, associated with students' social inclusion and academic functioning?
 - a. To what degree is teachers' reported support toward his students in general associated with students' social inclusion and academic functioning?
 - b. To what degree is teachers' reported support toward an individual student associated with the social inclusion and academic functioning of that specific student?

METHODS

Participants

We aim to recruit a sample of 15 teachers and their students (15 teachers × 25 students = 375) from Grade 4 (age 9–10

years) in public primary schools from each country selected for inclusion—Turkey and Romania.

Education systems in Turkey and Romania are similar in their structure and develop in a predominantly collectivist cultural background (Hofstede et al., 2010). In both countries, primary education is mandatory and free of charge for all citizens. The cycle develops over 4 years and the starting age is 6 years old in both cases. The pupil-teacher ratio is estimated at 18 pupils per teacher in Romania as opposed to 20 pupils per teacher in Turkey (UNESCO Institute for Statistics, 2016). In practice however, the National Law for Education (2011) in Romania assumes that classroom size varies across teachers but it is not meant to exceed 28 students and 25 is considered optimal. In Turkey, according to the data from Ministry of National Education (2016), the average number of students per teacher in the educational year 2015–2016 is 18. However, the hitherto recruited classrooms had an average size of 23. Having a mean of students per teacher larger than that of the country indicates the reliability of this number.

The ongoing data collection yields smaller classes in Romania as opposed to Turkey. We therefore aim to collect the estimated student sample across all teachers. Under determination of alpha at 0.05, power of 0.7 and medium Cohen's f^2 effect size of 0.15 a standard multiple regression sample size calculation in G*Power yielded 33 for the teacher sample size. In additional support of our sample aim we refer to a previous study (Marcus et al., 2011) which used a one-with-many design with a sample of 14 therapists and 398 substance use adolescents, making our teacher sample large enough to provide sufficient power. Limitations of applying standard power analysis on a one-with-many design are

discussed below to further elaborate on our teacher and student sample size aim.

Measures

Teacher–Student Relationship Scale (TSR)

The TSR (Gehlbach et al., Unpublished manuscript) includes teachers' and students' perspective of their relationship (see **Table 1**). Teachers and students items are correspondent hence suitable for a one-with-many design to assess both parties perceptions of their relationship. The scale measures both negative (five items) and positive (nine items) aspects of the relationship (Gehlbach et al., Unpublished manuscript). The positivity and negativity items are treated as two different subscales and as such will have their own score calculated, i.e., mean scores will be given for each teacher and each of his students. Examples of matching student and teacher items are “How motivating are the activities that <teacher's name> plans for class?” and “How motivating does <student's name> find the activities that you plan for class?” Gehlbach et al. (Unpublished manuscript) report on means and standard deviations for each of the subscales at two different time points. The provided standard deviations for the four subscales ranged from 0.52 to 1.01.

Social Inclusion

Students' social inclusion will be assessed using two measures and one peer nomination method.

Social Inclusion Assessment Instrument (SIAI)

The SIAI (Rinta et al., 2011) is a self-report, 26 item scale that measures social inclusion among students in the classroom. It

TABLE 1 | Psychometric properties of the scales used.

Name of Scale	Measures	Number of items	Reported by	Scoring system	Cronbach's α
Teacher–student relationship scale	Teachers' emotional support	14	Teacher-report and student self-report	5-point Likert scale	0.90 (teacher perspective on positivity) 0.78 (teacher perspective on negativity) 0.92 (student perspective on positivity) 0.78 (student perspective on negativity)
Social inclusion assessment instrument	Social inclusion	26	Student self-report	5-point Likert scale	0.87
Classroom peer context questionnaire	Classroom relations in the form of comfort, cooperation, conflict, mutual affection, and cohesion	31	Student self-report	5-point Likert scale	0.87 (comfort) 0.82 (cooperation) 0.87 (conflict) 0.76 (cohesion) 0.82 (isolation)
Peer nomination measure	Students relationship	10	Peer-nomination	Nomination by each student	N/A
Engagement vs. Disaffection with learning	Student engagement in the classroom	10	Student self-report	4-point Likert scale	0.72 (behavioral engagement) 0.82 (emotional engagement)

uses a 5-point Likert-type scale with smiley faces, ranging from a sad face (“I don’t agree”) to a happy face (“I agree”), with a neutral face in the middle (Rinta et al., 2011). This kind of response scale has been shown to work well in cross-cultural contexts (Islam and Rashid, 2012) and among special needs and migrant children as well (Rinta et al., 2011). Means and standard deviations will give scores for social inclusion in each classroom.

Classroom Peer Context Questionnaire (CPCQ)

The CPCQ (Boor-Klip et al., 2016) is a 5-point Likert scale measuring classroom climate with a total of 20 items. The five underlying factors are: comfort, cooperation, conflict, cohesion and isolation. An example item from the comfort factor is “In this class, I feel comfortable.” Means and standard deviations will be computed as a score for classroom climate.

All items in the questionnaire are either directed toward all classmates (class orientation) or individuals (personal orientation), which assesses student’s peer-contexts (Boor-Klip et al., 2016).

Peer nomination measure

Peer nominations measure classroom social relations (Cillessen and Marks, 2011). This will be assessed using 10 items measuring social inclusion and behavior. The questions and the different subscales can be seen in **Table 2**. These nominations have been chosen because they represent social positions relevant for the concept of inclusion.

The children will be presented with a peer nomination question (see **Table 2**), followed by nine numbered lines on which they can write the coded names of the peers they wish to nominate for that category. They will be given a list of codes for each peer additionally to this peer nomination measure. Children are allowed to nominate as many or as few of their classmates as they want, but not themselves or children outside of their classroom. The number of nominations each child receives per item will be summed up and standardized within classrooms, i.e., subsequently z-scores will be computed as a score for overall classroom social relations. Respectively z-scores less than -3 and bigger than $+3$ will be truncated (Tabachnick and Fidell, 2007).

Academic Functioning

Academic functioning will be assessed using two different measures. A student-report measure for academic engagement and a teacher report for academic performance.

Engagement vs. disaffection with learning

The engagement vs. disaffection with learning scale measures students engagement in the classroom and has scales for student and teacher perspective. The scale operationalizes engagement in learning into four distinct components: emotional engagement, behavioral engagement, emotional disaffection and behavioral disaffection (Skinner et al., 2009). For this study the emotional and behavioral engagement subscales will be used, gathered from the student perspective. This gives a total of ten items, scored on a four point Likert-type scale. The complete scores for students will be reported as means and standard deviation in each classroom.

TABLE 2 | Peer nomination items and subscale.

Items	Subscale
Who do you like best?	Acceptance
Who do you like least?	Rejection
Who are the most popular children in this class?	Popularity
Who are the least popular children in this class?	Popularity
Who are your best friends?	Friendship
Who would you say helps others a lot?	Cooperation
Who do you think cooperates well with others?	Cooperation
Who bullies others?	Bullying
Who is being bullied?	Victimization
Who would you say gets in fights often?	Aggression

Academic performance items

Two items were included in the teacher survey in order to measure students’ academic performance. These two items were “Compared to the other students, how well does this student do in language?” and “Compared to the other students, how well does this student do in maths?”

Procedure

Recruitment and data collection have commenced in April and developed over the months of May and June 2016. Data has been collected in classrooms, using paper questionnaires for students and online as well as paper questionnaires for teachers. All measures are in English and have been translated and back translate to both Turkish and Romanian.

Ethical approval has been obtained from the Ethics Committee Social Sciences (ECSS) from Radboud University in late January and it has been followed by approval from each university affiliated with the junior researchers collecting data. Additionally, the project was granted ethical approval in Turkey from Koç University and from the concerning department of the Ministry of Education for İstanbul. The application was made in December and the approval was obtained in late February. Consent forms were sent to schools in March, and data collection initiated in April. In Romania, the Regional Educational Division of the Ministry (RO: Inspectoratul Județean Argeș—ISJ) in county of Argeș, Romania granted approval to conduct the study. Primary schools located in Pitești (capital city of Argeș) were contacted right after. Consent forms were sent early in April and participants were given 3 weeks to return the completed forms. Data collection commenced on the 9th of May.

Overall, in both countries, data collection has been done in a 3-month process which has started in April and is scheduled to end in June. The time of data collection coincided with the end of the school year which was a change from our initial aim of aggregating data at the beginning of Semester 2. The difference in the times of the year could have had an effect on students’ level of enthusiasm toward school, hence, affect their need to communicate with their teacher. The students who might have completed the questionnaire in the middle of the school year might have felt more dedicated to their class and have a more

responsive relationship with their teachers than the students who completed it toward the end of the school year.

The recruitment process consisted in sending letters to selected schools, informing on the study's design, methods, procedure and information on privacy and confidentiality matters. Follow-up calls to teachers were made shortly after and active consent was sought from parents.

Data collection was scheduled for an hour for each classroom. The researcher distributed the questionnaires to all participating pupils and gave an introduction along with verbal instructions and reassurance of anonymity and the right to withdraw. A story about a secret mission of famous cartoon characters minions was introduced and participants were then encouraged to complete their questionnaire in silence. Teachers were also given their paper questionnaires about each participating child and were encouraged to complete them at the same time as children did.

PROPOSED ANALYSES

One-with-Many Design

In a reciprocal one-with-many design both the teacher and the student report on an outcome (e.g., emotional support). Variances can be estimated for both perspectives separately. Specifically four variances, two at the teacher level and two at the student level, will be estimated.

For teachers we will calculate the teachers' perceiver variance. This estimate indicates the degree to which a teacher reports to provide equal levels of emotional support across all of his students, thus the *assimilation* in his rating of provided emotional support across his students. Additionally the teachers' partner variance is obtained by calculating the means of ratings of each student. This indicates consistency in students' ratings of their teacher and thus their *consensus* as a group. Both measures will be utilized to give insight into the teacher–student relationship on a generalized, i.e., classroom level.

The students' variance estimates will give insight into the teacher–student relationship on a dyadic, i.e., individual level. The teacher relationship variance indicates *uniqueness*, i.e., the degree to which a teacher reports to provide an especially strong emotional support toward an individual student. The student relationship variance indicates *uniqueness*, i.e., the degree to which a student reports to obtain an especially strong emotional support from his teacher.

The single wave data can be estimated with multilevel modeling. The multilevel modeling framework takes into account the nestedness and non-independence of the data and will be used to estimate the different variance components introduced above (Kenny et al., 2006). Specifically a two level model, with teachers on the upper and students on the lower level, will be utilized. In the current demonstration version 21 of SPSS will be used to analyse the data.

Variance Partitioning

In the reciprocal one-with-many design both teachers and students provide scores for emotional support by completing the TSR questionnaire. To later be able to indicate which of them

provided a score the two intercept approach (Raudenbush et al., 1995) is used. To do this two dummy variables will be created to denote the provider of a score. Hence we create one dummy variable, T, which is coded 1 if the data are provided by the teacher and 0 if the data came from the student. Respectively a student dummy variable, S, will be created which is coded 0 if the data are provided by the teacher and 1 if the data is provided by the student. This way one intercept will be specific to teachers' ratings and one for students' ratings (Marcus et al., 2009).

Due to reciprocity and hence the introduced dummy variables variance partitioning is executed using a specific data structure. Specifically as each lower student level unit is embedded in a dyad and as each dyad includes two scores for emotional support there will be two rows per dyad indicating reciprocal emotional support scores including two columns for the newly created dummy variables, which indicate who provided the scores. The SPSS syntax to attain variance partitioning (Marcus et al., 2009) is provided in the **Appendix**.

Reciprocity

Research question 1a and 1b address correspondence of teachers' and students' perceptions of emotional support. In a second step, using the variance estimations, we will examine the correspondence between teachers' and students' report on emotional support at a generalized and dyadic level (see **Table 3** and **Figure 1**). We do this by correlating the variance components that we attained by the variance partitioning step. We first estimate the generalized reciprocity by correlating the teachers' partner variance (reflected in student reports) with the teacher perceiver variance (reflected in teachers' reports). This suggests whether teachers who report to provide strong emotional support are backed-up in their view by their students. Next, we will examine the dyadic reciprocity by correlating the two relationship variances (reflected in both teachers' as well as students' reports). This suggests whether a teacher who reports to provide a uniquely strong emotional support to a particular student is in turn seen as emotionally supportive by that particular student.

Emotional Support and Student Outcomes

Research question 2a and 2b address the influence of teachers' emotional support on students' social inclusion and academic functioning. In this last step of the analyses we will examine associations between the emotional support ratings given by teachers and students and four measures of outcome: social inclusion, classroom climate, hierarchy in classrooms and students' academic engagement. To analyse these student outcomes we will use linear multiple regression analyses.

The average of the outcomes across all students within each teacher will be predicted using the teacher variance components (e.g., teacher perceiver and partner effects; see **Table 3**). This way we can answer questions on a general classroom level like “if a teacher thinks s/he is generally more supportive (compared to other teachers), does s/he have students who feel more socially included, who have more egalitarian hierarchy and generally more academically engaged students?”

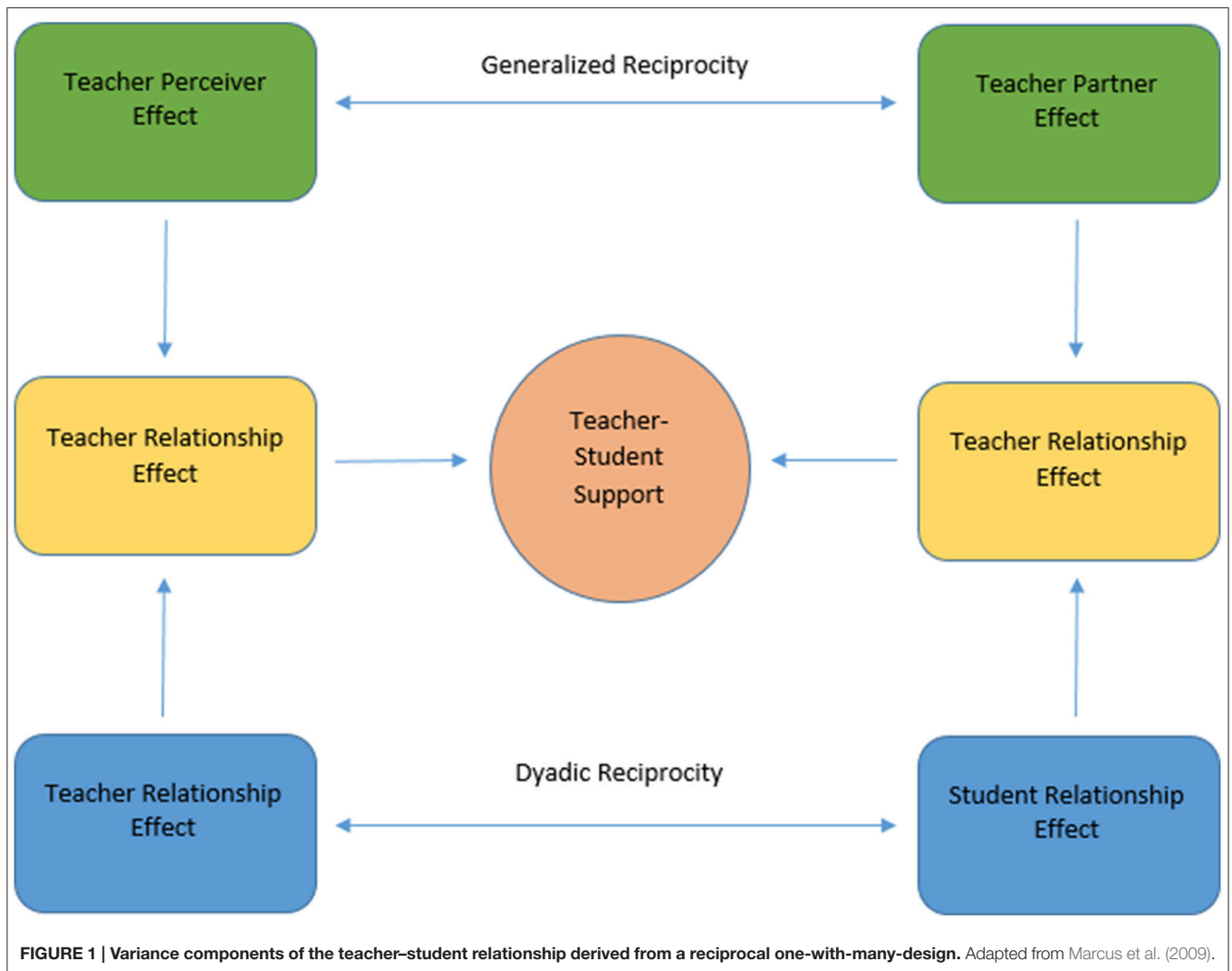


FIGURE 1 | Variance components of the teacher–student relationship derived from a reciprocal one-with-many-design. Adapted from Marcus et al. (2009).

TABLE 3 | Estimated effects by a one-with-many analysis of teacher emotional support.

Variance partitioning	Variance Source	Question answered by variance
TEACHER'S EMOTIONAL SUPPORT		
Teacher partner variance	Student report	Consensus: Do students report similar emotional support toward their teachers?
Student relationship variance	Student report	Uniqueness: Do students report unique emotional support provided by their teachers?
Teacher perceiver variance	Teacher report	Assimilation: Do teachers report to provide similar emotional support across their students?
Teacher relationship variance	Teacher report	Uniqueness: Do teachers report unique emotional support toward their students?
RECIPROCITY CORRELATIONS		
Generalized reciprocity	Teacher perceiver correlated with teacher partner	Are teachers who report strong support perceived to provide strong support by their students?
Dyadic reciprocity	Student relationship correlated with teacher relationship	If a student reports an especially strong support toward a teacher, does the teacher also report an especially strong support toward the student?

Based on scores for each outcome variable, individual student scores will be predicted using the relationship variance components (e.g., teacher relationship and student relationship effects; see Table 3). This way we can answer questions on an

individual dyadic level like “if a student thinks s/he is generally more emotionally supported by her/his teacher (compared to other students), does s/he feel more socially included and feel generally more academically engaged?”

Differences by Country

In all of the above analyses country will be added as a covariate in order to check for differences between countries.

PROSPECTIVE DISCUSSION

The proposed study will use a dyadic analysis called the one-with-many design in examining teacher–student relationships, which no other previous study has done before. This way of assessing teacher–student relationships will provide a wealth of information which have not yet been examined: By looking at the reciprocity of student and teacher reports of teacher emotional support, we can assess the importance of student's perceived emotional support on academic functioning and social inclusion. Therefore, this study will extend the literature on teacher–student relationships by including measures from the other, and equally important, part of the relationship, being the student.

Previous studies have shown that teacher support promotes academic competence and prevents problematic behaviors in the classroom (Tennant et al., 2014). However, the importance of teacher support for children's social well-being remains unknown (Farmer et al., 2011). With the dyadic analysis, perceptions of consensus between student and teacher reports could explain why certain children feel more socially included and why certain teachers establish especially inclusive social climates in their classrooms.

Accordingly it is anticipated that high teacher perceiver and partner variances—i.e., high generalized reciprocity—will predict social inclusion and classroom climate. In classrooms with low teacher perceiver and partner variances we expect to still see high relationship variances—i.e., high dyadic reciprocity—which would indicate specific dyads with uniquely strong emotional support and hence better student outcomes for the specific students involved in these dyads. Regarding Marcus et al. (2011) study in which high relationship variances were found for the therapists and substance use adolescents in regard to therapeutic alliance (Marcus et al., 2011), we do not expect such effects in the present study. As therapeutic alliances are dyadic by nature, classroom interactions between a teacher and multiple students are less likely to be denoted by such relationships as a teacher mostly interacts with the whole classroom at any given time. Hence generally we would expect higher teacher perceiver and partner variances and generalized reciprocity as opposed to relationship variances and dyadic reciprocity to reveal the general nature of classroom interactions. Due to the general similarity of the two countries where data collection has been conducted we do not expect any country differences.

Limitations exist for the one-with-many design. Since each student has one single teacher, it is not possible to completely isolate teacher partner effects. That means that it is not entirely clear whether students would report similar emotional support had they been educated by more than one teacher. Similarly it is not entirely clear whether different teachers would report similar emotional support had they all been educating the same student.

Still the partial variance partitioning provided by an one-with-many design is superior to analyses that ignore the nestedness of students. Future research that includes the perceptions of more than one teacher per classroom (potentially teachers of other subjects that also teach the same class) could make an even better use of such a design.

Assuming the current study finds effects of teacher's emotional support on children's social inclusion dependent on teacher–student relationships, it will highlight a new area of intervention. For instance, policies to improve teacher training or school interventions can be discussed to achieve inclusive classroom climates, which further can lead to better academic performance and increased well-being.

Sample size restrictions should be viewed in light of several practical limitations concerning recruitment, data collection and questionnaire administration. In a similar fashion, theoretical concerns regarding study design and power estimates need to be considered.

To begin with, recruitment and data collection were subject to ethical approval procedures. The imposed requirement to obtain active consent instead of passive parental consent resulted in delays in receiving ethical approval from the main investigator's University. Following this, a similar process of obtaining approval was undertaken and consequently delayed in each country participating in the study. In regards to parental consent, sample size was affected by factors such as refusal to participate or inability to return the completed forms. In some cases, children decided not to take part anymore or refused to complete the whole questionnaire. Consequently, these pitfalls affected the sample size at the lower level in particular.

Secondly, limitations regarding questionnaire administration have so far been identified and should be discussed. All participating pupils were asked to complete a 15-page long questionnaire. As previously mentioned, the tool had been previously piloted in English, Romanian as well as Turkish and the estimated time for completion ranged between 30 and 45 min for each language. In practice, the time allocated for data collection per classroom was set for approximately an hour. The time varied from one classroom to another, however it never exceeded the allocated time slot. Specific issues were identified with the sociometric questions. The coding system, which was set in place in order to anonymise the answers and facilitate peer nominations analysis, was difficult to understand for some participants. This is likely to have caused fatigue and boredom. Nonetheless, the questionnaire was designed in a way that would counter for the aforementioned boredom effects. In this sense, a storyline that was appealing to students was introduced prior to questionnaire completion and it was maintained throughout the whole process. Students were all told they were on a secret mission to help a team of minions find valuable information about their classroom. The response to the story was always positive and ensured students' commitment to and focus on the task. With the exception of peer nominations and the teacher–student relationship items, all questions employed a smiley-face rating scale which easy to understand and use. Moreover, each questionnaire was fairly short in length with

items varying in number from 10 (academic engagement; Skinner et al., 2008) to 26 (SIAI; Rinta et al., 2011). Finally, jokes and pictures of minions (famous cartoon characters) were included to break down the length of the questionnaire or the potential monotony.

The employed design is classified as cross-sectional as the study is conducted at one time point only. No causality could thus be inferred, however this type of design serves the aim of our research with regards to exploring the predictive nature of perceived emotional support on students' social inclusion. In addition to this, the study is designed to be multimodal and multi-informant. This is advantageous as multi-informant data is valuable in terms of obtaining a more accurate description of the studied phenomena. In practice, by gathering data about peer relationships, teacher-student relationships, classroom climate, social inclusion and academic engagement and performance, we are able to draw on a detailed depiction of classroom dynamics and their outcomes.

Finally, limitations also exist in the estimation of sample size. Generally the requirement with hierarchical data as opposed to leveled data is that the more levels there are the more parameters need to be estimated which make a priori estimations challenging as controversial discussions on the sample size estimation in MLM show (Field, 2013). Nevertheless the above standard power analyses has been run to demonstrate that sufficient power can be provided with our sample size aim. The neglect of multiple levels and slight underpower of 0.7 constitute a limitation of this simplified estimate. Though this calculation as well as any limitations of it need to be qualified by the complexity of sample size estimations in multilevel models in general, which caused many in practice to arrive to the rule of the more data, the better (Kreft and Leeuw, 1998; Field, 2013) as statistical power analyses are not traditionally being carried out for multilevel models. Another general rule with multilevel models that are reported in literature append more emphasis to the group level sample

size as opposed to the sample size of individuals within groups (Snijders, 2005; Twisk, 2006), i.e., the number on the group level is more important. More concrete estimates are provided by simulation studies stating that sample size greater than 30 have little impact on the accuracy of standard errors of fixed effects and advocate said number as normal in educational research to achieve sufficient power (Maas and Hox, 2005).

To conclude, this research will contribute to the use of psychological assessment in educational settings by introducing new methods for measuring emotional support, social inclusion and academic engagement from the view of the student and the teacher.

AUTHOR CONTRIBUTIONS

YV supervised the project, the paper and supported team members with expertise. ZH wrote the abstract, parts of the introduction, proposed analyses, and discussion. LH wrote the methods and proposed analyses. JB was in charge of the introduction and abstract. MÇ contributed to the introduction as well as prospective discussion. MCC recruitment in the UK and co-wrote abstract, introduction, methods and discussion AB was in charge of recruitment in Turkey and ethics.

FUNDING

This research was supported in part by the University of Oslo, Radboud University, and University of Worcester. The authors gratefully acknowledge the financial support which facilitated our data collection and dissemination plans.

ACKNOWLEDGMENTS

This article was supported by the Open Access Publishing Fund of the University of Vienna.

REFERENCES

- Balfanz, R., Herzog, L., and MacIver, D. (2007). Preventing student disengagement and keeping students on the graduation path in urban middle-grades schools: early identification and effective interventions. *Educ. Psychol.* 42, 223–235. doi: 10.1080/00461520701621079
- Baumeister, R. F., DeWall, C. N., Ciarocco, N. J., and Twenge, J. M. (2005). Social exclusion impairs self-regulation. *J. Pers. Soc. Psychol.* 88, 589–604. doi: 10.1037/0022-3514.88.4.589
- Boor-Klip, H. J., Segers, E., Hendrickx, M. M. H. G., and Cillessen, A. H. N. (2016). Development and psychometric properties of the classroom peer context questionnaire. *Soc. Dev.* 25, 370–389. doi: 10.1111/sode.12137
- Buyse, E., Verschueren, K., Verachtert, P., and Van Damme, J. (2009). Predicting school adjustment in early elementary school: impact of teacher-child relationship quality and relational classroom climate. *Elem. School J.* 110, 119–141. doi: 10.1086/605768
- Cillessen, A. H. N., and Marks, P. E. L. (2011). "Conceptualizing and measuring popularity," in *Popularity in the Peer System*, eds A. H. N. Cillessen, D. Schwartz, and L. Mayeux (New York, NY: Guilford Press), 25–56.
- Croninger, R. G., and Lee, V. E. (2001). Social capital and dropping out of high schools: benefits to at-risk students of teachers' support and guidance. *Teach. Coll. Rec.* 103, 548–558. doi: 10.1111/0161-4681.00127
- Farmer, T. W., McAuliffe Lines, M., and Hamm, J. V. (2011). Revealing the invisible hand: the role of teachers in children's peer experiences. *J. Appl. Dev. Psychol.* 32, 247–256. doi: 10.1016/j.appdev.2011.04.006
- Field, A. P. (2013). *Discovering Statistics Using SPSS, 3rd Edn.* London: SAGE.
- Fredricks, J. A., Blumenfeld, P., and Paris, A. (2004). School engagement: potential of the concept, state of the evidence. *Rev. Edu. Res.* 74, 59–109. doi: 10.3102/00346543074001059
- Hafen, C. A., Allen, J. P., Mikami, A. Y., Gregory, A., Hamre, B., and Pianta, R. C. (2012). The pivotal role of adolescent autonomy in secondary school classrooms. *J. Youth Adolesc.* 41, 245–255. doi: 10.1007/s10964-011-9739-2
- Hamre, B. K., and Pianta, R. C. (2006). "Student-teacher relationships," in *Children's Needs III: Development, Prevention, and Intervention*, eds G. G. Bear and K. M. Minke (Washington, DC: National Association of School Psychologists), 59–71.
- Hofstede, G., Hofstede, G. J., and Minkov, M. (2010). *Cultures and Organizations: Software of the Mind, 3rd Edn.* New York, NY: McGraw-Hill.

- Holz, G. (2004). *Lernen arme Kinder anders? Bildungsprozesse und soziale Herkunft*. Kompakt. Available online at: <http://www.katholische-kindergaerten.de/pdf/bensberger.pdf#page=39>
- Hughes, J. N. (2012). "Teachers as managers of students peer context," in *Peer Relationships and Adjustment at School*, eds A. M. Ryan and G. W. Ladd (Charlotte, NC: Information Age Publishing), 189–218.
- Hughes, J. N., and Kwok, O. (2006). Classroom engagement mediates the effect of teacher-student support on elementary students' peer acceptance: a prospective analysis. *J. Sch. Psychol.* 43, 465–480. doi: 10.1016/j.jsp.2005.10.001
- Islam, M. T., and Rashid, F. (2012). Social inclusion in secondary school classrooms in Bangladesh. *Bangladesh Educ. J.* 11, 31–39.
- Kenny, D. A., Kashy, D. A., and Cook, W. L. (2006). *Dyadic data analysis*. New York, NY: The Guildford Press.
- Kreft, I., and Leeuw, J. (1998). "Frequently Asked Questions" in *Introducing Multilevel Modeling*, ed D. B. Wright (London: SAGE), 105–137.
- Maas, C. J. M., and Hox, J. J. (2005). Sufficient Sample Sizes for multilevel modeling. *Methodology* 1, 86–92. doi: 10.1027/1614-2241.1.3.86
- Marcus, D. K., Kashy, D. A., and Baldwin, S. A. (2009). Studying psychotherapy using the one-with-many design: the therapeutic alliance as an exemplar. *J. Couns. Psychol.* 56, 537–548. doi: 10.1037/a0017291
- Marcus, D. K., Kashy, D. A., Wintersteen, M. B., and Diamond, G. S. (2011). The therapeutic alliance in adolescent substance abuse treatment: a one-with-many analysis. *J. Couns. Psychol.* 58, 449–455. doi: 10.1037/a0023196
- Martin, A. J. (2009). Age appropriateness and motivation, engagement, and performance in high school: effects of age within cohort, grade retention, and delayed school entry. *J. Educ. Psychol.* 101, 101–114. doi: 10.1037/a0013100
- Ministry of National Education (2016). *National Education Statistics: Formal Education 2015/16*. Ankara: Ministry of National Education.
- National Law for Education (2011). *Legea educatiei nationale Nr. 1/2011*. Available online at: <http://www.fsl.ro/Legislatie/Legi/LEN-Nr.1-din-2011-actualizata.pdf>
- Organisation for Economic Co-operation and Development (2013). *How long do students spend in classroom? Education at a glance 2013: Highlights*. Paris: OECD Publishing.
- Raudenbush, S. W., Brennan, R. T., and Barnett, R. C. (1995). A multivariate hierarchical model for studying psychological change within married couples. *J. Fam. Psychol.* 9, 161–174. doi: 10.1037/0893-3200.9.2.161
- Rinta, T., Purves, R., Welch, G., Elmer, S. S., and Bissig, R. (2011). Connections between children's feelings of social inclusion and their musical backgrounds. *J. Soc. Inclusion* 2, 34–57.
- Skinner, E. A., Kindermann, T. A., and Furrer, C. J. (2009). A motivational perspective of motivation and disaffection: A conceptualization and assessment of children's behavioral and emotional participation in academic activities in the classroom. *Educ. Psychol. Meas.* 69, 493–525. doi: 10.1177/0013164408323233
- Skinner, E. A., Furrer, C., Marchand, G., and Kindermann, T. (2008). Engagement and disaffection in the classroom: Part of a larger motivational dynamic? *J. Educ. Psychol.* 100, 765–781. doi: 10.1037/a0012840
- Snijders, T. A. B. (2005). "Power and sample size in multilevel linear models," in *Encyclopedia of Statistics in Behavioral Science*, Vol. 3, eds B. S. Everitt and D. C. Howell (Chichester: Wiley), 1570–1573.
- Tabachnick, G. G., and Fidell, L. S. (2007). *Experimental Designs Using ANOVA*. Belmont, CA: Duxbury.
- Tennant, J. E., Demaray, M. K., Malecki, C. K., Terry, M. N., Clary, M., and Elzinga, N. (2014). Students' ratings of teacher support and academic and social-emotional well-being. *Sch. Psychol. Q.* 30, 494–512. doi: 10.1037/spq0000106
- Twisk, J. W. R. (2006). *Applied Multilevel Analysis: A Practical Guide for Medical Researchers*. Cambridge: Cambridge University Press.
- UNESCO Institute for Statistics (2016). *Pupil-Teacher Ratio in Primary Education (Headcount Basis)*. Available online at: <http://data.worldbank.org/indicator/SE.PRM.ENRL.TC.ZS>
- United Nations Educational Scientific and Cultural Organization (2010). *Reaching the Marginalized*. Paris: Oxford University Press. Available online at: <http://unesdoc.unesco.org/images/0018/001866/186606E.pdf>

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2016 Hogekamp, Blomster, Bursaloğlu, Călin, Çetinçelik, Haastrup and van den Berg. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

APPENDIX

SPSS Syntax to Estimate the Variance Partitioning

MIXED

```

TSR WITH TEACHER STUDENT
/FIXED = TEACHER STUDENT|NOINT
PRINT = SOLUTION TESTCOV
RANDOM TEACHER STUDENT|SUBJECT(TEACHERID)
                                COVTYPE(UNR)
REPEATED = ROLE|SUBJECT(TEACHERID * STUDENTID)
                                COVTYPE(UNR)

```

Unbolded terms represent variable names. The two dummy codes for teacher (TEACHER) and student (STUDENT) are needed to estimate a two intercept model as stated in the proposed analysis above. Hence two separate intercepts are estimated for teacher and students by simultaneously suppressing the traditional intercept (i.e., NOINT).

The RANDOM statement results in estimation of the variance in the two intercepts. For teachers this variance estimates the teacher perceiver variance. For the students this variance estimates the teacher partner variance. Covariance is estimated by the COVTYPE term. UNR specification of that term means that instead a correlation between teacher perceiver and teacher partner is obtained.

The REPEATED statement is needed to specify the relationship variances. COVTYPE(UNR) obtains the correlations of between these variances.



“It’s Always the Judge’s Fault”: Attention, Emotion Recognition, and Expertise in Rhythmic Gymnastics Assessment

Lindsey G. van Bokhorst¹, Lenka Knapová^{2*}, Kim Majoranc³, Zea K. Szebeni⁴,
Adam Táborský², Dragana Tomić⁵ and Elena Cañadas⁶

¹ Maastricht University, Maastricht, Netherlands, ² Masaryk University, Brno, Czech Republic, ³ University of Ljubljana, Ljubljana, Slovenia, ⁴ Eötvös Loránd University, Budapest, Hungary, ⁵ University of Banja Luka, Banja Luka, Bosnia and Herzegovina, ⁶ University of Lausanne, Lausanne, Switzerland

OPEN ACCESS

Edited by:

Agnieszka Walczak,
Cambridge English Language
Assessment, UK

Reviewed by:

Antonio Calcagni,
Hungarian Academy of Science,
Hungary

Chris J. Gibbons,
University of Cambridge, UK

*Correspondence:

Lenka Knapová
knapova.l@gmail.com

Specialty section:

This article was submitted to
Quantitative Psychology
and Measurement,
a section of the journal
Frontiers in Psychology

Received: 08 March 2016

Accepted: 20 June 2016

Published: 05 July 2016

Citation:

van Bokhorst LG, Knapová L,
Majoranc K, Szebeni ZK,
Táborský A, Tomić D and Cañadas E
(2016) “It’s Always the Judge’s Fault”:
Attention, Emotion Recognition,
and Expertise in Rhythmic
Gymnastics Assessment.
Front. Psychol. 7:1008.
doi: 10.3389/fpsyg.2016.01008

In many sports, such as figure skating or gymnastics, the outcome of a performance does not rely exclusively on objective measurements, but on more subjective cues. Judges need high attentional capacities to process visual information and overcome fatigue. Also their emotion recognition abilities might have an effect in detecting errors and making a more accurate assessment. Moreover, the scoring given by judges could be also influenced by their level of expertise. This study aims to assess how rhythmic gymnastics judges’ emotion recognition and attentional abilities influence accuracy of performance assessment. Data will be collected from rhythmic gymnastics judges and coaches at different international levels. This study will employ an online questionnaire consisting on an emotion recognition test and attentional test. Participants’ task is to watch a set of videotaped rhythmic gymnastics performances and evaluate them on the artistic and execution components of performance. Their scoring will be compared with the official scores given at the competition the video was taken from to measure the accuracy of the participants’ evaluations. The proposed research represents an interdisciplinary approach that integrates cognitive and sport psychology within experimental and applied contexts. The current study advances the theoretical understanding of how emotional and attentional aspects affect the evaluation of sport performance. The results will provide valuable evidence on the direction and strength of the relationship between the above-mentioned factors and the accuracy of sport performance evaluation. Importantly, practical implications might be drawn from this study. Intervention programs directed at improving the accuracy of judges could be created based on the understanding of how emotion recognition and attentional abilities are related to the accuracy of performance assessment.

Keywords: rhythmic gymnastics, attention, emotion recognition, expertise, accuracy, judges

INTRODUCTION

Judges are often in the center of media coverage, repeatedly criticized, and sometimes undervalued part of the sport (Dosseville et al., 2014). Research concerning judges mostly focuses on the differences between novice and expert judges (Ste-Marie, 1999, 2000; Flessas et al., 2015) and the biases that occur during competitions (Ansorge and Scheer, 1988; Plessner, 1999;

Plessner and Schallies, 2005). However, no attention has been devoted to psychological factors such as attention, emotion recognition of judges and possible interventions which could reduce biases or stress. Mentioned factors are relevant and may influence the outcome in sports with subjective scoring, such as rhythmic gymnastics, where the results – scoring and ranking of performances – depend heavily on evaluations made by judges.

Typically the scoring, as a main tool of evaluation, used in rhythmic gymnastics is based on subjective decisions of the judges. These decisions might be affected by cognitive abilities of judges such as selective attention and vigilance. Lower cognitive abilities affect the processing of visual information (Glisky, 2007). During a competition, judges have to track multiple objects at the same time which requires high demands on their resources. Research examined that individuals can identify multiple objects simultaneously, and in case of professional judges, the number is significantly higher (Cavanagh and Alvarez, 2005). It seems training and expertise increase cognitive resources available to process visual information. Oksama and Hyönä (2004) showed that the capability for tracking multiple objects can be affected by factors such as increased speed of the objects (which decreases capabilities) and the novelty of the objects (familiar objects increase capabilities).

Throughout the evaluation of rhythmic gymnastics performances, the judges have to follow the performance of five or six gymnasts at once, while each of the gymnasts also holds an apparatus (hoops, balls, or ribbons) which becomes separated from them during throws and other elements. While paying attention to bodily movements, the skilful movement of the apparatus, coordination, etc., judges face several distracting factors such as the color of the gymnasts' dresses (Hagemann et al., 2008), the size of the crowd around them (Bai et al., 2007), the noise (Nevill et al., 2002), the disadvantageous angle from which they sometimes have to observe the performance (Oudejans et al., 2000), and even some psychological factors such as confirmation bias (Rodenberg, 2011).

Even though, rhythmic gymnastics settles rare conditions, it includes some human aspects which might be part of an evaluation and which cannot be avoided as emotions are present in every context of social interaction, including sport performance. They play an important part in the artistic evaluation of the performance, since the body movements and facial expressions have to be harmonized with the music. Emotion recognition is the ability to identify emotions of others through their nonverbal expression including face and body (Gabrielsson and Juslin, 1996). For gymnasts, emotions can influence sporting behaviors and actions through their effect on attention, and therefore concentration (Lazarus, 2000). Vast et al. (2010) studied the effect of perceived emotions on attention, especially concentration, and performance. They found that anger negatively affects concentration but not performance, whereas anxiety negatively affects performance but not concentration. Emotion recognition might play an important role during the evaluation of the performance in two ways. First, being able to accurately discriminate emotional intentions and expression of the gymnasts can improve assessment of the performance due to the amount of information that emotions

provide about the situation and the mental state of the gymnasts. Second, emotions play an important role in error detection. For instance, a brief expression of anger or disappointment by the gymnasts can be an alerting signal to judges that a mistake was committed. Without emotion expression, the mistake could have passed unnoticed.

Not only emotions play an important role in evaluation. The expertise of judges has been found to influence the evaluation of performance of gymnasts. Ste-Marie (2000) found that novice judges spend less time looking at the gymnasts and more time looking at the scoring paper than the judges with more expertise. Flessas et al. (2015) confirmed that expert judges have better error detection than novices. However, even the highest ranked international judges only reported 40% of true errors. Other research suggests that expert judges are better at perceptually anticipating upcoming gymnastic elements, which has a positive effect on their judging performance as compared to novice judges (Ste-Marie, 1999). A study, done with football referees, obtained positive correlation indices among hours practiced per week, the number of competitions judged, and the ability to evaluate a performance (Catteuw et al., 2009). Studies have also shown that both visual and motor experience with a specific sport account for correctly estimating the movement quality of other people (Loula et al., 2005; Blake and Shiffrar, 2007). However, according to a study concerning practical skills rhythmic gymnastics judges and their relation to the judging abilities was founded that judges mostly valued following skills: the knowledge of technical parameters of the sport and the capacity to adjust to any level of competition under self-assuredness and self-confidence circumstances, but not experience *per se* (Fernandez-Villarino et al., 2013).

Another factor that can greatly affect rhythmic gymnastics evaluation is chronotype of the judges. Chronotype is known as a trait that reflects individually preferred times for activity and sleep (Smith et al., 1989; Greenwood, 1994; Caci et al., 2000). According to the preferred times for activity, chronotypes are classified into “morningness”, “intermediate”, and “eveningness” types. Individuals of the morningness type prefer to be active in the early morning, whereas those of the eveningness type prefer to be active late in the evening. Chronotype is related to attention peaks at different times of day. Previous studies have shown that morning-types are more alert in the morning (Clarisse et al., 2010), while evening-types' alertness did not differ between morning and afternoon (Adan and Guàrdia, 1993) or they performed better in the evening (Schmidt et al., 2012). However, some studies found no association between chronotype and performance inattention tests (Adan, 1991; Gomes et al., 2011; Adan et al., 2012).

Due to mentioned factors it might be beneficial to use the Rasch model as it has been previously reported to be beneficial in assessing sport performances, and more concretely to evaluate consistency among the different evaluations of a judge and the different aspects of performance been evaluated (Looney, 2003).

Previous research on judged scoring in sports has mostly focused on the differences between novice and expert judges, but research on psychological factors (i.e., attention and emotion recognition) and their influence on performance assessment

have yet to be investigated. The present study investigates how attentional and emotion recognition abilities can affect accuracy of sport assessment in rhythmic gymnastics. The primary goal of this study is to evaluate attention and emotion recognition in rhythmic gymnastics judges and their relationship with accuracy in performance evaluation. Furthermore, the aim is to investigate possible differences between novice and expert judges. Finally, the results could help in improving the accuracy and future training of judges. We hypothesize that (i) attentional and emotion recognition abilities of judges are positively correlated with scoring accuracy, (ii) higher level of expertise is positively correlated with scoring accuracy, and (iii) expertise has a mediating role between attentional and emotion recognition abilities of the judges and scoring accuracy.

MATERIALS AND METHODS

Materials and Measures/Equipment Video

Participants will evaluate a set of videotaped group rhythmic gymnastics performances (2–3 min each) on the artistic and execution components of the performance score. Using recordings of past performances will ensure the same conditions for all participants and that will allow us to compare the differences in scoring between participants. Detailed information about the selected videos can be found in **Table 1**. These videos were selected in a two-step process. Firstly, all the videos from past Olympic Games freely available online were gathered. The Olympic Games were chosen for its best video coverage as well as broadest competition. It was decided that videos from the latest 2012 Olympic Games should be used with respect to the recording quality that ensures better visibility of performance details that is crucial for accurate performance evaluation. In the second step, the videos were studied regarding the obtained score, noticeability of errors, and dress conspicuousness. In the end, four videos including a range of performances from almost perfect performances to performances with more noticeable errors were selected. The resolution of the videos is 720p and the sound is muted to prevent the participants' evaluations from being influenced by the commentary.

A certain risk to the study is the fact that judges might remember the scores of the performances that were given at the Olympic Games. To check for this, the question: "Have you seen this performance before?" will be asked after each video. If their answer is yes, then they will be asked to write down the remembered scores. This will allow us to see whether they actually

remembered the score or whether the performance just looked familiar to them and control for this in the analysis.

Attentional Abilities

The attentional abilities will be measured by the Attention Network Test for Interactions and Vigilance (ANTI-V; Roca et al., 2011; including new measures of vigilance executive and vigilance of activation – ANTI-VEA) that takes about 15 minutes to complete. The test (**Figure 1**) is computer-based and measures participants' performance in five components of attention: phasic alertness, vigilance executive, vigilance of activation, orienting network, and executive control. ANTI-VEA measures these components through reaction time and accuracy (or percentage of errors). Participants have to indicate the direction of the central stimulus, by pressing C or M on the keyboard (typical flanker task). Direction of the flanker stimuli (either congruent or incongruent) tests the executive control functioning. When participants detect that the central stimulus is displaced, they have to press spacebar. This type of stimuli is unpredictable, infrequent, and unexpected and detection of such a stimulus measures vigilance executive, as Interactions and Vigilance (ANTI-V; Roca et al., 2011). Vigilance of activation is measured by the Psychomotor Vigilance Task (PVT; Correa et al., 2014), a numerical count down starting at 999 ms that participants have to stop as fast as possible by pressing any key. Visual cue assesses the functioning of the orienting network (valid, invalid, or no cue condition). Auditory warning signal (50 ms) that announces the appearance of the stimulus measures phasic alertness.

The advantage of ANTI-VEA is that it tests the five attentional components in an independent way and allows measuring of the interactions between them. Moreover, it measures the reaction time for every answer. Reliability of ANTI-V was 0.99 (Roca, 2012) and we expect the ANTI-VEA to be highly reliable as well as it is a new, slightly adjusted, version of ANTI-V.

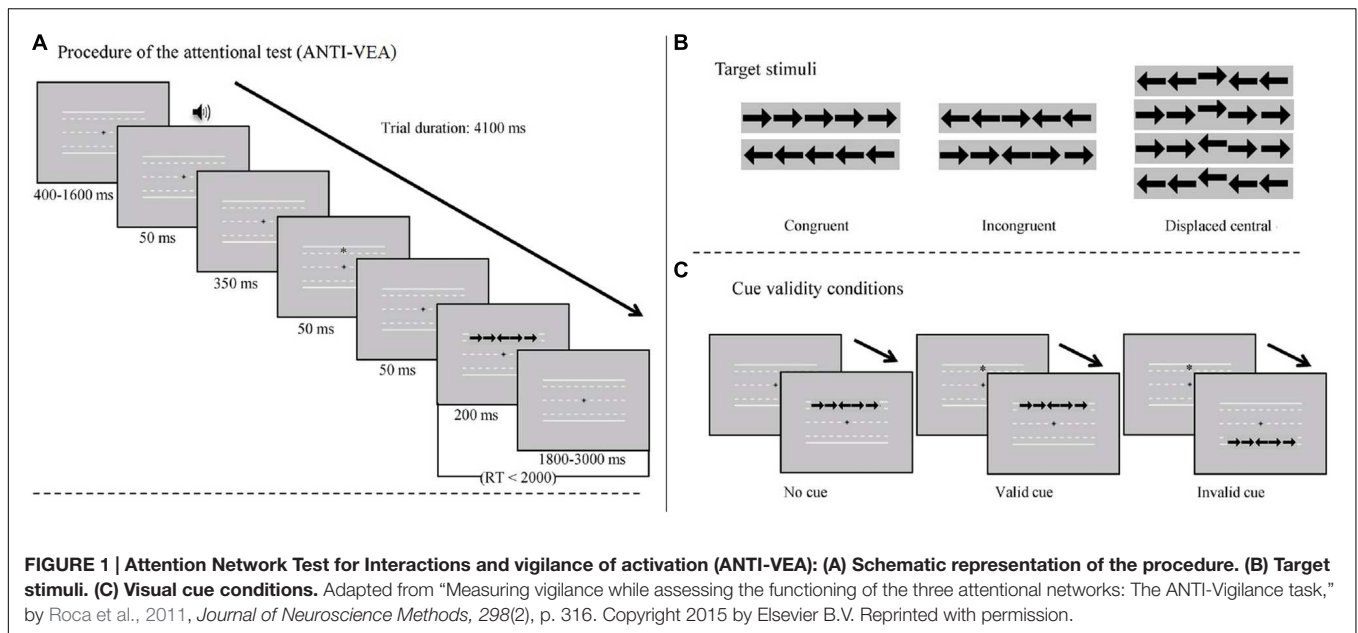
Emotion Recognition Ability

The Geneva Emotion Recognition Test-Short (GERT-S; Schlegel and Scherer, 2015) measures individual differences in the ability to recognize emotions of others. It is a shortened version of the Geneva Emotion Recognition Test (GERT; Schlegel et al., 2014). GERT-S consists of 42 items short audio-visual clips in which ten actors express 14 different emotions (e.g., anger, relief, joy, pride, and anxiety) through their face, voice, and body. After each clip, participants have to choose one of the 14 emotions which they believe was expressed by the actor. This test is dynamic and multimodal (short video

TABLE 1 | Chosen group rhythmic gymnastics performances from 2012 Olympic Games.

Competition	Discipline	Nation	Score artistic	Score execution	Place
Qualification	3 hoops, 2 ribbons	Russian Federation	9.550	9.100	1st
Qualification	3 hoops, 2 ribbons	Greece	8.900	8.450	9th
Final	3 hoops, 2 ribbons	Spain	9.300	8.800	4th
Final ^a	3 hoops, 2 ribbons	Bulgaria	9.450	8.850	6th

^aThis video will be used as a practice video to get familiar with the judging process.



clips with sound) and measures ERA in a more ecologically valid fashion. In comparison with other ERA tests which usually offer basic emotions, GERT-S features 14 different emotions, both positive and negative. Moreover, GERT-S internal consistency is between 0.80 and 0.83 (Schlegel and Scherer, 2015).

Expertise

Participants will be asked about their judging experience (judging category, years of judging at the international level, and number of judged competitions) to determine their expertise. There are four categories of FIG judges for all disciplines, Category IV being the lowest and Category I the highest. Category IV Judges are new international judges with no or little international experience. They are not assigned as judges in major competitions. Category III Judges are defined as experienced judges with good results in execution and artistic. They are designated to judge execution and artistic in major competitions. Category II Judges are experienced judges with very good results in difficulty. They are designated to judge difficulty as well as execution and artistic. Category I Judges are very experienced judges with excellency in difficulty and operate as members of the Superior Jury, Chair of Judges' Panel, and Difficulty Judges and are also designated to judge execution and artistry (Fédération Internationale de Gymnastique [FIG], 2013).

Additional Questions

Participants will further answer basic demographic questions (gender, age, nationality, etc.), questions about their chronotype (rMEQ; Adan and Almirall, 1991; example question: *At what time in the evening do you feel tired and as a result in need of sleep?*), and manipulation check questions (whether they have seen the presented performances before and whether they remember the scores).

STEPWISE PROCEDURE

Power Analysis

The software G*power (Faul et al., 2007) was used to calculate the required sample size. Values for α were set on 0.05 and power on 0.80. Based on previous literature and discussions between the authors, effect sizes were estimated on 0.15 (medium effect). In total, to reach the desired power, data from 55 participants has to be collected.

Qualtrics Survey

The study uses an online survey designed on Qualtrics which enables judges from different countries to participate. There is a great benefit of using videotaped performances that can be assessed on any computer over real-time evaluations at competitions. With regards to the different nationalities of the participants, the instructions will be translated into several languages to optimize the comprehensibility of the survey for judges from different cultural backgrounds and native languages. Although translating instructions into different languages might pose limitations to the internal validity of the study, the current study will use native speakers with a fluency in English to translate instructions into their native languages. Participants recruitment and the overall small research population can also cause a limitation in regards to the sample size of the research, but it is important to notice that the actual population of judges in rhythmic gymnastics at international level is also limited (376 judges approved by the FIG). Therefore, it would be interesting in future experiments to include judges from different disciplines, such as dance sport, to extent the research population. Finally, the current study applied a new approach to measure accuracy of performance scoring, as to the authors' knowledge such measuring tools have not been developed nor validated. However, the official scores given at international competitions

are considered as the most accurate attainable scores, and it can therefore be concluded that participants that gave similar scores have high scoring accuracy.

Participant Recruitment

This study aims to recruit 50 rhythmic gymnastics judges approved by the International Gymnastics Federation (13% of the total number of judges at this level) and a control group of 20 rhythmic gymnastics coaches unskilled in performance judging. Participants will be recruited via email, social media, personal contact, and with the help of national gymnastics federations. After confirming participation via a consent form following the ethical committee approval from the University of Lausanne, they will complete an online testing session. Participation will be anonymous and voluntary.

Procedure

Data will be collected using an online survey design based on Qualtrics including measures of personal characteristics and the assessment of videotaped performances of rhythmic gymnasts. The survey will have multiple language versions. The first part of the survey will focus on the evaluation of group rhythmic gymnastics performances. Initially, participants will be asked to read the International Gymnastics Federation's Code of Points on the artistic and execution components. Participants will then be given one practice video to get familiar with the judging process. After the practice video, participants will be shown three video clips of performances from the 2012 Olympic Games. Participants will watch the videos and score them according to the artistic and execution components. Their scores will be compared to the official scores given at the competition the videos were taken from to determine the scoring accuracy of each participant. The official scores given at the Olympic Games are considered to be the most accurate scores that are attainable for this purpose. The second part of the online survey will consist of two tests measuring attentional (ANTI-VEA) and emotion recognition abilities (GERT-S). Finally, participants will answer additional questions about their judging experience (expertise), chronotype (rMEQ), demographic data, and manipulation check.

STATISTICAL ANALYSIS

To test our hypothesis we will use statistical software including SPSS, R and Stata. We will follow the four-faceted Rasch model (gymnastic group ability, aspect difficulty, program difficulty, and judge severity) to assure that valid evaluation occurred between videos and within judges. To do that, we will follow the protocol recommended by Linacre (1999) and Wright and Mok (2000) by mean of the Mififac Rasch Software. Once confirmed that the data fit the model well enough so judges' performance evaluations are valid, we then will create a variable that indicates how much actual judges' evaluations deviate from Olympic judges' evaluations (accuracy).

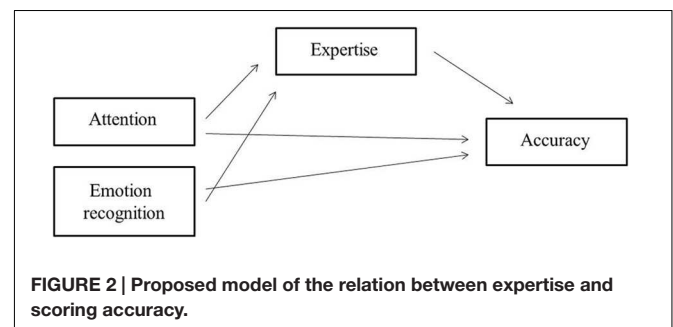
Next, we will perform correlational analysis to evaluate the hypothesized positive correlations between attentional abilities

and accuracy, ERA and accuracy as well as expertise and accuracy. When testing a more complex model that includes causality and interrelations between the variables, we encounter the problem of working with endogenous variables, in our case latent variables that are being estimated through their manifestations, i.e., attention and emotion recognition (Antonakis et al., 2010). Moreover, the relationship between these variables and the dependent variable (accuracy) is expected to partly function through other variables in the model (i.e., expertise). Therefore the data will be further analyzed using regression analysis and Structural Equation Modelling (SEM) to test a mediational role of the relationship between expertise and accuracy (**Figure 2**).

The evaluation of the fit of the model to the obtained data we will rely on the Root Mean Square Error of Approximation index (RMSEA; Steiger and Lind, 1980) that is based on the χ^2 value of the model as well as the number of observed cases and degrees of freedom, as depicted in Equation 1. RMSEA's values of 0.05 and lower indicate a good fit, while values up to 0.08 indicate an acceptable fit (Browne and Cudeck, 1993). We will use also *SWAIN* correction, as a measure to address sensitivities of χ^2 statistic to sample size and model complexity (Antonakis and Bastardo, 2013). RMSEA formula retrieved from <http://davidakenny.net/cm/fit.htm>,

$$\frac{\sqrt{(\chi^2 - df)}}{\sqrt{[df(N - 1)]}} \quad (1)$$

To address the above-mentioned issue of endogeneity and to deliver consistent parameter estimates, the analysis will follow Antonakis et al. (2010) recommendations: (a) include instrumental variables (variables uncorrelated with the error term of the dependent variable and correlated with the endogenous variables) that are necessary to provide a sufficient identification of the model, for instance, age, gender, etc., (b) be performed through the Two-Stage Least-Squares procedure (2SLS; James and Singh, 1978; Baltagi, 2011), a tool used in SEM to estimate path coefficients. In a first-stage, attention and ERA will be regressed on two instrumental variables and covariates, including fixed effects. In the second-stage, accuracy will be regressed on the predicted value of attention and ERA from the first stage and covariates, including fixed effects. We will verify that the instrumental variables used are supported by theoretical and empirical considerations.



ANTICIPATED RESULTS

We expect to support our hypotheses by testing the model presented in the methods section. The first hypothesis stated that attention and emotion recognition are positively related to scoring accuracy. It is assumed that higher levels of attention improve error detection of judges and therefore enable them to rate the performances more accurately, and higher levels of emotion recognition provide judges with the skill to detect changes in facial or bodily expressions that accompany mistakes in the performance. The second and third hypotheses propose a relationship between expertise and attention and emotion recognition and a relationship between expertise and scoring accuracy. By testing our anticipated model we expect to find that expertise is also associated with attention and emotion recognition, and expertise itself is associated with scoring accuracy. Better expertise, defined both as level and years of judging, provides judges with more experience and training to judge more accurately, and it may therefore be assumed that judges of higher levels would be significantly more accurate in their assessments.

The study is facing some potential pitfalls when collecting data. For example, as participants include judges at international levels, some may have already seen the performances to be

evaluated in the survey and therefore remember the scores that were given at the competition. We will control for positive answer to those questions. Another important factor that is included in the survey is whether participants are morning or evening types (i.e., chronotype), as different types reach their peak level of attention and concentration at different moments during the day. Not completing the survey within one's peak hours may possibly influence the person's accuracy and scores of the ANTI-VEA negatively. To take this into account, we control for both self-reported chronotype as well as the moment of the day the tasks were performed, to make sure everyone was at their full activity range.

AUTHOR CONTRIBUTIONS

EC is the author of the concept of the study, the supervisor of the project, and provided feedback when writing the manuscript. KM, ZS, and AT were in charge of writing the introductory part of the manuscript (including literature review). LvB, LK, and DT were in charge of writing the Materials and Methods, Stepwise Procedure, Statistical Analysis, and Anticipated Results. All of the authors contributed to frequent feedback sessions and the development of the project in general.

REFERENCES

- Adan, A. (1991). Influence of morningness–eveningness preference in the relationship between body temperature and performance: a diurnal study. *Pers. Individ. Dif.* 12, 1159–1169. doi: 10.1016/0191-8869(91)90080-U
- Adan, A., and Almirall, H. (1991). Horne & Östberg morningness-eveningness questionnaire: a reduced scale. *Pers. Individ. Dif.* 12, 241–253. doi: 10.1016/0191-8869(91)90110-W
- Adan, A., Archer, S. N., Hidalgo, M. P., Di Milia, L., Natale, V., and Randler, C. (2012). Circadian typology: a comprehensive review. *Chronobiol. Int.* 29, 1153–1175. doi: 10.3109/0742052.2012.719971
- Adan, A., and Guàrdia, J. (1993). Circadian variations of self-reported activation: a multidimensional approach. *Chronobiologia* 20, 233–244.
- Ansorge, C. J., and Scheer, J. K. (1988). International bias detected in judging gymnastic competition at the 1984 Olympic Games. *Res. Q. Exerc. Sport* 59, 103–107. doi: 10.1080/02701367.1988.10605486
- Antonakis, J., and Bastardo, N. (2013). *Swain: Stata Module to Correct the SEM Chi-Square Overidentification Test in Small Sample Sizes or Complex Models*. Chestnut Hill, MA: Boston College Department of Economics.
- Antonakis, J., Bendahan, S., Jacquart, P., and Lalive, R. (2010). On making causal claims: a review and recommendations. *Leadersh. Q.* 21, 1086–1120. doi: 10.1016/j.leaqua.2010.10.010
- Bai, L., Lao, S., Zhang, W., Jones, G. J., and Smeaton, A. F. (2007). “A semantic event detection approach for soccer video based on perception concepts and finite state machines,” in *Proceedings of the Eight International Workshop on Image Analysis for Multimedia Interactive Services* (Washington, DC: IEEE Computer Society), 30. doi: 10.1109/wiamis.2007.12
- Baltagi, B. H. (2011). *Econometrics*. New York, NY: Springer. doi: 10.1007/978-3-642-20059-5
- Blake, R., and Shiffrar, M. (2007). Perception of human motion. *Annu. Rev. Psychol.* 58, 47–73. doi: 10.1146/annurev.psych.57.102904.190152
- Browne, M. W., and Cudeck, R. (1993). “Alternative ways of assessing model fit,” in *Testing Structural Equation Models*, eds K. A. Bollen and J. S. Long (Beverly Hills, CA: Sage Publications), 136–162.
- Caci, H., Nadalet, L., Staccini, P., Myquel, M., and Boyer, P. (2000). The composite scale of morningness: further psychometric properties and temporal stability. *Eur. Psychiatry* 15, 278–281. doi: 10.1016/S0924-9338(00)00231-5
- Cartheeuw, P., Helsen, W., Gilis, B., and Wagemans, J. (2009). Decision-making skills, role specificity, and deliberate practice in association football refereeing. *J. Sports Sci.* 27, 1125–1136. doi: 10.1080/02640410903079179
- Cavanagh, P., and Alvarez, G. A. (2005). Tracking multiple targets with multifocal attention. *Trends Cogn. Sci.* 9, 349–354. doi: 10.1016/j.tics.2005.05.009
- Clarisse, R., Le Floch, N., Kindelberger, C., and Feunteun, P. (2010). Daily rhythmicity of attention in morning- vs. evening-type adolescents at boarding school under different psychosociological testing conditions. *Chronobiol. Int.* 27, 826–841. doi: 10.3109/07420521003794051
- Correa, Á., Molina, E., and Sanabria, D. (2014). Effects of chronotype and time of day on the vigilance decrement during simulated driving. *Accid. Anal. Prev.* 67, 113–118. doi: 10.1016/j.aap.2014.02.020
- Dosseville, F., Laborde, S., and Garnarczyk, C. (2014). “Current research in sports officiating and decision-making,” in *Contemporary Topics and Trends in the Psychology of Sports*, ed. C. Mohiyeddini (New York, NY: Nova Publishers), 13–38.
- Faul, F., Erdfelder, E., Lang, A. G., and Buchner, A. (2007). G*Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behav. Res. Methods* 39, 175–191. doi: 10.3758/BF03193146
- Fédération Internationale de Gymnastique [FIG] (2013). *Code of Points Rhythmic Gymnastics 2013–2016*. Lausanne: Fédération Internationale de Gymnastique.
- Fernandez-Villarino, M. A., Bobo-Arce, M., and Sierra-Palmeiro, E. (2013). Practical skills of rhythmic gymnastics judges. *J. Hum. Kinet.* 39, 243–249. doi: 10.2478/hukin-2013-0087
- Flessas, K., Mylonas, D., Panagiotaropoulou, G., Tsopani, D., Korda, A., Siettos, C., et al. (2015). Judging the judges' performance in rhythmic gymnastics. *Med. Sci. Sports Exerc.* 47, 640–648. doi: 10.1249/MSS.00000000000000425
- Gabrielsson, A., and Juslin, P. N. (1996). Emotional expression in music performance: between the performer's intention and the listener's experience. *Psychol. Music* 24, 68–91. doi: 10.1177/0305735696241007
- Glisky, E. L. (2007). “Changes in cognitive function in human aging,” in *Brain Aging: Models, Methods, and Mechanisms: Frontiers in Neuroscience*, ed. D. R. Riddle (Boca Raton, FL: CRC Press), 3–20.
- Gomes, A. A., Tavares, J., and deAzevedo, M. H. P. (2011). Sleep and academic performance in undergraduates: a multi-measure, multi-predictor approach. *Chronobiol. Int.* 28, 786–801. doi: 10.3109/0742052.2011.606518

- Greenwood, K. M. (1994). Long-term stability and psychometric properties of the Composite Scale of Morningness. *Ergonomics* 37, 377–383. doi: 10.1080/00140139408963653
- Hagemann, N., Strauss, B., and Leifßing, J. (2008). When the referee sees red. *Psychol. Sci.* 19, 769–771. doi: 10.1111/j.1467-9280.2008.02155.x
- James, L. R., and Singh, B. K. (1978). An introduction to the logic, assumptions, and basic analytic procedures of two-stage least squares. *Psychol. Bull.* 85, 1104–1122. doi: 10.1037/0033-2909.85.5.1104
- Lazarus, R. S. (2000). How emotions influence performance in competitive sport. *Sport Psychol.* 14, 229–252.
- Linacre, J. M. (1999). Estimation methods for Rasch measures. *J. Outcome Meas.* 3, 382–405.
- Looney, M. A. (2003). Evaluating judge performance in sport. *J. Appl. Meas.* 5, 31–47.
- Loula, F., Prasad, S., Harber, K., and Shiffrar, M. (2005). Recognizing people from their movement. *J. Exp. Psychol. Hum. Percept. Perform.* 31, 210–220. doi: 10.1037/0096-1523.31.1.210
- Nevill, A. M., Balmer, N. J., and Williams, A. M. (2002). The influence of crowd noise and experience upon refereeing decisions in football. *Psychol. Sport Exerc.* 3, 261–272. doi: 10.1016/S1469-0292(01)00033-4
- Oksama, L., and Hyönä, J. (2004). Is multiple object tracking carried out automatically by an early vision mechanism independent of higher-order cognition? An individual difference approach. *Vis. Cogn.* 11, 631–671. doi: 10.1080/13506280344000473
- Oudejans, R. R., Verheijen, R., Bakker, F. C., Gerrits, J. C., Steinbrückner, M., and Beek, P. J. (2000). Errors in judging ‘offside’ in football. *Nature* 404, 33. doi: 10.1038/35003639
- Plessner, H. (1999). Expectation biases in gymnastics judging. *J. Sport Exerc. Psychol.* 21, 131–144.
- Plessner, H., and Schallies, E. (2005). Judging the cross on rings: a matter of achieving shape constancy. *Appl. Cogn. Psychol.* 19, 1145–1156. doi: 10.1002/acp.1136
- Roca, J. (2012). *Measuring the Three Attentional Networks in Vigilance Context and their Relationship with Driving Behavior*. Doctoral dissertation, University of Granada, Granada.
- Roca, J., Castro, C., López-Ramón, M. F., and Lupiáñez, J. (2011). Measuring vigilance while assessing the functioning of the three attentional networks: the ANTI-Vigilance task. *J. Neurosci. Methods* 298, 312–324. doi: 10.1016/j.jneumeth.2011.04.014
- Rodenberg, R. (2011). Perception? Reality: analyzing specific allegations of NBA referee bias. *J. Quant. Anal. Sports* 7, 8. doi: 10.2202/1559-0410.1326
- Schlegel, K., Grandjean, D., and Scherer, K. R. (2014). Introducing the Geneva emotion recognition test: an example of Rasch-based test development. *Psychol. Assess.* 26, 666–672. doi: 10.1037/a0035246
- Schlegel, K., and Scherer, K. R. (2015). Introducing a short version of the Geneva Emotion Recognition Test (GERT-S): psychometric properties and construct validation. *Behav. Res. Methods* 47, 1–10.
- Schmidt, C., Peigneux, P., Leclercq, Y., Sterpenich, V., Vandewalle, G., Phillips, C., et al. (2012). Circadian preference modulates the neural substrate of conflict processing across the day. *PLoS ONE* 7:e29658. doi: 10.1371/journal.pone.0029658
- Smith, C. S., Reilly, C., and Midkiff, K. (1989). Evaluation of three circadian rhythm questionnaires with suggestions for an improved measure of morningness. *J. Appl. Psychol.* 74, 728–738. doi: 10.1037/0021-9010.74.5.728
- Ste-Marie, D. M. (1999). Expert-novice differences in gymnastic judging: an information-processing perspective. *Appl. Cogn. Psychol.* 13, 269–281. doi: 10.1002/(SICI)1099-0720(199906)13:3<269::AID-ACP567>3.0.CO;2-Y
- Ste-Marie, D. M. (2000). Expertise in women’s gymnastic judging: an observational approach. *Percept. Mot. Skills* 90, 543–546. doi: 10.2466/pms.2000.90.2.543
- Steiger, J. H., and Lind, J. (1980). Statistically-based tests for the number of common factors. *Paper Presented at the Annual Spring Meeting of the Psychometric Society*, Iowa City, IA.
- Vast, R. L., Young, R. L., and Thomas, P. R. (2010). Emotions in sport: perceived effects on attention, concentration, and performance. *Aust. Psychol.* 45, 132–140. doi: 10.1080/00050060903261538
- Wright, B. D., and Mok, M. (2000). Rasch models overview. *J. Appl. Meas.* 1, 83–106.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2016 van Bokhorst, Knapová, Majoranc, Szebeni, Táborský, Tomić and Cañadas. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Study Protocol on Ecological Momentary Assessment of Health-Related Quality of Life Using a Smartphone Application

Silvana Mareva^{1†}, David Thomson^{2†}, Pietro Marenco³, Víctor Estal Muñoz⁴,
Caroline V. Ott⁵, Barbara Schmidt⁶, Tobias Wingen⁷ and Angelos P. Kassianos^{8*}

¹ Department of Psychology, University of Edinburgh, Edinburgh, UK, ² School of Psychology, University of Glasgow, Glasgow, UK, ³ Department of Psychology, University of Bologna, Bologna, Italy, ⁴ Department of Personality, Evaluation and Psychological Treatment, Faculty of Psychology, Autonomous University of Madrid, Madrid, Spain, ⁵ Department of Psychology, University of Copenhagen, Copenhagen, Denmark, ⁶ Department of Ergonomics and Psychology, Budapest University of Technology and Economics, Budapest, Hungary, ⁷ Department of Psychology, University of Cologne, Cologne, Germany, ⁸ Department of Applied Health Research, University College London, London, UK

OPEN ACCESS

Edited by:

Sam Norton,
King's College London, UK

Reviewed by:

Antonio Calcagni,
Magyar Tudományos Akadémia,
Hungary

Andrew Robert Johnson,
Curtin University; University
of Sydney, Australia

*Correspondence:

Angelos P. Kassianos
angelos.kassianos@ucl.ac.uk

[†]These authors have contributed
equally to this work.

Specialty section:

This article was submitted to
Quantitative Psychology
and Measurement,
a section of the journal
Frontiers in Psychology

Received: 07 March 2016

Accepted: 04 July 2016

Published: 18 July 2016

Citation:

Mareva S, Thomson D, Marenco P,
Estal Muñoz V, Ott CV, Schmidt B,
Wingen T and Kassianos AP (2016)
Study Protocol on Ecological
Momentary Assessment
of Health-Related Quality of Life Using
a Smartphone Application.
Front. Psychol. 7:1086.
doi: 10.3389/fpsyg.2016.01086

Health-Related Quality of Life (HRQoL) is a construct of increasing importance in modern healthcare, and has typically been assessed using retrospective instruments. While such measures have been shown to have predictive utility for clinical outcomes, several cognitive biases associated with human recall and current mood state may undermine their validity and reliability. Retrospective tools can be further criticized for their lack of ecology, as individuals are usually assessed in less natural settings such as hospitals and health centers, and may be obliged to spend time and money traveling to receive assessment. Ecological momentary assessment (EMA) is an alternative, as mobile assessment using mobile health (mHealth) technology has the potential to minimize biases and overcome many of these limitations. Employing an EMA methodology, we will use a smartphone application to collect data on real-time HRQoL, with an adapted version of the widely used WHOQOL-BREF questionnaire. We aim to recruit a total of 450 healthy participants. Participants will be prompted by the application to report their real-time HRQoL over 2 weeks together with information on mood and current activities. At the end of 2 weeks, they will complete a retrospective assessment of their HRQoL and they will provide information about their sleep quality and perceived stress. The psychometric properties of real-time HRQoL will be assessed, including analysis of the factorial structure, reliability and validity of the measure, and compared with retrospective HRQoL responses for the same 2-week testing period. Further, we aim to identify factors associated with real-time HRQoL (e.g., mood, activities), the feasibility of the application, and within- and between-person variability in real-time HRQoL. We expect real-time HRQoL to have adequate validity and reliability, and positive responses on the feasibility of using a smartphone application for routine HRQoL assessment. The direct comparison of real-time and retrospective measures in this study will provide important novel insight into the efficacy of mHealth applications for HRQoL assessment. If shown to be valid, reliable and feasible for the collection

of HRQoL data, mHealth applications may have future potential for facilitating clinical assessment, patient-physician communication, and monitoring individual HRQoL over course of treatment.

Keywords: mobile health, health-related quality of life, ecological momentary assessment, sleep quality, real-time assessment, smartphone application

INTRODUCTION

Health-Related Quality of Life (HRQoL) constitutes a multidimensional construct for the interpretation of health states of individuals or groups. Health-Related Quality of Life explains variation in survival of chronic conditions such as cancer (Steel et al., 2014) and is associated with outcomes in non-clinical populations, such as better sleep quality (Ratcliff et al., 2014), activity levels (Bize et al., 2007) and exercise capacity (Lindholm et al., 2003). Further, routine assessment of HRQoL has been shown to improve patient-physician communication (Velikova et al., 2004).

The increasing importance of measuring HRQoL, particularly in clinical settings (Catania et al., 2015), has precipitated greater demand for the development of standardized measurement tools. Typically HRQoL is assessed using retrospective self-reports, which rely on participants' ability to recall information from episodic memory. As episodic memory declines over time, individuals develop greater reliance on semantic memory to complete the resultant 'gaps' in recall (Maes et al., 2015). Constructive mental processes recombine elements of past events, and are prone to cognitive biases (Schacter, 2012). Specifically, when individuals respond to questions regarding their HRQoL, they estimate the intensity and frequency of experiences based on a set of highly subjective heuristics (Solhan et al., 2009).

Several cognitive biases compromise the validity of retrospective HRQoL assessment. Recall bias creates inaccuracies during retrospective assessment (Blome and Augustin, 2015) and undermines the statistical power and validity of HRQoL tools (Schwartz et al., 2004). The peak-end phenomenon is another cognitive bias involving the tendency to recall the most extreme and recent instances of an experience or feeling. The mood congruency effect refers to the employment of personalized heuristics to reconstruct memories. Therefore, individuals often use their current mood as a reference point rather than accurately recalling specific instances of moods (Solhan et al., 2009), resulting in better recall for states congruent with current mood, and potentially generating recall bias. Further, individuals with greater fluctuations in momentary experiences (e.g., pain, mood) recall instances less accurately than individuals with more stable feelings, upon weekly retrospective assessment (Stone et al., 2005).

The limitations of retrospective assessment necessitate the development of more robust tools. Modern advances in mobile health (mHealth) have facilitated ecological momentary assessment (EMA), the repeated collection of information about participants' real-time experiences in their natural environments

(Shiffman et al., 2008). EMA encapsulates many modes of assessment such as transactional diaries (Freedman et al., 2006) or the use of palm-top computers (Shiffman et al., 2008). EMA has the potential to overcome barriers of HRQoL assessment in clinical practice such as time consumption, expensive resources, paper filling and data management (Wright et al., 2003).

The primary benefit of EMA is that real-time experiential measurement circumvents the previously described cognitive biases faced when using retrospective assessment. Experiential variance and fluctuation become informative factors, as EMA seeks to provide a clear picture of subjective experience over the course of time. Indeed, using the electronic beep device *PsyMate*, Maes et al. (2015) administered HRQoL assessment 10 times a day during a 6 days period to both clinical and healthy populations. Their results revealed that real-time reports of moods and symptoms predicted within-person variation in real-time, but not retrospective HRQoL. This finding provides further evidence to suggest that retrospective assessments may provide a biased account of the impact of health problems on the lives of those affected. Moreover, this bias may differ across different conditions. Thus, the EMA promises to provide a valuable improvement to the measurement of HRQoL.

Ecological momentary assessment can also be convenient in clinical practice: remote assessment eliminates time and traveling costs, and allows individuals more flexibility in daily routine (Mehl and Holleran, 2007). The idiographic nature of EMA enables assessment in specific situations. For example, the PedsQL Visual Analog Scale, a momentary HRQoL assessment intended for young children, was found to be reliable for recording their experiences (Sherman et al., 2006).

In light of such potential benefits of the EMA approach, here we provide a protocol that seeks to extend the work of Maes et al. (2015). In particular, we aim to improve the feasibility of the EMA assessment by implementing it in a more accessible device (i.e., mobile phone) and by collecting reports at four time points during a 2-week period aiming to thus minimize respondent's time-burden. Further, we shall test the psychometric properties and the perceived feasibility of this EMA approach. While some studies employing similar methodologies have reported good ease-of-use and responder satisfaction (Maes et al., 2015), these feasibility analyses have not been comprehensive. Similarly, the validity of developing EMA measures is a key concern: while many EMA studies report their methodology as useful for experiential assessment, few have explicitly validated their measure with direct comparison to traditional measures.

These considerations highlight the need for more evidence on the validity of EMA measures and HRQoL assessed using mHealth applications. Hence, the primary aim of this study is to determine the validity and reliability of using a mHealth application to collect real-time HRQoL. This population is used to identify whether the application is valid in order to determine if there is any merit in testing the new method with a clinical population in the future. This feasibility study aims to test a modality of measuring HRQoL using an established, valid and reliable questionnaire (WHOQOL-BREF). The secondary aims are to investigate individual factors associated with HRQoL variation and to examine the feasibility of this EMA method. These aims will be explored through the following research questions:

1. How do specific domains of HRQoL correlate between real-time and retrospective measurement?
2. Does real-time HRQoL have the same factorial structure as retrospective HRQoL?
3. What is the convergent validity of real-time HRQoL?
4. How reliable is real-time HRQoL, across different time points, compared to retrospective HRQoL?
5. To what extent do mood and current activities account for variation in real-time HRQoL?
6. How do participants judge the feasibility of the mobile application?

In this protocol we provide details about the materials and procedures necessary for EMA of HRQoL using a mobile application. Further, we outline a potential data analysis strategy and prospective discussion of the protocol's implications and limitations.

MATERIALS AND EQUIPMENT

Literature Search and Choice of Measures

To identify suitable research measures a literature search of electronic databases (PubMed, PsycNet) was performed for literature relevant to HRQoL assessment and mHealth applications. The tools outlined below were selected for their relevance to the research question and their good psychometric properties.

Demographic Questionnaire

Participants will be asked for information on gender, occupation (field and level of study, if students), family status, socio-economic status, country of residence, living arrangements, number of children, frequency of smartphone usage, and major life events.

HRQoL

The WHOQOL-BREF will be used (The WHOQOL Group, 1994) to assess HRQoL. It contains 26 items comprising four domains: physical health, mental health, social relationships and environment, and two general health items (one for overall quality of life and one for overall health). The instrument

has satisfactory validity and reliability in clinical and healthy samples (Lin et al., 2007; Krägeloh et al., 2011). Further, the instrument was developed through a cross-cultural collaboration and its dimensions have been found reliable and valid across many different cultures (Power et al., 1999). This allows for scores obtained in different countries to be combined. For EMA, the wording of the original WHOQOL-BREF questionnaire was modified to be appropriate for real-time responses (e.g., instructing participants to think about their experiences “at this exact moment in time” rather than “over the last 2 weeks”). The original retrospective questionnaire will be used at the end of the 2 weeks and the modified real-time version will be administered during the 2-week assessment. Further, we will only use the physical and mental domains of the WHOQOL-BREF, as the social and environmental domain items were considered less flexible for real-time modification (i.e., people tend not to evaluate social relationships or living conditions on a real-time basis). The questionnaire scoring procedure will be followed. For this study, two domain scores will be provided (physical health and mental health) whilst the two general health items will be scored separately. The mean score of items of each domain will be used for the domain score. Following this, the scores will be converted into a scale for each domain ranging from 0 to 100.

Mood and Current Activities

Mood will be assessed in real-time using the Brief Mood Introspection Scale (BMIS; Mayer and Gaschke, 1988) which tests two main components – individuals' direct experience of specific moods, and the overall “pleasantness” of their mood. The scale has satisfactory reliability and has sufficient sensitivity to distinguish between individuals in low and high mood (Mayer and Gaschke, 1988). The tool will be administered along with the real-time HRQoL questionnaire to assess a potential mood-congruency effect on reports of HRQoL. Participants will be asked to rate their mood on seven mood items (lively, happy, grouchy, sad, tired, nervous, content) on a four-item Likert scale rating from ‘definitely do not feel’ to ‘definitely feel.’ Then, they will be asked to rate their current mood on a scale from –10 to 10 ranging from ‘very unpleasant’ to ‘pleasant.’ The item responses will be summed to obtain a score for each specific mood and total mood score. To further appreciate the context of mood-congruency judgments, participants will also provide information about their current activities prior to reporting their HRQoL.

Sleep Quality

Sleep quality will be measured retrospectively at the end of the 2 weeks using the Pittsburgh Sleep Quality Index (PSQI; Buysse et al., 1989). The PSQI asks participants to rate series of items to generate seven component scores: subjective sleep quality, sleep latency, sleep duration, sleep efficiency, sleep disturbance, use of sleep medication and daytime dysfunction. To make the assessment more feasible bed partner ratings will not be recorded. The tool has high test–retest reliability and good validity with both clinical and healthy populations (Backhaus et al., 2002). The association between PSQI and retrospective WHOQOL-BREF scores has been frequently reported (e.g., Meiavia et al., 2013).

Here we will investigate whether this relationship is replicated when HRQoL is reported in real-time. If successful we can demonstrate that real-time HRQoL can overcome other biases of retrospective assessment but measuring the same construct.

Perceived Stress

At the end of the 2 weeks, participants will complete an online version of the 10-item Perceived Stress Scale (PSS; Cohen et al., 1983; Cohen and Williamson, 1988). The PSS is a widely used questionnaire for measuring the perception of stress. It mainly assesses the unpredictability, uncontrollability and overload of an individual's life and was designed for use in community samples. The validity and reliability of the scale is well established (Cohen and Williamson, 1988; Roberti et al., 2006). The items request responders to rate how often they experience various feelings and thoughts during the last month on a 5-point Likert scale ranging from *Never* to *Very often*. After reversing the four positively valence items, a sum score is calculated using the 10 items.

Social Class

At the end of the 2 weeks, participants will complete an online version of the 10-step ladder social class measurement (Adler et al., 2000), the Sense of Power Scale ($\alpha = 0.90$; Anderson and Galinsky, 2006) and the Sense of Status Scale ($\alpha = 0.83$; Dubois et al., 2015). These measures are employed to assess the relationship between social class and HRQoL, as well as the potential mediating role of social power and/or status. In general, higher socio-economic status is associated with higher HRQoL (Huguet et al., 2008). Such associations will be pursued with the aim to acquire an understanding of the underlying determinants of variation in HRQoL.

Feasibility

The Mobile App Rating Scale (MARS; Stoyanov et al., 2015) will be used to assess the feasibility of the mHealth application. The MARS is a multidimensional assessment of mobile application quality, and will be used to reveal both subjective and recurring issues with the app. The MARS has been reliably used by end-users to assess the quality of mHealth apps and it has good internal consistency and test-retest reliability (Stoyanov et al., 2016). For the purposes of the current feasibility evaluation, items not relevant to our application were excluded from the scale (e.g., items about participants' willingness to pay for the application, as the study has no commercial interest).

STEPWISE PROCEDURE

Translation Process

Once all research tools have been identified, translation of all materials to six target languages (Danish, German, Greek, Hungarian, Italian, and Spanish; chosen for researcher's fluency in these languages) was pursued to maximize the accessibility of the mobile application. The WHOQOL-BREF, the PSQI and the PSS have previously been translated and validated in all study languages. Measures that were not available in the study languages (Demographic Questionnaire, Major Life

Events [validated version available in German, Hungarian, Italian, and Spanish], the BMIS, Current Activities, Social Class Questionnaires, MARS [validated translation available in Italian]) were translated using the forward-backward translation method and cognitive debriefing (Wild et al., 2005). Within this method, a native speaker of the target language, who was also fluent in English, translated the material into the target language (forward translation). A second native speaker, similarly fluent in English, re-translated the native language translation back to English (backward translation). All discrepancies between the versions were discussed and resolved between the two translators, thereby creating a consensus version of the questionnaire. Finally, the consensus version was administered to two or three native speakers of the target language who were asked to assess its comprehensibility. Any issues raised within this process were brought to the attention of the whole research team and were collectively discussed and resolved. The same translation procedure was followed for the adapted real-time version of the WHOQOL-BREF, which was first devised in English.

The mHealth Application Development and Data Collection Strategy

The mHealth application was developed in close collaboration between the researchers and external collaborators with expertise in the development of such software. The researchers had the opportunity to test and provide feedback on early versions of the application. In order to assess areas of improvement the application was piloted with two or three participants in each of the study's languages. During the study period, participants will be asked to use the mHealth application for a period of 2 weeks, during which the application will send four prompts (at four different time points), asking participants to report their real-time HRQoL, current mood and activities. Prompts will be sent at random times during the day (between an earliest and latest time for notification, defined by participants for their convenience). Participants will be asked to respond within a 6-h interval. Subsequently, at the end of the 2 weeks they will complete a questionnaire assessing retrospective HRQoL, major life events, perceived stress, social class; sleep quality and feasibility (see **Figure 1**).

Selecting the Target Audience (Participants)

As a means to assess the feasibility of data collection procedure, we aim to recruit 450 healthy participants. Similar sample sizes were used in previous real-time assessment studies (e.g., Maes et al., 2015). This number is more than sufficient for our analysis; an *a priori* power analysis revealed, for example, that only $N = 46$ participants are needed to detect a significant relationship between real-time HRQoL and sleep quality ($1 - \beta = 0.95$, $\alpha = 0.05$, $r = 0.446$). Participants must be over 18 years old and they must own an Android or IOS phone with Internet access. Participants will be excluded if they have a serious mental health condition compromising their ability to respond or their memory. Participants will be recruited through the study's website, which was designed within a further external

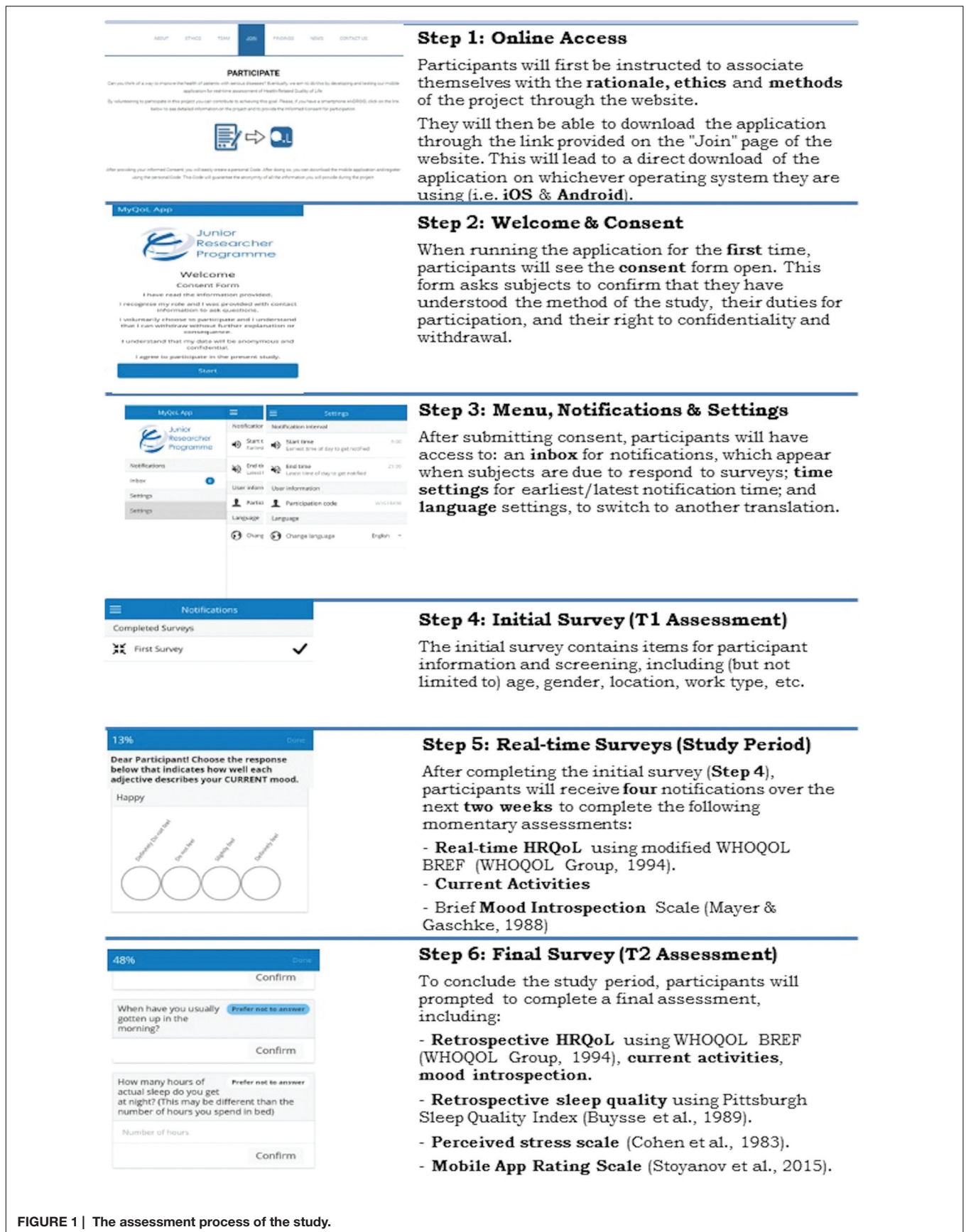


FIGURE 1 | The assessment process of the study.

collaboration. The link to this website will be distributed via email lists and social media. All participation will be voluntary. On the study website participants will be provided with a web link for downloading the application. The application will contain the study's Information Sheet and Consent Form. Once Consent is obtained participants will be able to use the mHealth application.

Proposed Analysis

Psychometric Properties of the Real-Time HRQoL Measure

The four real-time HRQoL scores (four time points) will be combined to obtain aggregated real-time HRQoL scores, in order to examine measurement invariance across assessment methods. Measurement invariance will also be examined across time points. Pearson's correlation coefficient will be calculated between real-time and retrospective domains of HRQoL. The reliability of the real-time and retrospective HRQoL tools will be assessed using Cronbach's alpha and omega coefficients. Finally, to assess whether the relationship between sleep quality and HRQoL is present when HRQoL is measured in real-time, Pearson's correlation coefficient will be calculated between the aggregated real-time HRQoL and PSQI score. A two-parameter item response model of real-time HRQoL aggregated data will be used to determine the difficulty and discrimination of questions.

Variability in Real-Time HRQoL

Multilevel modeling will be used to obtain estimates of within- and between- person variability in real-time HRQoL. At the first level, coefficients will be estimated for an equation for each person who expresses real-time HRQoL as a function of momentary mood. Subsequently, individual parameters will be used as dependent variables in the level 2 equations to evaluate whether within-person patterns differ across individuals and whether between-person variables (demographics, level of perceived stress, social class, sense of social power and status) and life events might account for the variance.

Feasibility

Feasibility will be assessed through analysis of the responses to the MARS questionnaire; percentages will be calculated for the close-ended questions and content analysis will be conducted on responses to the additional open-ended questions.

Ethics Statement and Current Status of Project

We went through the typical processes for meeting the ethics requirements for each participating University. We are currently working on finalizing the development of the mobile application and we are planning our pilot study and recruitment procedure.

ANTICIPATED RESULTS

We expect the project to contribute to the evidence on the validity and reliability of measuring HRQoL using an mHealth application, and to further our knowledge on the development of similar applications. Furthermore, we expect the project to provide insight about the nature of real-time HRQoL data

aiming to overcome the cognitive bias and feasibility issues of retrospective assessment. Crucially, there are a number of methodological considerations which merit discussion.

Firstly, the limited assessment time in this study may be problematic. However, we chose this short-time period to minimize missing data and respondent's time-burden; prolonged assessment periods could exacerbate potential problems with participant commitment and retention. We aim to minimize dropouts by making the application attractive, intuitive and navigable and by allowing participants to skip uncomfortable questions (Reips, 2002). Further, we will engage with participants through the companion website. The study's website was designed to provide participants with accessible information about the project's aims. It further contains a step-by-step overview of the participation process, which we expect will aid participants' engagement. Moreover, the website allows participants to directly contact the research team with arising questions and/or issues. In this way, it further constitutes an important troubleshooting tool.

Another important challenge could be participant compliance; in ecological, natural settings such as the home, it may be expected that participants feel less obliged to provide a complete response within given time-frames. However, promising recent EMA research has shown compliance levels comparable to traditional measures (Mehl and Holleran, 2007; Maes et al., 2015). Participant reactivity is another important consideration and refers to the potential for behavior to be affected by the act of assessing it (Shiffman et al., 2008). Particularly for EMA, one might anticipate frequent prompts to be irritating and impact on response quality. However, studies investigating participant reactivity in EMA have typically found effects to be non-significant (Peters et al., 2000; Stone et al., 2003). Crucially, these concerns are recognized and they shall be evaluated as part of the planned feasibility analysis.

Finally, since this is a feasibility study no power calculation or methods of Limits of Detection (LOD) or Limits of Quantification (LOQ) were established as we view the feasibility findings as the avenue to establish these for future studies and larger trials. Such subsequent investigations would further allow validating the translations that were prepared for the current study. Whilst we employed a rigorous translation procedure, a larger sample size would allow an estimate of measurement invariance across language versions.

Overall, we anticipate that our results will elucidate the relevance of such potential limitations. Further, we consider that our results, along with the current protocol, will aid future research exploring the potential of using EMA of HRQoL with a mobile application in clinical settings. Routine assessment of HRQoL in such settings is known to benefit patient-physician communication (Velikova et al., 2004), may facilitate clinical assessment and can be crucial for monitoring individual HRQoL over course of treatment. However, such assessments are also associated with substantial expenses and patient's time-burden. Future cost-effectiveness analyses can shed a light in this issue but we consider that EMA delivery in a mobile application may hold the potential to overcome these feasibility drawbacks.

AUTHOR CONTRIBUTIONS

This study was conceived and initially designed by APK. All of the authors further contributed to the research design, methodology, analysis plan and prospective discussion. First authors SM and DT drafted the first manuscript and were assisted by PM, VEM, CVO, BS, TW, and APK who contributed with additional writing and critical commentary. All authors approved the final manuscript.

REFERENCES

- Adler, N. E., Epel, E. S., Castellazzo, G., and Ickovics, J. R. (2000). Relationship of subjective and objective social status with psychological and physiological functioning: preliminary data in healthy white women. *Health Psychol.* 19, 586–592. doi: 10.1037/0278-6133.19.6.586
- Anderson, C., and Galinsky, A. D. (2006). Power, optimism, and risk-taking. *Eur. J. Soc. Psychol.* 36, 511–536. doi: 10.1002/ejsp.324
- Backhaus, J., Junghanns, K., Broocks, A., Riemann, D., and Hohagen, F. (2002). Test–retest reliability and validity of the Pittsburgh Sleep Quality Index in primary insomnia. *J. Psychosom. Res.* 53, 737–740. doi: 10.1016/S0022-3999(02)00330-6
- Bize, R., Johnson, J. A., and Plotnikoff, R. C. (2007). Physical activity level and health-related quality of life in the general adult population: a systematic review. *Prev. Med.* 45, 401–415. doi: 10.1016/j.ypmed.2007.07.017
- Blome, C., and Augustin, M. (2015). Measuring change in quality of life: bias in prospective and retrospective evaluation. *Value Health* 18, 110–115. doi: 10.1016/j.jval.2014.10.007
- Buysse, D. J., Reynolds, C. F., Monk, T. H., Berman, S. R., and Kupfer, D. J. (1989). The Pittsburgh sleep quality index: a new instrument for psychiatric practice and research. *Psychiatry Res.* 28, 193–213. doi: 10.1016/0165-1781(89)90047-4
- Catania, G., Beccaro, M., Costantini, M., Ugolini, D., De Silvestri, A., Bagnasco, A., et al. (2015). Effectiveness of complex interventions focused on quality-of-life assessment to improve palliative care patients' outcomes: a systematic review. *Palliat. Med.* 29, 5–21. doi: 10.1177/0269216314539718
- Cohen, S., Kamarck, T., and Mermelstein, R. (1983). A global measure of perceived stress. *J. Health Soc. Behav.* 24, 386–396. doi: 10.2307/2136404
- Cohen, S., and Williamson, G. (1988). "Perceived stress in a probability sample of the united states," in *The Social Psychology of Health*, eds S. Spacapan and S. Oskamp (Newbury Park, CA: Sage).
- Dubois, D., Rucker, D. D., and Galinsky, A. D. (2015). Social class, power, and selfishness: when and why upper and lower class individuals behave unethically. *J. Pers. Soc. Psychol.* 108, 436–439. doi: 10.1037/pspi0000008
- Freedman, M. J., Lester, K. M., McNamara, C., Milby, J. B., and Schumacher, J. E. (2006). Cell phones for ecological momentary assessment with cocaine-addicted homeless patients in treatment. *J. Subst. Abuse Treat.* 30, 105–111. doi: 10.1016/j.jsat.2005.10.005
- Huguet, N., Kaplan, M. S., and Feeny, D. (2008). Socioeconomic status and health-related quality of life among elderly people: results from the Joint Canada/United States Survey of Health. *Soc. Sci. Med.* 66, 803–810. doi: 10.1016/j.socscimed.2007.11.011
- Krägeloh, C., Henning, M., Hawken, S., Zhao, Y., Shepherd, D., and Billington, R. (2011). Validation of the WHOQOL-BREF quality of life questionnaire for use with medical students. *Educ. Health* 24, 1–5.
- Lin, M.-R., Hwang, H.-F., Chen, C.-Y., and Chiu, W.-T. (2007). Comparisons of the brief form of the World Health Organization Quality of Life and Short Form-36 for persons with spinal cord injuries. *Am. J. Phys. Med. Rehabil.* 86, 104–113. doi: 10.1097/01.phm.0000247780.64373.0e
- Lindholm, E., Brevinge, H., Bergh, C. H., Körner, U., and Lundholm, K. (2003). Relationships between self-reported health related quality of life and measures of standardized exercise capacity and metabolic efficiency in a middle-aged and aged healthy population. *Qual. Life Res.* 12, 575–582. doi: 10.1023/A:1025034919526
- Maes, I. H., Delespaul, P. A., Peters, M. L., White, M. P., van Horn, Y., Schruers, K., et al. (2015). Measuring health-related quality of life by

ACKNOWLEDGMENTS

The study was conducted as part of the Junior Researcher Programme (JRP). We would like to thank Niki Karakonstanti, Elena Garrone, the Happiness Research Organization and Anthony Bukhalana for their time and effort, working externally from JRP. Also, we would like to thank all the native speakers who helped with the forward and backward translations of the questionnaires.

- experiences: the experience sampling method. *Value Health* 18, 44–51. doi: 10.1016/j.jval.2014.10.003
- Mayer, J. D., and Gaschke, Y. N. (1988). The experience and meta-experience of mood. *J. Pers. Soc. Psychol.* 55, 102–111. doi: 10.1037//0022-3514.55.1.102
- Mehl, M. R., and Holleran, S. E. (2007). An empirical analysis of the obtrusiveness of and participants' compliance with the electronically activated recorder (EAR). *Eur. J. Psychol. Assess.* 23, 248–257. doi: 10.1027/1015-5759.23.4.248
- Meiavia, A., Marques, D., and Gomes, A. A. (2013). Quality of sleep and quality of life in higher education students. *Sleep Med.* 14:e140. doi: 10.1016/j.sleep.2013.11.317
- Peters, M. L., Sorbi, M. J., Kruse, D. A., Kerssens, J. J., Verhaak, P. F., and Bensing, J. M. (2000). Electronic diary assessment of pain, disability and psychological adaptation in patients differing in duration of pain. *Pain* 84, 181–192. doi: 10.1016/s0304-3959(99)00206-7
- Power, M., Bullinger, M., and Harper, A. (1999). The World Health Organization WHOQOL-100: tests of the universality of quality of life in 15 different cultural groups worldwide. *Health Psychol.* 18, 495–505. doi: 10.1037/0278-6133.18.5.495
- Ratcliff, C. G., Lam, C. Y., Arun, B., Valero, V., and Cohen, L. (2014). Ecological momentary assessment of sleep, symptoms, and mood during chemotherapy for breast cancer. *Psychooncology* 23, 1220–1228. doi: 10.1002/pon.3525
- Reips, U. D. (2002). Standards for Internet-based experimenting. *Exp. Psychol.* 49, 243–256. doi: 10.1027//1618-3169.49.4.243
- Roberti, J. W., Harrington, L. N., and Storch, E. A. (2006). Further psychometric support for the 10-item version of the perceived stress scale. *J. Coll. Couns.* 9, 135–147. doi: 10.1002/j.2161-1882.2006.tb00100.x
- Schacter, D. L. (2012). Constructive memory: past and future. *Dialogues Clin. Neurosci.* 14, 7–18.
- Schwartz, C. E., Sprangers, M. A., Carey, A., and Reed, G. (2004). Exploring response shift in longitudinal data. *Psychol. Health* 19, 51–69. doi: 10.1080/0887044031000118456
- Sherman, S. A., Eisen, S., Burwinkle, T. M., and Varni, J. W. (2006). The PedsQL™ present functioning visual analogue scales: preliminary reliability and validity. *Health Qual. Life Outcomes* 4:75. doi: 10.1186/1477-7525-4-75
- Shiffman, S., Stone, A. A., and Hufford, M. R. (2008). Ecological momentary assessment. *Annu. Rev. Clin. Psychol.* 4, 1–32. doi: 10.1146/annurev.clinpsy.3.022806.091415
- Solhan, M. B., Trull, T. J., Jahng, S., and Wood, P. K. (2009). Clinical assessment of affective instability: comparing EMA indices, questionnaire reports, and retrospective recall. *Psychol. Assess.* 21, 425–436. doi: 10.1037/a0016869
- Steel, J. L., Geller, D. A., Robinson, T. L., Savkova, A. Y., Brower, D. S., Marsh, J. W., et al. (2014). Health-related quality of life as a prognostic factor in patients with advanced cancer. *Cancer* 120, 3717–3721. doi: 10.1002/cncr.28902
- Stone, A. A., Broderick, J. E., Schwartz, J. E., Shiffman, S., Litcher-Kelly, L., and Calvanese, P. (2003). Intensive momentary reporting of pain with an electronic diary: reactivity, compliance, and patient satisfaction. *Pain* 104, 343–351. doi: 10.1016/s0304-3959(03)00040-x
- Stone, A. A., Schwartz, J. E., Broderick, J. E., and Shiffman, S. S. (2005). Variability of momentary pain predicts recall of weekly pain: a consequence of the peak (or salience) memory heuristic. *Pers. Soc. Psychol. Bull.* 31, 1340–1346. doi: 10.1177/0146167205275615
- Stoyanov, S. R., Hides, L., Kavanagh, D. J., and Wilson, H. (2016). Development and validation of the User Version of the Mobile Application Rating Scale (uMARS). *JMIR Mhealth Uhealth* 4:e72. doi: 10.2196/mhealth.5849

- Stoyanov, S. R., Hides, L., Kavanagh, D. J., Zelenko, O., Tjondronegoro, D., and Mani, M. (2015). Mobile app rating scale: a new tool for assessing the quality of health mobile apps. *JMIR Mhealth Uhealth* 3:e27. doi: 10.2196/mhealth.3422
- The WHOQOL Group (1994). "The development of the World Health Organization quality of Life Assessment instrument (the WHOQOL)," in *Quality of Life Assessment: International Perspectives*, eds J. Orley and W. Kuyken (Heidelberg: Springer). doi: 10.1007/978-3-642-79123-9_4
- Velikova, G., Booth, L., Smith, A. B., Brown, P. M., Lynch, P., Brown, J. M., et al. (2004). Measuring quality of life in routine oncology practice improves communication and patient well-being: a randomized controlled trial. *J. Clin. Oncol.* 22, 714–724. doi: 10.1200/jco.2004.06.078
- Wild, D., Grove, A., Martin, M., Eremenco, S., McElroy, S., Verjee-Lorenz, A., et al. (2005). Principles of good practice for the translation and cultural adaptation process for patient-reported outcomes (PRO) measures: report of the ISPOR Task Force for Translation and Cultural Adaptation. *Value Health* 8, 94–104. doi: 10.1111/j.1524-4733.2005.04054.x
- Wright, E. P., Selby, P. J., Crawford, M., Gillibrand, A., Johnston, C., Perren, T. J., et al. (2003). Feasibility and compliance of automated measurement of quality of life in oncology practice. *J. Clin. Oncol.* 21, 374–382. doi: 10.1200/JCO.2003.11.044

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2016 Mareva, Thomson, Marengo, Estal Muñoz, Ott, Schmidt, Wingen and Kassianos. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Study Protocol on Cognitive Performance in Bulgaria, Croatia, and the Netherlands: The Normacog Brief Battery

Lea Jakob¹, Lana Bojanić², Desislava D. Tsvetanova³, Eike K. Buabang⁴, Nienke J. de Bles⁴, Alexandra Sarafoglou⁵, Annet Dijkzeul⁶ and Rocio Del Pino^{7*}

¹ Department of Psychology, Centre for Croatian Studies, University of Zagreb, Zagreb, Croatia, ² Department of Psychology, Faculty of Humanities and Social Sciences, University of Zagreb, Zagreb, Croatia, ³ Department of Psychology, Sofia University "St. Kliment Ohridski", Sofia, Bulgaria, ⁴ Institute of Psychology, Leiden University, Leiden, Netherlands, ⁵ Department of Psychology, University of Amsterdam, Amsterdam, Netherlands, ⁶ Department of Psychology, Utrecht University, Utrecht, Netherlands, ⁷ Department of Methods and Experimental Psychology, University of Deusto, Bilbao, Spain

OPEN ACCESS

Edited by:

Pietro Cipresso,
Istituto Auxologico Italiano, Italy

Reviewed by:

Evgueni Borokhovski,
Concordia University, Canada
Eva M. Palacios,
University of California,
San Francisco, USA

*Correspondence:

Rocio Del Pino
rociodelpino@deusto.es

Specialty section:

This article was submitted to
Quantitative Psychology
and Measurement,
a section of the journal
Frontiers in Psychology

Received: 08 March 2016

Accepted: 11 October 2016

Published: 27 October 2016

Citation:

Jakob L, Bojanić L, Tsvetanova DD, Buabang EK, de Bles NJ, Sarafoglou A, Dijkzeul A and Del Pino R (2016) Study Protocol on Cognitive Performance in Bulgaria, Croatia, and the Netherlands: The Normacog Brief Battery. *Front. Psychol.* 7:1658. doi: 10.3389/fpsyg.2016.01658

The Normacog Brief Battery (NBB) provides a comprehensive overview of an individual's cognitive functioning within a short amount of time. It was originally developed for the Spanish population in Spain. However, there is a considerable need for brief batteries in clinical neuropsychological assessment, especially in eastern European countries. Cultural background and other individual characteristics—such as age, level of education, and sex—are shown to influence both cognition and patients' performance on neuropsychological tests. Therefore, it is important to develop understanding of how and why culture impacts on cognitive testing and determine which sociodemographic variables affect cognitive performance. The current study aims to translate, adapt, and standardize the NBB in Bulgaria, Croatia, and the Netherlands, and to analyze the effect of sex, age, and education level on cognitive performance between these three countries. This brief battery assesses eleven cognitive domains, including those most currently relevant in cognition such as premorbid intelligence, attention, executive function, processing speed, and memory. The translation and adaptation of the battery for different cultures will be done using the back-translation process. After exclusion criteria, the current study will include a total sample of 300 participants (≥ 18 years old). The samples of 100 participants per country will be balanced through the consideration of their age and level of education. Effects of the sociodemographic variables (age, level of education, and sex) on cognitive performance are expected. Furthermore, this relationship is expected to differ across countries. A multivariate hierarchical linear regression will be used and exploratory analysis will be carried out to investigate further effects. The results will be particularly valuable for future research and assessment in cognitive performance. The growing demand for accurate and fast neuropsychological assessment shows the importance of creating a universal brief assessment tool for wider cross-cultural application.

Keywords: brief battery, cognitive performance, cross-cultural, neuropsychological assessment, Normacog

INTRODUCTION

Neuropsychological assessment is a performance-based method that is used to obtain information about a person's cognitive functioning in different domains, such as memory, attention, processing speed, executive functions, spatial, and language functions (Harvey, 2012). A neuropsychological brief battery is a set of cognitive tests that provides a more complete profile of such cognitive functions. These batteries are used for several different purposes, including the acquisition of differential diagnostic information, assessment of treatment response, prediction of functional potential, and functional recovery (Harvey, 2012; Lezak et al., 2012). Therefore, it is of great importance that a person's assessment is conducted and interpreted correctly. In order to correctly interpret the outcome of a cognitive test, a participant's score must be compared to scores of a similar group. However, the availability of standardized tests and measures as well as norms is very limited when different populations have to be assessed (Nell, 1999). Many neuropsychological tests and measures have been developed for Caucasian, well-educated people, native English-speakers, and middle to upper class citizens, and consequently do not have the same diagnostic accuracy when used within other populations (Manly, 2008). In order to supply the lack of standardized brief batteries, the Normacog Brief Battery-NBB was created as part of the Spanish Normacog project using 700 healthy participants for validation (Del Pino, 2014; Del Pino et al., 2015a). The NBB presented high internal consistency in providing standardized norms for cognitive performance for Spanish speakers in Spain. These norms were adapted according to the sociodemographic characteristics of the country, and the results of the study showed a significant effect of age and level of education on the cognitive performance with no significant effect of sex. Furthermore, it is known that cultural background and related factors influence cognition and outcome of neuropsychological tests (Nell, 1999; Pérez-Arce, 1999; Ardila, 2007). According to Nell (1999), Ardila (2005), and Del Pino et al. (2015b), it is important to use calibrated norms for neuropsychological measures in different linguistic and cultural groups, and to account for the effect of culture on cognitive testing in order to provide correct interpretation of results. Sociodemographic characteristics—such as age, level of education, and sex—have been shown to influence cognitive performance (Mann et al., 1990; Collins and Kimura, 1997; Park et al., 1999; Reilly, 2012; Ojeda et al., 2014). However, the influence of these variables in cognitive performance seems to vary across cultures.

Regarding the influence of age on cognitive performance, Park et al. (1999) argued that differences in cognitive performance across cultures might disappear with increasing age, because of an overall cognitive decline. However, Ojeda et al. (2014) found systematic differences in cognitive performance between elderly Spanish and American individuals. The authors suggested that historical experience of political oppression and cultural background (like the attitude to make as few errors as possible) may have influenced the variation. Only older cohorts showed these discrepancies (Ojeda et al., 2014).

Considering education level, it has been shown that people with higher education perform better on neuropsychological tests than those from lower educational groups. This effect is most prominent in verbal neuropsychological tests (Lezak et al., 2012). Although, performance on non-verbal neuropsychological tests is also influenced by educational level when individuals with different educational levels within the same cultural group are compared (Rosselli and Ardila, 2003). However, despite the fact that educational level correlates with performance on some neuropsychological tests, it is not systematically related to everyday problem solving, which is a functional criterion of intelligence (Cornelius and Caspi, 1987). According to Rosselli and Ardila (2003), individuals with different levels of education have developed different ways of learning. Education could thus be considered a type of subculture. The development of different types of skills is influenced by culture, which results in different learning styles (Ostrosky-Solis et al., 2004). Furthermore, there seems to be an interaction between age and education. Groups with lower level of education start showing cognitive decline earlier in life while the cognitive functioning of better-educated groups tend to show decline at later age (Joao et al., 2016). Specifically, education seems to have a protective function against cognitive decline for general mental status but not for more complex tasks that require verbal abilities or working memory (Alley et al., 2007). While more years of education might be associated with higher scores in verbal abilities, working memory, and processing speed, there are no long-term effects of education regarding cognitive decline in any of those domains (Zahodne et al., 2011). In general, there seems to be no moderating role of education on cognitive decline directly; instead people with a higher level of education show a delayed in their cognitive impairment due to a higher baseline performance (Wilson et al., 2009; Lenehan et al., 2015). Another possible explanation for this phenomenon is that groups with lower educational attainment tend to have less intellectually stimulating jobs and lifestyles, resulting in a more rapid decline in cognitive ability (Heaton et al., 2009).

Another variable connected with cognitive performance is the participant's sex. However, sex differences vary in magnitude across countries (Kimura, 1999). For instance, mental rotation and line angle judgment performance were assessed in more than 90,000 women and 111,000 men from 53 nations: males from wealthier nations demonstrated greater spatial abilities (Lippa et al., 2009). Another study conducted by Weber et al. (2014) in Europe showed that the magnitude of sex differences varies systematically across birth cohorts and regions. These variations were associated with changes in living conditions and cognitive stimulation over time. Weber et al. (2014) suggested that females benefit more than males from these societal improvements because females start from a more disadvantaged level than males.

In addition, socioeconomic status has already been proven to influence cognitive performance (Bradley and Corwyn, 2002; Haan et al., 2011; Eryigit-Madzwamuse et al., 2014). According to the literature reviewed, its effects begin as early as the prenatal period and continue throughout life. The strongest association

between socioeconomic status and neurocognitive performance is found for language (Noble et al., 2007). According to Noble et al. (2007), this association could be due to the fact that the brain regions involved in language processing have a longer maturation period *in vivo* than any other brain region, which makes them more susceptible to environmental factors that covary with socioeconomic status. Another influential factor is individual professional activity. The literature suggests that the longer the period an individual has not been professionally active, the greater their decline in cognitive functioning is (Adam et al., 2013).

Exactly how age, level of education, and sex may influence cognitive performance is still an open research question. The effects are either contradictory or not stable. However, the authors of the aforementioned studies do agree that differences in cognitive performance, if they occur, are due to exposure to different educational opportunities, living standards, and historical backgrounds (Ojeda et al., 2014). Therefore, there is consensus that variations in cultural environment drive the significance of variables such as age and sex. For instance, cultural differences have been identified between Eastern and Western European countries, primarily in terms of collectivism and individualism, respectively (Kolman et al., 2003; Lykes and Kemmelmeier, 2014). Studies have shown that countries from Eastern Europe, such as Bulgaria and Croatia, are more socially interdependent than those from Western Europe, and place less importance on values such as mastery and autonomy while countries from Western Europe, such as the Netherlands, are thought to be more independent (Kolman et al., 2003). Furthermore, research has linked cultural variation to differences in cognitive styles. It is suggested that social orientations leads to different patterns of cognition. Independent societies tend to be more analytic while interdependent societies are more holistic (Nisbett and Miyamoto, 2005). This indicates that Eastern Europeans have a more holistic cognitive style compared to Western Europeans, who are more analytical (Varnum et al., 2008). Considering the aforementioned results, this study aims to further analyze whether intercultural differences in cognitive functioning exist not only between the countries from Eastern and Western Europe but also within Eastern European countries. Therefore, comparison of the countries of Bulgaria, Croatia, and the Netherlands are considered as adequate for the purposes of this study.

This study will explore aforementioned characteristics to analyze the cross-cultural differences in cognitive performance in three different countries (Bulgaria, Croatia, and the Netherlands). However, considering the state of current scientific literature, it is relevant to analyze not only cultural differences between the countries investigated, but also the effects of sex, age, and level of education on cognitive performance. More precisely, it is of interest whether there is an interaction between sex and age and if this interaction is influenced by the level of education of the participants, when controlled for sociodemographic covariates.

Therefore, the goals of the study are specified as follows: Firstly, the current project aims to translate and adapt the NBB (Del Pino et al., 2015a) in Bulgaria, Croatia, and the Netherlands; Secondly, the study aims to test both cross-cultural differences

and the interaction between sex, age, and level of education on cognitive performance.

MATERIALS AND EQUIPMENT

Measures

Structured Interview

A structured interview was designed to collect the sociodemographic characteristics from participants and to make an informed decision regarding inclusion and exclusion criteria. The Hollingshead four-factor index of socioeconomic status (SES) will be used to measure each participant's SES based on four domains: marital status, retirement/employment status, educational attainment, and occupational prestige (Hollingshead, 1975, unpublished).

Normacog Brief Battery

Neuropsychological data will be obtained by administering the NBB (Del Pino et al., 2015a) to participants from Bulgaria, Croatia and the Netherlands that meet the inclusion criteria (see paragraph Participants). The battery assesses eleven cognitive domains using eight subtests, listed below. The process of translation and back-translation will be carried out for several subtests, since they were not available in specific languages. The eight subtests forming the NBB are as follows (**Table 1**).

- The Prospective Memory Test (PMT) (Einstein and McDaniel, 1990) aims to assess prospective memory. Participants are instructed to remember performing an intended action (asking the examiner to return their keys or other personal item) at a particular time in the future (at the end of the testing). The participants' ability to recall the instruction is scored according to the level of help from the examiner (from 0: no help; to 4: the examiner asks some

TABLE 1 | Tests included in the Normacog Brief Battery (Del Pino et al., 2015a).

Cognitive domain	Normacog Brief Battery
Prospective memory	Prospective Memory Test (Einstein and McDaniel, 1990)
Premorbid functioning	Word Accentuation Test (Gomar et al., 2011)
General cognitive status	Montreal Cognitive Assessment (Nasreddine et al., 2005)
Attention and interference resistance	University of Deusto Interference Test (Ojeda et al., 2013. Based on Stroop Test, Stroop, 1935; Golden, 2001)
Visuoconstructive abilities and visual memory	Taylor Complex Figure (Taylor, 1969)
Executive function and mental flexibility	Modified Wisconsin Card Sorting Test (Schretlen, 2010)
Processing and perceptual speed	Salthouse Perceptual Comparison Test (Salthouse, 1991)
Semantic fluency	Semantic fluency subtest of the CIFA Test (Schretlen and Vannorsdall, 2010)

- questions to help but the examinee does not remember the assigned task).
- The Word Accentuation Test (WAT) (González Montalvo, 1991; Del Ser et al., 1997; Gomar et al., 2011) is the Spanish adaptation of the National Adult Reading Test (NART) (Nelson and O'Connell, 1978; Nelson, 1982). It aims to assess premorbid cognitive functioning by correctly reading aloud 30 low frequency words whose graphic accents have been removed. The total score is the number of words correctly read (from 0 to 30). The Dutch Adult Reading Test (DART) (Schmand et al., 1992) is a Dutch-language version of the NART that is being used for the Dutch sample. This test aims to determine premorbid intelligence by asking the participant to correctly read aloud 50 words that do not follow regular pronunciation rules. The total score is the sum of words pronounced correctly and can range from 0 to 100, with two points awarded for every correct word. Note that for the Bulgarian and Croatian sample premorbid functioning will not be assessed, as no scales adapted for respective populations are available.
 - The Montreal Cognitive Assessment (MoCA) (Nasreddine et al., 2005) is a 30-point brief screening tool for evaluating mild cognitive impairment. It aims to assess the general cognitive status of the participant in the following cognitive domains: attention and concentration, executive functions, memory, language, conceptual thinking, computing and orientation. The performance of the test takes between 5 and 7 min. A score of 26 points or more is considered normal.
 - The Animal naming fluency subtest from the Calibrated Ideational Fluency Assessment (CIFA) (Schretlen and Vannorsdall, 2010), was used to measure semantic fluency. The subtest used in the NBB assesses semantic fluency through determining the number of animals that a participant can name in 1 min.
 - The Taylor Complex Figure (TCF) (Taylor, 1969) was initially constructed as an alternate form of the Rey-Osterrieth Complex Figure test. It consists of two parts: first, the participant is asked to copy the TCF, which provides information on visuoconstructive abilities. The second part aims to evaluate visual memory by asking the participant to reproduce the figure on a blank sheet of paper with a 3-min delay. Both drawing tasks are timed and have time limits for completion: 4 min for copying the figure and 2 min for recall. The score in both tasks ranges from 0 to 36 points according to the accuracy of the drawn elements and their correct placement.
 - The University of Deusto Interference Test (Ojeda et al., 2013), based on the Stroop Test, (Stroop, 1935; Golden, 2001) assesses attention and interference resistance. It consists of three parts, each lasting 30 s. Much like the Stroop Test, it requires participants to name colors and incongruently colored names of colors aloud to assess level of interference. This new version overcomes the limitations that the original version presents related to people with color blindness and reading difficulties in elderly people. This new version includes different colors to the original one (blue, black, and pink). In addition, this version is shorter (90 s in total). The time limit for each section is 30 s (15 s less than the original version). In order to improve the reading of elderly people, the stimulus in this new version are bigger than the original one, and there are less stimulus (64 instead of 100).
 - The Salthouse Perceptual Comparison Test (SPCT) (Salthouse, 1991) assesses processing and perceptual speed. This is done with two 30-s trials where the participant's task is to mark whether a pair of letter strings (three- or six-letter strings, respectively) is same or different.
 - The Modified Wisconsin Card Sorting Test (M-WCST) (Schretlen, 2010) is a shorter, modified version of the Wisconsin Card Sorting Test (Grant and Berg, 1948) that assesses executive functioning and mental flexibility. This version differs from the standard one in many ways (e.g., the reduction in number of cards from the original 128-card deck to only 48 cards). This and other modifications reduce the risk of frustration and quitting among elderly and impaired individuals. The test requires participants to sort a deck of 48 cards according to one of several implicit rules while adapting the rule used depending on examiner feedback about whether the last card sorted was correct or incorrect.

STEPWISE PROCEDURES

Participants

Data will be collected from adults in three countries: Bulgaria, Croatia, and the Netherlands. The aim is to assess at least 300 participants in total, 100 from each country, after exclusion. Participants will be recruited by "word of mouth" from different geographical locations in each country. In addition, universities, companies, and retirement homes will be contacted to ensure a comprehensive sample of the population in each country.

The sample size will be chosen taking into account the suggested sample size by the epidemiologic program "EPI INFO" and it will be based on realistic time and location constraints that authors will face when obtaining participants. According to the size of population older than 18 years for each country (Bulgaria: 6,179,026; Croatia: 3,468,429; and the Netherlands: 13,563,456), the sample size will require at least 100 participants per country. The samples will be balanced considering two main demographic characteristics: age and education. Stratification will be done according to eight levels of age (18–25, 26–35, 36–45, 46–55, 56–65, 66–75, 76–80, >80 years old) and four levels of education (0–6, 7–10, 11–12, and >12 years) (Ivnik et al., 1997; Peña-Casanova et al., 2009; Del Pino et al., 2015b). The age ranges will be chosen considering the dynamics of cognitive performance throughout the lifetime. This demographic characteristic is also crucial for the purposes of this study, which is why the researchers will aim to collect as diverse a sample as possible. Levels of formal education were chosen considering the differences in the education systems in the three countries. The inclusion criteria for participants in the study will be the following: (1) people of both sexes; (2) at least 18 years old; (3) sufficiently developed reading and writing skills; (4) voluntary participation; (5) signed informed consent.

Exclusion criteria include: (1) having medical history of physical or mental illness that can interfere with cognitive functioning; (2) severe cognitive impairment; (3) having sensory impairment that cannot be corrected using aids; (4) being addicted to drugs or alcohol; (5) not being a native speaker of the language in which the assessment is being carried out; (6) and being functionally illiterate.

Ethics

The study was approved by the Ethics Committee at the University of Deusto, Bilbao, Spain, which is the coordinator of the study. The study has also been approved by Sofia University “St. Kliment Ohridski,” Bulgaria, Faculty of Humanities and Social Sciences, and the Centre for Croatian Studies, University of Zagreb, Croatia. The ethical approval has been submitted for Utrecht University, Netherlands, to which the evaluators in this study are affiliated. All subjects will be volunteers and will provide written informed consent prior to their participation in the study, in accordance with the Declaration of Helsinki.

Design and Procedure

The main goal of the study is to translate and adapt the NBB into Dutch, Bulgarian, and Croatian. For this purpose, copyright for all subtests included in the original battery was obtained. The battery has been translated and back-translated from English into each language by proficient individuals (who have studied English at least the bachelor level and are native speakers of language for which the test is being adapted). In the Netherlands, most of the subtests were already available. Once the instructions and answer sheets for each subtest have been translated and back-translated, an instruction manual will be developed in each of the three languages. The examiners for each country were trained individually in neuropsychological assessment. The first author of the NBB, Rocio Del Pino, executed this training. There will be five assessing examiners who have been trained in the NBB; two examiners from the Netherlands, two examiners from Croatia, and one examiner from Bulgaria. A detailed plan was developed for sample recruitment, which will be closely followed in order to make the sample comparable across countries (see **Table 2**). The sampling plan was developed following the proportions of the recruitment plan by Del Pino et al. (2015b).

- (1) The population older than 18 years old of each country (Bulgaria, Croatia, and the Netherlands) was reviewed in their corresponding National Institute of Statistics (Bulgarian National Institute of Statistics, Croatian Bureau of Statistics, and Centraal Bureau Statistiek the Netherlands).
- (2) The sample size was calculated with “EPI INFO” program for a total sample of 23,210,911. According to this, at least 270 participants should be included in the study for a confidence level of 90% (Barlett et al., 2001).

$$n = \frac{Z^2 * N * p * q}{N * d^2 + Z^2 * p * q}$$

n = sample size

Z = Z -score value for the selected confidence level

N = Population

p = probability of success

q = probability of failure

d = acceptable margin error

- (3) Adhering to previous studies (Ivnik et al., 1997; Peña-Casanova et al., 2009; Del Pino et al., 2015b), the sample will be stratified according to eight levels of age and four levels of education.
- (4) The final sample will consist of 100 participants per country; that is the adjusted sample after excluding participants that fail to meet the inclusion criteria.

The examiners will assess all participants in a quiet, neutral environment with minimum distractions (e.g., outdoor noises). The assessment itself will consist of two parts: a structured interview and the guided completion of the NBB (Del Pino et al., 2015a). A schematic illustration of the testing procedure is displayed in **Figure 1**. At the beginning of the study, all participants will have to read and sign the informed consent. They will then be interviewed by the examiner in order to gather demographic data and data related to their medical history, use of drugs, alcohol, etc. Once the interview is completed, the examiner will present all of the tests from the battery. The whole assessment procedure should last about 20 min with the majority of the participants, but with consideration of the fact that older participants are more likely take longer in completing the full battery. Neither the names nor any other information that may lead to the identification of the participants collaborating in the study will be published in any of the work resulting from this investigation. Therefore, each evaluator will have a code, as well as each participant. This code will be written in each sheet of the assessment in order to meet Data Protection Law and in accordance with the Declaration of Helsinki. To ensure that the coding scheme is comparable across examiners, the inter-rater reliability as well as the internal consistency will be checked. Finally, the complete data from the assessment will be codified and included in a data collection sheet. This data will then be processed and analyzed. A flowchart of the analysis plan is illustrated in **Figure 2**. All analyses will be performed using SPSS (version 19.0, 2010).

Proposed Analysis

Controlling for Covariates

In order to identify possible covariates, it will be checked whether sociodemographic variables, such as SES, are equally distributed across countries. To do this, a Chi-squared test of homogeneity will be conducted for each of the sociodemographic variables. If significant deviations across countries are found, the corresponding sociodemographic variables will be included as covariates in the model.

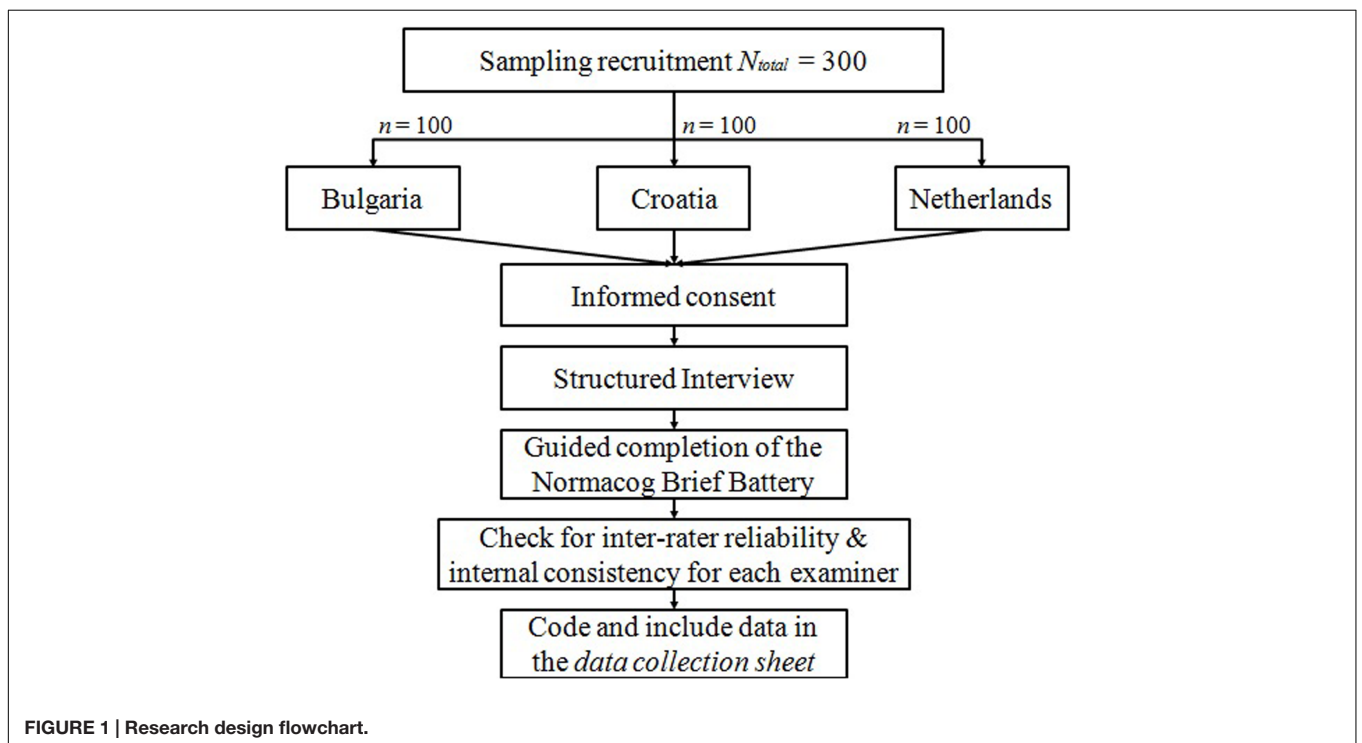
Main Analysis

A hierarchical model will be used for the analysis of participant and country effects because the participants are hierarchically nested within countries (participants within each country perform more similar than participant between countries). Cultural differences as well as the interaction between sex, age,

TABLE 2 | Plan for sample recruitment for the whole sample.

Age	Years of education				n
	0–6	7–10	11–12	>12	
18–25 years old	0 (0%)	6 (~16%)	15 (~40.5%)	16 (~43%)	37
26–35 years old	0 (0%)	6 (~16%)	19 (~24.5%)	12 (~32.5%)	37
36–45 years old	6 (~16%)	8 (~22%)	12 (~32.5%)	11 (~30%)	37
46–55 years old	6 (~16%)	8 (~22%)	11 (~30%)	12 (~32.5%)	37
56–65 years old	8 (~21%)	9 (~23.5%)	9 (~23.5%)	12 (~31.5%)	38
66–75 years old	17 (~45%)	12 (~31.5%)	3 (~8%)	6 (~16%)	38
76–80 years old	20 (~52.5%)	12 (~31.5%)	3 (~8%)	3 (~8%)	38
>80 years old	18 (~47%)	14 (~37%)	3 (~8%)	3 (~8%)	38
	75	75	75	75	N = 300

The proportions are relative to the previous sampling recruitment used for the Normacog Brief Battery (Del Pino et al., 2015b, p. 62).

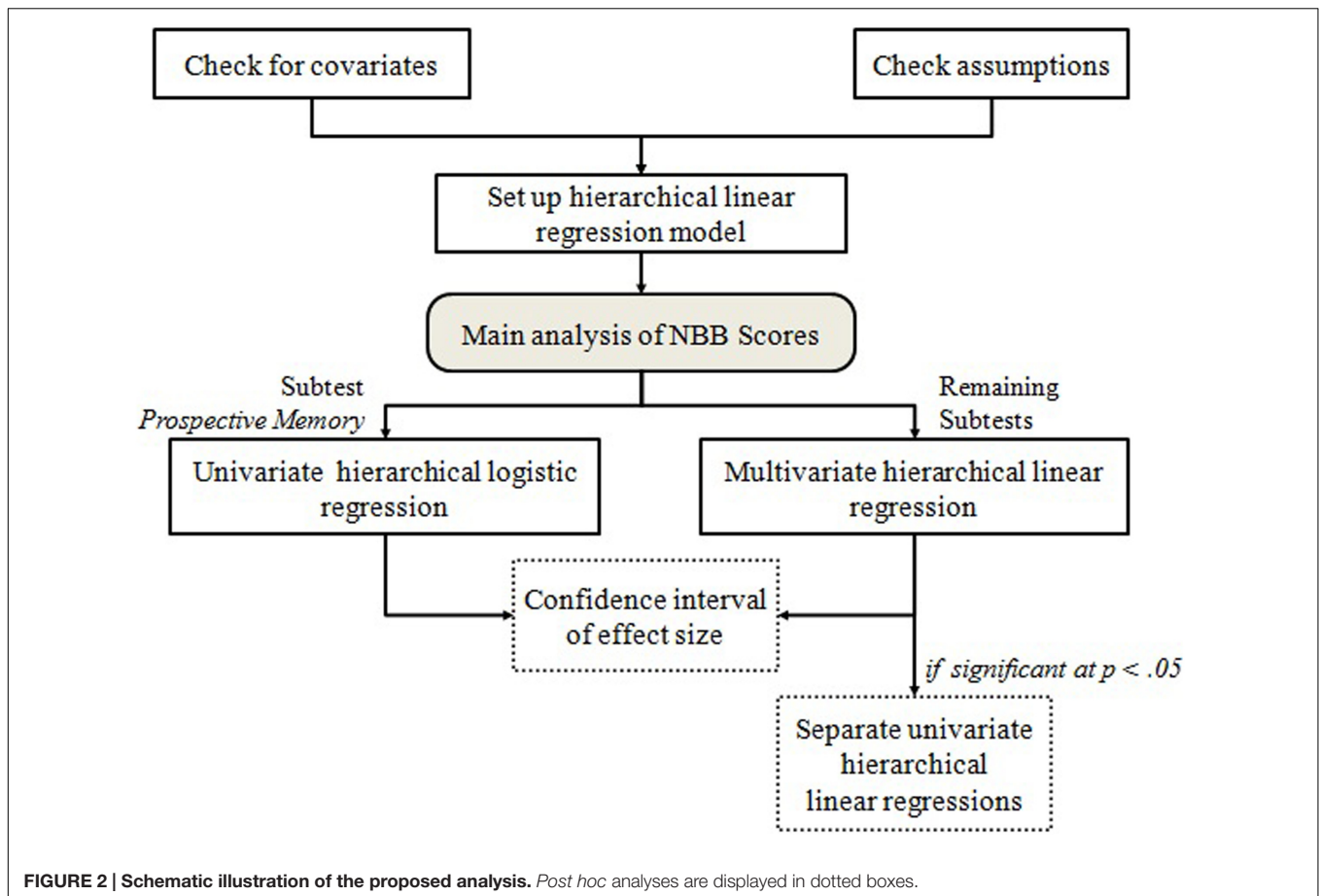


and level of education on cognitive performance will be tested with a hierarchical multivariate linear regression with multiple predictors. The proposed model includes the country of living as a predictor as well as the interaction between the predictors age, sex and education level. To control for their associated intraclass correlation country of living will be specified as a random factor. The predictors are expected to affect participants' scores in the NBB. In addition, all sociodemographic variables that are unequally distributed across countries will be included as covariates in the model. A separate analysis will be carried out for the categorical variable resultant of the prospective memory test. A hierarchical logistic regression will be conducted to predict the probability that the participant needs help in the prospective memory task. For the logistic regression the same model is used as with the multivariate multiple regression. To validate

the proposed model, a likelihood ratio test will be conducted to assess whether country of living and the proposed interaction contributes significantly to the model fit. The likelihood ratio test compares the model containing the proposed effect with the restricted model without the effect. This analysis will be executed both for the hierarchical multivariate regression as well as for the hierarchical logistic regression.

Proposed *Post hoc* Tests

Two *post hoc* tests are proposed for the main analysis. First, a *post hoc* power analysis will be conducted by analyzing the width of the confidence interval for the effect sizes. This analysis will indicate the likelihood of the real effect size being (non)zero. Second, the multivariate main analysis includes the standardized scores of all subscales in the NBB – except for the prospective



memory – as an outcome variable. To obtain more detailed information about the location of the possible effect, separate univariate regressions will be conducted for each independent subscale score. The predictors will be the same as in the main analysis and corrections for multiple testing will be applied.

Additional Exploratory Analysis

In order to acknowledge the complexity of possible interactions, we will conduct additional exploratory analysis of the data, with the aim of gaining deeper insight into the interaction effects.

ANTICIPATED RESULTS

Considering there is a serious lack of standardized neuropsychological instruments, we aim to ameliorate that situation by providing the interested parties with a readily translated and standardized brief battery. A review of currently available brief neuropsychological batteries has clearly demonstrated the demand for brief cognitive batteries in respective countries, especially in eastern European countries. A careful examination of the state of available neuropsychological instruments available for clinicians and other uses in Bulgaria and Croatia yielded no availability, while the Netherlands has limited availability of such resources. This will change once this battery is made available

for use in the respective countries, providing professionals with a much-needed instrument.

With the growing understanding of the importance of accurate and fast neuropsychological assessment, the making of a universal brief assessment tool for a wider cross-cultural application will be put to test. We expect to demonstrate that cultural aspects—such as language, education, and age—affect individual cognitive performance and each country should take into account its own characteristics to provide accurate interpretation guidelines crucial for making appropriate clinical decisions. We expect our project to have impact across Europe, making the issue of assessment instrument availability acknowledged by professionals, encouraging parties involved in such assessment to tackle the issue by joining the Normacog development initiative in respective countries that will be supported from the original authors. This Protocol should serve as a starting point for any researcher who is interested in adaptation for their language.

Besides these strengths, some limitations have emerged so far. Obtaining an adapted test for premorbid functioning in both Bulgaria and Croatia has already proven to be problematic, which may complicate the comparability of results within the given domain. Further, the test for premorbid functioning used in the Netherlands consists of 50 words, while the same test used in Spain consists of 30 words, which might cause difficulties

comparing the results. Secondly, another test, the Salthouse Perceptual Comparison Test, where a participant has to compare a series of letters with each other, may also be problematic in our study. This test has been made with the Latin alphabet while people in Bulgaria use the Cyrillic alphabet. For this reason, Latin letters were replaced with Cyrillic ones, potentially impeding comparison between people in Europe who did this test. Thirdly, the current study involves the minimum adequate sample size required for analyzing cross-cultural differences; therefore, if future studies aim to test a representative sample of each country, a larger sample size should be included.

Previous research suggests that cultural differences exist in cognitive functioning between countries from Eastern and Western Europe (Varnum et al., 2008). Therefore, the authors anticipate that such differences will also be found in the current research project. Considering the historical background of Eastern European countries, combined with better educational systems in the Western European countries, the authors expect that the overall cognitive performance in the Netherlands might be better than the one in Bulgaria and Croatia. However, the main goal of this study is not to measure and rank performance of these countries, but to determine if cultural differences between them actually exist. This finding would prove that it is necessary to create standardized culture-specific norms for each country in order to be able to validly interpret the results of the instrument.

To summarize, the translation, adaptation, and standardization of the NBB for the Netherlands, Croatia, and Bulgaria have been completed. Our next step is to gather

participants and start the analysis of cross-cultural differences in cognitive performance.

AUTHOR CONTRIBUTIONS

All authors contributed equally to and have approved the final manuscript. RDP originated the study design as part of the Normacog project and was the supervisor of the study. LJ, LB, DT, NdB, and AD were responsible for collection of data. AS was responsible for the proposed analysis EB was responsible for writing the introduction and leading the copyright process. LJ was the team's communication officer.

ACKNOWLEDGMENTS

This research was made possible by the Junior Researcher Programme (<http://jrp.psycholars.org/>). We would like to thank everyone involved in the organization of the Programme for their assistance. Furthermore, we thank the Normacog team for providing materials and support during the course of this study. In particular, we thank: Natalia Ojeda, Ph.D., Javier Peña, Ph.D., and Naroa Ibarretxe, Ph.D. from the Department of Methods and Experimental Psychology, University of Deusto, Bilbao, Spain; and David J. Schretlen, Ph.D. from the Department of Psychiatry and Behavioral Sciences, Johns Hopkins University School of Medicine, Baltimore, MD, USA.

REFERENCES

- Adam, S., Bonsang, E., Grotz, C., and Perelman, S. (2013). Occupational activity and cognitive reserve: implications in terms of prevention of cognitive aging and Alzheimer's disease. *Clin. Interv. Aging* 8, 377–390. doi: 10.2147/CLIA.S39921
- Alley, D., Suthers, K., and Crimmins, E. (2007). Education and cognitive decline in older americans: results from the AHEAD sample. *Res. Aging* 29, 73–94. doi: 10.1177/0164027506294245
- Ardila, A. (2005). Cultural values underlying psychometric cognitive testing. *Neuropsychol. Rev.* 15, 185–195. doi: 10.1007/s11065-005-9180-y
- Ardila, A. (2007). "The impact of culture on neuropsychological test performance," in *International Handbook of Cross-Cultural Neuropsychology*, eds B. P. Uzzell, M. Ponton, and A. Ardila (Mahwah, NJ: Lawrence Erlbaum Associates), 23–44.
- Barlett, J. E., Kotrlik, J. W., and Higgins, C. C. (2001). Organizational research: determining appropriate sample size in survey research. *Inf. Technol. Learn. Perform. J.* 19, 43–50.
- Bradley, R. H., and Corwyn, R. F. (2002). Socioeconomic status and child development. *Annu. Rev. Psychol.* 53, 371–399. doi: 10.1146/annurev.psych.53.100901.135233
- Collins, D. W., and Kimura, D. (1997). A large sex difference on a two-dimensional mental rotation task. *Behav. Neurosci.* 111, 845–849. doi: 10.1037/0735-7044.111.4.845
- Cornelius, S. W., and Caspi, A. (1987). Everyday problem solving in adulthood and old age. *Psychol. Aging* 2, 144–153. doi: 10.1037/0882-7974.2.2.144
- Del Pino, R. (2014). *Normalización y Estandarización de la Bateria Breve Normacog (BBN) de Evaluación Neuropsicológica Para Adultos*. Ph.D. Dissertation, University of Deusto, Bilbao.
- Del Pino, R., Ojeda, N., Peña, J., and Ibarretxe-Bilbao, N. (2015a). *Normacog Brief Battery (NBB). Intellectual Property Application N°: BI-79-15*.
- Del Pino, R., Peña, J., Schretlen, D. J., Ibarretxe-Bilbao, N., and Ojeda, N. (2015b). Multisite study for norming and standardizing neuropsychological instruments in healthy people for Spanish population: methods and characteristics of Normacog project. *Rev. Neurol.* 61, 57–65.
- Del Ser, T., González-Montalvo, J., Martínez-Espinosa, S., Delgado-Villapalos, C., and Bermejo, F. (1997). Estimation of premorbid intelligence in Spanish people with the word accentuation test and its application to the diagnosis of dementia. *Brain Cogn.* 33, 343–356. doi: 10.1006/brcg.1997.0877
- Einstein, G. O., and McDaniel, M. A. (1990). Normal aging and prospective memory. *J. Exp. Psychol. Learn. Mem. Cogn.* 16, 717–726. doi: 10.1037/0278-7393.16.4.717
- Eryigit-Madzwamuse, S., Baumann, N., Jaekel, J., Bartmann, P., and Wolke, D. (2014). Neuro-cognitive performance of very preterm or very low birth weight adults at 26 years. *J. Child Psychol. Psychiatry.* 56, 857–864. doi: 10.1111/jcpp.12358
- Golden, C. J. (2001). *STROOP: Test de Colores y Palabras: Manual*. Seville: TEA ediciones.
- Gomar, J. J., Ortiz-Gil, J., McKenna, P. J., Salvador, R., Sans-Sansa, B., Sarró, S., et al. (2011). Validation of the Word Accentuation Test (TAP) as a means of estimating premorbid IQ in Spanish speakers. *Schizophr. Res.* 128, 175–176. doi: 10.1016/j.schres.2010.11.016
- González Montalvo, J. I. (1991). *Creación y Validación de un Test de Lectura Para el Diagnóstico del Deterioro Mental en el Anciano*. Ph.D. Dissertation, Complutense University of Madrid, Madrid.
- Grant, D. A., and Berg, E. (1948). A behavioral analysis of degree of reinforcement and ease of shifting to new responses in a Weigl-type card-sorting problem. *J. Exp. Psychol.* 38, 404–411. doi: 10.1037/h0059831
- Haan, M. N., Al-Hazzouri, A. Z., and Aiello, A. E. (2011). Life-span socioeconomic trajectory, nativity, and cognitive aging in Mexican Americans: The Sacramento area Latino study on aging. *J. Gerontol. Ser. B Psychol. Sci. Soc. Sci.* 66, 102–110. doi: 10.1093/geronb/gbq071

- Harvey, P. D. (2012). Clinical applications of neuropsychological assessment. *Dialogues Clin. Neurosci.* 14, 91–99.
- Heaton, R. K., Ryan, L., and Grant, I. (2009). “Demographic influences and use of demographically corrected norms in neuropsychological assessment,” in *Neuropsychological Assessment of Neuropsychiatric and Neuromedical Disorders*, 3rd Edn, eds I. Grant and K. M. Adams (New York, NY: Oxford University Press), 127–155.
- Ivnik, R. J., Smith, G. E., Lucas, J. A., Tangalos, E. G., Kokmen, E., and Petersen, R. C. (1997). Free and cued selective reminding test: MOANS norms. *J. Clin. Exp. Neuropsychol.* 19, 676–691. doi: 10.1080/01688639708403753
- Joao, A. A., Maroco, J., Gino, S., Mendes, T., de Mendonca, A., and Martins, I. P. (2016). Education modifies the type of subjective memory complaints in older people. *Int. J. Geriatr. Psychiatry* 31, 153–160. doi: 10.1002/gps.4305
- Kimura, D. (1999). *Sex and Cognition*. Cambridge, MA: The MIT Press.
- Kolman, L., Noorderhaven, N. G., Hofstede, G., and Dienes, E. (2003). Cross-cultural differences in Central Europe. *J. Manag. Psychol.* 18, 76–88. doi: 10.1108/02683940310459600
- Lenahan, M. E., Summers, M. J., Saunders, N. L., Summers, J. J., and Vickers, J. C. (2015). Relationship between education and age-related cognitive decline: a review of recent research. *Psychogeriatrics* 15, 154–162. doi: 10.1111/psyg.12083
- Lezak, M., Howieson, D. B., Bigler, E. D., and Tranel, D. (2012). *Neuropsychological Assessment*, 5th Edn. New York, NY: Oxford University Press.
- Lippa, R. A., Collaer, M. L., and Peters, M. (2009). Sex differences in mental rotation and line angle judgments are positively associated with gender equality and economic development across 53 nations. *Arch. Sex. Behav.* 39, 990–997. doi: 10.1007/s10508-008-9460-8
- Lykes, V. A., and Kemmelmeier, M. (2014). What predicts loneliness? Cultural difference between individualistic and collectivistic societies in Europe. *J. Cross Cult. Psychol.* 45, 468–490. doi: 10.1177/0022022113509881
- Manly, J. J. (2008). Critical issues in cultural neuropsychology: profit from diversity. *Neuropsychol. Rev.* 18, 179–183. doi: 10.1007/s11065-008-9068-8
- Mann, V. A., Sasanuma, S., Sakuma, S., and Masaki, S. (1990). Sex differences in cognitive abilities: a cross-cultural perspective. *Neuropsychologia* 28, 1063–1077. doi: 10.1016/0028-3932(90)90141-a
- Nasreddine, Z. S., Phillips, N. A., Bédirian, V., Charbonneau, S., Whitehead, V., Collin, I., et al. (2005). The Montreal cognitive assessment, MoCA: a brief screening tool for mild cognitive impairment. *J. Am. Geriatr. Soc.* 53, 695–699. doi: 10.1111/j.1532-5415.2005.53221.x
- Nell, V. (1999). *Cross-cultural Neuropsychological Assessment: Theory and Practice*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Nelson, H. E. (1982). *National Adult Reading Test (NART): For the Assessment of Premorbid Intelligence in Patients with Dementia: Test Manual*. Windsor, ON: NFER-Nelson.
- Nelson, H. E., and O’Connell, A. (1978). Dementia: the estimation of premorbid intelligence levels using the new adult reading test. *Cortex* 14, 234–244. doi: 10.1016/S0010-9452(78)80049-5
- Nisbett, R. E., and Miyamoto, Y. (2005). The influence of culture: holistic versus analytic perception. *Trends Cogn. Sci.* 9, 467–473. doi: 10.1016/j.tics.2005.08.004
- Noble, K. G., McCandliss, B. D., and Farah, M. J. (2007). Socioeconomic gradients predict individual differences in neurocognitive abilities. *Dev. Sci.* 10, 464–480. doi: 10.1111/j.1467-7687.2007.00600.x
- Ojeda, N., Aretouli, E., Pena, J., and Schretlen, D. J. (2014). Age differences in cognitive performance: a study of cultural differences in historical context. *J. Neuropsychol.* 10, 104–115. doi: 10.1111/jnp.12059
- Ojeda, N., Del Pino, R., and Peña, J. (2013). *University of Deusto Interference Test. Alternative and Adapted Version of the Stroop Test for the Elderly Population. Intellectual Property Application N°: BI-895–813*.
- Ostrosky-Solis, F., Ramirez, M., and Ardila, A. (2004). Effects of culture and education on neuropsychological testing: a preliminary study with indigenous and nonindigenous population. *Appl. Neuropsychol.* 11, 186–193. doi: 10.1207/s15324826an1104_3
- Park, D. C., Nisbett, R., and Hedden, T. (1999). Aging, culture, and cognition. *J. Gerontol. Ser. B Psychol. Sci. Soc. Sci.* 54, 75–84. doi: 10.1093/geronb/54b.2.p75
- Peña-Casanova, J., Blesa, R., Aguilar, M., Gramunt-Fombuena, N., Gomez-Anson, B., Oliva, R., et al. (2009). Spanish multicenter normative studies (NEURONORMA project): methods and sample characteristics. *Arch. Clin. Neuropsychol.* 24, 307–319. doi: 10.1093/arclin/acp027
- Pérez-Arce, P. (1999). The influence of culture on cognition. *Arch. Clin. Neuropsychol.* 14, 581–592. doi: 10.1016/S0887-6177(99)00007-4
- Reilly, D. (2012). Gender, culture, and sex-typed cognitive abilities. *PLoS ONE* 7:e39904. doi: 10.1371/journal.pone.0039904
- Rosselli, M., and Ardila, A. (2003). The impact of culture and education on non-verbal neuropsychological measurements: a critical review. *Brain Cogn.* 52, 326–333. doi: 10.1016/s0278-2626(03)00170-2
- Salthouse, T. A. (1991). Mediation of adult age differences in cognition by reductions in working memory and speed of processing. *Psychol. Sci.* 2, 179–183. doi: 10.1111/j.1467-9280.1991.tb00127.x
- Schmand, B., Lindeboom, J., and Van Harskamp, F. (1992). *Dutch Adult Reading Test*. The Netherlands: Swets & Zeitlinger, Lisse.
- Schretlen, D. J. (2010). *Modified Wisconsin Card Sorting Test professional Manual*. Lutz, FL: Psychological Assessment Resources.
- Schretlen, D. J., and Vannorsdall, T. D. (2010). *Calibrated Ideational Fluency Assessment (CIFA) Professional Manual*. Odessa, FL: Psychological Assessment Resources.
- Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *J. Exp. Psychol.* 18, 643–662. doi: 10.1037/h0054651
- Taylor, L. (1969). Localization of cerebral lesions by psychological testing. *Clin. Neurosurg.* 16, 269–287.
- Varnum, M. E., Grossmann, I., Katunar, D., Nisbett, R. E., and Kitayama, S. (2008). Holism in a European Cultural Context: differences in cognitive style between Central and East Europeans and Westerners. *J. Cogn. Cult.* 8, 321–333. doi: 10.1163/156853708x358209
- Weber, D., Skirbekk, V., Freund, I., and Herlitz, A. (2014). The changing face of cognitive gender differences in Europe. *Proc. Natl. Acad. Sci. U.S.A.* 111, 11673–11678. doi: 10.1073/pnas.1319538111
- Wilson, R. S., Hebert, L. E., Scherr, P. A., Barnes, L. L., De Leon, C. M., and Evans, D. A. (2009). Educational attainment and cognitive decline in old age. *Neurology* 72, 460–465. doi: 10.1212/01.wnl.0000341782.71418.6c
- Zahodne, L. B., Glymour, M. M., Sparks, C., Bontempo, D., Dixon, R. A., MacDonald, S. W., et al. (2011). Education does not slow cognitive decline with aging: 12-year evidence from the Victoria Longitudinal Study. *J. Int. Neuropsychol. Soc.* 17, 1039–1046. doi: 10.1017/S1355617711001044

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2016 Jakob, Bojanić, Tsvetanova, Buabang, de Bles, Sarafoglou, Dijkzeul and Del Pino. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Formal and Informal Learning and First-Year Psychology Students' Development of Scientific Thinking: A Two-Wave Panel Study

Demet Soyylmaz¹, Laura M. Griffin², Miguel H. Martín^{3,4}, Šimon Kucharský⁵, Ekaterina D. Peycheva⁶, Nina Vaupotič⁷ and Peter A. Edelsbrunner^{8*}

¹ Department of Psychology, Istanbul Bilgi University, Istanbul, Turkey, ² Faculty of Film, Art and Creative Technologies, Dún Laoghaire Institute of Art, Design and Technology, Dún Laoghaire, Ireland, ³ Faculty of Psychology, Pontifical University of Salamanca, Salamanca, Spain, ⁴ Faculty of Psychology, Ghent University, Ghent, Belgium, ⁵ Department of Psychology, University of Amsterdam, Amsterdam, Netherlands, ⁶ Department of General, Experimental and Genetic Psychology, Sofia University St. Kliment Ohridski, Sofia, Bulgaria, ⁷ Department of Psychology, University of Ljubljana, Ljubljana, Slovenia, ⁸ Research on Learning and Instruction, Department of Humanities, Social and Political Sciences, ETH Zurich, Zurich, Switzerland

OPEN ACCESS

Edited by:

Kristina Egumenovska,
International School for Advanced
Studies, Italy

Reviewed by:

Tom Rosman,
Leibniz Institute for Psychology
Information, Germany
Caitlin Drummond,
Carnegie Mellon University, USA

*Correspondence:

Peter A. Edelsbrunner
peter.edelsbrunner@ifv.gess.ethz.ch

Specialty section:

This article was submitted to
Quantitative Psychology
and Measurement,
a section of the journal
Frontiers in Psychology

Received: 07 November 2016

Accepted: 18 January 2017

Published: 10 February 2017

Citation:

Soyylmaz D, Griffin LM, Martín MH,
Kucharský Š, Peycheva ED,
Vaupotič N and Edelsbrunner PA
(2017) Formal and Informal Learning
and First-Year Psychology Students'
Development of Scientific Thinking:
A Two-Wave Panel Study.
Front. Psychol. 8:133.
doi: 10.3389/fpsyg.2017.00133

Scientific thinking is a predicate for scientific inquiry, and thus important to develop early in psychology students as potential future researchers. The present research is aimed at fathoming the contributions of formal and informal learning experiences to psychology students' development of scientific thinking during their 1st-year of study. We hypothesize that informal experiences are relevant beyond formal experiences. First-year psychology student cohorts from various European countries will be assessed at the beginning and again at the end of the second semester. Assessments of scientific thinking will include scientific reasoning skills, the understanding of basic statistics concepts, and epistemic cognition. Formal learning experiences will include engagement in academic activities which are guided by university authorities. Informal learning experiences will include non-compulsory, self-guided learning experiences. Formal and informal experiences will be assessed with a newly developed survey. As dispositional predictors, students' need for cognition and self-efficacy in psychological science will be assessed. In a structural equation model, students' learning experiences and personal dispositions will be examined as predictors of their development of scientific thinking. Commonalities and differences in predictive weights across universities will be tested. The project is aimed at contributing information for designing university environments to optimize the development of students' scientific thinking.

Keywords: epistemic cognition, informal learning, need for cognition, self-efficacy, scientific thinking, psychology students

INTRODUCTION

Scientific thinking encompasses purposeful thinking with the aim to enhance knowledge, using the abilities to generate, test and revise theories as well as being able to reflect on how knowledge is acquired and changed (Kuhn, 2002). It is a prerequisite for engagement in scientific inquiry (Kuhn et al., 2000). Little is known about the development of scientific thinking during aspiring researchers' development, especially at the early undergraduate level (Parent and Oliver, 2015).

The 1st-year at university is a particularly critical educational period for students' development of skills, interests, and aspirations (Jenert et al., 2015). For those students who want to become researchers, 1st-year education gives a first impression of science, with core courses found in psychology programs such as research methods and statistics (Stoloff et al., 2009). These educational experiences contribute to students' future motivation and understanding of science (Manning et al., 2006; Jenert et al., 2015). Not all students know in their 1st-year of university education if they want to engage in research and become researchers. But if they decide so at any point in the future, it is necessary to equip them with advanced scientific thinking as early as possible. Thus, the quality of 1st-year education can influence students' further aspirations and development as students and researchers. In this study, we examine the influence of learning experiences in 1st-year psychology on students' development of scientific thinking. One aim of the present study is to identify learning experiences related to the development of scientific thinking in the 1st-year of higher education as well as to pinpoint those that are most prevalent among successful scientific thinkers. This way, we try to capture the experiential profile of budding psychology researchers. Such findings could be of vital service in the development of psychology curricula that better reflect the learning needs of aspiring researchers as well as motivate students to become such.

A facet of scientific thinking that we consider central for potential future researchers is scientific reasoning¹. Scientific reasoning delineates the skills needed to conduct scientific inquiry, such as argumentation, drawing inferences from data, and engaging in experimentation (Zimmerman, 2007). It includes understanding and identifying relevant variables and how to interpret the information obtained from an experiment and various other research designs (Klahr, 2000).

Related to scientific reasoning skills, students should understand basic statistical concepts to evaluate the strength and uncertainty of scientific evidence. Previous research has shown that misconceptions about common statistical indicators, such as *p*-values and confidence intervals, are prevalent in student and teacher populations (Hoekstra et al., 2014; Morey et al., 2015), and in published research literature (Gelman and Stern, 2006; Nieuwenhuis et al., 2011). It is not surprising that students develop misconceptions about statistics, given that classical statistical methods violate common sense (Wagenmakers, 2007; Duffy, 2010). Psychology students commonly learn about null hypothesis significance testing in their statistics courses, which leads them to make interpretations based on arbitrary *p*-values previously set by researchers. This method violates common sense and leads to misunderstandings because the results are actually based on observations that have not occurred (Wagenmakers, 2007). For example, researchers intuitively tend to think that hypotheses tests inform them about the probability that the alternative hypothesis is true, but hypotheses

tests based on *p*-values have a different aim; they only inform about the long-term frequency of possible data given the null hypothesis. Crucially, those misconceptions can lead to wrong inferences, both in the conduct of research but also in the evaluation of published literature, which might have contributed to a current crisis in the confidence of psychological science (Pashler and Wagenmakers, 2012). With this in mind, we regard the absence of statistics misconceptions as a relevant aspect of scientific thinking in today's psychology students as developing researchers.

Another core facet of scientific thinking is epistemic cognition, which encompasses beliefs about knowledge, knowing, and the processes by which those beliefs are formed and influence further learning (Hofer and Pintrich, 1997; Kitchener, 2002; Greene et al., 2008). Greene et al. (2010) developed a model of epistemic and ontological cognition that integrates prior models by positing positions and dimensions. Each of four positions (Realism, Dogmatism, Skepticism, and Rationalism) corresponds to a distinctive pattern of individuals' beliefs along the three dimensions of simplicity and certainty, justification by authority, and personal justification of knowledge (Hofer, 2000). Simple and certain knowledge refers to the opinion that knowledge is isolated, simple and constant over time, justification by authority reflects a belief that knowledge can be ambiguous but holds greater weight when presented by an authority figure, and personal justification is a belief that all information presented must be engaged with critically before judging it to be true, and even then, it may not remain true over time. The first developmental stage, realism, represents strong beliefs in simple and certain knowledge, justification by authority, and personal justification. The position of dogmatism is demonstrated through strong justification by authority. The position of skepticism reflects strong personal justification. Lastly, rationalism presents moderate agreement with justification by authority and personal justification, but strong disagreement with the concept of simple and certain knowledge. Epistemic cognition can influence critical thinking, scientific argumentation, and learning (Kuhn et al., 2000; Nussbaum et al., 2008; Franco et al., 2012), and it is related to students' self-regulated learning (Bråten and Strømsø, 2005). Several studies thus far have linked epistemic cognition with students' learning achievements. A study conducted by Muis and Franco (2009) illustrates that epistemic cognition directly influences achievement goals of students in educational psychology course, which, in turn influenced their engagement in their tasks and final course achievement. In the similar vein, Bråten and Ferguson (2014) showed that epistemic beliefs contribute to achievement over and above cognitive capacity and personality traits of students. Moreover, Chen et al. (2014) showed that students who are self-efficacious about learning science, approach a task by examining arguments from several sources to make a decision, thus indicating a moderating role of self-efficacy in how epistemic cognition is related to academic outcomes. The results of these studies therefore support the idea that epistemic cognition plays a major role in students' further engagement and development of scientific thinking and we consider it a facet of scientific thinking that should develop early in psychology students.

¹Scientific reasoning is sometimes used interchangeably with scientific thinking (and sometimes with the related but broader construct critical thinking, Halpern et al., 2012) but in the context of the present research we deem it useful to conceptualize scientific reasoning and epistemic beliefs as facets of scientific thinking.

The development of scientific thinking begins in the first stages of life and continues throughout childhood and adolescence up into adulthood (Sodian et al., 1991; Kuhn et al., 2000, 2015; Zimmerman, 2007). However, this development does not occur automatically but through steady input from deliberate learning experiences (Kuhn, 2002, 2009; Klahr et al., 2011). Perhaps the most conspicuous way of improving the level of scientific reasoning is through formal education. In fact, research shows that students with higher level of education as well as students who were exposed to a research methodology course are more likely to exhibit better scientific reasoning skills (Lehman and Nisbett, 1990; Lawson et al., 2000). Demonstrating and engaging students in quantitative methodology and scientific inference improves their skills in theoretical modeling and experimentation (Duschl and Grandy, 2013; Holmes et al., 2015). Further learning experiences that predict university students' scientific thinking include collaborative learning (e.g., Gokhale, 1995; Johnson et al., 1998), social media use (Ebner et al., 2010; Dabbagh and Kitsantas, 2012; Kassens-Noor, 2012; Vivian et al., 2014), taking research methods and statistics courses (Lehman and Nisbett, 1990; VanderStoep and Shaughnessy, 1997), passive and active participation in research projects (Wayment and Dickson, 2008; VanWormer et al., 2014) and taking laboratory modules that include interpreting the results of an experiment (Coleman et al., 2015). Thus, university environments offer varied experiences that can help students develop scientific thinking.

In order for students to develop scientific thinking, it is not sufficient that relevant learning opportunities are offered at university. It is necessary that students show high engagement in formal activities and beyond. In the current study, we therefore look into students' engagement in relevant formal and informal learning activities. Informal learning at university can be defined as self-directed learning in the sense that the student chooses the topic, curriculum, and contents, and learning and assessment modalities, with the aim to develop knowledge, skills, or competences (Hofstein and Rosenfeld, 1996; Laurillard, 2009). It is closely related to conceptions of student engagement (Krause and Coates, 2008), as well as self-sustained, self-initiated, and free choice learning (Falk, 2001; Barron, 2006; Yang, 2015). Based on this definition, examples of informal learning experiences are attending science conferences, reading scientific books, and engaging in science-related discussions with peers. Formal learning, in comparison, is highly structured through university bodies in its curriculum, fixed learning activities, and assessment, with a course achievement or qualification as an end product (Resnick, 1987; Eshach, 2007; Patrick, 2010). This distinction posits informal learning as interest-driven, in comparison to formal learning as curriculum-based, assessment-driven, and qualification-oriented activities.

What factors predict students' engagement in learning activities that are likely to foster their scientific thinking? A characteristic that we take into account is students' science self-efficacy, that is, the confidence they have in their ability to do science (Beißert et al., 2014). Self-efficacy, defined as the belief in own capability to succeed (Bandura, 1997), is a major predictors of university students' cognitive engagement,

academic persistence in science-related courses, and career choices (Chemers et al., 2001; Walker et al., 2006; Chen and Usher, 2013). Along with self-efficacy, need for cognition has been shown to predict academic success (Elias and Loomis, 2002). It is a stable tendency to engage in and enjoy effortful thinking (Cacioppo and Petty, 1982). Need for cognition is related to intellectual engagement and positive attitudes toward effortful tasks, and thus with a richer personal history of gaining knowledge on a variety of topics (Woo et al., 2007).

In the current study, we assess psychology students twice to examine the contributions of formal and informal learning experiences to their development of scientific reasoning including statistics misconceptions, and epistemic cognition during the 1st-year of study. The assessments take place at the beginning and again at the end of their second semester. We aim to examine interrelations in the development of scientific reasoning and epistemic cognition during the semester, and the contribution of students' engagement in both types of learning experiences to this development. This includes their additive effects, and the involvement of students' self-efficacy and need for cognition in these effects. Our definition of informal learning posits that it is self-guided and goes beyond the mere aim of finishing courses and obtaining grades. It thus related informal learning strongly with interest-driven student engagement. Student engagement in educationally purposeful activities is positively related to academic outcomes in 1st-year students as well as students' persistence at the same institution (Kuh et al., 2008). Similarly, student engagement has been linked to desirable learning outcomes such as critical thinking and academic achievement (Carini et al., 2006). Whereas engagement in formal learning could stem from internal or external motivational factors, engagement in informal activities represents only intrinsically motivated behavior, which is derived from interest and performed for pleasure and desire. For these reasons we assume that informal learning contributes to students' development of scientific reasoning and epistemic cognition, beyond formal learning. The overarching aim of the design is to establish the circumstances under which potential future researchers in psychological science are able to develop scientific thinking during the early stages of their studies. We therefore examine specific patterns of scientific thinking and its predictors in students who identify themselves as aspiring researchers.

MATERIALS AND METHODS

Design and Sample

The study has a two-wave correlational panel design. Participants will be drawn from the 1st-year psychology courses of 11 universities from eight countries across Europe. We collaborate with 1st-year professors from each university. The choice of universities was based on personal affiliations and on the aim of gathering students from diverse backgrounds across Europe. Psychology student cohorts at the universities span between 40 and 700 students. Students from eight universities will participate during a regular class lesson and the remaining three

universities online. At three of the universities, students will receive assessment credits for participating in the study.

Sample size planning based on power analysis is not relevant because we will use Bayesian estimation and hypothesis testing for statistical analysis (Etz et al., 2016; Wagenmakers et al., 2016). In this statistical framework, power is not conceptualized because hypothesis testing is not based on an inferential framework but on continuous evaluation of evidence (Schönbrodt and Wagenmakers, 2016).

Materials and Equipment

Choice of Measures

For every construct that we aim to assess, a literature search was done in the PsycInfo and Scopus databases to identify available measures. The choice of the instruments was based on psychometric quality, appropriateness for university context, administration time, translation feasibility, and meaningfulness of usage in a variety of international universities. Regarding psychometric quality, we ensured that basic analysis such as factor analysis, estimation of reliability or internal consistency had been conducted and achieved at least moderate results. Appropriateness for the 1st-year of university was taken into account insofar as we tried to estimate on which level Psychology students develop during their 1st-year. For example, scientific reasoning is a broad construct, and we chose an instrument that assesses skills which we think are critical for students' further development, and likely to show at least some development already during their first university year. The chosen instrument assesses principles of experimental design that we deem relevant for understanding the critical quality characteristics of any research the students learn about (Drummond and Fischhoff, 2015).

Demographics Questionnaire

Students' demographic characteristics will include their age, gender, former university education, career aspirations, grades in high school, the grade of first university examination and family socioeconomic status (see Appendix A). For the latter, we will ask students about their parents' highest achieved education, bedroom availability and the number of books at home in their adolescence (Evans et al., 2010). Socioeconomic status is assessed to examine its influence on the main study variables and to estimate other variables' influence while controlling for it. We will assess family socioeconomic status because university students are still in education, which constrains their own educational level and also their working situation, the most common indicators of personal socioeconomic status. Family socioeconomic status is thus commonly assessed for research in academic contexts (Caro and Cortes, 2012). Students' estimated score from the first principal component of the four variables will be used as an indicator of their family socioeconomic status. Finally, we will assess the quantity of formal education relevant for developing scientific thinking (number of methodology and statistics-related courses, number of philosophy of science and epistemology-related courses).

Scientific Reasoning

As a measure of scientific reasoning, the *Scientific Reasoning Scale* developed and validated by Drummond and Fischhoff (2015) will be used. It contains eleven true or false items in which hypothetical research scenarios are described and the participant has to decide whether the scenario can lead to proposed inferences. Each of the items relates to a specific concept crucial for the ability to come to valid scientific conclusions. The concepts include understanding the importance of control groups and random assignment, identifying confounding variables, and distinguishing between correlation and causation. Scores on the SRS show adequate internal consistency (Cronbach's $\alpha = 0.70$) and correlate positively with cognitive reflection, numeracy, open minded thinking, and the ability to analyze scientific information (Drummond and Fischhoff, 2015). Following this scale, we added an additional item assessing students' understanding of sample representativeness (Appendix B). Students' mean score on the scale will be used in descriptive analysis as an indicator of their scientific reasoning. Whether the item on sample representativeness can be added to the scale will be decided based on a confirmatory factor analysis: It will be added in case its factor loading is within the range of the other items.

Statistics Misconceptions

We developed a questionnaire encompassing five questions that deal with common statistical misconceptions (Appendix B). Items dealing with p -value and confidence interval misinterpretations were taken directly from Gigerenzer (2004) and Morey et al. (2015). We chose the item with the highest prevalence of wrong answers among university students from each article to achieve high variance in our sample of 1st-year students. We further developed items similar in structure dealing with the interpretation of non-significant results, the equivalence of significant and non-significant results (Gelman and Stern, 2006; Nieuwenhuis et al., 2011), and sample representativeness. The items share structure and answer format with the scientific reasoning scale by Drummond and Fischhoff (2015). We added the items after the end of the scientific reasoning scale. Participants are also asked whether they have ever learned about p -values, confidence intervals, and sample representativeness. In case they check "no," their answers on the respective questions will be treated as missing values. Students' mean value across the four questions dealing with p -values and confidence intervals will be used as an indicator of their statistics misconceptions. The question on sample representativeness, as described above, will be used as an additional item of the scientific reasoning scale.

Validation Questions

For the Scientific Reasoning Scale and the added statistics misconceptions items, we will add one open-answer validation question. Each student will receive the following question at one random item of the 16 items that the two scales encompass: "Why did you choose this answer? Please provide an explanation.", followed by two lines on which the students are supposed to provide a short rationale for their multiple choice-answer. The question to which this additional open answer is added will

differ randomly between students, so that a random subsample of the students will deal as validation sample for each question. We implement this validation measure because the SRS to the best of our knowledge has not yet been translated into our sampled languages and not been used in the sampled countries. It is therefore necessary to examine whether 1st-year psychology students' answers on these questions reflect the target construct. The statistics misconceptions to the best of our knowledge have not yet been thoroughly validated but rather used to assess the prevalence of wrong answers among students and academics, and we developed three of the questions on our own, therefore we include them in this validation procedure.

Epistemic Cognition

To assess epistemic cognition we will administer the Epistemic and Ontological Cognition Questionnaire (EOCQ; Greene et al., 2010). It contains 13 items and a 6-point item response scale ranging from 1 (completely disagree) to 6 (completely agree). The instrument takes into account the contextuality of epistemic cognition by providing the opportunity to insert a domain into the item stems (Greene et al., 2008). We insert *Psychology* and *Psychological science* for the domain that the students should rate the items about. Five items represent simple and certain knowledge (example: “*in psychological science, what is a fact today will be a fact tomorrow*”), four items represent justification by authority (“*I believe everything I learn in psychology class*”), and four items represent personal justification (“*in psychological science, what's a fact depends upon a person's view*”). Higher ratings of ten items indicate stronger beliefs and high ratings of three items indicate weaker beliefs. Reliability estimates (*H* coefficient) range from 0.45 to 0.90 depending on facet and context (Greene et al., 2010). Mean scores on all three subscales will undergo mixture modeling analysis, which will yield an epistemic cognition-profile for each student that will be used for further analysis (Greene et al., 2010).

Need for Cognition

We will use the Need for Cognition Short Scale (NFC-K; BeiBert et al., 2014) to measure the tendency to engage in and enjoy thinking. The short scale is a modified 4-item version of the 18-item Need for Cognition Scale created by Cacioppo and Petty (1982). On a 7-point scale the students are asked to rate to which extent they agree with four simple statements. An example item is “*I would prefer complex to simple problems.*” Mean scores from this scale will be used for descriptive analysis, with higher scores indicating that students are more motivated to apply their thinking skills. Test retest reliability is $r = 0.78$, Cronbach's $\alpha = 0.86$ (BeiBert et al., 2014). The score will be used to predict students' development of scientific thinking, and also as a control variable to examine which variables predict students' development beyond need for cognition.

Science Self-Efficacy

The Science Self-Efficacy (SSE) scale, which consists of 10-items used by Moss (2012) will be used (Cronbach's $\alpha > 0.80$). It is a modified version of a vocational self-efficacy survey designed by Riggs et al. (1994). It particularly aims to measure confidence in

skills to engage in scientific inquiry. The items are rated on a scale from 1 to 10 (1 = *not able or not true at all*, 10 = *completely able or completely true*). An example item is “*I have all the skills needed to perform science tasks very well.*” Students' mean score on the scale will be used for statistical modeling. The score will be used to predict students' development of scientific thinking, and also as a control variable to examine which variables predict students' development beyond science self-efficacy.

Formal and Informal Learning Experiences

We developed a survey to assess students' engagement in learning experiences that we presume relevant for the development of scientific thinking (Appendix C). The selection of experiences is based on the discussed literature, and it will be further informed and adapted based on the pilot study interviews (Appendix D). Our definitions of formal and informal learning imply a continuum of formality within and across learning activities. For example, a frequent formal learning activity is the studying of a text that is mandatory reading for a research methods course. When students gain interest in the text contents, they might initiate further voluntary reading to inform themselves beyond the course requirements, which in our definition is then an informal learning experience. Our assessment method encompasses a wide variety of prescribed and non-prescribed scientific learning experiences: For each of the assessed activities that can be either formal or informal, we ask students how often they engaged in these as part of mandatory course activities, or for reasons going beyond these. Specifically, we let students rate subjectively for experiences where this applied how much they engaged in them because it was obligatory for course requirements (formal engagement), because it was obligatory but they were also interested (formal and informal engagement), or merely out of own interest (informal engagement).

In the second part of the survey, we ask students about the most relevant three courses they took that were related to research methods, statistics, science, history of science or other similar concepts. We ask for up to three courses because we studied the official bachelor curricula from the targeted universities and most students will not have more highly relevant courses during their first and second semester. Therefore, reporting on further courses might make it strongly subjective which courses the students deem relevant to this question, and it might take rather long and be exhausting to report details on any relevant courses they could think of. To check that they did not have many more relevant courses we, however, ask in the demographics for the absolute numbers of relevant courses. Thus, for up to three most relevant courses, they first list the names of the courses and whether the courses were mandatory or elective. Then, we ask students about their general engagement in these courses (student presence, devoted working time), and course quality (ratings of overall course quality, teaching quality, frequency of inquiry and reflective course elements). Finally, reflecting informal engagement, they rate how much they engaged in each of these courses out of their motivation or interest, beyond the course requirements. Estimating principal components, we will weigh general course engagement across courses with course content ratings to yield an indicator of formal engagement,

and informal (out of own motivation or interest) engagement with course quality ratings to yield an indicator of informal engagement.

Translations and Pilot Study

Considering students from different countries' levels of competence in English may not be sufficiently high, the materials and instruments have been translated into Spanish, Slovenian, Turkish, Bulgarian, and Czech by the local researchers from these countries. Then, they have been back translated by bilingual speakers to enable reconciliation of the translated texts with the original. During a pilot study, the materials and instruments were administered to a small number (10 from each country) of 2nd–4th year psychology students with cognitive surveying and interviewing to identify problematic passages in terms of ambiguous or confusing instructions and translations (Ziegler et al., 2015). During the cognitive surveying, participants were asked to read the instructions and items aloud. After each passage, they were instructed to report everything that came to their mind when thinking about the instruction or item and what they were thinking while answering the items. In the end, they were asked to reflect freely on the purpose, comprehensibility, and quality of the instrument. This data were used to alter potential problematic passages. Proposed changes were again translated back to English for comparison with the original. Pilot participants were also requested to respond to several interview questions regarding their formal and informal educational experiences throughout their lives that they believe might have contributed to their scientific reasoning and epistemic cognition (Appendix D). Their responses served to improve the formal and informal experiences survey, so that it would more adequately reflect students' relevant learning experiences.

Stepwise Procedures

The data will be collected at two time points. The first assessment will be conducted during the first 2 weeks of the second semester (between January and March) and the second will take place within the last 2 weeks of the academic schedule before exams (May and June), depending on each university's calendar. For universities at which collaborators agree to in-class assessments, these will take place directly in the classrooms or other provided university space. Professors will be asked to reward students with course credits for research participation, depending on the ethical policy of the institution. Ideally, with the professors' prior consent, the entire first-year courses will be assessed during a lecture. The local researchers in each country will distribute the questionnaires before the assessment starts and collect them afterward. In case an in-class administration of our instruments is not possible, we will ask the students to participate in an online version of our assessments. An online version has been prepared in the Qualtrics (Qualtrics, Provo, UT, USA) environment with a similar structure to the pen and paper version. For the online version, students will be provided with a hyperlink and encouraged to fill it in at their convenience within a week. If they have not yet finished the survey, they will receive a reminder email 2 days before this time limit.

The questionnaires will have the same structure in the in-class and in the online version. In both cases, participants will be given a short explanation of what the research is for, and what their participation will entail, which will be read aloud by the experimenter in class. They will then be asked to read an information sheet and read and sign a consent form, prior to proceeding with the assessment. One administration process is expected to last for about 35 min. The scales more strongly related to cognitive skills will be presented in the beginning of the assessment and the learning experiences will be assessed in the end to prevent the experiential questions from influencing later answers. The structure of the assessments is depicted in **Figure 1**.

The participants will first be asked to compose an identification code consisting of their mother's and father's initials, and the month that they were born (in mm format). They will then be asked to complete the demographics information about themselves. They will subsequently proceed to complete five scales measuring scientific reasoning, statistical misconceptions, epistemic cognition, need for cognition, and scientific self-efficacy. This will reflect their skills and attitudes after one semester studying psychology. In addition, they will be asked to complete the survey regarding their formal and informal learning experiences during the first study semester. In the end of the assessment, the students will be thanked for their participation, and informed that the assessment will take place again in the end of the semester. They will be asked to refrain from discussing the assessment with each other or to look up the contents.

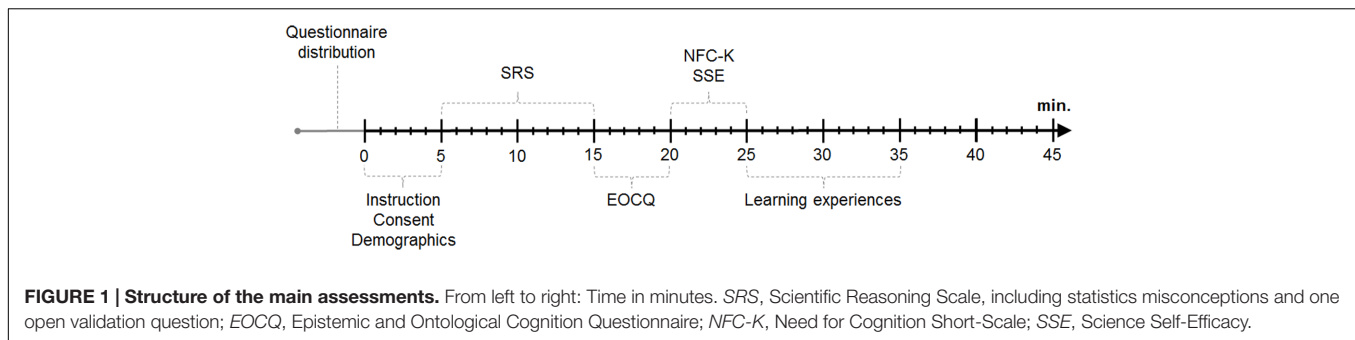
At the end of the second semester, participants will be approached as before to participate in the study. They will be asked to write their identification number as before, and to again complete the same scales as at the first assessment. On the demographics sheet, this time they will be asked additionally whether and to which extent they discussed the contents of the first assessment with peers or looked up the contents. The survey on formal and informal learning experiences will this time be referring to experiences during their second study semester. At this point participants will be given a debriefing form and thanked for their participation.

PROPOSED ANALYSIS AND ANTICIPATED RESULTS

Qualitative Data Analysis

Qualitative data will stem from the pilot study interviews. Transcriptions of the interviews will be analyzed using content analysis. The analysis will be aimed at identifying relevant formal and informal learning experiences else than those known from available literature. Insights from these data will be used to refine the learning experiences questionnaire for the main assessments.

We will also analyze the open validation questions about scientific reasoning and statistical misconceptions to see whether the correct multiple choice-answers on the items reflect the intended concepts (Drummond and Fischhoff, 2015). Given the fact that the items are forced choice between true or false, we of course expect that some of the correct answers will be a



result of guessing. We will encode whether the rationale for the answer is sufficient to come to the correct choice given the specific question. Then, we will try to group answers that did not have the correct rationale for the answer to see common misunderstandings. This will inform us about the validity of the items. We will also seek for common misconceptions leading to erroneous answers and try to categorize them with open followed by axial coding to get insight into why students make mistakes regarding the specific concepts. This will inform us what aspects of given concepts are hard to grasp and which misconceptions should be deliberately targeted by university lecturers.

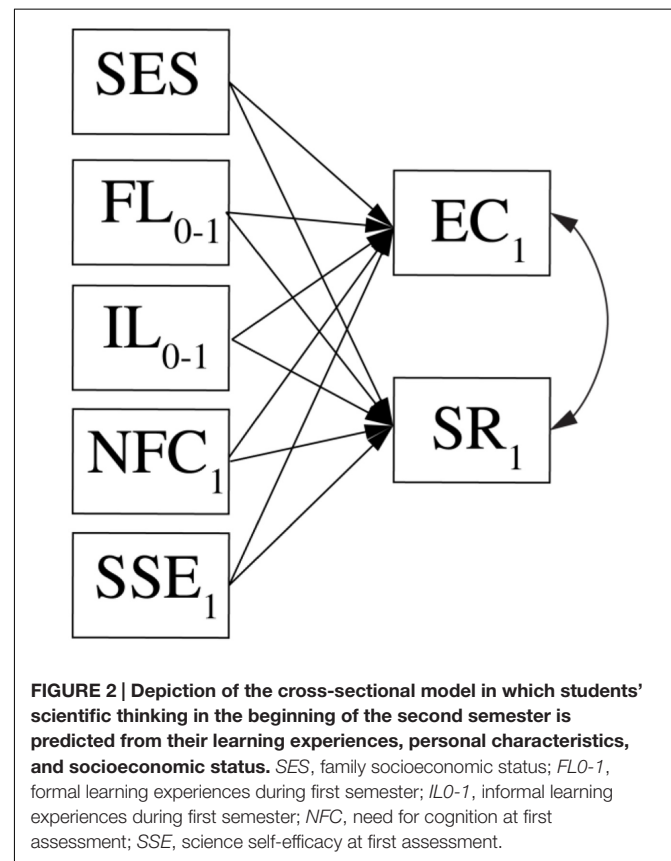
Confirmatory Statistical Analysis

Bayesian structural equation modeling will be applied to examine our main research questions. The models will be written using the *r2jags* (Su and Yajima, 2012) and *rjags* (Plummer, 2013) packages in the R software (R Core Team, 2013) to be estimated in the JAGS software (Plummer, 2003). There will be two main models. In a cross-sectional model, we will predict students' scientific thinking and research aspirations after the first half year from their educational experiences during the first half year, science self-efficacy and need for cognition, and from their learning experiences and family socioeconomic status. A depiction of the structural relations in this model is provided in **Figure 2**.

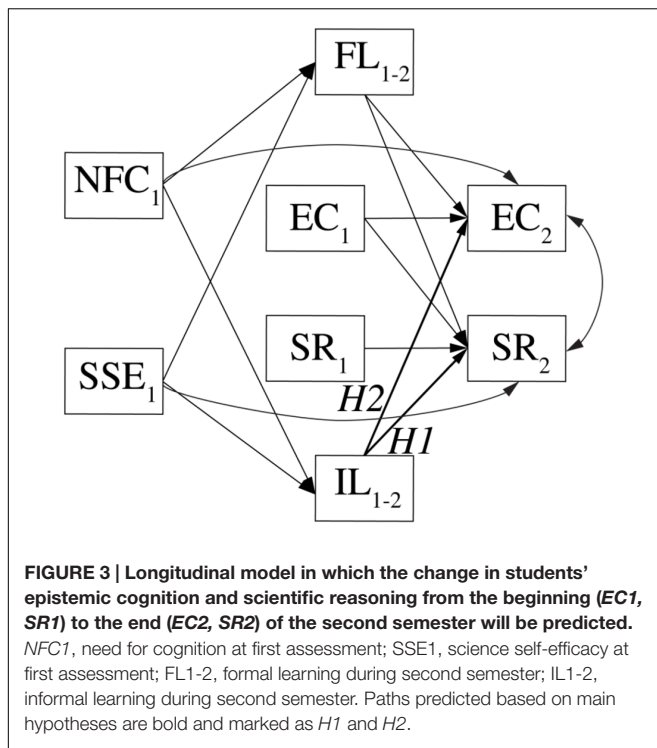
In a longitudinal model, we will examine developmental interrelations between students' scientific reasoning and epistemic cognition, and predict their development from students' formal and informal learning experiences during the second half year, and how these are influenced by students' personal characteristics. A depiction of the structural relations in this model is provided in **Figure 3**.

Bayes factors will be used for hypothesis testing. The tested predictions based on our hypotheses include that informal learning experiences predict the development of scientific reasoning and epistemic cognition, controlling for formal learning experiences, need for cognition, and science self-efficacy.

The models will be estimated separately for students from each university to examine commonalities and differences in predictive weights. Since sample sizes vary strongly between countries and institutions, samples from some universities might not be sufficiently big to ensure convergence and precision of model estimation. Data from the biggest samples will therefore be used to slightly inform the parameter priors from the smaller samples. Only priors of non-focal (i.e., non-hypothesis-testing)



parameters will be informed in this way (for an overview of related techniques, see McElreath, 2016, pp. 424–430). This strategy is similar to hierarchical modeling but implies weaker partial pooling. Scripts for all confirmatory analyses will be uploaded to the Open Science Framework prior to data analysis. We will interpret the magnitude of obtained Bayes factors based on the accruing samples from the different universities. The Bayes factors will be computed as a ratio of likelihoods of two models that describe the theoretical alternatives we put to the test (Jarosz and Wiley, 2014). We derive those models for two focal hypotheses (depicted in the **Figure 3** as H1 and H2) here. Given we control for formal learning experiences (FL in **Figure 3**), need for cognition (NFC), and science self-efficacy (SSE), we expect that parameters for informal learning experiences (IL in **Figure 3**)



have a positive value. This expectation is equivalent to a one-sided hypothesis. We will therefore use so called one-sided hybrid Bayes factors (Morey and Rouder, 2011). As a null model, we will combine a point nil with part of the Cauchy distribution from the range of values of 0 up to the point where the effect size becomes important (equivalence region). As an alternative model, we will use the remaining of the Cauchy distribution. The Cauchy will have a scaling factor of 0.5, the equivalence region will be defined as $<0,0.1>$ and the mixture probability of the two parts of the null model will be equal to 0.5 (the point nil and equivalence region will have the same weights). This type of Bayes factors have been shown to possess desirable properties (Morey and Rouder, 2011). They asymptotically converge toward support for null or alternative if the true parameter lies in the area of one of the respective models, and remain indifferent if the true parameter lies on the boundary of null and alternative (0.1 in this case).

Missing data stemming from attrition or single non-answered items will be dealt with in the Bayesian analysis. Specifically, students' self-reported interest in research and becoming researchers and all other study variables that might be associated with participation willingness and missingness will be used to estimate students' missing data (see McElreath, 2016).

Exploratory Statistical Analysis

These analyses serve mainly to find specific patterns between all variables and trying to identify students interested in becoming future researchers. In addition, they will serve to develop potential hypotheses about profiles of psychology students interested in becoming researchers. For this research question, we do not have specific hypotheses and we will use exploratory analyses to examine how the intention to become a researcher

or not is associated with the other study variables. Specifically, we will use two methods. We will use network modeling to explore relations between the main study variables at the two assessments. For this analysis the *mgm*-package will be used, which can handle different distributions of the exponential family and applies regularization for sparse solutions (Haslbeck and Waldorp, 2016). The estimated networks at the two assessments will provide a concise and informative overview of interrelations between the study variables in the beginning and end of students' second semester. We will also estimate finite mixture models (Hickendorff et al., unpublished), to extract profiles of scientific thinking. We will examine profiles including scientific reasoning, the three epistemic cognition facets, and students' research aspirations as profile indicator variables, the two dispositional scales as profile predictors, and statistics misconceptions as a distal variable.

To further substantiate comparisons between universities and countries, we will examine measurement invariance, which shows whether the assessment instruments have comparable structure across different samples. Measurement invariance is not of critical importance for our hypotheses but it is informative for exploratory purposes, to see for example whether the instruments function similarly in the different languages. In our Bayesian framework, we will be able to handle small deviations from invariance by modeling approximate invariance (Van De Schoot et al., 2012).

LIMITATIONS

The design will allow estimating the predictive value of formal and informal learning experiences but causal conclusions are not fully warranted for various reasons. Controlling for students' general maturation in higher education would be possible by adding a control group, for example 1st-year students from a different field. Such a comparison would, however, be biased by self-selection effects because we cannot assign students randomly to different fields of study.

Another limitation is that measurement will take place twice within one semester, specifically within the second semester. This might be early for expecting students to develop in scientific reasoning, epistemic cognition, and also statistical misconceptions, depending on when these topics are part of students' courses. Not all students might learn about these topics in their 1st-year. We looked into students' official curricula at the target universities and in all of them there are courses that might be relevant but this is not clear for all universities. Regarding assessing twice within one semester, more change might be expected during a longer time period. We will, however, ask the students also for relevant experiences during the first semester, which overall will yield a picture of the whole 1st-year, which covers the focus of our study.

Also, because the collecting data will take place on two different occasions during the semester, there may be attrition between time one and time two. However, by aiming to collect data in class, we hope to maximize initial potential participation and minimize potential attrition rate by time two. Missing data

will also be minimized by ensuring that participants use the same identification code on both occasions. To maximize participation and minimize attrition, assessments will be conducted in-class at most universities. To minimize the missing data due to unintended skipping of the responses within the questionnaires, we will encourage students to thoroughly review their responses to ensure that they answered every question. In the online version, we will use automatized options to check missing responses to alert participants that they did not answer the question.

Another issue regards that we use self-report measures, particularly retrospective measures for students' learning experiences. These might be biased because remembering and subjectively judging the quality of courses from the last semester is error prone. Averaging ratings across students will hopefully lead to averaging out some error and the magnitude of between-student variance might indicate how error-prone retrospection is in this case. Also, for epistemic cognition it has been pointed out that self-report measures only allow quite superficial assessment (Mason, 2016). Alternatively, the idea of incorporating both quantitative and qualitative research methods upon EC has been supported (e.g., Greene and Yu, 2014). For the aims of our study we deem the EOCQ self-report measure appropriate but it will not allow a comprehensive look into the processes underlying students' development.

With reference to the country comparisons, even if it is not our main focus, they might be biased by the fact that at some universities all students participate within courses, at some they participate voluntarily without incentive, and at some voluntarily with incentive. Consequently, we acknowledge that these differences in recruitment may influence the results. In fact, voluntary participation in this research topic about scientific thinking can be seen as an indicator of student engagement and interest in science (and then, be categorized as an informal experience). In addition, it could happen that samples where students participate voluntarily with no incentive are overrepresented in the levels of

informal experiences' engagement. We will check to which extent students' research interest differs between these three groups.

Finally, since our research question is aimed at 1st-year students, we assess scientific thinking on a level that might be seen as quite basic for a higher education level. Measures on a more advanced level could be added, for example assessing justification in multiple sources as another facet of epistemic cognition (Ferguson and Bråten, 2013). More advanced measures might add relevant information, but they might potentially also mostly reveal floor effects due to 1st-year students' limited experiences.

ETHICS STATEMENT

The study materials and consent forms have been developed in accordance with ethical norms and guidelines from all participating countries and universities. The study was approved by the Institutional Review Board of Istanbul Bilgi University and IADT Institute Research Ethics Committee.

AUTHOR CONTRIBUTIONS

All authors listed, have made substantial, direct and intellectual contribution to the work, and approved it for publication.

ACKNOWLEDGMENT

This project is conducted under Junior Researcher Programme.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fpsyg.2017.00133/full#supplementary-material>

REFERENCES

- Bandura, A. (1997). *Self-Efficacy: The Exercise of Control*. New York, NY: Freeman.
- Barron, B. (2006). Interest and self-sustained learning as catalysts of development: a learning ecology perspective. *Hum. Dev.* 49, 193–224. doi: 10.1159/000094368
- Beißert, H., Köhler, M., Rempel, M., and Beierlein, C. (2014). *Eine Deutschsprachige Kurzsкала zur Messung des Konstrukts Need for Cognition: Die Need for Cognition Kurzsкала (NFC-K). GESIS-Working Papers 2014/32*. Available at: <http://nbn-resolving.de/urn:nbn:de:0168-ssoar-403157>
- Bråten, I., and Ferguson, L. E. (2014). Investigating cognitive capacity, personality, and epistemic beliefs in relation to science achievement. *Learn. Individ. Differ.* 36, 124–130. doi: 10.1016/j.lindif.2014.10.003
- Bråten, I., and Strømso, H. I. (2005). The relationship between epistemological beliefs, implicit theories of intelligence, and self-regulated learning among norwegian postsecondary students. *Br. J. Educ. Psychol.* 75, 539–565. doi: 10.1348/000709905X25067
- Cacioppo, J. T., and Petty, R. E. (1982). The need for cognition. *J. Pers. Soc. Psychol.* 42, 116–131. doi: 10.1037/0022-3514.42.1.116
- Carini, R. M., Kuh, G. D., and Klein, S. P. (2006). Student engagement and student learning: testing the linkages. *Res. High. Educ.* 47, 1–32. doi: 10.1007/s11162-005-8150-9
- Caro, D., and Cortes, D. (2012). Measuring family socioeconomic status: an illustration using data from PIRLS 2006. *IERI Monogr. Ser.* 5, 9–33.
- Chemers, M. M., Hu, L.-T., and Garcia, B. F. (2001). Academic self-efficacy and first-year college student performance and adjustment. *J. Educ. Psychol.* 93, 55–64. doi: 10.1037/0022-0663.93.1.55
- Chen, J. A., Metcalf, S. J., and Tutwiler, M. S. (2014). Motivation and beliefs about the nature of scientific knowledge within an immersive virtual ecosystems environment. *Contemp. Educ. Psychol.* 39, 112–123. doi: 10.1016/j.cedpsych.2014.02.004
- Chen, J. A., and Usher, E. L. (2013). Profiles of the sources of science self-efficacy. *Learn. Individ. Differ.* 24, 11–12. doi: 10.1016/j.lindif.2012.11.002
- Coleman, A. B., Lam, D. P., and Soowal, L. N. (2015). Correlation, necessity, and sufficiency: common errors in the scientific reasoning of undergraduate students for interpreting experiments. *Biochem. Mol. Biol. Educ.* 43, 305–315. doi: 10.1002/bmb.20879
- Dabbagh, N., and Kitsantas, A. (2012). Personal learning environments, social media, and self-regulated learning: a natural formula for connecting formal

- and informal learning. *Internet High. Educ.* 15, 3–8. doi: 10.1016/j.iheduc.2011.06.002
- Drummond, C., and Fischhoff, B. (2015). Development and validation of the scientific reasoning scale. *J. Behav. Decis. Mak.* 30, 26–38. doi: 10.1002/bdm.1906
- Duffy, S. (2010). Random numbers demonstrate the frequency of Type I errors: three spreadsheets for class instruction. *J. Stat. Educ.* 18, 1–16.
- Duschl, R. A., and Grandy, R. (2013). Two views about explicitly teaching nature of science. *Sci. Educ.* 22, 2109–2139. doi: 10.1007/s11191-012-9539-4
- Ebner, M., Lienhardt, C., Rohs, M., and Meyer, I. (2010). Microblogs in higher education - a chance to facilitate informal and process-oriented learning? *Comput. Educ.* 55, 92–100. doi: 10.1016/j.compedu.2009.12.006
- Elias, S. M., and Loomis, R. J. (2002). Utilizing need for cognition and perceived self-efficacy to predict academic performance. *J. Appl. Soc. Psychol.* 32, 1687–1702. doi: 10.1111/j.1559-1816.2002.tb02770.x
- Eshach, H. (2007). Bridging in-school and out-of-school learning: formal, non-formal, and informal education. *J. Sci. Educ. Technol.* 16, 171–190. doi: 10.1007/s10956-006-9027-1
- Etz, A., Gronau, Q. F., Dablander, F., Edelsbrunner, P. A., and Baribault, B. (2016). *How to Become a Bayesian in Eight Easy Steps: An Annotated Reading List*. Available at: <http://tinyurl.com/eightstepsV1>
- Evans, M. D. R., Kelley, J., Sikora, J., and Treiman, D. J. (2010). Family scholarly culture and educational success: books and schooling in 27 nations. *Res. Soc. Stratif. Mobil.* 28, 171–197. doi: 10.1016/j.rssm.2010.01.002
- Falk, J. H. (2001). *Free-Choice Science Education: How We Learn Science outside of School. Ways of Knowing in Science and Mathematics Series*. Available at: <https://eric.ed.gov/?id=ED465540>
- Ferguson, L. E., and Bråten, I. (2013). Student profiles of knowledge and epistemic beliefs: changes and relations to multiple-text comprehension. *Learn. Instr.* 25, 49–61. doi: 10.1016/j.learninstruc.2012.11.003
- Franco, G. M., Muis, K. R., Kendeou, P., Ranellucci, J., Sampasivam, L., and Wang, X. (2012). Examining the influences of epistemic beliefs and knowledge representations on cognitive processing and conceptual change when learning physics. *Learn. Instr.* 22, 62–77. doi: 10.1016/j.learninstruc.2011.06.003
- Gelman, A., and Stern, H. (2006). The difference between ‘significant’ and ‘not significant’ is not itself statistically significant. *Am. Stat.* 60, 328–331. doi: 10.1198/000313006X152649
- Gigerenzer, G. (2004). Mindless statistics. *J. Socio Econ.* 33, 587–606. doi: 10.1016/j.jsocec.2004.09.033
- Gokhale, A. A. (1995). Collaborative learning enhances critical thinking. *J. Technol. Educ.* 7, 22–30. doi: 10.21061/jte.v7i1.a.2
- Greene, J. A., Azevedo, R., and Torney-Purta, J. (2008). Modeling epistemic and ontological cognition: philosophical perspectives and methodological directions. *Educ. Psychol.* 43, 142–160. doi: 10.1080/00461520802178458
- Greene, J. A., Torney-Purta, J., and Azevedo, R. (2010). Empirical evidence regarding relations among a model of epistemic and ontological cognition, academic performance, and educational level. *J. Educ. Psychol.* 102, 234–255. doi: 10.1037/a0017998
- Greene, J. A., and Yu, S. B. (2014). Modeling and measuring epistemic cognition: a qualitative re-investigation. *Contemp. Educ. Psychol.* 39, 12–28. doi: 10.1016/j.cedpsych.2013.10.002
- Halpern, D. F., Millis, K., Graesser, A. C., Butler, H., Forsyth, C., and Cai, Z. (2012). Operation ARA: a computerized learning game that teaches critical thinking and scientific reasoning. *Think. Skills Creat.* 7, 93–100. doi: 10.1016/j.tsc.2012.03.006
- Haslbeck, J. M. B., and Waldorp, L. J. (2016). *Mgm: Structure Estimation for Time-Varying Mixed Graphical Models in High-Dimensional Data*. Available at: <http://arxiv.org/abs/1510.06871>
- Hoekstra, R., Morey, R. D., Rouder, J. N., and Wagenmakers, E.-J. (2014). Robust misinterpretation of confidence intervals. *Psychon. Bull. Rev.* 21, 1157–1164. doi: 10.3758/s13423-013-0572-3
- Hofer, B. K. (2000). Dimensionality and disciplinary differences in personal epistemology. *Contemp. Educ. Psychol.* 25, 378–405. doi: 10.1006/ceps.1999.1026
- Hofer, B. K., and Pintrich, P. R. (1997). The development of epistemological theories: beliefs about knowledge and knowing and their relation to learning. *Rev. Educ. Res.* 67, 88–140. doi: 10.3102/00346543067001088
- Hofstein, A., and Rosenfeld, S. (1996). Bridging the gap between formal and informal science learning. *Stud. Sci. Educ.* 28, 87–112. doi: 10.1080/03057269608560085
- Holmes, N. G., Wieman, C. E., and Bonn, D. A. (2015). Teaching critical thinking. *Proc. Natl. Acad. Sci. U.S.A.* 112, 11199–11204. doi: 10.1073/pnas.1505329112
- Jarosoz, A. F., and Wiley, J. (2014). What are the odds? A practical guide to computing and reporting bayes factors. *J. Probl. Solving* 7:2. doi: 10.7771/1932-6246.1167
- Jenert, T., Postareff, L., Brahm, T., and Sari, L.-Y. (2015). Editorial: enculturation and development of beginning students. *Zeitschrift Für Hochschulentwicklung* 10, 9–21.
- Johnson, D. W., Johnson, R. T., and Smith, K. A. (1998). Cooperative learning returns to college what evidence is there that it works? *Change* 30, 26–35. doi: 10.1080/00091389809602629
- Kassens-Noor, E. (2012). Twitter as a teaching practice to enhance active and informal learning in higher education: the case of sustainable tweets. *Active Learn. High. Educ.* 13, 9–21. doi: 10.1177/1469787411429190
- Kitchener, R. F. (2002). Folk epistemology: an introduction. *New ideas Psychol.* 20, 89–105. doi: 10.1016/S0732-118X(02)00003-X
- Klahr, D. (2000). *Exploring science: The Cognition and Development of Discovery Processes*. Cambridge, MA: MIT press.
- Klahr, D., Zimmerman, C., and Jirout, J. (2011). Educational interventions to advance children’s scientific thinking. *Science* 333, 971–975. doi: 10.1126/science.1204528
- Krause, K.-L., and Coates, H. (2008). Students’ engagement in first-year university. *Assess. Eval. High. Educ.* 22, 493–505. doi: 10.1080/02602930701698892
- Kuh, G. D., Cruce, T. M., Shoup, R., Kinzie, J., and Gonyea, R. M. (2008). Unmasking the effects of student engagement on first-year college grades and persistence. *J. High. Educ.* 79, 540–563. doi: 10.1353/jhe.0.0019
- Kuhn, D. (2002). “What Is Scientific Thinking and How Does It Develop?,” in *Blackwell Handbook of Childhood Cognitive Development*, ed. U. Goswami (Hoboken, NJ: Blackwell Publishers Ltd), 371–393.
- Kuhn, D. (2009). Do students need to be taught how to reason? *Educ. Res. Rev.* 4, 1–6. doi: 10.1016/j.edurev.2008.11.001
- Kuhn, D., Cheney, R., and Weinstock, M. (2000). The development of epistemological understanding. *Cogn. Dev.* 15, 309–328. doi: 10.1016/S0885-2014(00)00030-7
- Kuhn, D., Ramsey, S., and Arvidsson, T. S. (2015). Developing multivariable thinkers. *Cogn. Dev.* 35, 92–110. doi: 10.1016/j.cogdev.2014.11.003
- Laurillard, D. (2009). The pedagogical challenges to collaborative technologies. *Int. J. Comput. Support. Collab. Learn.* 4, 5–20. doi: 10.1007/s11412-008-9056-2
- Lawson, A. E., Clark, B., Cramer-Meldrum, E., Falconer, K. A., Sequist, J. M., and Kwon, Y.-J. (2000). Development of scientific reasoning in college biology: Do two levels of general hypothesis-testing skills exist? *J. Res. Sci. Teach.* 37, 81–101.
- Lehman, D. R., and Nisbett, R. E. (1990). A longitudinal study of the effects of undergraduate training on reasoning. *Dev. Psychol.* 26, 952–960. doi: 10.1037/0012-1649.26.6.952
- Manning, K., Zachar, P., Ray, G., and LoBello, S. (2006). Research methods courses and the scientist and practitioner interests of psychology majors. *Teach. Psychol.* 33, 194–196.
- Mason, L. (2016). “Psychological perspectives on measuring epistemic cognition,” in *Handbook of Epistemic Cognition*, eds J. A. Greene, W. A. Sandoval, and I. Bråten (New York, NY: Routledge), 375–392. doi: 10.4324/9781315795225
- McElreath, R. (2016). *Statistical Rethinking: A Bayesian Course with Examples in R and Stan*. Boca Raton, FL: CRC Press.
- Morey, R. D., Hoekstra, R., Rouder, J. N., Lee, M. D., and Wagenmakers, E.-J. (2015). The fallacy of placing confidence in confidence intervals. *Psychon. Bull. Rev.* 23, 103–123. doi: 10.3758/s13423-015-0947-8
- Morey, R. D., and Rouder, J. N. (2011). Bayes factor approaches for testing interval null hypotheses. *Psychol. Methods* 16, 406–419. doi: 10.1037/a0024377
- Moss, E. (2012). *Assessing Understanding of the Nature of Science and Science Self-Efficacy in Undergraduates Involved in Research in an Introductory Geology Course*. Available at: <http://lib.dr.iastate.edu/etd/12825>
- Muis, K. R., and Franco, G. M. (2009). Epistemic profiles and metacognition: support for the consistency hypothesis. *Metacogn. Learn.* 5, 27–45. doi: 10.1007/s11409-009-9041-9

- Nieuwenhuis, S., Forstmann, B. U., and Wagenmakers, E.-J. (2011). Erroneous analyses of interactions in neuroscience: a problem of significance. *Nat. Neurosci.* 14, 1105–1107. doi: 10.1038/nn.2886
- Nussbaum, E. M., Sinatra, G. M., and Poliquin, A. (2008). Role of epistemic beliefs and scientific argumentation in science learning. *Int. J. Sci. Educ.* 30, 1977–1999. doi: 10.1080/09500690701545919
- Parent, M. C., and Oliver, J. A. (2015). Mentoring the earliest-career psychologists: role models, knowledge of internship issues, and attitudes toward research and science. *Prof. Psychol. Res. Pr.* 46, 55–61. doi: 10.1037/a0038839
- Pashler, H., and Wagenmakers, E.-J. (2012). Editors' introduction to the special section on replicability in psychological science a crisis of confidence? *Perspect. Psychol. Sci.* 7, 528–530. doi: 10.1177/1745691612465253
- Patrick, W. (2010). *Recognising Non-Formal and Informal Learning Outcomes, Policies and Practices: Outcomes, Policies and Practices*. Paris: OECD Publishing.
- Plummer, M. (2003). "JAGS: a program for analysis of Bayesian graphical models using Gibbs sampling," in *Proceedings of the 3rd International Workshop on Distributed Statistical Computing*, Vienna, Vol. 124, 125.
- Plummer, M. (2013). *rjags: Bayesian Graphical Models Using MCMC. R Package Version 3*.
- R Core Team (2013). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing, 409.
- Resnick, L. B. (1987). The 1987 presidential address learning in school and out. *Educ. Res.* 16, 13–54. doi: 10.3102/0013189X016009013
- Riggs, M. L., Warka, J., Babasa, B., Betancourt, R., and Hooker, S. (1994). Development and validation of self-efficacy and outcome expectancy scales for job-related applications. *Educ. Psychol. Meas.* 54, 793–802. doi: 10.1177/0013164494054003026
- Schönbrodt, F. D., and Wagenmakers, E.-J. (2016). *Bayes Factor Design Analysis: Planning for Compelling Evidence*. SSRN Scholarly Paper ID 2722435. Rochester, NY: Social Science Research Network.
- Sodian, B., Zaitchik, D., and Carey, S. (1991). Young children's differentiation of hypothetical beliefs from evidence. *Child Dev.* 62, 753–766. doi: 10.1111/j.1467-8624.1991.tb01567.x
- Stoloff, M., McCarthy, M., Keller, L., Varfolomeeva, V., Lynch, J., Makara, K., et al. (2009). The undergraduate psychology major: an examination of structure and sequence. *Teach. Psychol.* 37, 4–15. doi: 10.1080/00986280903426274
- Su, Y.-S., and Yajima, M. (2012). *R2jags: A Package for Running jags from R. R Package Version 0.03-08*. Available at: <http://CRAN.R-project.org/package=R2jags>
- Van De Schoot, R., Kluytmans, A., Tummers, L., Lugtig, P., Hox, J., and Muthén, B. (2012). Facing off with scylla and charybdis: a comparison of scalar, partial, and the novel possibility of approximate measurement invariance. *Front. Psychol.* 4:770. doi: 10.3389/fpsyg.2013.00770
- VanderStoep, S. W., and Shaughnessy, J. J. (1997). Taking a course in research methods improves reasoning about real-life events. *Teach. Psychol.* 24, 122–124. doi: 10.1207/s15328023top2402_8
- VanWormer, L. A., Jordan, E. F., and Blalock, L. D. (2014). Assessing the perceived value of research participation. *Teach. Psychol.* 41, 233–236. doi: 10.1177/0098628314537974
- Vivian, R., Barnes, A., Geer, R., and Wood, D. (2014). The academic journey of university students on facebook: an analysis of informal academic-related activity over a semester. *Res. Learn. Technol.* 22:24681. doi: 10.3402/rlt.v22.24681
- Wagenmakers, E.-J. (2007). A practical solution to the pervasive problems of p values. *Psychon. Bull. Rev.* 14, 779–804. doi: 10.3758/BF03194105
- Wagenmakers, E.-J., Morey, R. D., and Lee, M. D. (2016). Bayesian benefits for the pragmatic researcher. *Curr. Dir. Psychol. Sci.* 25, 169–176. doi: 10.1177/0963721416643289
- Walker, C. O., Greene, B. A., and Mansell, R. A. (2006). Identification with academics, intrinsic/extrinsic motivation, and self-efficacy as predictors of cognitive engagement. *Learn. Individ. Differ.* 16, 1–12. doi: 10.1016/j.lindif.2005.06.004
- Wayment, H. A., and Dickson, K. L. (2008). Increasing student participation in undergraduate research benefits students, faculty, and department. *Teach. Psychol.* 35, 194–197. doi: 10.1177/009862830803500307
- Woo, S. E., Harms, P. D., and Kuncel, N. R. (2007). Integrating personality and intelligence: typical intellectual engagement and need for cognition. *Pers. Individ. Dif.* 6, 1635–1639. doi: 10.1016/j.paid.2007.04.022
- Yang, M. (2015). *Promoting Self-Sustained Learning in Higher Education: The ISEE Framework*. Available at: <http://repository.lib.eduhk.hk/jspui/handle/2260/17647>
- Ziegler, M., Kemper, C. J., and Lenzner, T. (2015). The issue of fuzzy concepts in test construction and possible remedies. *Eur. J. Psychol. Assess.* 31, 1–4. doi: 10.1027/1015-5759/a000255
- Zimmerman, C. (2007). The development of scientific thinking skills in elementary and middle school. *Dev. Rev.* 27, 172–223. doi: 10.1016/j.dr.2006.12.001

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2017 Soyylmaz, Griffin, Martín, Kucharský, Peycheva, Vaupotič and Edelsbrunner. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Developing a Taxonomy of Dark Triad Triggers at Work – A Grounded Theory Study Protocol

Annika Nübold^{1*}, Josef Bader², Nera Bozin³, Romil Depala⁴, Helena Eidast⁵, Elisabeth A. Johannessen⁶ and Gerhard Prinz^{7*}

¹ Department of Work and Social Psychology, Maastricht University, Maastricht, Netherlands, ² Faculty of Psychology and Educational Sciences, University of Coimbra, Coimbra, Portugal, ³ Department of Psychology, University of Ljubljana, Ljubljana, Slovenia, ⁴ Department of Experimental Psychology, University of Oxford, Oxford, UK, ⁵ Independent Researcher, Tallinn, Estonia, ⁶ Department of Psychology, University of Winchester, Winchester, UK, ⁷ Department of Basic Psychological Research and Research Methods, University of Vienna, Vienna, Austria

OPEN ACCESS

Edited by:

Rocio Del Pino,
University of Deusto, Spain

Reviewed by:

Serge Brand,
University of Basel, Switzerland
Peter Karl Jonason,
Western Sydney University, Australia

*Correspondence:

Annika Nübold
a.nubold@maastrichtuniversity.nl
Gerhard Prinz
gerhard.prinz@univie.ac.at

Specialty section:

This article was submitted to
Quantitative Psychology
and Measurement,
a section of the journal
Frontiers in Psychology

Received: 11 June 2016

Accepted: 15 February 2017

Published: 07 March 2017

Citation:

Nübold A, Bader J, Bozin N, Depala R, Eidast H, Johannessen EA and Prinz G (2017) Developing a Taxonomy of Dark Triad Triggers at Work – A Grounded Theory Study Protocol. *Front. Psychol.* 8:293. doi: 10.3389/fpsyg.2017.00293

In past years, research and corporate scandals have evidenced the destructive effects of the dark triad at work, consisting of narcissism (extreme self-centeredness), psychopathy (lack of empathy and remorse) and Machiavellianism (a sense of duplicity and manipulateness). The dark triad dimensions have typically been conceptualized as stable personality traits, ignoring the accumulating evidence that momentary personality expressions – personality *states* – may change due to the characteristics of the situation. The present research protocol describes a qualitative study that aims to identify triggers of dark triad *states* at work by following a grounded theory approach using semi-structured interviews. By building a comprehensive categorization of dark triad triggers at work scholars may study these triggers in a parsimonious and structured way and organizations may derive more effective interventions to buffer or prevent the detrimental effects of dark personality at work.

Keywords: dark triad, personality states, workplace triggers, taxonomy, grounded theory

INTRODUCTION

The study of dark personality and its impact in the workplace has gained increasing attention in the past years (Spain et al., 2014). Dark personality traits are defined as characteristics that reflect a motivation to elevate the self and harm others (Paulhus and Williams, 2002). Amongst other conceptualizations (e.g., Hogan and Hogan, 2001) the dark triad consisting of narcissism, psychopathy, and Machiavellianism (Paulhus and Williams, 2002) represents the most popular operationalization of dark personality at work (Spain et al., 2014).

Narcissism is characterized by feelings of grandiosity, entitlement, dominance, and superiority (Spain et al., 2014). People who show this trait tend to be charming or pleasant in the short term while in the long run presenting difficulty in maintaining successful interpersonal relationships, lacking trust and care for others (Morf and Rhodewalt, 2001). *Psychopathy* involves feelings of impulsivity, thrill-seeking, low empathy and anxiety. Those that present a psychopathic trait seek immediate gratification of their needs, lack guilt and conscience, being less prone to experience embarrassment and failing to learn from punishment for misdeeds (Hare, 1985). *Machiavellianism* is associated with cynicism, low affect, an unconventional view of morality and a focus exclusively on personal goals (Christie and Geis, 1970). Thus, those who exemplify this trait tend to be exceedingly willing to manipulate others and take a certain pleasure in successfully deceiving them (Jones and Paulhus, 2014).

Importantly, despite similarities and overlap, the dark triad is not identical with clinically relevant personality disorders nor does it reflect simply extreme forms of normal personality traits (Harms et al., 2014). Both research and past corporate scandals have evidenced the dark triad's effects on counterproductive work behaviors (e.g., O'Boyle et al., 2012; Spain et al., 2014) and a variety of other destructive outcomes, such as heightened competitiveness, dysfunctional job crafting, and corruption (Carter et al., 2014; Roczniowska and Bakker, 2016; Zhao et al., 2016).

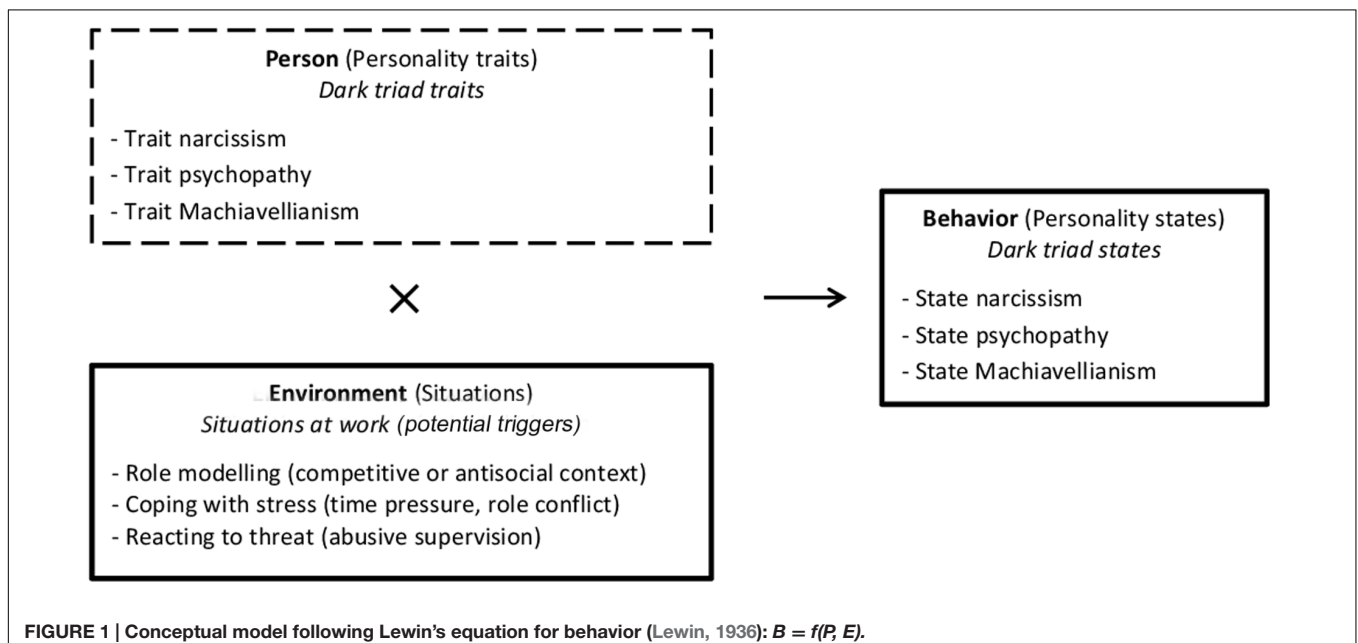
To date, research has mainly focused on the detrimental *outcomes* of dark personality. Although initial efforts have been made in past years to discover the psychological underpinnings of the dark triad (Paulhus and Williams, 2002), research has failed to address the role of situational factors in eliciting momentary expressions of dark personality characteristics. To date, the dark triad has exclusively been conceptualized and investigated in its trait-like form, ignoring evidence for the malleability and short-term fluctuation of personality states, the expression of one's personality in a specific moment (e.g., Fleeson, 2001). Due to the predominant view of the dark triad as stable traits, situational cues eliciting dark triad behavior have not been of concern so far. Attempts to identify the roots of dark triad personality have thus focused on very broad, generic explanations, such as evolutionary (e.g., Jonason et al., 2009), behavioral genetic (e.g., Vernon et al., 2008), socio-ecological (Jonason et al., 2016), neuro-biological (Jonason and Jackson, 2016), and motivational (Harms et al., 2014; Jonason and Ferrell, 2016) foundations.

In our study we apply a more dynamic approach to dark personality. Drawing on interactionist models of personality, such as the cognitive affective personality system (CAPS; Mischel and Shoda, 1995, 1998) or whole trait theory (Fleeson and Jayawickreme, 2015) that build upon Lewin's (1936) equation for predicting behavioral reactions [$B = f(P \times E)$], we assume

that stable personality traits (P) interact with environmental characteristics (E) to produce a specific behavioral response (B), or in other words, a specific personality state (see **Figure 1** for our conceptual model; please note that the listed situations at work merely represent examples of potential triggers and don't reflect empirical findings). Mischel and Shoda (1995, 1998) called this complex interplay *if-then contingencies* describing the idea that specific situational cues make people reliably react in a specific way, based on their personality traits. Empirical studies (e.g., Judge et al., 2013) and entire special issues (Beckmann and Wood, 2016) have evidenced that the behavioral expressions (i.e., personality states) in relation to changing situations present a potentially predictable reflection of personality (Mischel and Shoda, 1995, 1998).

Despite the growing evidence that personality expressions are dependent on situational cues, research on the triggering function of job characteristics is still in its infancy and has only focused on adaptive personality states so far (Judge et al., 2013; Dóci and Hofmans, 2015). The usual approach for adaptive personality states to conceptualize situations mainly as opportunities to express one's personality (Ten Berge and De Raad, 1999) may, however, be problematic for the concept of the dark triad. Whereas some situations may allow for or even encourage the expression of dark personality characteristics because they may be functional in that moment (e.g., being self-aggrandizing in a selection interview), other situations may rather trigger dark personality states because specific needs and motives are not fulfilled (e.g., the need for power). Thus, although "a psychology of situations has begun to take shape" (Funder, 2016, p. 203), a comprehensive taxonomy of situational *triggers* of *specific* personality characteristics, including the dark triad, has yet to be established.

In this study, we aim to identify the underlying *situational* antecedents in the work environment (E) that lead to



within-person variation in momentary dark triad expressions, i.e., in dark triad personality states (B), that is, *state* narcissism, *state* psychopathy and *state* Machiavellianism, resulting in a comprehensive taxonomy of triggers. We use a grounded theory approach (Strauss and Corbin, 1998; Corbin and Strauss, 2008) as this qualitative methodology is particularly suited for complex social processes about which little is known yet (Glaser and Strauss, 1967; Willig, 2009). Identifying potential triggers of dark personality expressions at work and building taxonomy of these triggers is important in several ways. On a theoretical level, a more comprehensive understanding of the triggers of dark personality states and their potential interconnections is crucial in order to acquire knowledge on the nomological network and common mechanisms that may be associated with specific groups or categories of triggers. Moreover, identifying triggers of dark personality expressions at work is practically important, as organizations strongly benefit from detailed knowledge on situations that may “make people snap.” Organizations and Human Resource (HR) professionals will be able to design evidence-based interventions that help prevent employees from expressing their dark tendencies which may bring harm to organizations and their members over the short and long term. These actions may not only include job design and selective placement, but also training and coaching interventions that sensitize employees for potentially dangerous situations and helping them to either themselves better manage and regulate their dark impulses or increase their ability to cope with others’ dark behaviors, enabling better relationships and work ethics within their organizations.

With the present study, we contribute to the literature in three important ways: Firstly, identifying triggers of *dark* personality expressions at work adds to the evolving personality state literature by broadening the domain of situational predictors and allowing us to better understand the complex interplay between person and situation characteristics ($P \times E$) jointly leading to the expression of personality states (B) in a specific situation. Secondly, our taxonomy may be used to further advance research in the field of dark personality at work by stimulating the creation of instruments such as questionnaires or situational judgment tests that enable scholars to study these triggers in quantitative studies in a more standardized and parsimonious way. In addition, our taxonomy may further add to the study of long-term personality development, helping to further explore how individuals’ personality turns dark over time by investigating how short-term dynamics add up in a longitudinal fashion (Hogan et al., 1994). Finally, editors of top tier journals (e.g., Suddaby, 2006; Bansal and Corley, 2011) explicitly acknowledge the value of grounded theory and call for more qualitative research in the organizational literature. By following this call we promote a methodology that is particularly useful for examining situated processes like employee interactions in complex organizational settings (Locke, 2001). Qualitative research complements the many quantitative studies in the organizational literature by offering the reader a close-up of the phenomenon being studied and providing the opportunity to raise new research questions, revealing “deeper insights into management, organizations, and society, which are critical to

understanding and potentially shaping our world” (Bansal and Corley, 2011, p. 235).

In sum, our research questions are as follows:

- (1) Which situational characteristics at work trigger expressions of state narcissism?
- (2) Which situational characteristics at work trigger expressions of state psychopathy?
- (3) Which situational characteristics at work trigger expressions of state Machiavellianism?

MATERIALS AND EQUIPMENT

Semi-structured In-depth Interviews

In order to gather information on momentary experiences of dark triad states and their eliciting factors at work, we will conduct semi-structured in-depth interviews with employees. Semi-structured interviews enable the interviewer to flexibly adapt and add further questions depending on the answers given by the interviewee allowing for more in-depth explanations of how the person experienced the situation, thereby increasing the validity of the interview (McLeod, 2014).

Interviews will be conducted with jobholders of all seniority levels reporting about their own experiences (i.e., self-reports) or about the behavior of someone else (i.e., observer-reports), to account for different perspectives and to create a multifaceted view on the phenomenon of interest (Glaser and Strauss, 1967; Bluhm et al., 2011). As we expect it to be more difficult to obtain answers from participants about their own dark behaviors due to the possibility of socially desirable responding, we will first invite participants to talk about their own behaviors before also offering them the option of reporting the behaviors of a significant other (e.g., a colleague, supervisor or subordinate). Although self-reports are particularly valuable as they can target internal emotions and cognitions, it is important to note that research has evidenced that dark triad characteristics (e.g., features of psychopathy) can also be reliably and validly detected by lay raters, particularly if they involve interpersonal behaviors (Fowler et al., 2009). Therefore, employees reporting on their impressions and observations of a relevant situation involving *another* person are valuable sources of information as well (Fowler et al., 2009).

Although it has also been recommended to apply several methods of data collection (Bluhm et al., 2011), conducting observations and analyzing existing written materials will not be the focus of our research. As our study’s purpose is to identify the (to date unknown) triggering factors of dark triad states, it will neither be possible to reasonably plan specific observation periods nor to conduct observations spanning many hours or even days hoping for a potential triggering situation to occur.

The actual interview will consist of three main parts:

- (1) An open and generic question about the interviewees’ experiences at work and their job in general in order to allow interviewees to get comfortable with reporting about their experiences while we introduce them to the topic.

- (2) The second part will include questions about experiences of specific dark triad expressions at work. Descriptions of these dark personality expressions will be based on items from validated instruments, such as the Mach IV (Christie and Geis, 1970), the Self-Report Psychopathy Scale III (SRP-III; Hare, 1985), and the Narcissistic Personality Inventory (NPI; Raskin and Hall, 1979). These descriptions will ensure a common understanding of the focal behaviors and will stimulate interviewees to reflect on situations in which they have shown these behaviors themselves (or observed them in another person). A sample question for state Machiavellianism would be: “When thinking about your work experiences in the past 12 months – can you think about a specific situation in which you have done certain things to get someone to do what you want?”. In cases where participants experience difficulties in answering the question, a prompt could be: “For instance, have you said something particularly to someone to get what you want?”. Interviewees will then be asked to describe a *specific situation* in which they expressed this behavior in more detail, for example in terms of where the situation took place, the specific individuals involved in the interaction, and the specific actions.
- (3) The third step will focus on the aim of the interviews, the potential triggering situations of dark triad expressions at work. Accordingly, interviewees will be asked their opinion on what they believe may have *evoked* their specific behavior, that is, the preceding situation and the specific trigger of that behavior. Interviewees will also be asked about the broader context of the situation (e.g., if there have been prior incidents with someone), as the meaning of situational triggers may change from one context to another. Participants will be asked to describe this triggering situation in as much detail as possible. As suggested by previous research (Judge et al., 2013; Jonason and Ferrell, 2016; Jonason et al., 2016), we will ask participants reporting their own experiences for both, more objective environmental events (i.e., something someone did or said) and internal events (e.g., subjective feelings, motives, or cognitions), as both events may serve as antecedents to dark triad personality states. Participants reporting the behaviors of someone else will only be asked about the more objective features of the situation and will not be asked to speculate about the potential thoughts and feelings of the target person. As noted by Funder (2016) with reference to Reis (2008), situations are always filtered through the perception of the individual who experiences them, but should nonetheless be conceptualized separately from individual construal. Likewise, Rauthmann et al. (2015) emphasize in the so called processing principle that people’s experiences are based upon objective environmental events which are explicitly and/or implicitly processed by individuals, thereby providing situations with meaning. The perspective that both internal and external events may be meaningful as triggers of dark personality states is in line with a post-positivist perspective and the aim to come closer to a singular truth while acknowledging that evidence is not confined to what

can be physically observed and that subjective influences exist.

METHODS

Design

In the present study we use a grounded theory approach to answer our research questions. Grounded theory captures the complexity of social processes like no other methodology, it reveals content that is highly embedded in practice, and gives researchers the possibility to describe a phenomenon in great detail (Martin and Turner, 1986; Strauss and Corbin, 1998). Most importantly, it supports theorizing in “new” areas of research (Strauss and Corbin, 1998; Birks and Mills, 2015) and allows researchers to revise the direction and framework of research in real time as soon as new information and findings emerge. Our research questions are particularly appropriate for a grounded theory approach as there is a lack of research on dark personality states and therefore also a lack of essential information on their triggers.

There are multiple philosophies regarding grounded theory methodology (e.g., Glaser, 1978; Strauss and Corbin, 1998; Corbin and Strauss, 2008; Charmaz, 2014). In the present study, we follow the epistemological approach of *postpositivism* which assumes that there is one truth that can be discovered, but acknowledges that individuals’ perceptions are influenced by the context and that information gathered in the research process is not a neutral reflection of the truth. Thus, we approach grounded theory with the understanding that reality exists and that objectivity can be reached by discovering an emergent theory that represents this reality as accurately as possible. In line with the post-positivist approach, we follow Strauss and Corbin’s (1998) assumption that a theory is *discovered* in the data instead of being fully constructed.

Although cross-cultural research has many benefits, conducting interviews in two different languages (English and German) may be an area of concern in qualitative studies (Squires, 2009; Nurjannah et al., 2014). As recommended for multilingual research projects (Van Nes et al., 2010), we aim to make use of the original language for as long as possible. Specifically, interviews that will be conducted in English will be transcribed and coded in English while the interviews conducted in German will be transcribed and coded in German. Only after the coding procedure, we will translate the codes and the respective text passages and sample quotes derived from the German interviews into English. In order to ensure equivalence of meaning of these translations, we will follow the translation back-translation procedure by Brislin (1970), while making use of a translator moderator (Van Nes et al., 2010). The translator moderator will be the first author who will conduct the interviews in German and at the same time is highly proficient in English. Translating the codes instead of the original transcripts ensures the authenticity of the data and quality of analysis by minimizing potential misinterpretation and loss of participants’ intended meanings (Larkin et al., 2007) while at the same time being more economic and feasible (Chen and Boore, 2009).

Participants

In order to identify triggers of dark triad expressions at work, we will approach jobholders that are either willing to report on their own behavior at work or on behavioral observations of someone else, for example a colleague, their supervisor or a subordinate. In addition, we will approach subject matter experts, such as HR consultants, who may be able to report on dark triad expressions of clients and potential triggering situations based on their work with organizations. As the dark triad (Paulhus and Williams, 2002) refers to *subclinical* or *everyday versions* of maladaptive personality, in contrast to clinically relevant disorders, dark triad characteristics may be well represented in normal populations. Thus, our target sample will consist of regular employees and professionals. Potential participants will initially be approached via the interviewers' networks (e.g., via professional business and employment-oriented social networking service).

In order to achieve high heterogeneity of data sources (i.e., a maximum amount of variance in the target behaviors and situational triggers), we plan to approach male and female employees of different ages from a wide variety of branches, jobs, positions, and hierarchy levels. Through this approach, we also aim to identify more severe situations (that may be needed to trigger individuals with low levels in the dark triad) as well as less critical situations (that may function as triggers for those with high dark triad levels). We will not determine the number of interviews (our sample size) *a priori* (Eisenhardt and Graebner, 2007), but will continue to collect data until a *theoretical saturation point* has been reached and no new relevant categories of triggers emerge (Glaser and Strauss, 1967). As qualitative research handles non-numerical information and because the right sample size depends on a number of factors, such as the variety and content of answers and the scope of the study, power calculations that are appropriate in quantitative research are not applicable here (Morse, 2000; Leung, 2015). Nonetheless, scholars have for example suggested 20–30 interviews for grounded theory (Creswell, 1998), a sample size that has also been confirmed in grounded theory studies conducted in an organizational context (Seivwright and Unsworth, 2016; Wilhelmy et al., 2016). Additionally, we will document the specific steps of the theoretical sampling process to make the choice of our eventual final sample size as transparent as possible (Nelson, 2016).

Procedure

A model of the procedural steps in our study is depicted in **Figure 2**.

Interviews will be conducted in English and German. All interviews will be audio recorded after obtaining permission from participants to tape-record the session. Interviews will take place face-to-face at the facilities of the university the respective interviewer is affiliated with or via telephone or videoconference calls (e.g., Skype). Telephone and videophone calls have been found to be a solid substitute for face-to-face interviews, especially in semi-structured interviews, while at the same time allowing for more efficient data collection (Sturges and Hanrahan, 2004; Berg, 2007; Sullivan, 2012).

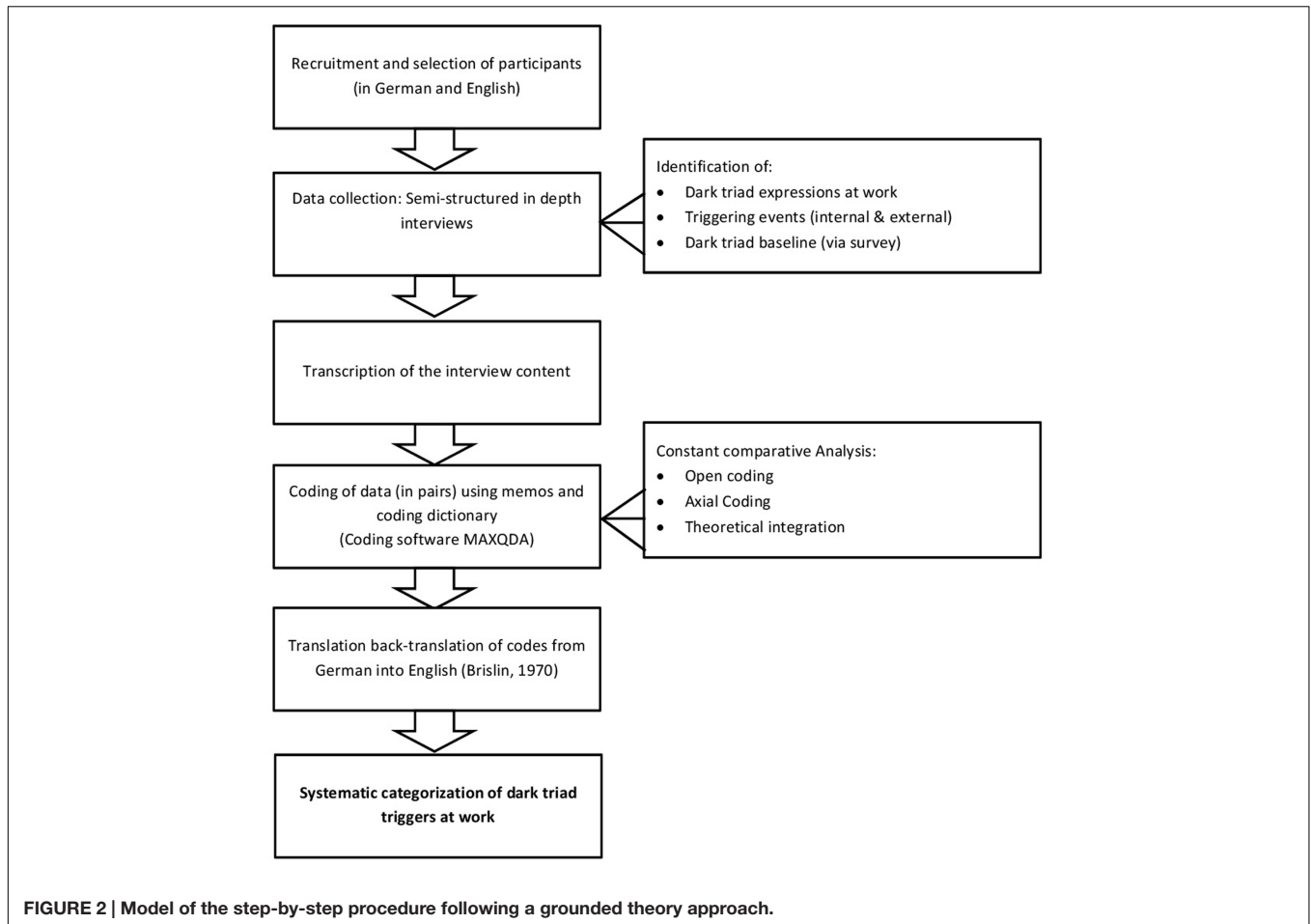
At the beginning of each in-depth interview, participants will be informed about the purpose and general context of the study in order to ensure transparency and allow for proper consideration of participation (e.g., Wilhelmy et al., 2016). Further, individuals will be ensured of the anonymity and confidentiality of their answers and the right to withdraw from the study at any point. In order to limit recall bias, participants will be instructed to respond to our interview questions based on their work experiences within the past 12 months. At the end of each interview, participants will be encouraged to complete a short questionnaire of the dark triad traits, the SD3 (Jones and Paulhus, 2014), in order to capture their baseline level of these characteristics. The assessment of their baseline will allow for a more detailed understanding of the distribution and level of dark personality characteristics in our sample and will allow us to link this information with the descriptions of the triggering situations, enabling us to control for possible moderating effects ($P \times E$). Research has frequently documented that personality characteristics (P), such as neuroticism, moderate the perception of and reaction to daily experiences (E) (e.g., stressful events; Bolger and Zuckerman, 1995; Suls and Martin, 2005) and systematically influence the level and variability of personality states (B) (Fleeson and Jayawickreme, 2015). Likewise, individuals with different levels of dark triad traits may perceive and react to situations differently (either in terms of the quality and character or the severity of situational cues). In order to prevent sensitization effects, we decided to administer the SD3 (Jones and Paulhus, 2014) after rather than before the interview.

Participants will be asked to sign the informed consent form and give general demographic related information. Although it is possible that participants will find some content of the interview upsetting or disturbing, this is very unlikely (Corbin and Morse, 2003). Reporting on sensitive topics can even benefit participants as it gives them an opportunity to be heard and to express their thoughts and feelings (Corbin and Morse, 2003). Participants will also have the opportunity to request that their interviews are not used in our study and the option to withdraw at any stage. Finally, individuals will be asked if they can recommend further potential interview candidates who would be suited and willing to participate in the study while bringing a benefit to our project (i.e., snowballing procedure).

Information gathered throughout the interviews will be used to develop new and more detailed questions to be added to the interview guide as well as to adjust our sampling strategy by, for example, targeting additional branches (Glaser and Strauss, 1967; Eisenhardt and Graebner, 2007). Through this procedure we will be able to further verify ideas that emerged from previous interviews and to ensure that we gather a rich and comprehensive view on the situations and identify interrelations between triggers, common trends, as well as their respective validity and importance (Glaser and Strauss, 1967).

Proposed Analysis

As recommended by Corbin and Strauss (2008), data will be analyzed and discussed by multiple researchers (i.e., the authors). All interviews will be coded by pairs of two researchers to ensure



multiple perspectives on the data (Corbin and Strauss, 2008) while minimizing personal bias and increasing reliability of our coding procedure. For the coding of the transcripts we will use the coding software (MAXQDA, 1995–2017). The transcriptions will be coded in three different stages – open coding, axial coding and theoretical integration (Corbin and Strauss, 2008). Coding is a procedure through which researchers create meaningful labels for sections of texts. Using coding to make sense of the data is not a step-by-step procedure, but rather a very flexible, iterative process (Rich, 2012). The different stages do not have to be followed in a strict manner but the data may take the analysis back and forth until *theoretical saturation* is reached and no new coding categories emerge (Pandit, 1996; Strauss and Corbin, 1998). In order to facilitate the coding procedure, we will make use of a so-called coding dictionary, a document including the evolving system of categories that will be continually modified through constant comparative analysis of existing and newly evolving codes (Kreiner et al., 2009). In order to maintain a high theoretical sensitivity, it is essential not to emerge oneself in the existing literature, as this may lead to biased results; researcher may unconsciously search for specific information to “fit” previous findings and fail to identify other important concepts inherent in the data (Strauss and Corbin, 1998).

The first stage of coding is open coding where the data is fractured or broken down into discrete parts which are closely examined and compared and then provided with codes which at a later stage can be grouped into categories (Glaser and Strauss, 1967). In general, codes can be given to the data word-by-word, phrase-by-phrase, sentence-by-sentence, line-by-line, or paragraph-by-paragraph (Strauss and Corbin, 1998). We will use line-by-line coding which allows researchers to be an active reader while writing memos on particularly interesting codes (Strauss and Corbin, 1998; Charmaz, 2014). Memos are written records of the researchers thought process and will be taken throughout the entire study (Birks and Mills, 2015). They are essential to the discovery of the theory as they keep track of and explain the thought process of the researcher during the coding process (e.g., why the data was coded in a certain way; Birks and Mills, 2015). In order to establish consensus on the proper use of a code, pairs of coders will meet up to compare their individual codings and discuss potential discrepancies.

In the second stage we will use axial coding in order to identify links and relationships between the concepts and in order to create main and subcategories of the codes (Pandit, 1996). Explanatory and conceptual patterns and relationships are identified by looking for recurring phenomena, incidents,

actions and interactions and by putting them in either main or subcategories (Strauss and Corbin, 1998). For example, all violence related codes could be categorized with the main code *violence* while subcategories could be named *behavioral violence* and *verbal violence*. In order to achieve consensus also on the more abstract categories or concepts, pairs of coders will once more meet to discuss their reasoning and approach of categorization. The emergence of new categories or changes in categories as well as their potential relation to existing literature (Locke, 2001) will be documented in the coding dictionary. Constantly comparing the data and the codes during the coding process ensures that the codes are congruent (Strauss and Corbin, 1998). In order to further explain the relationships between the concepts, we will start writing a story line, “a strategy for facilitating integration, construction, formulation and presentation of research findings through the production of a coherent grounded theory” (Birks and Mills, 2015, p. 176).

The final stage is the one of theoretical integration. In this stage a core category will be identified and the theory will be consolidated (Corbin and Strauss, 2008). A core category is defined by its ability to include all codes, sub and main categories, tying everything together to discover the theory (Corbin and Strauss, 2008). In this integrative process the storyline is developed further and results in a theory that is grounded in the data. In this process it is important to not have any preconceptions about the results and to look at the data objectively (Corbin and Strauss, 2008). Also in this final stage, pairs of coders will aim to reach consensus and will document all core categories and identified links in the coding dictionary.

Ethics Statement

This study (ECP-164_14_03_2016) has been approved by the Ethical Review Committee Psychology and Neuroscience (ERCPN) of the Faculty of Psychology and Neuroscience of Maastricht University, The Netherlands. The review was done according to Dutch law and also in the light of the highest ethical standards in the Dutch, Anglo-American and European (Union) context.

ANTICIPATED RESULTS

As very little is known about situational triggers of dark triad states at work, we can anticipate results only based on research on influencing factors of dark triad traits (Jonason et al., 2016) and maladaptive behaviors that are clinically relevant (i.e., schema modes; Keulen-de Vos et al., 2014). Research suggests for example that antisocial behavior at work (Robinson and O’Leary-Kelly, 1998; Lubit, 2002), competitive environments (Greenberg, 1990) and environments that can reduce a sense of behavioral accountability (e.g., cyberspace; Nevin, 2015) could facilitate dark personality expressions at work via social learning processes. In addition, scholars suggested that dark characteristics are more likely to emerge under periods of stress because this leads to a lack of cognitive resources that are needed to inhibit these dark impulses and motives in order to fulfill social expectations (Hogan and Hogan, 2001). Furthermore, research hints to the

detrimental effects of traumatic experiences at work, such as victimization, threat, manipulation, bullying, and destructive leadership, which may potentially trigger expressions of dark personality states (e.g., Jonason et al., 2012; Sharf et al., 2014; Cheang and Appelbaum, 2015; Nevin, 2015; Smith et al., 2016). This is in line with research on cluster B personality disorders (e.g., antisocial personality disorder), showing that violent and delinquent behavior can be explained by an unfolding sequence of schema modes, feelings of vulnerability and abandonment or loneliness, which then lead to violent psychopathic behaviors, such as bullying and manipulation (Keulen-de Vos et al., 2014).

In sum, we expect to derive a taxonomy of triggers that represents an initial first step for conducting further quantitative research on these dynamics. For example, based on our taxonomy scholars could conduct field studies (e.g., diary studies with a cross-sectional or lagged design) as well as laboratory experiments to verify the factor structure of the triggers that we hope to identify as well as to test their causal impact on dark personality states. Further, we also call for additional qualitative studies on this topic in order to test if our taxonomy can be identified with other samples or data sources as well (i.e., proving consistency of our findings; Leung, 2015).

Limitations

Although our study has a number of strengths, it also comes with several limitations and challenges. Firstly, grounded theory remains to some extent a subjective process bearing the risk of confirmation bias (Leung, 2015). Although subjectivity is considered an undesirable confounder in quantitative research, it is considered essential and even treasurable in qualitative research as it enriches the content of the findings (Leung, 2015). However, to make our findings more valid and generalizable (Johnson, 1997), we have interviewers/coders train their interviewing and coding skills and conduct pilot interviews, try to ensure that they are aware of the preconceptions they bring to the data coding process (Strauss and Corbin, 1998) and make our philosophical stance as transparent as possible. Importantly, the generalizability of grounded theory is partially achieved through the process of abstraction applied in the entire research process via the creation of codes, categories, and core categories (Strauss and Corbin, 1998).

Secondly, a major challenge concerns the attainment of high quality responses from participants in our interviews. As the interview topic touches the personal experiences of individuals and may also potentially explore deviant or illegal activities, it is of major importance to ensure that participants feel comfortable to report about this sensitive topic and to explicitly ensure anonymity and confidentiality (Larossa et al., 1981). We aim to build in *a priori* strategies for evaluating and terminating an interview should participants become severely distressed (Lee and Renzetti, 1990). These may include calling participants several days after the interview or providing them with a list of local counselors should the need arise (Corbin and Morse, 2003). In any case, we will try to make very explicit when approaching participants that speaking about this topic in the interview will not have any therapeutic implications and does not compensate seeking professional help to deal with stress or trauma.

Finally, as interviews will be conducted in German and English it is possible that cross-lingual and cross-cultural matters may arise, for instance data or categories derived in one language may not match the information derived in the other language (Squires, 2009). Although we will follow the recommendations of several authors with regard to a proper translation procedure (Temple, 2002; Van Nes et al., 2010) it is nonetheless important to acknowledge the cultural contexts people are situated in. To ensure a high quality of the project interviewers will receive training prior to conducting the interviews so as to allow them to master their interviewing skills and to minimize personal biases (Anderson, 2010). This also includes competence in cultural awareness; in other words, to be mindful of their own culture and how this may shape the interview process (Fontes, 2008).

Implications and Conclusion

The results of our study will be of theoretical as well as of practical relevance. On a theoretical level, our study will add to research on the dark triad at work which has ignored more malleable conceptualizations of personality characteristics (i.e., personality states) until now. By investigating personality dynamics as they have occurred in a specific situation, we provide a more proximal and fine-grained perspective on dark triad expressions and their eliciting factors. Further, as research on the effect of work characteristics on personality states is still in its infancy, shedding light on *dark* personality dynamics at work broadens the domain of triggers and personality states that are of importance in organizational settings.

Identifying the impact of situational characteristics at work on individuals' dark personality expressions is also highly relevant on a practical level. Research has shown that dark triad traits significantly relate to workplace deviance and counterproductive work behaviors (CWBs), such as workplace aggression, theft, and absenteeism (O'Boyle et al., 2012). These

unethical organizational behaviors cause extreme damage to organizations. The Association of Certified Fraud Examiners estimated that globally, businesses suffer annual losses of U\$2.9 trillion as a result of fraudulent activity (Moore et al., 2012). Building a categorization of dark triad triggers at work that will help organizations to design interventions that can prevent people from expressing their dark impulses (e.g., through appropriate job design or placement decisions) is therefore of great (economic) value. With our work, we want to support organizations to help their employees work on their dark side, thereby significantly improving people's lives in the long run.

AUTHOR CONTRIBUTIONS

All authors listed have made substantial, direct and intellectual contributions to the work, and approved it for publication. AN came up with the idea for this project, contributed to and managed all steps of the design and writing process and took the lead in writing the protocol. JB, NB, RD, HE, EAJ, and GP contributed equally to this protocol in the areas of study design and support in manuscript writing. All co-authors are listed in alphabetical order.

ACKNOWLEDGMENTS

This research was made possible by the *Junior Researcher Programme* (<http://jrp.psycholars.org/>). We would like to thank Dr. Kai Ruggeri and Elisa Haller for their feedback and support and the whole organizing team for their effort and dedication. This article was supported by the Open Access Publishing Fund of the University of Vienna, Austria. The conduction of the project will be supported by the University Fund Limburg.

REFERENCES

- Anderson, C. (2010). *Presenting and Evaluating Qualitative Research*. Available at: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2987281/>
- Bansal, P., and Corley, K. (2011). The coming of age for qualitative research: embracing the diversity of qualitative methods. *Acad. Manag. J.* 54, 233–237. doi: 10.5465/AMJ.2011.60262792
- Beckmann, N., and Wood, R. E. (eds). (2016). *Dynamic personality science. Integrating between-person stability and within-person change* [Research topic]. *Front. Psychol.* Available at: <http://journal.frontiersin.org/researchtopic/3551/dynamic-personality-science-integrating-between-person-stability-and-within-person-change>
- Berg, B. L. (2007). *Qualitative Research Methods for Social Sciences*, 6th Edn. Boston, MA: Pearson Education.
- Birks, M., and Mills, J. (2015). *Grounded Theory: A Practical Guide*. Thousand Oaks, California: Sage Publications.
- Bluhm, D. J., Harman, W., Lee, T. W., and Mitchell, T. R. (2011). Qualitative research in management: a decade of progress. *J. Manag. Stud.* 48, 1866–1891. doi: 10.1111/j.1467-6486.2010.00972.x
- Bolger, N., and Zuckerman, A. (1995). A framework for studying personality in the stress process. *J. Pers. Soc. Psychol.* 69, 890–902. doi: 10.1037/0022-3514.69.5.890
- Brislin, R. W. (1970). Back-translation for cross-cultural research. *J. Cross Cult. Psychol.* 1, 185–216. doi: 10.1177/135910457000100301
- Carter, G. L., Campbell, A. C., and Muncer, S. (2014). The dark triad: beyond a 'male' mating strategy. *Pers. Individ. Dif.* 56, 159–164. doi: 10.1016/j.paid.2013.09.001
- Charmaz, K. (2014). *Constructing Grounded Theory: A Practical Guide through Qualitative Analysis*. London: SAGE.
- Cheang, H. S., and Appelbaum, S. H. (2015). Corporate psychopathy: deviant workplace behaviour and toxic leaders – part one. *Ind. Commer. Train.* 47, 165–173. doi: 10.1108/ICT-12-2013-0086
- Chen, H. Y., and Boore, J. R. P. (2009). Translation and back-translation in qualitative nursing research: methodological review. *J. Clin. Nurs.* 19, 234–239. doi: 10.1111/j.1365-2702.2009.02896.x
- Christie, R., and Geis, F. L. (1970). *Studies in Machiavellianism*. New York, NY: Academic Press.
- Corbin, J., and Morse, J. M. (2003). The unstructured interactive interview: issues of reciprocity and risks when dealing with sensitive topics. *Qual. Inq.* 9, 335–354. doi: 10.1177/1077800403009003001
- Corbin, J., and Strauss, A. (2008). *Basics of Qualitative Research: Techniques and Procedures for Developing Grounded Theory*, 3rd Edn. Thousand Oaks, CA: SAGE.
- Creswell, J. W. (1998). *Qualitative Inquiry and Research Design: Choosing Among Five Traditions*. Thousand Oaks, CA: Sage Publications.
- Dóci, E., and Hofmans, J. (2015). Task complexity and transformational leadership: the mediating role of leaders' state core self-evaluations. *Leadersh. Q.* 26, 436–447. doi: 10.1016/j.leaqua.2015.02.008

- Eisenhardt, K. M., and Graebner, M. E. (2007). Theory building from cases: opportunities and challenges. *Acad. Manag. J.* 50, 25–32. doi: 10.5465/AMJ.2007.24160888
- Fleeson, W. (2001). Toward a structure- and process-integrated view of personality: traits as density distributions of states. *J. Pers. Soc. Psychol.* 80, 1011–1027. doi: 10.1037//0022-3514.80.6.1011
- Fleeson, W., and Jayawickreme, E. (2015). Whole trait theory. *J. Res. Pers.* 56, 82–92. doi: 10.1016/j.jrp.2014.10.009
- Fontes, L. A. (2008). *Interviewing Clients Across Cultures: A practitioner's Guides*. New York, NY: Guilford Press.
- Fowler, K. A., Lilienfeld, S. O., and Patrick, C. J. (2009). Detecting psychopathy from thin slices of behavior. *Psychol. Assess.* 21, 68–78. doi: 10.1037/a0014938
- Funder, D. C. (2016). Taking situations seriously: the situation construal model and the riverside situational Q-sort. *Curr. Dir. Psychol. Sci.* 25, 203–208. doi: 10.1177/0963721416635552
- Glaser, B. G. (1978). *Theoretical Sensitivity: Advances in the Methodology of Grounded Theory*. Mill Valley, CA: Sociology Press.
- Glaser, B. G., and Strauss, A. (1967). *The Discovery of Grounded Theory*. Chicago, IL: Aldine.
- Greenberg, J. (1990). Employee theft as a reaction to underpayment inequity: the hidden cost of pay cuts. *Journal of Applied Psychology* 75, 561–568. doi: 10.1037/0021-9010.75.6.667
- Hare, R. D. (1985). Comparison of procedures for the assessment of psychopathy. *J. Consult. Clin. Psychol.* 53, 7–16. doi: 10.1037/0022-006X.53.1.7
- Harms, P. D., Spain, S. M., and Wood, D. (2014). Mapping personality in dark places. *Ind. Organ. Psychol.* 7, 114–117. doi: 10.1111/iops.12117
- Hogan, R., Curphy, G. J., and Hogan, J. (1994). What we know about leadership: effectiveness and personality. *Am. Psychol.* 49, 493–504. doi: 10.1037//0003-066X.49.6.493
- Hogan, R., and Hogan, J. (2001). Assessing leadership: a view from the dark side. *Int. J. Sel. Assess.* 9, 40–51. doi: 10.1111/1468-2389.00162
- Johnson, J. L. (1997). “Generalizability in Qualitative Research,” in *Completing a Qualitative Project*, ed. J. M. Morse (Thousand Oaks, CA: SAGE Publications), 191–210. doi: 10.1604/9780761906018
- Jonason, P. K., and Ferrell, J. D. (2016). Looking under the hood: the psychogenic motivational foundations of the Dark Triad. *Pers. Individ. Dif.* 94, 324–331. doi: 10.1016/j.paid.2016.01.039
- Jonason, P. K., Icho, A., and Ireland, K. (2016). Resources, harshness, and unpredictability: the socioeconomic conditions associated with the Dark Triad traits. *Evol. Psychol.* 14, 11. doi: 10.1177/1474704915623699
- Jonason, P. K., and Jackson, C. J. (2016). The Dark Triad traits through the lens of Reinforcement Sensitivity Theory. *Pers. Individ. Dif.* 90, 273–277. doi: 10.1016/j.paid.2015.11.023
- Jonason, P. K., Li, N. P., Webster, G. D., and Schmitt, D. P. (2009). The dark triad: facilitating a short-term mating strategy in men. *Eur. J. Pers.* 23, 5–18. doi: 10.1002/per.698
- Jonason, P. K., Slomski, S., and Partyka, J. (2012). The Dark Triad at work: how toxic employees get their way. *Pers. Individ. Dif.* 52, 449–453. doi: 10.1016/j.paid.2011.11.008
- Jones, D. N., and Paulhus, D. L. (2014). Introducing the short dark triad (SD3) a brief measure of dark personality traits. *Assessment* 21, 28–41. doi: 10.1177/1073191113514105
- Judge, T. A., Simon, L. S., Hurst, C., and Kelley, K. (2013). What I experienced yesterday is who I am today: relationship of work motivations and behaviors to within-individual variation in the five-factor model of personality. *J. Appl. Psychol.* 99, 199–221. doi: 10.1037/a0034485
- Keulen-de Vos, M. E., Bernstein, D. P., Vanstipelen, S., Vogel, V., Luckner, T. P., Slaats, M., et al. (2014). Schema modes in criminal and violent behaviour of forensic cluster B PD patients: a retrospective and prospective study. *Legal Criminol. Psychol.* 21, 56–76. doi: 10.1111/lcrp.12047
- Kreiner, G. E., Hollensbe, E. C., and Sheep, M. L. (2009). Balancing borders and bridges: negotiating the work-home interface via boundary work tactics. *Acad. Manag. J.* 52, 704–730. doi: 10.5465/amj.2009.43669916
- Larkin, P. J., Casterle, B. D., and Schotsmans, P. (2007). Multilingual translation issues in qualitative research: reflections on a metaphorical process. *Qual. Health Res.* 17, 468–476. doi: 10.1177/1049732307299258
- Larossa, R., Bennett, L. A., and Gelles, R. J. (1981). Ethical dilemmas in qualitative family research. *J. Marriage Fam.* 43, 303–313. doi: 10.2307/351382
- Lee, R. M., and Renzetti, C. M. (1990). The problems of researching sensitive topics: an overview and introduction. *Am. Behav. Sci.* 33, 510–528. doi: 10.1177/0002764290033005002
- Leung, L. (2015). Validity, reliability, and generalizability in qualitative research. *J. Fam. Med. Prim. Care* 4, 324–327. doi: 10.4103/2249-4863.161306
- Lewin, K. (1936). *Principles of Topological Psychology*. New York, NY: McGraw Hill.
- Locke, K. (2001). *Grounded Theory in Management Research*. London: Sage.
- Lubit, R. (2002). The long-term organizational impact of destructively narcissistic managers. *Acad. Manag. Exec.* 16, 127–138. doi: 10.5465/ame.2002.6640218
- Martin, P. Y., and Turner, B. A. (1986). Grounded theory and organizational research. *J. Appl. Behav. Sci.* 22, 141–157. doi: 10.1177/002188638602200207
- MAXQDA (1995–2017). (Version 12) [Computer Software]. VERBI Software GmbH. Available at: <http://www.maxqda.com>
- McLeod, S. A. (2014). *The Interview Method*. Available at: www.simplypsychology.org/interviews.html
- Mischel, W., and Shoda, Y. (1995). A cognitive-affective system theory of personality: reconceptualizing situations, dispositions, dynamics, and invariance in personality structure. *Psychol. Rev.* 102, 246–268. doi: 10.1037/0033-295X.102.2.246
- Mischel, W., and Shoda, Y. (1998). Reconciling processing dynamics and personality dispositions. *Annu. Rev. Psychol.* 49, 229. doi: 10.1146/annurev.psych.49.1.229
- Moore, C., Detert, J. R., Klebe Treviño, L., Baker, V. L., and Mayer, D. M. (2012). Why employees do bad things: moral disengagement and unethical organizational behavior. *Pers. Psychol.* 65, 1–48. doi: 10.1111/j.1744-6570.2011.01237.x
- Morf, C. C., and Rhodewalt, F. (2001). Expanding the dynamic self-regulatory processing model of narcissism: research directions for the future. *Psychol. Inq.* 12, 243–251. doi: 10.1207/S15327965PLI1204_3
- Morse, J. M. (2000). Determining sample size. *Qual. Health Res.* 10, 3–5. doi: 10.1177/104973200129118183
- Nelson, J. (2016). Using conceptual depth criteria: addressing the challenge of reaching saturation in qualitative research. *Qual. Res.* doi: 10.1177/1468794116679873
- Nevin, A. D. (2015). *Cyber-Psychopathy: Examining the Relationship between Dark E-Personality and Online Misconduct*. Master's thesis, The University of Western Ontario, London, ON.
- Nurjannah, I., Mills, J., Park, T., and Usher, K. (2014). Conducting a grounded theory study in a language other than English: procedures for ensuring the integrity of translation. *SAGE Open* 4, 1–10. doi: 10.1177/2158244014528920
- O'Boyle, E. H., Forsyth, D. R., Banks, G. C., and McDaniel, M. A. (2012). A meta-analysis of the Dark Triad and work behaviour: a social exchange perspective. *J. Appl. Psychol.* 97, 557–579. doi: 10.1037/a0025679
- Pandit, N. R. (1996). *The Creation of a Theory: A Recent Application of the Grounded Theory Method*. Available at: <http://www.nova.edu/ssss/QR/QR2-4/pandit.html>
- Paulhus, D. L., and Williams, K. M. (2002). The dark triad of personality: Narcissism, Machiavellianism, and psychopathy. *J. Res. Pers.* 36, 556–563. doi: 10.1016/s0092-6566(02)00505-6
- Raskin, R., and Hall, C. S. (1979). A narcissistic personality inventory. *Psychol. Rep.* 45:590. doi: 10.2466/pr0.1979.45.2.590
- Rauthmann, J. F., Sherman, R. A., and Funder, D. C. (2015). Principles of situation research: towards a better understanding of psychological situations. *Eur. J. Pers.* 29, 363–381. doi: 10.1002/per.1994
- Reis, H. T. (2008). Reinvigorating the concept of situation in social psychology. *Pers. Soc. Psychol. Rev.* 12, 311–329. doi: 10.1177/1088868308321721
- Rich, P. (2012). Inside the black box: revealing the process in applying a grounded theory analysis. *Qual. Rep.* 17, 1–23.
- Robinson, S. L., and O'Leary-Kelly, A. M. (1998). Monkey see, monkey do: the influence of work groups on the antisocial behavior of employees. *Acad. Manag. J.* 41, 658–672. doi: 10.2307/256963
- Roczniewska, M., and Bakker, A. B. (2016). Who seeks job resources, and who avoids job demands? The link between dark personality traits and job crafting. *J. Psychol.* 150, 1026–1045. doi: 10.1080/00223980.2016.1235537
- Seivwright, A. N., and Unsworth, K. L. (2016). Making sense of corporate social responsibility and work. *Front. Psychol.* 7:443. doi: 10.3389/fpsyg.2016.00443

- Sharf, A., Kimonis, E., and Howard, A. (2014). Negative life events and posttraumatic stress disorder among incarcerated boys with callous-unemotional traits. *J. Psychopathol. Behav. Assess.* 36, 401–414. doi: 10.1007/s10862-013-9404-z
- Smith, M. B., Craig Wallace, J., and Jordan, P. (2016). When the dark ones become darker: how promotion focus moderates the effects of the dark triad on supervisor performance ratings. *J. Organ. Behav.* 37, 236–254. doi: 10.1002/job.2038
- Spain, S. M., Harms, P. D., and LeBreton, J. (2014). The dark side of personality at work. *J. Organ. Behav.* 35, S41–S60. doi: 10.1002/job.1894
- Squires, A. (2009). Methodological challenges in cross-language qualitative research: a research review. *Int. J. Nurs. Stud.* 46, 277–287. doi: 10.1016/j.ijnurstu.2008.08.006
- Strauss, A. L., and Corbin, J. M. (1998). *Basics of Qualitative Research: Techniques and Procedures for Developing Grounded Theory*, 2nd Edn. Thousand Oaks, CA: SAGE.
- Sturges, J. E., and Hanrahan, K. J. (2004). Comparing telephone and face-to-face qualitative interviewing: a research note. *Qual. Res.* 4, 107–118. doi: 10.1177/1468794104041110
- Suddaby, R. (2006). From the editors: what grounded theory is not. *Acad. Manag. J.* 49, 633–642. doi: 10.5465/AMJ.2006.22083020
- Sullivan, J. R. (2012). Skype: an appropriate method of data collection for qualitative interviews? *Hilltop Rev.* 6, 54–60.
- Suls, J., and Martin, R. (2005). The daily life of the garden-variety neurotic: reactivity, stressor exposure, mood spillover, and maladaptive coping. *J. Pers.* 73, 1485–1510. doi: 10.1111/j.1467-6494.2005.00356.x
- Temple, B. (2002). Crossed wires: interpreters, translators, and bilingual workers in cross-language research. *Qual. Health Res.* 12, 844–854. doi: 10.1177/104973230201200610
- Ten Berge, M. A., and De Raad, B. (1999). Taxonomies of situations from a trait psychological perspective: a review. *Eur. J. Pers.* 13, 337–360. doi: 10.1002/(sici)1099-0984(199909/10)13:53.0.co;2-f
- Van Nes, F., Abma, T., Jonsson, H., and Deeg, D. (2010). Language differences in qualitative research: is meaning lost in translation? *Eur. J. Ageing* 7, 313–316. doi: 10.1007/s10433-010-0168-y
- Vernon, P. A., Villani, V. C., Vickers, L. C., and Harris, J. A. (2008). A behavioral genetic investigation of the Dark Triad and the Big 5. *Pers. Individ. Dif.* 44, 445–452. doi: 10.1016/j.paid.2007.09.007
- Wilhelmy, A., Kleinmann, M., König, C. J., Melchers, K. G., and Truxillo, D. M. (2016). How and why do interviewers try to make impressions on applicants? A qualitative study. *J. Appl. Psychol.* 101, 313–332. doi: 10.1037/apl0000046
- Willig, C. (2009). *Introducing Qualitative Research in Psychology: Adventures in Theory and Method*, 2nd Edn. Maidenhead: Open University Press.
- Zhao, H., Zhang, H., and Xu, Y. (2016). Does the Dark Triad of personality predict corrupt intention? The mediating role of belief in good luck. *Front. Psychol.* 7:608. doi: 10.3389/fpsyg.2016.00608

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2017 Nübold, Bader, Bozin, Depala, Eidast, Johannessen and Prinz. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Communicating the Neuroscience of Psychopathy and Its Influence on Moral Behavior: Protocol of Two Experimental Studies

Robert Blakey^{1*}, Adrian D. Askelund², Matilde Boccanera³, Johanna Immonen⁴, Nejc Plohl⁵, Cassandra Popham⁶, Clarissa Sorger⁷ and Julia Stuhldreier⁸

¹ Centre for Criminology, University of Oxford, Oxford, UK, ² Department of Psychology, University of Oslo, Oslo, Norway, ³ Department of Psychology, King's College London, London, UK, ⁴ Psychology Unit, University of Helsinki, Helsinki, Finland, ⁵ Department of Psychology, University of Maribor, Maribor, Slovenia, ⁶ Department of Experimental Psychology, University of Oxford, Oxford, UK, ⁷ Division of Psychology and Language Sciences, University College London, London, UK, ⁸ Department of Psychology, Leiden University, Leiden, Netherlands

OPEN ACCESS

Edited by:

Rocio Del Pino,
University of Deusto, Spain

Reviewed by:

Daniel M. Barros,
University of São Paulo, Brazil
Mei Chang,
University of North Texas, USA

*Correspondence:

Robert Blakey
robert.blakey@crim.ox.ac.uk

Specialty section:

This article was submitted to
Quantitative Psychology
and Measurement,
a section of the journal
Frontiers in Psychology

Received: 08 November 2016

Accepted: 15 February 2017

Published: 14 March 2017

Citation:

Blakey R, Askelund AD, Boccanera M, Immonen J, Plohl N, Popham C, Sorger C and Stuhldreier J (2017) Communicating the Neuroscience of Psychopathy and Its Influence on Moral Behavior: Protocol of Two Experimental Studies. *Front. Psychol.* 8:294. doi: 10.3389/fpsyg.2017.00294

Neuroscience has identified brain structures and functions that correlate with psychopathic tendencies. Since psychopathic traits can be traced back to physical neural attributes, it has been argued that psychopaths are not truly responsible for their actions and therefore should not be blamed for their psychopathic behaviors. This experimental research aims to evaluate what effect communicating this theory of psychopathy has on the moral behavior of lay people. If psychopathy is blamed on the brain, people may feel less morally responsible for their own psychopathic tendencies and therefore may be more likely to display those tendencies. An online study will provide participants with false feedback about their psychopathic traits supposedly based on their digital footprint (i.e., Facebook likes), thus classifying them as having either above-average or below-average psychopathic traits and describing psychopathy in cognitive or neurobiological terms. This particular study will assess the extent to which lay people are influenced by feedback regarding their psychopathic traits, and how this might affect their moral behavior in online tasks. Public recognition of these potential negative consequences of neuroscience communication will also be assessed. A field study using the lost letter technique will be conducted to examine lay people's endorsement of neurobiological, as compared to cognitive, explanations of criminal behavior. This field and online experimental research could inform the future communication of neuroscience to the public in a way that is sensitive to the potential negative consequences of communicating such science. In particular, this research may have implications for the future means by which neurobiological predictors of offending can be safely communicated to offenders.

Keywords: psychopathy, belief in free will, utilitarian moral judgment, neuroscience communication, dishonesty, attributions, belief in determinism, self-control

INTRODUCTION

Since the time of Aristotle it has been argued that all human behavior can be described in terms of deterministic causality, and that there is no such thing as free will. Although philosophical arguments challenging free will have existed for centuries, these arguments do not appear to have filtered into the lay mind. However, there has been much recent lay interest in the rise of neuroscience as a means of explaining complex behaviors (Legrenzi and Umiltà, 2011). Therefore, one might predict that in the future, lay belief in free will could be challenged through the communication of neuroscience. Hence empirical research has begun to test whether people believe that free will could exist in a world where all events were products of brain activity (Nahmias et al., 2007).

This study is concerned with the behavioral implications of such beliefs. One of the greatest consumers of the neuroscience behind behavior could be people who might benefit from a neurobiological understanding of their mental condition. People may be especially receptive to neuroscience if the explanation is construed as a scientific means of excusing the socially disapproved symptoms of their condition. One such condition is psychopathy. Psychopaths have been shown to differ from ordinary people in both neurobiological and cognitive terms. For example, previous research has shown that psychopaths differ from lay people in moral dilemmas, such that they choose utilitarian reasoning more often. One focus of our study is therefore whether a neuroscientific explanation of typical psychopathic behavior will affect behavior in this sort of task, perhaps by excusing the behavior as not a result of free will.

In our field study, we will test whether lay people are more likely to return a postcard that contains a cognitive rather than a neurobiological explanation of criminal behavior, and whether they are more likely to return the postcard when it is directed to prisoners or non-prisoners. Subsequently, we will conduct an online study in which participants will be given false feedback about having above-average or below-average psychopathic traits; we are investigating the effects of communicating either a neurobiological or cognitive explanation of psychopathy on reasoning in moral dilemmas and behavior in a measure of actual cheating (Shalvi et al., 2012).

Behavioral Attributes of Psychopaths – the Lack of Empathy and Utilitarian Reasoning

Most established definitions of psychopathy emphasize two main characteristics of psychopaths: emotional impairment (e.g., reduced empathy and guilt) and behavioral disturbance (e.g., criminal activity) (Hare, 1991). Of particular importance to the current study, psychopathy is considered to be one of the prototypical disorders associated with empathic dysfunction, an absence of the appropriate empathic response to the suffering of another (Aniskiewicz, 1979; Hare, 1991).

The psychopath's lack of affective empathy plays an important role in moral reasoning. Many studies support a dual-process model of moral judgment (Greene et al., 2008), in which

both automatic emotional processes and controlled cognitive processes drive moral judgment. According to this theory, some moral judgments are driven primarily by social-emotional responses, while other moral judgments are driven less by social-emotional responses and more by cognitive processes (Greene et al., 2004).

Automatic emotional processes normally dominate for deontological decisions, while controlled cognitive processes drive utilitarian decisions (Duke and Bègue, 2015). This distinction is evident in moral dilemmas; a prototypical utilitarian favors performing actions in the name of the greater good, while a prototypical deontologist regards this actions as an unacceptable violation of rights and duties (Greene et al., 2008).

One such moral dilemma is the footbridge dilemma, in which a trolley threatens to kill five people, who can only be saved if you decide to push a stranger off the bridge, onto the tracks below. The stranger will die if you push him, but in the process, his body will prevent the trolley killing the five others (Thomson, 1985). Automatic emotional responses tend to drive people to disapprove of pushing the man off the footbridge, while controlled cognitive processes tend to drive people to approve of this action (Greene et al., 2008). Normally, in this particular dilemma, the automatic emotional response prevails; most people do not decide to push the man off the bridge (Greene et al., 2001). However, in the case of psychopathy, one would expect psychopaths to push the man given their lack of empathic concern. Studies indeed show that psychopathic personality characteristics, especially decreased levels of empathy, correlate with utilitarian choices (e.g., Bartels and Pizarro, 2011; Conway and Gawronski, 2013; Gleichgerrcht and Young, 2013).

Although it's relatively clear that there's a strong relationship between empathic concern and utilitarian reasoning, studies that actually measure the utilitarian reasoning of psychopaths are very scarce. In a recent study by Koenigs et al. (2012), psychopathic and non-psychopathic participants made judgments on 24 moral dilemmas. Results indicated that across all moral scenarios, psychopaths endorsed a significantly greater proportion of the proposed utilitarian actions than did the non-psychopaths. However, another recent study found no differences in utilitarian moral judgment between psychopaths and non-psychopaths (Cima et al., 2010). This lack of significant differences could be attributed to the smaller sample size and more lenient criteria for classifying participants as psychopaths (Koenigs et al., 2012).

These studies present participants with a variety of moral dilemmas, which can be distinguished by the extent to which the dilemma engages cognitive and affective processes respectively (Greene et al., 2001). The footbridge dilemma is considered a "personal dilemma"; it involves direct, intimate, physical contact (Greene et al., 2004). This type of dilemma engages emotional processing to a greater extent than other dilemmas (Greene et al., 2001). Previous studies show that some personal dilemmas, such as the footbridge dilemma, can be considered relatively easy (Wiegmann et al., 2013), while others can be considered more difficult; the latter bring cognitive and emotional factors into a very balanced tension. An example of a difficult personal dilemma is the crying baby dilemma, in which participants must decide whether it is appropriate to smother a child in order to save

oneself and other townspeople. In response to this dilemma, participants tend to answer more slowly and show less consensus.

In contrast to personal moral dilemmas, there are also impersonal dilemmas that involve more indirect, remote actions or rule violations (Greene et al., 2004) and engage emotional processing to a lesser extent (Greene et al., 2001). A classic example of an impersonal moral dilemma is the standard trolley dilemma (Foot, 1978), in which a runaway trolley is approaching five railway workmen and the only way to avoid their deaths is to hit a switch that will cause the trolley to change the path and kill one single workman instead.

Many previous studies have shown that *personal* moral dilemmas, like the footbridge dilemma, elicit increased activity in brain regions associated with emotion and social cognition (Greene et al., 2001). Mendez et al. (2005) found that patients with frontotemporal dementia, who are also known for their emotional blunting, were disproportionately likely to approve of the action in the footbridge dilemma. Koenigs et al. (2007) found similar results studying patients with emotional deficits due to ventromedial prefrontal lesions. A recent study by Koenigs et al. (2012) showed that only low-anxious psychopaths were significantly more likely to endorse personal harms in moral dilemmas. Compared to non-psychopaths, both types of psychopaths were significantly more likely to endorse the impersonal actions. The differences between low and high anxious psychopaths are less relevant to our study, but the findings of Koenigs et al. (2012) show that, in order to be thorough, this study should measure reasoning in both personal and impersonal dilemmas.

Hence, in our study, we will ask participants to complete three types of moral dilemmas: an easy personal (the footbridge dilemma), a difficult personal (the crying baby dilemma) and an impersonal (the standard trolley dilemma) dilemma; these tasks will form part of our dependent variables. At the start of the study, participants will be presented with one explanation regarding why psychopaths exhibit the low levels of empathy required to make utilitarian choices in these tasks. Importantly, only one of these explanations will refer to the neurobiological features of psychopathy in order that we can isolate the effect of making a biological attribution for the behavior. We will now consider, more broadly, the effect of describing mental conditions in biological terms.

Biological Attributions

Belief in biological explanations of behavior affects the perception of people suffering from a number of psychiatric disorders (Hyman, 2007). The comprehension of biological explanations of mental illnesses depends on the lay solution to the dualistic mind-body problem (Kendler, 2005). This raises the question of how lay people might view the brain relative to the mind and how this could influence the inferences that are drawn from neuroscience.

The effects of biological attributions represent a double-edged sword (Aspinwall et al., 2012). On one hand, biological explanations can have positive effects on lay conceptions of mental disorder. If the disorder is deemed biological, people may view sufferers of the disorder as less responsible for having the disorder, thereby blaming and stigmatizing sufferers to

a lesser extent (Corrigan and Watson, 2004). On the other hand, biological attributions may also have negative effects; a biological disorder may be viewed as less changeable, also as a result of the perception that biological causes are uncontrollable. Consequently, patients, their families and friends may be less likely to believe in the efficacy of treatment, thereby reducing any placebo effect of treatment (Angermeyer et al., 2011). Hence biological attributions represent a double-edged sword.

Lebowitz (2014) reviewed studies assessing the impact of biological explanations of mental illnesses. Observational studies indicate that individuals who ascribe their mental illness to biological causes are more pessimistic about the success of their treatment. Moreover, belief in biological explanations was often related to greater stigmatization, given the perception that biological disorders are unchangeable. In contrast, experimental studies suggest that pessimistic views about the success of treatment are reduced when people receive information about the changeability of biological components of illnesses. However, individuals who have a mental illness and believe in a biological explanation of that illness do not show reduced compliance with treatment programs (Lebowitz, 2014). Consequently, belief in biological explanation has an impact on how lay people perceive their own psychiatric disorders and on others, thus having an effect on perceived blame for the condition and thereby potentially influencing treatment success.

Advances in Science Communication

Today neuroscience appears to be particularly popular in the public eye as a means of explaining behavior. Indeed, evidence suggests that people find explanations of behavior more persuasive if those explanations feature circular references to the brain (Weisberg et al., 2008; Fernandez-Duque et al., 2015). Given its capacity to explain multiple aspects of the mind in a seemingly objective way, people have increasingly sought neuroscientific explanations of complex behaviors (Satel and Lilienfeld, 2015). The term 'neuromania' describes the tendency of the public to place greater faith in psychological explanations that are supplemented with references to the brain (Legrenzi and Umiltà, 2011). Given its power to draw attention to scientific explanations of behavior, neuroscience could indeed be presented in various professional settings, such as the criminal justice system.

In this regard, Greene and Cohen (2004) predict that neurobiological explanations of criminal behavior will, and should, change lay attributions of free will and moral responsibility to offenders by rendering the physical mechanisms of human behavior more visible. Indeed, our increasing knowledge of the behavioral consequences of deficits in brain regions implicated in decision-making, morality and empathy may 1 day be integrated into the criminal justice system (Umbach et al., 2015). In accordance with such reasoning, we believe that in the future, the criminal justice system will be informed by science that is far more advanced than currently exists. Specifically, we predict that one day offenders may receive direct personalized feedback regarding the presence or absence of cognitive, genetic and neurobiological predictors of different mental illnesses and criminal behaviors. This may be useful in multiple contexts, such

as prior to receiving a sentence in court and upon entering and departing prison grounds. For instance, criminal psychopaths could be shown how certain parts of their brain, specifically the limbic structures, exhibit less affect-related activity (Kiehl et al., 2001). Such procedures would grant offenders an understanding of the otherwise hidden scientific reasons behind their criminal behavior.

Given the practical and ethical issues implicated in measuring the response of real offenders to personalized scientific feedback, in our study, we are interested in analyzing how lay people respond to such feedback. In the current age of technology, social media has generated major new opportunities to analyze behavior online; in particular, by capturing the so-called 'digital footprints' left by millions of people on social networks. Using these sources of big data, researchers are generating opportunities for people to receive personalized data-driven feedback about their psychological and physical health. For example, Kosinski et al. (2015) analyzed the data of millions of Facebook users to create an algorithm capable of predicting users' gender, sexuality, age, personal interests and political views, only based on their Facebook profiles (including statuses, likes, etc). Such algorithms have also been used to identify the possible psychopathic traits of ordinary people (Garcia and Sikström, 2014).

The method of the current study is based on this idea that trait information can be inferred from an individual's Facebook profile. Specifically, participants will be given false feedback about having high or low psychopathic traits after entering their Facebook login details; the effect of providing such feedback on their moral behavior will then be measured. If individual scientific feedback is capable of changing the moral behavior of lay people, one might also expect this feedback to influence the moral behavior of offenders who receive such feedback in the future. Hence the findings of our study will pose implications for the real world, in which personalized neuroscience might one day influence how offenders are treated after trial, how offenders explain their own criminal behavior and therefore their own likelihood of reoffending (Maruna and Copes, 2005).

The Impact of Belief in Free Will on Behavior

Previous research has shown that attributions of free will can influence behavior on many different levels: studies have documented effects of belief in free will versus disbelief in free will on well-being (Crescioni et al., 2015), self-control (Rigoni et al., 2012), cheating (Vohs and Schooler, 2008), aggression (Baumeister et al., 2009) or conformity (Alquist et al., 2013). Therefore, belief in free will poses important implications for how people behave. Hence, we will begin this section by considering the behavioral consequences of adopting different perspectives on the causes of behavior, where neuroscience could induce a change in such perspectives.

Mindsets

In order to contextualize the hypothesized effects of attributions for psychopathy, we draw upon the analogy of attributions for intelligence, which have received far more empirical attention. In

the study of intelligence, two different views about the nature of intelligence have emerged: the view of intelligence as a fixed part of a person's personality that cannot be changed, and the view of intelligence as incremental (i.e., as always having the potential to be improved through exercise and effort). Dweck (1999) labeled these implicit theories as 'growth' and 'fixed' mindsets (or incremental theory and entity theory), and applied these theories to her research in self-theories, motivation, and personality.

A growth mindset refers to the belief that a person's abilities are not predetermined, but can develop, improve and change over time through practice. In contrast, the 'fixed mindset' implies that a person's abilities are static and cannot be changed as they are predetermined. These terms can be linked to the concepts of determinism and free will: a growth mindset implies the potential to change through the exercise of free will or a change in environments, whereas a fixed mindset implies belief in genetic and fatalistic determinism, such that any conscious motivation to change is futile.

Whether people believe in growth or fixed mindsets poses important implications for their behavior: inducing a growth mindset as compared to a fixed mindset greatly influences people's levels of intrinsic motivation (Dweck, 1999, 2006). Dweck's studies indicated, for example, that people who learnt about growth mindsets reacted in a far more positive way to failures than people who were taught about fixed mindsets. While those with a belief in the growth mindset used their failure as a reason and motivation to improve in the future, those with a belief in a fixed mindset reacted in a much more negative way. Specifically, those with the fixed mindset belief blamed others for their failure, made excuses or even became depressed; as they believed that their abilities were predetermined and could not be changed over time. Hence, it appears that the way in which people respond to feedback about their learning depends on the extent to which they perceive intelligence to be controllable.

Similarly, we hypothesize that the way in which people respond to feedback about their psychopathic traits depends on the extent to which they perceive psychopathic traits to be controllable. Participants will read either a neurobiological or a cognitive description of their psychopathic traits. We hypothesize that the neurobiological explanation of psychopathy will undermine the perception that psychopathic traits are controllable and therefore undermine the perceived moral responsibility of the participant. In the terms of Dweck (1999), we expect the neurobiological and cognitive attributions respectively to promote a fixed (uncontrollable) and growth (controllable) mindset toward psychopathic traits.

The effect of neuroscience communication on attributions of control and moral responsibility to the self has yet to be tested. Hence our predictions are based on the emerging body of research that considers the impact of neuroscience communication on attributions of moral responsibility to other people. Specifically, researchers have tested the effect of describing mental illnesses in neurobiological terms on the attributions of moral responsibility to criminal behaviors that are related to those illnesses. In mock court scenarios, people attribute less moral responsibility to an offender whose mental illness is described in neurobiological, rather than

solely cognitive, terms (Gurley and Marcus, 2008; Greene and Cahill, 2011; Schweitzer et al., 2011; Schweitzer and Saks, 2011; Aspinwall et al., 2012).

The net mitigating effect of neuroscience has also been found with real judges engaged in mock sentencing (Aspinwall et al., 2012) and real sentencing (Denno, 2015). Similarly, students recommend shorter prison sentences for a mock offender after taking a cognitive neuroscience module and after reading an article about brain stimulation or the neuroscientific predictors of conscious intent (Shariff et al., 2014). Collectively this research lends support to Greene and Cohen's (2004) prediction: people may recognize neurobiological dispositions to offend as undermining the culpability of offenders and their deservingness of punishment, unlike social dispositions to offend (Dar-Nimrod and Heine, 2011).

Researchers have considered the effects of presenting neuroscience not only as an explanation for the mental illness inflicting a particular defendant but also as a complete explanation of all behaviors in general (Nahmias et al., 2007). In this context, far fewer participants believed that people had free will (and could be held responsible) in the neurobiologically (relative to cognitively) determined world (38% vs. 85%, excluding responses of 'I don't know').

As replicated by Nahmias et al. (2005), the vast majority of participants continued to attribute responsibility to the cognitively determined actor, thereby demonstrating a 'compatibilist' perspective on free will: the philosophical position that people are morally responsible for their actions even if those actions are the inevitable outcome of a chain of preceding events (Kane, 1999). Hence, neuroscience may challenge belief in free will not by highlighting the chain of preceding causal events but by suggesting that, as a neurobiological phenomenon, the cause of behavior must be somewhat unconscious; somewhat beyond the control of conscious thought. This dualist perception of neurobiological phenomena as unconscious might grant neurobiological determinism greater opportunity to challenge belief in free will than cognitive determinism. In other words, neuroscience might challenge attributions of responsibility by reducing the perceived availability or causality of conscious cognition rather than by promoting belief in determinism. In respect to our study, therefore, we expect the neurobiological explanation of psychopathy to reduce belief in free will to a greater extent than can be explained by any corresponding increase in the acceptance of determinism.

Regardless of the mechanism, the findings of Nahmias et al. (2007) suggest that, for judgments of people in general, neurobiological causation is granted more exculpatory power than conscious causation. Our study seeks to extend this finding to perceptions of the self in particular, rather than people in general, by applying the theory of fixed and growth mindsets beyond attributions for intelligence to attributions for psychopathic traits. First, we expect a neurobiological explanation of psychopathy to promote a fixed mindset toward psychopathic traits; a perception of psychopathic traits as uncontrollable, unchangeable and therefore beyond the moral responsibility of the individual. Second and in contrast, we expect a cognitive explanation of psychopathy to promote a growth

mindset toward psychopathic traits; a perception of psychopathic traits as controllable, changeable and therefore within the moral responsibility of the individual. In order to support the hypothesis that neuroscience will reduce attributions of moral responsibility, we will now consider a proposed mediator of this relationship; that is the effect of neuroscience communication on how the mind and brain are perceived to relate to each other.

Dualism

Dualism and physicalism are the two opposing philosophical solutions to the problem of how the mind and the brain are connected. Dualism corresponds to the belief that mind and brain are separate, whereas physicalism assumes that the subjective experience of humans is a function of brain activity. Forstmann and Burgmer (2015) found that adults intuitively believe in mind-body dualism and that dualism is the default mindset of lay people. In the current study, we are interested in how communicating neuroscience might influence this default mindset and behaviors that are affected by dualist intuitions. Since neuroscientific explanations of human behavior assume that our thinking and thus the mind are represented in the brain, we predict that neuroscience communication could challenge intuitive lay belief in dualism.

There is some evidence that whether people believe in physicalism or dualism poses implications for their choices in real-life. Specifically, Forstmann et al. (2012) considered the impact of dualist beliefs on health behaviors: participants who were primed with dualistic beliefs reported less commitment to healthy behaviors and made less healthy real-life decisions compared to participants primed with physicalism. Although Forstmann et al. (2012) observed that priming physicalist beliefs promoted healthy behaviors, we predict that physicalism would actually promote immoral behavior. Their study only documents the effect of dualism on health behaviors rather than moral behaviors: eating unhealthy food does not represent an act of aggression toward oneself and choosing a healthy lunch is not a moral behavior even though it has implications for one's well-being. Forstmann et al. (2012) reason that physicalist beliefs promote health behaviors through their implication that the state of the body influences the state of the mind. We do not expect physicalist beliefs to promote moral behaviors in this way, since the behaviors measured in our study – cheating and utilitarian reasoning – bear no implications for bodily health.

Nevertheless, there is another mechanism by which physicalist beliefs might influence moral behavior. This mechanism concerns the potential relationship between dualistic beliefs and belief in free will, where the latter has been found to influence various forms of behavior linked to morality; those are self-control (e.g., Rigoni et al., 2012), cheating (e.g., Vohs and Schooler, 2008), aggression (e.g., Baumeister et al., 2009) and conformity (Alquist et al., 2013).

There are two mechanisms by which physicalist beliefs could challenge belief in free will. First, the perception of the mind as brain activity might highlight the causal chain of events that generates any behavior: people may more readily represent brain activity as a closed loop, in which present brain activity is the necessary and sufficient result of preceding brain activity

in an unbreakable and inevitable chain of events. In contrast, people may more readily represent mental activity, in which present thoughts are not the necessary and sufficient result of previous thoughts. In other words, physicalism, as promoted by neuroscience, may illustrate the philosophy of determinism more effectively than the perception of mental activity independent of brain activity. Second, the perception of the mind as brain activity might bolster the belief that the mind – or cognitive influences on behavior – are largely unconscious and therefore beyond the control of conscious thought. Given the perceived compatibility of cognitive, yet not neurobiological, determinism with free will (Nahmias et al., 2005), we predict the second mechanism to constitute the means by which dualistic beliefs are reduced in the current study.

In their study, Forstmann et al. (2012) report preliminary data indicating that measures of mind-body dualism, free will and determinism are largely uncorrelated. We find this result most surprising and in fact predict a positive relationship between beliefs in dualism and free will. If the brain is conceived to constitute the mind, causal influences may subsequently appear to exert their effects beneath the scrutiny of conscious awareness. Hence we expect belief in physicalism to undermine belief in free will. Likewise, ‘libertarian views about free will [, that is belief in an independent free will, are] . . . likely rooted in some kind of dualism about mind (or soul) and brain’ (Kolber, 2016, p. 8). Therefore, we conclude that neuroscience could promote immoral behavior by undermining lay belief in dualism, the causal contribution of conscious thought and consequently free will; hence, we now consider the effects of belief in free will on immoral behavior.

Cheating

In initiating this line of research, Vohs and Schooler (2008) investigated the relationship between belief in determinism and cheating behavior. As hypothesized, reading a passage on neurobiological determinism and the non-existence of free will by Crick (1994) led to a significant increase in cheating as compared to the control group. The findings were replicated in a second study that measured a more proactive form of cheating. However, the results failed to replicate in a third study that was part of the collaborative ‘Estimating the Reproducibility of Psychological Science’ project (Open Science Collaboration, 2015).

While cheating will also be measured in our study, we intend to use a far less explicit means of manipulating belief in free will than previous research. Specifically, we intend to manipulate belief in free will by giving participants either a neurobiological or a cognitive explanation of psychopathic traits. This approach extends beyond previous research by separating the two phenomena of determinism and free will rather than conflating them, as was common in previous manipulations (e.g., Crick, 1994). The manipulation in our study is also more representative of the means by which lay belief in free will could be challenged in the future. People will arguably be informed increasingly about neuroscience not only in the media but also in the use of neuroimaging. This could help to inform individuals about their neurobiological health and to modify brain states using neurofeedback and brain stimulation.

There is also reason to believe that people will be persuaded more by neuroscience than the personalized cognitive feedback that they receive from self-assessment questionnaires today and the philosophical arguments presented in previous research (Greene and Cohen, 2004).

In fact, studies have shown that psychological information appears to be more appealing and salient if accompanied by additional, and frequently superfluous or irrelevant, neuroscientific explanations (Weisberg et al., 2008). This neuroscientific bias is due to lay theories and reverence for the natural sciences that consequently are regarded more than social science explanations (Fernandez-Duque et al., 2015).

The current study measures cheating using the ‘die-under-cup’ task (taken from Shalvi et al., 2012), where people can reap benefits by misreporting the outcome of a die roll. Certain factors including number of times the die is rolled, the outcomes of other rolls, and time pressure, have been shown to increase dishonesty in this die-roll test (Shalvi et al., 2011, 2012; Gino and Ariely, 2012; Lewis et al., 2012). For our research, the die-under-cup paradigm is adapted to suit into an online questionnaire and to include conditions that increase likelihood of dishonesty.

Aggression and Helpfulness

Baumeister et al. (2009) investigated Vohs and Schooler’s (2008) findings further by assessing the effects of belief/disbelief in free will on pro- and anti-social behavior in three experiments. In their research, disbelief in free will increased aggression and reduced helpfulness, while belief in free will resulted in more pro-social behavior such as the willingness to help. One might speculate therefore that promoting belief in free will generates a greater sense of personal responsibility and accountability for one’s actions, which arguably promotes socially desirable behavior. The finding that belief in free will motivates pro-social behavior is particularly relevant to our research, since we will test the effects of communicating neuroscientific explanations of psychopathy on the moral behavior of lay people.

Self-control

Theoretically, telling a person that free will does not exist (directly or indirectly) could lead to that person being less willing or able to exercise self-control, which might actually explain the effects of disbelief in free will on cheating and aggression. If you believe that you can not control your life in any ultimate way, you may feel that there is no point in trying to control each of your actions, including impulses to act immorally. Several studies now confirm the idea that belief in free will is linked to self-control, both when operationalized at the levels of conscious perceptions and preconscious neural activity. In one study, weakening belief in free will reduced both perceived self-control and intentional inhibition (Rigoni et al., 2012). The authors interpreted these results as indicating that reduced self-control could be the mechanism by which disbelief in free will leads to antisocial tendencies.

The finding that disbelief in free will reduces self-control has also been documented at the level of basic neurocognitive processes. In one study, inducing disbelief in free will attenuated neural reactions to error, which are implicated in the very early

phases of exerting self-control (Rigoni et al., 2015). Moreover, brain correlates of preconscious motor preparation were shown in the first study to be altered by inducing a belief in determinism, as compared to a belief in free will (Rigoni et al., 2012). In the context of the current study, self-control at the behavioral level will be included as a potential mechanism by which the manipulation influences moral reasoning and cheating.

Conformity

While disbelief in free will may reduce self-control, it may increase social control; that is the influence that other people have on the behavior one exhibits. Indeed, research by Alquist et al. (2013) has shown that independently, less belief in free will and greater belief in determinism resulted in greater conformity to the judgments of other participants. It was suggested that a belief in free will contributes to more autonomous decisions and actions and therefore less conformity (to group norms).

This finding bears relevance to our online study, since participants will be provided with a supposedly scientific judgment about their degree of psychopathic traits. Different participants may conform to this judgment of themselves to differing extents; some participants may exhibit the psychopathic tendencies that they are described as having, while others may not. Given its expected effect on belief in free will, the neurobiological explanation of psychopathy might promote the conformity of participants to the psychopathy feedback. In contrast, since we do not expect the cognitive explanation of psychopathy to challenge belief in free will, participants who read this explanation may conform less to the feedback about their degree of psychopathy. Therefore, by reducing belief in free will, neuroscience may increase the receptivity of participants to external opinions, including the personalized science that we present. Hence, the persuasiveness of the opinion represents an additional factor that could explain the greater effect of neuroscience. We intend to capture and control for this effect by measuring the perceived believability of the presented explanations of psychopathy; the neurobiological explanation is hypothesized to be more believable.

Summary and Hypotheses

Considering all of the above, there are three ways in which our study will add to the literature in this field. First, we will be testing the effects of specifically presenting neuroscience to lay people, rather than a generic passage about free will and determinism (e.g., Vohs and Schooler, 2008). Second, we will be looking at the effects of presenting neuroscience to explain a particular set of traits – psychopathic traits – among lay people rather than presenting explanations of a mental illness in a clinical population (see Lebowitz, 2014). Third, our study will examine the effects of providing personal feedback about psychopathic traits that was allegedly generated from a digital footprint (i.e., Facebook ‘likes’) rather than a survey measure of psychopathy.

The field study and the online study bear relations to each other, since our field study will test whether the public are sensitive to the hypotheses we propose for the online study. While the online study tests how the communication of the basis of psychopathy affects moral behavior, the field study is intended to

capture the general public’s attitudes toward this communication. This will be done using the lost-letter technique, comparing return rates of postcards describing neurobiological or cognitive explanations of criminal behavior intended for prisoners or non-prisoners. We hypothesize that people will be sensitive to the potential negative behavioral consequences of communicating a neurobiological explanation of criminal behavior, as reflected by reduced return rates of the postcards. Specifically, we predict the return rate, indicating endorsement of the postcard’s content, to be higher for the cognitive (than neurobiological) explanation (Hypothesis 1) and higher in the non-prisoner (than the prisoner) condition (Hypothesis 2), and that these effects will interact (Hypothesis 3). Lay people may anticipate that neurobiological explanations of behavior undermine attributions of responsibility and hence seek to avoid the communication of neuroscience to offenders.

In comparison, the online study will measure whether this anticipation is justified; specifically, whether feedback about the neurobiological or cognitive psychopathic traits (specifically the strength of their moral alarm) of the participant influences utilitarian reasoning in moral dilemmas and dishonesty in a die-under-cup test, and whether this is mediated by self-control, and beliefs in dualism, free will and determinism. We hypothesize that participants who are led to believe they have a weak moral alarm (associated with higher levels of psychopathy) will act in ways consistent with psychopathic tendencies, i.e., use more utilitarian reasoning and cheat more (Hypothesis 4), especially after reading a neurobiological explanation of psychopathy (Hypothesis 5). Our final hypothesis (Hypothesis 6) is that our manipulation will influence self-control and belief in dualism, free will, and determinism, and that these will mediate the relationships outlined in Hypotheses 4 and 5.

STUDY 1 – FIELD STUDY

Materials and Equipments

The Lost Letter Technique

The lost letter technique (LLT) was first adopted by Merritt and Fowler (1948) as a means of assessing the public’s attitudinal approach to an undelivered letter (Stern and Faber, 1997). By distributing a large number of apparently lost letters referring to a particular topic, the return rate of the letters can be used to measure the public’s compliance with such issue (Milgram et al., 1965). This method has been deemed as valid and can be implemented conveniently: participants are unaware of their participation in this unique sociological survey, whereby natural behaviors are recorded, possibly reflecting concrete attitudes (Milgram et al., 1965; Cahill and Sherrets, 1979). This technique will be used to evaluate the public’s approval of disseminating the neuroscience of criminal behavior to both lay people and prisoners.

Stepwise Procedures Participants

Data will be collected from the responses given by a convenience sample of participants, whereby no recruitment or selection

criteria is required. Therefore, age and gender and other individual factors cannot be selected. Those who decide to pick up a lost postcard and either mail, ignore or purposely destroy it will be considered participants (Milgram et al., 1965). As this field study is non-obtrusive, the number of participants taking part in the study cannot be determined. However, 832 postcards will be scattered around the city streets, thus authorizing approximately the same number of people to unconsciously take part in the study. This sample size is sufficiently large given the moderate response rates recorded by prior research; for example, from 37% in poorer neighborhoods to 87% in richer neighborhoods (Holland et al., 2012).

Ethics Statement

The study has been approved by the Ethics Committee of the University of Oxford, and is fully compliant with the Declaration of Helsinki.

Design and Procedure

A total of 832 printed, stamped and addressed postcards will be dropped throughout London. This large number of postcards will be dropped to increase the probability of gathering a large number of participants, thus increasing the sensitivity of the measure to the independent variable and the reliability of the obtained results (Milgram, 1969; Cherulnik, 1975). The postcards will be distributed face-up in proximity of parked cars, in shops and on pavements throughout a random selection of London boroughs.

Boroughs of London

The 832 postcards will be distributed in boroughs of London with different socio-economic status (SES) by four members of the research team. The SES of the borough will be calculated from the combined average degree of inequality, homelessness, housing quality, unemployment, income, benefits, and education (Trust for London and New Policy Institute, 2017). Within each of the four categories of SES, postcards will be distributed in the following boroughs:

- Poorest: Barking and Dagenham, Newham, Brent, Ealing
- Poor: Enfield, Haringey, Waltham Forest, Lewisham
- Rich: Hackney, Southwark, Tower Hamlets, Croydon
- Richest: Islington, Lambeth, Camden, Kensington and Chelsea, Merton

Distribution Process

Two hundred and eight postcards will be distributed in the poorest, 208 in poor, 208 in rich, and 208 in the richest boroughs. This decision has been made with the aim of reducing the probability of reaching a floor or ceiling effect of the manipulation: if the return rate is already high or low as a result of the borough SES (Holland et al., 2012; **Table 1B**; Kraus and Keltner, 2013), the scope for our manipulation to exert a supplementary effect may be limited. Therefore, we will distribute the postcards in boroughs of differing SES to ensure there remains this scope for the manipulation, while also increasing the generalisability of the findings from a more representative sample of boroughs.

The distribution will take place on 4 days (Monday, Tuesday, Thursday, and Friday) across four different time slots, such that each of the distributors will drop 208 postcards (52 per day of all four types of postcards). Each distributor will rotate through the four conditions, such that postcard type B is dropped after type A, C after B, D after C, A after D etc. On Monday the distribution will take place at 9–11 am, on Tuesday at 11 am–1 pm, on Thursday at 1–3 pm and on Friday 3–5 pm. A day of distribution will, however, be skipped until the following weekday if it is raining, since rainy weather could severely reduce the response rate. The distributors will drop the postcards on the same days and at the same times so that any effects of external factors between boroughs (e.g., the weather, time of day) are minimal. Thus, one person will distribute the postcards in one of the poorest boroughs, the second person in a poor borough, the third person in a rich borough, and the fourth person in one of the richest boroughs. On the second day the first person will distribute the postcards in a poor borough, the second person in a rich borough, the third person in one of the richest boroughs, and the fourth person in one of the poorest boroughs etc. Consequently, every distributor will drop postcards in every category of SES. No borough will be visited twice.

In the process of distribution the distributors will drop the postcards not too close to each other so that one person will not find two postcards. Furthermore, the distributors will drop the postcards in place that are visible and accessible to the general public so that the postcards can be found easily. In addition, the postcards will be dropped carefully and secretly.

Content

The postcards will be addressed to a PO BOX address (to avoid the use of a real traceable address), with the manipulation bolstered by supplementing the first line of this address with the ‘Organisation for Educating Prisoners/Students in Crime.’ All postcards will have the same front cover, as to avoid different images or colors biasing the participant’s subsequent response. An exception will be made for the front-cover wording, in so far as brain-based postcards will contain the emboldened word “brain,” and mind-based ones will contain the emboldened word

TABLE 1A | Conditions of the field study.

	Non-prisoner	Prisoner
Brain-based	208	208
Mind-based	208	208

TABLE 1B | Total number of dispersed postcards in London boroughs.

	Poor	Mildly poor	Mildly rich	Rich	Total
Brain-based to non-prisoner	52	52	52	52	208
Brain-based to prisoner	52	52	52	52	208
Mind-based to non-prisoner	52	52	52	52	208
Mind-based to prisoner	52	52	52	52	208
Total	208	208	208	208	832

“the person”. Thus far, research has documented only that adults do not perceive specific mental traits (e.g., memory) to be entirely physical (Forstmann and Burgmer, 2015). This suggests that people perceive ‘the person’ to consist of physical and non-physical causes of behavior. It remains possible, however, that the same people still equate the linguistic labels of ‘mind’ and ‘brain’. So people may be more dualistic in their implicit beliefs (when judging specific traits) than in their explicit beliefs (when judging ‘the mind’ as a concrete label). Since this study intended to test the effects of dualistic reasoning, the manipulation was designed to engage implicit beliefs about the person as a whole, rather than explicit beliefs about ‘the mind’. Hence the manipulation was oriented around ‘the person’. Therefore, this difference will guide subjects in understanding the explanations given in the postcard.

The main body of the postcard will comprise a brief description detailing the causes of criminal behavior, written by an imaginary person who has supposedly bought the postcard as part of a scientific campaign, the latter aiming to spread a particular message about the causes of criminal behavior. Half of the postcards will present a neurobiological (brain-based) explanation of criminal behavior (see Appendix), while the rest will present a cognitive (mind-based) explanation (see Appendix). Additionally, 50% of the postcards will be directed to prisoners (see Appendix), while the remaining half will be directed to non-prisoners (see Appendix). In both cases, the recipient will be an alleged friend of the writer. The writer will ask his friend to pass the postcard onto the ‘prisoners’ or (non-imprisoned) ‘students’ that he supposedly teaches. By comparing the response rates of all four conditions, one may infer different evaluations and conclusions about how neuroscientific and cognitive descriptions of criminal behavior influence the public’s decision to spread such information.

Consequently, participants will be randomly divided into four conditions, depending on the type of information and recipient reported on their postcard (Table 1A). As a result, a total of 208 postcards will be dispersed for each condition. This sample size was selected on the basis of power analysis assuming a normal distribution of the data (the power calculator we used can be found at <https://www.stat.ubc.ca/~rollin/stats/ssize/n2.html>). In this independent-samples analysis, we set the probability of Type I error to 5% and the probability of Type II error to 20%, and assumed that the effect size would be small (Cohen’s $d = 0.2$).

Proposed Analysis and Anticipated Results

Statistical analysis will involve analyzing the response rates of all four conditions and so binary data (did not return the postcard = 0; returned the postcard = 1) will be collected. We expect to obtain an explanation effect supporting Hypothesis 1, by which a larger number of mind-based postcards than brain-based ones will be posted. A chi-square test of association will determine any significant differences between observed and expected response rates: χ^2 (1, $N =$ number of returned postcards) > 3.841 , $p < 0.05$ (Tables 2A,B). Furthermore, an effect of the recipient is predicted, whereby we expect to receive a larger number of postcards addressed to non-prisoners than prisoner, supporting Hypothesis 2 (Tables 3A,B). A second chi-square test of association will be carried out to assess whether

TABLE 2A | Expected frequencies of returned postcards due to given explanation.

Chi-square test of association: Expected frequencies (hypothetical $N = 400$) in returned mind-based and brain-based postcards

	Mind-based	Brain-based	Total
Returned postcards	200	200	400

TABLE 2B | Observed frequencies of returned postcards.

Hypothetical Chi-square test of association: Observed frequencies (hypothetical $N = 400$) in returned mind-based and brain-based postcards

	Mind-based	Brain-based	Total
Returned postcards	300	100	400

TABLE 3A | Expected frequencies of returned postcards due to receiver.

Hypothetical Chi-square test of association: Expected frequencies (hypothetical $N = 400$) in returned non-prisoner and prisoner addressed postcards

	Non-prisoner	Prisoner	Total
Returned postcards	200	200	400

TABLE 3B | Observed frequencies of returned postcards due to receiver.

Hypothetical Chi-square test of association: Observed frequencies (hypothetical $N = 400$) in returned non-prisoner and prisoner addressed postcards

	Non-prisoner	Prisoner	Total
Returned postcards	300	100	400

such difference in returned postcards exists: χ^2 (1, $N =$ number of returned postcards) > 3.841 $p < 0.05$. In particular, an interaction effect is expected, whereby we expect the effect of the recipient to be particularly strong for the brain-based postcards (Hypothesis 3). Therefore, we predict a larger difference between non-prisoners and prisoners in the brain-based condition than in the mind-based condition (Tables 4A,B). The software IBM SPSS Statistics Version 23.0 will be employed.

Limitations

The LLT has a number of limitations. First, one might question whether the technique is sufficiently sensitive to document subtle manipulations. For the manipulation to be successful, participants must attend to the address and text on the postcard and their decision to return the postcard (or not) must reflect their approval of this specific message. The participants, however, might not pay sufficient attention to the manipulation.

Second, even more so that the online study, the field study cannot document the mechanisms that mediate the decision to return the postcard. Whilst we interpret the return rate as indicating the degree of acceptance for the presented explanation of offending, the return rate will also be sensitive to unpredictable events (e.g., street cleaners who throw away

TABLE 4A | Interaction effect of expected frequencies of returned postcards.

Hypothetical Chi-square of contingency tables: Expected frequencies (hypothetical $N = 400$)			
	Non-prisoner	Prisoner	Total
Mind-based	100	100	200
Brain-based	100	100	200
Total	200	200	400

TABLE 4B | Interaction effect of observed frequencies of returned postcards.

Hypothetical Chi-square of contingency tables: Observed frequencies (hypothetical $N = 400$)			
	Non-prisoner	Prisoner	Total
Mind-based	215	85	300
Brain-based	85	15	100
Total	300	100	400

different numbers of postcards of different conditions). We aim to overcome this problem by implementing a strict plan for dropping the postcards: each distributor will rotate between dropping cards from all four conditions in boroughs of every SES category at all dropping times, spread across the day (see distribution of postcards). Nevertheless, the LLT has been shown to be reliable (Milgram et al., 1965; Cherulnik, 1975). In addition, we have chosen to feature relatively extreme statements on the postcards in order to strengthen our manipulation. Thus this explicit manipulation may be strong, especially since participants will be unaware of their participation, thereby removing potentially overshadowing Hawthorne effects.

STUDY 2 – ONLINE STUDY

Materials/Equipment

Neurobiological vs. Cognitive Manipulation

Our manipulation was adapted to focus on a neurobiological vs. cognitive understanding of psychopathy, based on a study by Aspinwall et al. (2012) where the explanation of psychopathy was drawn from James Blair's neurocognitive model (Blair, 2006). We removed any direct references to genetics from the original stimuli to increase the scientific equivalence of the two explanations. The neurobiological details in the brain-based explanation were deliberately superfluous; in reality, these details contributed very little substance to the argument. This decision was based on findings that superfluous neuroscience increases the perceived credibility of psychological science, even when the neuroscience itself is a circular repetition of the psychological science (Weisberg et al., 2008). Here, we present the material for the two conditions in the same paragraphs, emphasizing the equivalence of the conditions independent of the manipulation:

The brain's/mind's moral alarm

Here, we present the material for the two conditions in the same paragraphs, emphasizing the equivalence of the conditions independent of the manipulation: Extensive research shows that human brains/minds have a moral alarm. The moral alarm is the physical/psychological system that produces feelings of anxiety when you behave badly. When humans behave badly, their brain/mind normally generates particular electrical signals and chemical reactions/thoughts and emotions that produce feelings of anxiety. The purpose of this anxiety is to physically/psychologically reduce your desire to behave badly."

Your brain/mind

We would now like to tell you more about people like you, who have an 18-22% stronger/weaker moral alarm than the average person.

The moral alarm is the physical/psychological system in the brain/mind that produces feelings of anxiety when you behave badly. The purpose of this anxiety is to physically/psychologically reduce your desire to behave badly. Since your moral alarm is 18-22% stronger/weaker than the average moral alarm, you are 18-22% less/more likely to behave badly than the average person. This is true of anyone with an 18-22% stronger/weaker moral alarm.

People have moral alarms of different strengths because of physical/psychological differences in how their brains/minds work. When people with a brain/mind like yours behave badly, their brain/mind generates more/less of the electrical signals and chemical reactions/thoughts and emotions that produce feelings of anxiety.

Therefore, people with a brain/mind like yours feel 18-22% more/less anxious when they behave badly. Consequently, people with a brain/mind like yours are 18-22% less/more likely to behave badly.

The Short Dark Triad Scale

The Short Dark Triad scale (SD3; Jones and Paulhus, 2014) is a brief measure of three socially aversive traits – Machiavellianism, narcissism and psychopathy. The whole scale normally consists of 27 items, rated on a five-point scale from 1 (disagree strongly) to 5 (strongly agree). As we are only interested in one element of the dark triad constellation, psychopathy, we will only use the psychopathy subscale of this instrument. This subscale includes 9 items (e.g., "Payback needs to be quick and nasty") and provides an efficient, valid and reliable way of measuring psychopathy, with Cronbach's alpha ranging somewhere from 0.77 to 0.79 (Buckels et al., 2014; Jones and Paulhus, 2014). This scale will be used to assess the participants' real psychopathic traits.

The Dualism Scale

We will measure dualistic beliefs with a modified version of the thought experiment used by Forstmann and Burgmer (2015). Participants are asked to imagine that scientists have developed a device capable of duplicating any person in a matter of seconds, using highly advanced technology. Participants are told that after placing a person into a chamber, a computer scans the entire person (i.e., the entire content of the chamber), its every molecule and atom, and stores the information digitally. The information

is then used to recreate the scanned person from basic chemical elements in a second chamber, resulting in a 100% identical copy of the scanned object, with a 100% success rate. In contrast to the original task, our participants will be asked to imagine that they are placed in the first chamber and are duplicated. After the process is complete and a 100% perfect duplicate emerges, the participants will indicate on 7-point Likert-type scales ranging from 'definitely no' to 'definitely yes' the extent to which six properties of themselves also describe their duplicate. Three of the properties will be mental and relate to the manipulation text, e.g., "Is the moral alarm in your duplicate the same strength as the moral alarm in you?". The remaining three items will be physical, e.g., "Does your duplicate have the same eye color as you?". If people do separate minds from bodies, there will be a difference in the mental and physical properties ascribed to the duplicate.

The Determinism Subscale

The Determinism subscale of the Free Will Inventory (Nadelhoffer et al., 2014) consists of five items that make different deterministic statements. For example, "Every event that has ever occurred, including human decisions and actions, was completely determined by prior events." Participants are asked to rate their agreement on a seven-point Likert rating scale with anchors ranging from 1 (strongly disagree) to 7 (strongly agree). The Determinism subscale has an acceptable to good internal consistency, with Cronbach's $\alpha = 0.772$ (Nadelhoffer et al., 2014).

The Free Will Subscale

The Free Will subscale of the Free Will Inventory (Nadelhoffer et al., 2014) consists of five items stating in different ways that free will exists. For example, one of the items states that "People ultimately have complete control over their decisions and their actions." The scale was chosen over the FAD+ scale (Paulhus and Carey, 2011) because it avoids religious terms. Participants are asked to score each item on a scale from 1 (strongly disagree) to 7 (strongly agree). This subscale has a good internal consistency, with Cronbach's $\alpha = 0.803$ (Nadelhoffer et al., 2014).

Die-under-Cup Measure of Dishonesty

Dishonesty/cheating will be measured using an online version of the die-under-cup test (Shalvi et al., 2011). Participants will be asked to press the 'next page' button to roll a virtual die within the online questionnaire in place of the physical die and cup. The 'die' will be rolled three times, the results of which will be fixed to show a two, a six, and a three respectively. Participants will report the first outcome by typing the number into a box and this response must be made within a 30 s window. The die-under-the-cup appears to be a valid measure of dishonesty, as found in a study conducted by Halevy et al. (2013), whereby high scores on this task are caused by the participant cheating rather than luck.

Crying Baby Dilemma

The crying baby dilemma (Greene et al., 2001) involves participants deciding how to behave with their child when enemy soldiers have taken over their village. In order to save their own lives and every village townspeople's lives, they must smother their crying baby to death, in order to avoid the attention of

the enemy soldiers. Alternatively, saving the child would mean putting the whole village at risk and letting all townspeople face death.

Standard Trolley Dilemma

The standard trolley dilemma (Foot, 1978) involves the participant being at the wheel of a runaway trolley. The latter is approaching a track, at the end of which five railway workmen are standing. Participants are given the option to switch a lever on the dashboard so that the trolley proceeds off toward a right-hand track, where only one workman is standing. The participant is left to decide whether to take no responsibility for the situation and let the trolley proceed straight toward the five men, or change the trolley's direction in order to save as many lives as possible.

Footbridge Dilemma

In the Footbridge dilemma (Thomson, 1985) a personal moral violation can be authorized in order to justify a precise utilitarian reasoning (Valdesolo and DeSteno, 2006). Individuals are presented with a scenario in which a trolley is moving toward five workmen who have no way to escape. The participants are asked to imagine that they are on a footbridge next to a large stranger, whom they may push off the bridge in order to stop the trolley from hitting the five workmen. By doing so, only one person would be actively killed and five people would be saved. Through a replication of the study by Greene et al. (2001), this task has been demonstrated to measure a different construct to personal moral dilemmas (Nakamura, 2013).

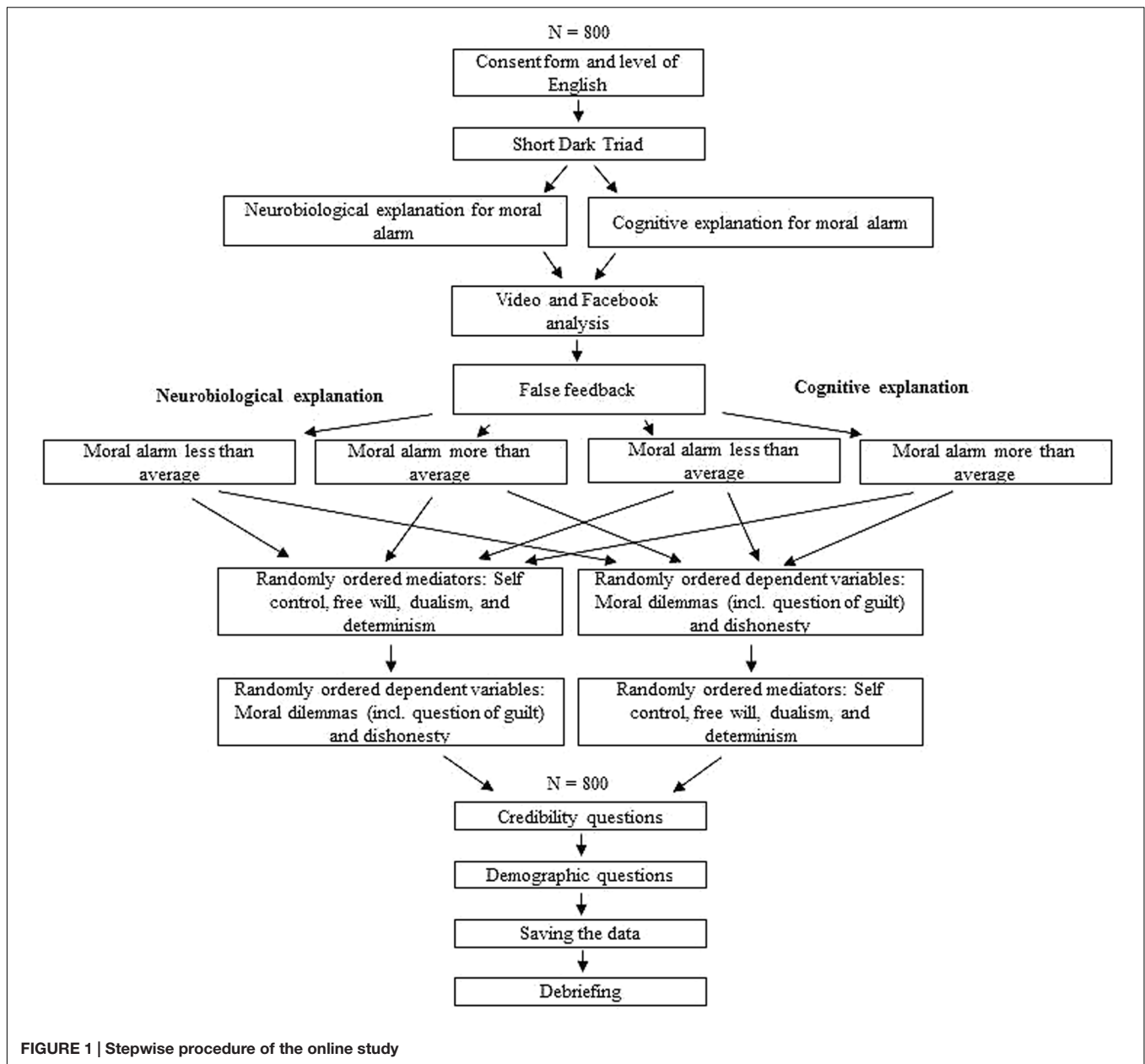
Utilitarian reasoning will be assessed by administering participants all three moral dilemmas.

Stepwise Procedures Participants

Participants will be recruited through adverts for the study posted on social media sites, such as Facebook. We will also contact universities in Austria, Germany, Italy, Slovenia, Finland, Norway, and the UK to promote the online study to their English-speaking students. Therefore, both students and members of the public aged 18 years of age or older will be able to take part in our study. The participants must be able to speak and comprehend English in order to fully understand all the information presented to them. Hence, we will ask participants to rate their English competence before completing the study. As we expect a small effect size, we aim to recruit at least 800 participants. This sample size was selected on the basis of power analysis. In this analysis, we set the probability of Type I error to 5% and the probability of Type II error to 20%, and assumed that the effect size would be small (Cohen's $d = 0.2$). Upon completing the study, participants will be entered into a lottery, giving them the chance to win a sum of money ranging from €75 – €200.

Ethics Statement

The study has been submitted to the Ethics Committee of the University of Oxford, and is fully compliant with the Declaration of Helsinki.



Design and Procedure

Participants will be asked to complete a series of online tasks in a single session (see **Figure 1**), administered via the Qualtrics platform. The online study will be divided into three sections, in order to facilitate the completion and understanding of the study. They will firstly be instructed that the purpose of this study is to investigate different ways of measuring personality traits, thus comparing traditional and newly developed means of measuring a certain psychopathic trait, specifically normal variation in the anxiety people feel when committing an immoral act. In addition, an abbreviated online version of the Short Dark Triad (SD3) will be administered to determine participants' psychopathic traits. We will therefore be able to compare the influence of

actual psychopathy with the influence of the false feedback about psychopathy presented to each participant.

In the second part of the online study, participants will be given either a cognitive or a neurobiological explanation of moral alarm; that is the anxiety produced during immoral behavior. This description will explain the function of the moral alarm in producing feelings of anxiety when people behave badly. A single multiple choice question will be included at the end of the description, in order to make sure that participants are reading thoroughly the online questionnaire and that our manipulations are effective. Participants will also be informed of why it is difficult to assess the moral alarm through a self-report questionnaire, thus justifying the purpose of the Facebook

analysis. Again, a single multiple choice question will assess their understanding of the information given.

As a way of persuading participants that assessing moral alarm through online personal data is valid and reliable, they will all be asked to watch an online video reviewing Kosinski's research (Kosinski et al., 2013) into the prediction of personality traits from digital footprints. The video will consider how researchers can predict personality traits, intelligence, ethnicity, political views and in particular psychopathic traits, by simply taking into account Facebook likes. A question measuring their accuracy in comprehending the text given to them will also be included. Accordingly, participants will be asked to provide a shortened link (through URL shortener) of their Facebook account for the purpose of analyzing their Facebook likes. None of the entered login details will be saved.

After entering their shortened URL to their Facebook page, participants will receive false feedback about their psychopathic traits. Note these traits will be described without actually referring to psychopathy in order to avoid triggering the popular negative perception of psychopathy. Specifically, participants will be randomly allocated to read one of four types of feedback: half of the participants will read that they have a 18–22% weaker-than-average moral alarm, while the other half will read that they have a 18–22% stronger-than-average moral alarm. Additionally, within each of these groups, half of the feedback messages will refer to a neurobiological (brain-based) explanation of moral alarm, while the remaining half will refer to a cognitive (mind-based) explanation of moral alarm. Both types of explanations were adapted from a subsection of the explanations presented by Aspinwall et al. (2012), who illustrated the power of biological explanations of psychopathic behavior, including moral alarm, to shape the sentencing decisions of judges. In sum, participants will read one of four different types of feedback that differ along two dimensions: the degree of personal moral alarm and the neurobiological or cognitive nature of this trait.

The third section of the study will require participants to complete a series of brief tasks. The measurements for the mediators and dependent variables will be counterbalanced. Therefore, half of the participants will complete the (randomly ordered) mediators, followed by the (randomly ordered) dependent variables. Additionally half of the participants will complete the (randomly ordered) dependent variables followed by the (randomly ordered) mediators. Consequently, the participants will complete scales intended to measure the proposed mediators, that is the Determinism Subscale, the Free Will Subscale of the Free Will Inventory, and a measure of dualistic beliefs.

Subsequently, self-control will be measured through a modified online version of the famous marshmallow test (Mischel et al., 1972). As all participants will be entered into a final lottery, they will be asked when they would prefer to discover the outcome of the draw. They will have the choice to either find out immediately after the completion of the study if they have won their specific amount of money, or whether they would prefer to receive an increment of €100 but wait 3 months to find out the lottery's outcomes. Participants will be given the measure of belief in dualism. This measure concerns a futuristic device that enables

scientists to precisely duplicate any person; participants are asked to answer questions about their hypothetical duplicate.

Participants will also be required to respond to three different moral dilemmas designed to measure utilitarian reasoning: the difficult personal dilemma, the easy personal dilemma and the impersonal dilemma. For the first type of dilemma, we will use the crying baby dilemma. The footbridge dilemma will be used to test the easy moral dilemma, while the standard trolley dilemma will be administered in order to test the impersonal dilemma. All three dilemmas will be counterbalanced, in order to avoid any first response interfering or influencing the remaining responses. At the end of each response, the participants will be asked whether they felt guilty about their virtual actions, through a 6-point Likert scale.

The die-under-the-cup test will also be administered to participants to measure their willingness to lie. They will be asked to press a button on the screen to roll a die three times to decide the amount of money they could potentially win. Finally, they will be asked to select the outcome of their first roll (from 1 to 6) on-screen; they will be given 30 s to enter the outcome before the page progresses. Participants will be warned that if they fail to type the outcome down within 30 s, they will only be awarded the minimum amount of money. The time limit will be visible from a ticking counter.

The roll outcome that participants report will determine the value of the lottery prize: the higher the outcome, the greater the value of the prize. Hence participants may misreport the outcome of their first roll in order to increase the value of their potential prize. In reality, the prize will be fixed at the maximum value. In order that participants can receive their potential prize for entry into the lottery, we will lastly ask for their email address. However, this email address will be stored separately to all other data to ensure their responses remain anonymous. Finally, participants will be debriefed about the false feedback.

It is important to note that a counterbalancing procedure will be included, whereby the tasks measuring self-control, dishonesty and utilitarian reasoning will be presented in a randomly generated order. This is important because we hypothesize that certain conditions, such as the stronger-than-average moral alarm condition, will promote more inhibitory, honest and empathic responses on the first measure of any psychopathic behavior. Consequently, this could reduce the *willingness* of participants to exhibit inhibitory, honest and empathic responses on subsequent measures of these behaviors.

According to moral licensing theory (Merritt et al., 2010), individuals who show moral behaviors initially, tend to display immoral, unethical or problematic behaviors later (Blanken et al., 2015). This may be attributed to the fact that such individuals feel authorized to award themselves moral credits, believe that all temptations wear down their self-control, or simply become desensitized to the thought of cheating. Participants in our stronger-than-average moral alarm condition may be more likely to cheat in later tasks than earlier task due to this confounding effect of moral licensing.

In contrast, we hypothesize that participants who read about their neurobiologically weak moral alarm may exhibit less inhibitory, less honest and less empathic responses during initial

tasks, given the perception that their psychopathic traits are independent of their free will and are ultimately due to their brain. Consequently, the very act of behaving immorally may induce subsequent guilt and remorse, thereby reducing the perceived appropriateness of continuing to respond immorally. Therefore, participants may exhibit more moral behaviors in the later tasks.

For example, participants in the neurobiologically weak moral alarm condition who first receive the dishonesty task may feel that cheating is acceptable, given a reduced attribution of their actions to free will. However, participants may then believe that enough cheating has been done and therefore respond more morally in the subsequent tasks, such as the self-control one. In order to control all these possible outcomes, randomly changing the order of presentation of these tasks could minimize any possible confounding effects of completing each task on responses to subsequent tasks.

Before concluding the study, participants will be asked to what extent they thought the feedback they had received (i.e., the false feedback) was true about themselves. Furthermore, the participants will be asked to rate the degree to which they believed the presented explanation of psychopathy was true. Furthermore, participants will be asked to provide personal demographic information, including their age, gender, nationality and field of studies/work. At the end of the study, the participants will be comprehensively debriefed.

Proposed Analysis and Anticipated Results

We will test our hypotheses using hierarchical multiple linear regression, according to the recommendations of Hayes (2013). Hierarchical multiple linear regression is an appropriate procedure because we want to see how the average values of the dependent variables change as the independent variable is varied through our manipulation, while at the same time several demographic variables are held fixed. Hierarchical regression was selected instead of multivariate analysis of variance because we want to test hypotheses of mediation. Hayes' procedure for mediation analyses involves bootstrapping confidence intervals of the indirect effects; this procedure was considered preferable over the "causal steps" model of Baron and Kenny (1986), due to several shortcomings of this model (for detailed coverage, see Hayes, 2013). The software IBM SPSS Statistics Version 23.0 will be employed. For the mediation analyses, the PROCESS macro for SPSS will be used¹.

Before testing the model, we will check the assumptions of linear regression. If the observations are normally distributed, then parametric regression is appropriate. Outliers will be removed systematically. There will be no missing data points, as the form does not allow continuing without selecting an option. Nevertheless, participants may terminate their participation early: where participants discontinue their participation after at least one dependent variable has been measured and have not withdrawn their consent, the data will be used in the analysis of that particular variable. Categorical dependent variables will be dummy coded as whole numbers.

¹processmacro.org

In the first step of the analysis, the dependent variables Dishonesty and Utilitarian reasoning 1 (Crying Baby), 2 (Footbridge) and 3 (Trolley) will be entered into the model. Then the independent variables will be entered in a fixed order of steps or blocks. In the first block of the hierarchical regression model, the demographic variables Age, Gender, Nationality and Education level will be included. This means that these variables are held constant in the further analyses. In the second block, the independent variables of Psychopathy and Neuroscience will be added, first separately and then together; the direct and interactive effects can be estimated in this manner. Next the mediation analysis of indirect effects through Free will, Determinism, Dualism, Guilt and Self-control will be carried out using the PROCESS macro.

We hypothesize that the dependent variables will be significantly predicted by the independent variables but also that the mediation analyses will show significant indirect effects. Specifically, in accordance with Hypothesis 4, we expect participants in the weak moral alarm condition to show more dishonesty and utilitarian reasoning compared to those in the strong moral alarm condition. We also expect to see a stronger demonstration of this in the neurobiological explanation condition (Hypothesis 5). Finally, we expect the indirect effects observed to support Hypothesis 6, and show that the measures of free will, dualism, guilt, and self-control mediate the relationships between the independent and dependent variables.

Limitations

The primary limitation of the online study will be our inability to identify the precise mechanism of the effects, e.g., which type of belief in free will has been challenged by the manipulation: a compatibilist or incompatibilist notion of choice? It is impossible to control for all the differences between the neurobiological and cognitive conditions. Specifically, the neurobiological and cognitive conditions might induce differences in lay perceptions of the availability and causal efficacy of the conscious mind over our feelings of moral alarm (compatibilist choice), or the scope for free will to exist before the brain/mind and therefore the scope to attribute ultimate control to our actions (incompatibilist choice). Given a more nuanced understanding of compatibilist choice, the neurobiological and cognitive conditions could also induce differences in the lay perception that the degree of moral alarm experienced is a feature of our Deep Self – our stable self – or merely our Acting Self – our temporary self in a particular situation (Sripada, 2009).

This limitation in our ability to specify the mechanism could only be overcome by measuring more mediators and including more control conditions, which would be impractical due to the number of participants and length of survey then required. Despite having no control condition in which participants perform all the tasks without reading about their own moral alarm, we can still establish effects of describing neurobiology relative to describing cognition – the purpose of our study. Our goal is to document effects of giving people personal feedback in neurobiological terms, not to document effects of giving people personal feedback relative to no feedback. Note, with this design,

we can still document effects of giving above-average feedback relative to below-average feedback.

Given the nature of the manipulation, we can only recruit Facebook users for the online study. Although Facebook is very widely used, it is more popular among younger (and other types of) people. The cross-cultural design of our study, however, promotes the generalisability of our findings in a different direction: across the countries. In order to assess the generalisability of our sample, we are of course collecting demographic information in order to know if and how our sample could be biased.

One might also contest whether our findings can be generalized to real life examples of immoral behavior, since, for example, cheating was only measured online and people are more likely to lie online (Naquin et al., 2010). On the other hand, the potential for the researcher to record cheating is clearer online – this potential might therefore discourage cheating. Consequently, there is also reason to suggest the cheating observed online may not be any more frequent than the cheating observed face-to-face. One might also argue that measures of online cheating are gaining ecological validity with the increasing tendency for people to spend their time online.

The basis of our manipulation in a false analysis of Facebook Likes creates a potential pitfall for the credibility of the manipulation. Participants might not believe that we have analyzed their Likes and that their Likes reveal they have below/above-average levels of moral alarm. Also, the participants might suspect whether the cheating task is a genuine means of determining the amount of money available in the prize draw rather than a means of determining cheating. These

potential artifacts will be monitored by asking participants about any suspicions and the believability of the manipulation at the end.

AUTHOR CONTRIBUTIONS

All authors listed have actively contributed to this work, and given approval for its publication.

FUNDING

This research will be funded by an Economic and Social Research Council 1 + 3 studentship.

ACKNOWLEDGMENTS

This research was made possible by the Junior Researcher Programme (<http://jrp.psycholars.org/>). We would like to thank the entire Program team for their impeccable and consistent assistance.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fpsyg.2017.00294/full#supplementary-material>

REFERENCES

- Alquist, J. L., Ainsworth, S. E., and Baumeister, R. F. (2013). Determined to conform: disbelief in free will increases conformity. *J. Exp. Soc. Psychol.* 49, 80–86. doi: 10.1016/j.jesp.2012.08.015
- Angermeyer, M. C., Holzinger, A., Carta, M. G., and Schomerus, G. (2011). Biogenetic explanations and public acceptance of mental illness: systematic review of population studies. *Br. J. Psychiatry* 199, 367–372. doi: 10.1192/bjp.bp.110.085563
- Aniskiewicz, A. S. (1979). Autonomic components of vicarious conditioning and psychopathy. *J. Clin. Psychol.* 35, 60–67. doi: 10.1002/1097-4679(197901)35:1<60::AID-JCLP2270350106<3.0.CO;2-R
- Aspinwall, L. G., Brown, T. R., and Tabery, J. (2012). The double-edged sword: does biomechanism increase or decrease judges' sentencing of psychopaths? *Science* 337, 846–849. doi: 10.1126/science.1219569
- Baron, R. M., and Kenny, D. A. (1986). The moderator–mediator variable distinction in social psychological research: conceptual, strategic, and statistical considerations. *J. Pers. Soc. Psychol.* 51, 1173–1182. doi: 10.1037/0022-3514.51.6.1173
- Bartels, D. M., and Pizarro, D. A. (2011). The mismeasure of morals: antisocial personality traits predict utilitarian responses to moral dilemmas. *Cognition* 121, 154–161. doi: 10.1016/j.cognition.2011.05.010
- Baumeister, R. F., Masicampo, E. J., and Dwall, C. N. (2009). Prosocial benefits of feeling free: disbelief in free will increases aggression and reduces helpfulness. *Personal. Soc. Psychol. Bull.* 35, 260–268. doi: 10.1177/0146167208327217
- Blair, R. (2006). The emergence of psychopathy: implications for the neuropsychological approach to developmental disorders. *Cognition* 101, 414–442. doi: 10.1016/j.cognition.2006.04.005
- Blanken, I., Van De Ven, N., and Zeelenberg, M. (2015). A meta-analytic review of moral licensing. *Personal. Soc. Psychol. Bull.* 41, 540–558. doi: 10.1177/0146167215572134
- Buckels, E. E., Trapnell, P. D., and Paulhus, D. L. (2014). Trolls just want to have fun. *Personal. Individ. Differ.* 67, 97–102. doi: 10.1016/j.paid.2014.01.016
- Cahill, N., and Sherrets, S. (1979). The lost-letter technique: a measure of social and political attitudes in three socio-economic groups. *Psychol. Rep.* 45, 144–146. doi: 10.2466/pr0.1979.45.1.144
- Cherulnik, P. D. (1975). An independent validation of the lost-letter technique. *J. Soc. Psychol.* 96, 299–300. doi: 10.1080/00224545.1975.9923300
- Cima, M., Tonnaer, F., and Hauser, M. D. (2010). Psychopaths know right from wrong but don't care. *Soc. Cogn. Affect. Neurosci.* 5, 59–67. doi: 10.1093/scan/nsp051
- Conway, P., and Gawronski, B. (2013). Deontological and utilitarian inclinations in moral decision making: a process dissociation approach. *J. Pers. Soc. Psychol.* 104, 216–235. doi: 10.1037/a0031021
- Corrigan, P. W., and Watson, A. C. (2004). At issue: stop the stigma: call mental illness a brain disease. *Schizophr. Bull.* 30, 477–479. doi: 10.1093/oxfordjournals.schbul.a007095
- Crescioni, A. W., Baumeister, R. F., Ainsworth, S. E., Ent, M., and Lambert, N. M. (2015). Subjective correlates and consequences of belief in free will. *Philos. Psychol.* 29, 41–63. doi: 10.1080/09515089.2014.996285
- Crick, F. (1994). *The Astonishing Hypothesis: The Scientific Search for the Soul*. New York, NY: Scribner.
- Dar-Nimrod, I., and Heine, S. J. (2011). Genetic essentialism: on the deceptive determinism of DNA. *Psychol. Bull.* 137, 800–818. doi: 10.1037/a0021860
- Denno, D. W. (2015). Myth of the double-edged sword: an empirical study of neuroscience evidence in criminal cases. *Boston Coll. Rev.* 56, 493–551.

- Duke, A. A., and Bègue, L. (2015). The drunk utilitarian: blood alcohol concentration predicts utilitarian responses in moral dilemmas. *Cognition* 134, 121–127. doi: 10.1016/j.cognition.2014.09.006
- Dweck, C. S. (1999). *Self-theories: Their Role in Motivation, Personality, and Development*. Philadelphia, PA: Psychology Press.
- Dweck, C. S. (2006). *Mindset: The New Psychology of Success*. New York, NY: Random House.
- Fernandez-Duque, D., Evans, J., Christian, C., and Hodges, S. D. (2015). Superfluous neuroscience information makes explanations of psychological phenomena more appealing. *J. Cogn. Neurosci.* 27, 926–944. doi: 10.1162/jocn_a_00750
- Foot, P. (1978). *Virtues and Vices and Other Essays in Moral Philosophy*. Berkeley, CA: University of California Press.
- Forstmann, M., and Burgmer, P. (2015). Adults are intuitive mind-body dualists. *J. Exp. Psychol.* 144, 222–235. doi: 10.1037/xge0000045
- Forstmann, M., Burgmer, P., and Mussweiler, T. (2012). “The Mind Is Willing, but the Flesh Is Weak”: the effects of mind-body dualism on health behavior. *Psychol. Sci.* 23, 1239–1245. doi: 10.1177/0956797612442392
- Garcia, D., and Sikström, S. (2014). The dark side of facebook: semantic representations of status updates predict the dark triad of personality. *Personal. Individ. Differ.* 67, 92–96. doi: 10.1016/j.paid.2013.10.001
- Gino, F., and Ariely, D. (2012). The dark side of creativity: original thinkers can be more dishonest. *J. Pers. Soc. Psychol.* 102, 445–459. doi: 10.1037/a0026406
- Gleichgerricht, E., and Young, L. (2013). Low levels of empathic concern predict utilitarian moral judgment. *PLoS ONE* 8:e60418. doi: 10.1371/journal.pone.0060418
- Greene, E., and Cahill, B. S. (2011). Effects of neuroimaging evidence on mock juror decision making. *Behav. Sci. Law* 30, 280–296. doi: 10.1002/bsl.1993
- Greene, J., and Cohen, J. (2004). For the law, neuroscience changes nothing and everything. *Philos. Trans. R. Soc. B Biol. Sci.* 359, 1775–1785.
- Greene, J. D., Morelli, S. A., Lowenberg, K., Nystrom, L. E., and Cohen, J. D. (2008). Cognitive load selectively interferes with utilitarian moral judgment. *Cognition* 107, 1144–1154.
- Greene, J. D., Nystrom, L. E., Engell, A. D., Darley, J. M., and Cohen, J. D. (2004). The neural bases of cognitive conflict and control in moral judgment. *Neuron* 44, 389–400. doi: 10.1016/j.neuron.2004.09.027
- Greene, J. D., Sommerville, R. B., Nystrom, L. E., Darley, J. M., and Cohen, J. D. (2001). An fMRI investigation of emotional engagement in moral judgment. *Science* 293, 2105–2108.
- Gurley, J. R., and Marcus, D. K. (2008). The effects of neuroimaging and brain injury on insanity defenses. *Behav. Sci. Law* 26, 85–97. doi: 10.1002/bsl.797
- Halevy, R., Shalvi, S., and Verschuere, B. (2013). Being honest about dishonesty: correlating self-reports and actual lying. *Hum. Commun. Res.* 40, 54–72. doi: 10.1111/hcre.12019
- Hare, R. D. (1991). *The Hare Psychopathy Checklist-Revised: Manual*. Toronto, ON: Multi-Health Systems.
- Hayes, A. F. (2013). *Introduction to Mediation, Moderation and Conditional Process Analysis: A Regression-Based Approach*. New York, NY: Guilford Press.
- Holland, J., Silva, A. S., and Mace, R. (2012). Lost letter measure of variation in altruistic behavior in 20 neighbourhoods. *PLoS ONE* 7:e43294. doi: 10.1371/journal.pone.0043294
- Hyman, S. E. (2007). Can neuroscience be integrated into the DSM-V? *Nat. Rev. Neurosci.* 8, 725–732. doi: 10.1038/nrn2218
- Jones, D. N., and Paulhus, D. L. (2014). Introducing the short dark triad (SD3): a brief measure of dark personality traits. *Assessment* 21, 28–41. doi: 10.1177/1073191113514105
- Kane, R. (1999). Responsibility, luck, and chance. *J. Philos.* 96, 217–240. doi: 10.5840/jphil199996537
- Kendler, K. S. (2005). Toward a philosophical structure for psychiatry. *Am. J. Psychiatry* 162, 433–440. doi: 10.1176/appi.ajp.162.3.433
- Kiehl, K. A., Smith, A. M., Hare, R. D., Mendrek, A., Forster, B. B., Brink, J., et al. (2001). Limbic abnormalities in affective processing by criminal psychopaths as revealed by functional magnetic resonance imaging. *Biol. Psychiatry* 50, 677–684. doi: 10.1016/s0006-3223(01)01222-7
- Koenigs, M., Kruepke, M., Zeier, J., and Newman, J. P. (2012). Utilitarian moral judgment in psychopathy. *Soc. Cogn. Affect. Neurosci.* 7, 708–714. doi: 10.1093/scan/nsr048
- Koenigs, M., Young, L., Adolphs, R., Tranel, D., Cushman, F., Hauser, M., et al. (2007). Damage to the prefrontal cortex increases utilitarian moral judgements. *Nature* 446, 908–911. doi: 10.1038/nature05631
- Kolber, A. J. (2016). “Free Will as a Matter of Law,” in *Philosophical Foundations of Law and Neuroscience*, Chap. 2, eds M. Pardo and D. Patterso (Oxford: Oxford University Press).
- Kosinski, M., Matz, S. C., Gosling, S. D., Popov, V., and Stillwell, D. (2015). Facebook as a research tool for the social sciences: opportunities, challenges, ethical considerations, and practical guidelines. *Am. Psychol.* 70, 543–556. doi: 10.1037/a0039210
- Kosinski, M., Stillwell, D., and Graepel, T. (2013). Private traits and attributes are predictable from digital records of human behavior. *Proc. Natl. Acad. Sci. U.S.A.* 110, 5802–5805. doi: 10.1073/pnas.1218772110
- Kraus, M. W., and Keltner, D. (2013). Social class rank, essentialism, and punitive judgment. *J. Pers. Soc. Psychol.* 105, 247–261. doi: 10.1037/a0032895
- Lebowitz, M. S. (2014). Biological conceptualizations of mental disorders among affected individuals: a review of correlates and consequences. *Clin. Psychol.* 21, 67–83. doi: 10.1111/cpsp.12056
- Legrenzi, P., and Umiltà, C. A. (2011). *Neuromania: On the Limits of Brain Science*. Oxford: Oxford University Press.
- Lewis, A., Bardis, A., Flint, C., Mason, C., Smith, N., Tickle, C., et al. (2012). Drawing the line somewhere: an experimental study of moral compromise. *J. Econ. Psychol.* 33, 718–725. doi: 10.1016/j.joep.2012.01.005
- Maruna, S., and Copes, H. (2005). What have we learned from five decades of neutralization research? *Crime Justice* 32, 221–320. doi: 10.1086/655355
- Mendez, M. F., Anderson, E., and Shapira, J. S. (2005). An investigation of moral judgement in frontotemporal dementia. *Cogn. Behav. Neurol.* 18, 193–197. doi: 10.1097/01.wnn.0000191292.17964.bb
- Merritt, A. C., Efron, D. A., and Monin, B. (2010). Moral self-licensing: when being good frees us to be bad. *Soc. Personal. Psychol. Compass* 4, 344–357. doi: 10.1111/j.1751-9004.2010.00263.x
- Merritt, C. B., and Fowler, R. G. (1948). The pecuniary honesty of the public at large. *J. Abnorm. Soc. Psychol.* 43, 90–93. doi: 10.1037/h0061846
- Milgram, S. (1969). Comment on “A Failure to Validate the Lost Letter Technique.” *Public Opin. Q.* 33, 263–264. doi: 10.1086/267698
- Milgram, S., Mann, L., and Harter, S. (1965). The lost-letter technique: a tool of social research. *Public Opin. Q.* 29, 437–438. doi: 10.1086/267344
- Mischel, W., Ebbesen, E. B., and Zeiss, A. R. (1972). Cognitive and attentional mechanisms in delay of gratification. *J. Pers. Soc. Psychol.* 21, 204–218.
- Nadelhoffer, T., Shepard, J., Nahmias, E., Sripada, C., and Ross, L. T. (2014). The free will inventory: measuring beliefs about agency and responsibility. *Conscious. Cogn.* 25, 27–41. doi: 10.1016/j.concog.2014.01.006
- Nahmias, E., Justin Coates, D., and Kvaran, T. (2007). Free will, moral responsibility, and mechanism: experiments on folk intuitions. *Midwest Stud. Philos.* 31, 214–242. doi: 10.1111/j.1475-4975.2007.00158.x
- Nahmias, E., Morris, S., Nadelhoffer, T., and Turner, J. (2005). Surveying freedom: folk intuitions about free will and moral responsibility. *Philos. Psychol.* 18, 561–584. doi: 10.1080/09515080500264180
- Nakamura, K. (2013). A closer look at moral dilemmas: latent dimensions of morality and the difference between trolley and footbridge dilemmas. *Think. Reason.* 19, 178–204. doi: 10.1080/13546783.2013.768551
- Naquin, C. E., Kurtzberg, T. R., and Belkin, L. Y. (2010). The finer points of lying online: e-mail versus pen and paper. *J. Appl. Psychol.* 95, 387–394.
- Open Science Collaboration (2015). Estimating the reproducibility of psychological science. *Science* 349:aac4716. doi: 10.1126/science.aac4716
- Paulhus, D. L., and Carey, J. M. (2011). The FAD-Plus: measuring lay beliefs regarding free will and related constructs. *J. Pers. Assess.* 93, 96–104. doi: 10.1080/00223891.2010.528483
- Rigoni, D., Kühn, S., Gaudino, G., Sartori, G., and Brass, M. (2012). Reducing self-control by weakening belief in free will. *Conscious. Cogn.* 21, 1482–1490. doi: 10.1016/j.concog.2012.04.004
- Rigoni, D., Pourtois, G., and Brass, M. (2015). “Why Should I Care?” challenging free will attenuates neural reaction to errors. *Soc. Cogn. Affect. Neurosci.* 10, 262–268. doi: 10.1093/scan/nsu068
- Satel, S., and Lilienfeld, S. O. (2015). *Brainwashed: The Seductive Appeal of Mindless Neuroscience*. New York, NY: Basic Books.
- Schwartz, N. J., and Saks, M. J. (2011). Neuroimage evidence and the insanity defense. *Behav. Sci. Law* 29, 592–607. doi: 10.1002/bsl.995

- Schweitzer, N. J., Saks, M. J., Murphy, E. R., Roskies, A. L., Sinnott-Armstrong, W., and Gaudet, L. M. (2011). Neuroimages as evidence in a mens rea defense: no impact. *Psychol. Public Policy Law* 17, 357–393. doi: 10.1037/a0023581
- Shalvi, S., Dana, J., Handgraaf, M. J. J., and De Dreu, C. K. W. (2011). Justified ethicality: observing desired counterfactuals modifies ethical perceptions and behavior. *Organ. Behav. Hum. Decis. Process.* 115, 181–190. doi: 10.1037/e722992011-056
- Shalvi, S., Eldar, O., and Bereby-Meyer, Y. (2012). Honesty requires time (and lack of justifications). *Psychol. Sci.* 23, 1264–1270. doi: 10.1177/0956797612443835
- Shariff, A. F., Greene, J. D., Karremans, J. C., Luguri, J. B., Clark, C. J., Schooler, J. W., et al. (2014). Free will and punishment: a mechanistic view of human nature reduces retribution. *Psychol. Sci.* 25, 1563–1570. doi: 10.1177/0956797614534693
- Sripada, C. S. (2009). *The Deep Self Model and Asymmetries in Folk Judgments about Intentional Action*. Berlin: Springer Science + Business Media. doi: 10.1007/s11098-009-9423-5
- Stern, S. E., and Faber, J. E. (1997). The lost e-mail method: Milgram's lost-letter technique in the age of the internet. *Behav. Res. Methods Instrum. Comput.* 29, 260–263. doi: 10.3758/bf03204823
- Thomson, J. J. (1985). "The Trolley Problem," in *Ethics: Problems and Principles*, eds J. M. Fischer and M. Ravizza (San Diego, CA: Harcourt Brace Jovanovich), 67–76.
- Trust for London and New Policy Institute (2017). "Overview of London boroughs." *Overview of London boroughs | London's Poverty Profile*. September 23, 2016. Available at: <http://www.londonpovertyprofile.org.uk/key-facts/overview-of-london-boroughs/> [accessed January 11, 2017].
- Umbach, R., Berryessa, C. M., and Raine, A. (2015). Brain imaging research on psychopathy: implications for punishment, prediction, and treatment in youth and adults. *J. Crim. Justice* 43, 295–306. doi: 10.1016/j.jcrimjus.2015.04.003
- Valdesolo, P., and DeSteno, D. (2006). Manipulations of emotional context shape moral judgment. *Psychol. Sci.* 17, 476–477. doi: 10.1111/j.1467-9280.2006.01731.x
- Vohs, K. D., and Schooler, J. W. (2008). The value of believing in free will: encouraging a belief in determinism increases cheating. *Psychol. Sci.* 19, 49–54. doi: 10.1111/j.1467-9280.2008.02045.x
- Weisberg, D. S., Keil, F. C., Goodstein, J., Rawson, E., and Gray, J. R. (2008). The seductive allure of neuroscience explanations. *J. Cogn. Neurosci.* 20, 470–477. doi: 10.1162/jocn.2008.20.3.470
- Wiegmann, A., Lippold, M., and Grigull, R. (2013). "On the robustness of intuitions in the two best-known trolley dilemmas," in *Proceedings of the 35th Annual Conference of the Cognitive Science Society*, eds M. Knauff, M. Pauen, N. Sebanz, and I. Wachsmuth (Austin, TX: Cognitive Science Society), 3759–3764.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2017 Blakey, Askelund, Boccanera, Immonen, Plohl, Popham, Sorger and Stuhleyer. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Exploring Musical Activities and Their Relationship to Emotional Well-Being in Elderly People across Europe: A Study Protocol

Jennifer Grau-Sánchez^{1,2*}, Meabh Foley³, Renata Hlavová⁴, Ilkka Muukkonen⁵, Olatz Ojinaga-Alfageme⁶, Andrijana Radukic⁷, Melanie Spindler⁸ and Bodil Hundevad^{9*}

¹ Department of Cognition, Development and Educational Psychology, University of Barcelona, Barcelona, Spain, ² Cognition and Brain Plasticity Unit, Bellvitge Biomedical Research Institute, Barcelona, Spain, ³ Department of Psychology, National University of Ireland, Galway, Ireland, ⁴ Department of Psychology, Masaryk University, Brno, Czechia, ⁵ Department of Psychology, University of Helsinki, Helsinki, Finland, ⁶ Department of Psychology, University of Deusto, Bilbao, Spain, ⁷ Department of Psychology, University of Banja Luka, Banja Luka, Bosnia and Herzegovina, ⁸ Department of Psychology, University of Oldenburg, Oldenburg, Germany, ⁹ Department of Psychology, University of Vienna, Vienna, Austria

OPEN ACCESS

Edited by:

Kristina Egumenovska,
Scuola Internazionale di Studi
Superiori Avanzati (SISSA), Italy

Reviewed by:

Fabian Gander,
University of Zurich, Switzerland
Gregor Sočan,
University of Ljubljana, Slovenia

*Correspondence:

Jennifer Grau-Sánchez
jenny_grau@ub.edu
Bodil Hundevad
bodil@hundevad.de

Specialty section:

This article was submitted to
Quantitative Psychology
and Measurement,
a section of the journal
Frontiers in Psychology

Received: 14 November 2016

Accepted: 21 February 2017

Published: 20 March 2017

Citation:

Grau-Sánchez J, Foley M,
Hlavová R, Muukkonen I,
Ojinaga-Alfageme O, Radukic A,
Spindler M and Hundevad B (2017)
Exploring Musical Activities and Their
Relationship to Emotional Well-Being
in Elderly People across Europe:
A Study Protocol.
Front. Psychol. 8:330.
doi: 10.3389/fpsyg.2017.00330

Music is a powerful, pleasurable stimulus that can induce positive feelings and can therefore be used for emotional self-regulation. Musical activities such as listening to music, playing an instrument, singing or dancing are also an important source for social contact, promoting interaction and the sense of belonging with others. Recent evidence has suggested that after retirement, other functions of music, such as self-conceptual processing related to autobiographical memories, become more salient. However, few studies have addressed the meaningfulness of music in the elderly. This study aims to investigate elderly people's habits and preferences related to music, study the role music plays in their everyday life, and explore the relationship between musical activities and emotional well-being across different countries of Europe. A survey will be administered to elderly people over the age of 65 from five different European countries (Bosnia and Herzegovina, Czechia, Germany, Ireland, and UK) and to a control group. Participants in both groups will be asked about basic sociodemographic information, habits and preferences in their participation in musical activities and emotional well-being. Overall, the aim of this study is to gain a deeper understanding of the role of music in the elderly from a psychological perspective. This advanced knowledge could help to develop therapeutic applications, such as musical recreational programs for healthy older people or elderly in residential care, which are better able to meet their emotional and social needs.

Keywords: elderly population, music, well-being, emotion, cross-cultural

INTRODUCTION

Many important changes such as retirement, changes in social ties and decline in physical and cognitive capabilities, occur during the later years of life. They can affect the psychological well-being of the elderly, and lead to loneliness (Luanaigh and Lawlor, 2008; Golden et al., 2009) and depression (Alexopoulos, 2005). Although defining well-being is challenging due to

its multidimensional nature, there have been several attempts focusing mainly on the constructs rather than on the definition of well-being itself. In this regard, well-being can be understood as the balance point between an individual's resources and challenges (Dodge et al., 2012). The presence of positive emotions and absence of negative ones as well as satisfaction with life and functioning are components of well-being (Diener et al., 2000; Ryff and Keyes, 1995). Several factors, such as personality and health, are known to correlate with well-being (Kahneman and Deaton, 2010). Moreover, studies addressing the role of leisure activities have found that they contribute to subjective well-being, and that the effect holds in cross-sectional, longitudinal, and experimental studies (Kuykendall et al., 2015). Specifically in the context of the elderly, participation in leisure activities seems to increase well-being and can reduce the risk of dementia (Herzog et al., 1998; Silverstein and Parker, 2002; Verghese et al., 2003; Adams et al., 2011; Chang et al., 2014).

Musical engagement, as part of everyday life, can positively influence and contribute to well-being. To evoke emotions, is one of the primary reasons for listening to or producing music (Juslin and Västfjäll, 2008; Croom, 2012; Koelsch, 2014). Cross-cultural research investigating the role of music on emotion suggests that emotional cues in music transcend both language and culture (Kim et al., 2010). Music can be used for emotional self-regulation through different strategies, such as diverting from contemplation over negative emotions, maintenance of positive mood, or relaxation; these strategies seem to be stable across the lifespan (Saarikallio, 2011). Musical activities are also an important source for social contact, promoting interaction and the sense of belonging with others (Rilling et al., 2002; Koelsch, 2014). Therefore, activities involving music such as listening to music, playing, singing, and dancing have been shown to have great impact on well-being, increasing the person's life satisfaction.

To date, most research on the importance of music has focused on its cognitive and emotional functions, without consideration of the collective features of the musical experience. This is despite several researchers demonstrating the central importance of music to social and cultural settings (DeNora, 2000). Research has shown that the use of music for the self-regulation of mood is broadly similar across several disparate cultures, including Latin American, African, and European states (Saarikallio, 2008; Boer and Fischer, 2010). Some culture specific variations in the function of music were identified among sub-groups of these studies; this is in keeping with the wide scope of cultures examined. However, there have been no studies investigating the role of music in well-being in more similar cultures, such as between different European nations.

Despite the significant contribution of music to well-being, it is still not clear whether the importance of music remains stable across the lifespan (Lonsdale and North, 2011). Especially in senior years, musical activities can be accompanied by overcoming many barriers, such as loss of hearing, decline in memory, and an aging body. Musical importance in general is shown to be independent of level of mental competence and is also correlated with musical involvement (Cohen et al., 2002), however, no major differences between musicians and

those without musical education have been found (Hays and Minichiello, 2005a). Compared to younger generations, a small decrease of importance of music in the elderly has been reported, one possible reason might be that the elderly simply do not have as many opportunities to listen to music they value as the younger generations do (Cohen et al., 2002; Laukka, 2007).

Nevertheless, the majority of elderly people listen to music on a daily basis (Cohen et al., 2002). Research indicates that the primary motive for engaging in activities involving music during the elderly is maintaining the identity and agency, and regulating mood (Laukka, 2007). Moreover, music helps the elderly understand their emotions, maintain their sense of well-being as well as to give them hope or meaning to live (Hays and Minichiello, 2005b). Music can be a source of relaxation and enjoyment (Laukka, 2007). Further, it can be a useful tool for expressing spirituality and can sometimes serve as an escape from everyday living through imagination or the evocation of memories (Hays and Minichiello, 2005b; Schäfer et al., 2012). Although these functions of music become more salient after retirement, only few studies have addressed the meaningfulness of music in the elderly.

The present study aims to (i) investigate elderly people's habits and preferences related to music in a cross-cultural European sample, (ii) study the role music plays in their everyday life, (iii) and explore the relationship between music and emotional well-being. In relation to these aims, a survey will be administered to people over the age of 65 from five different European countries (Bosnia and Herzegovina, Czechia, Germany, Ireland, and UK) and to a control group. Participants will answer a survey about basic sociodemographic questions, habits and preferences in their participation in musical activities and a well-being questionnaire. The study will explore the daily use of music among elderly people, and we expect to find a positive correlation between frequency of participation in musical activities and emotional well-being regardless of the country of origin of the participants.

MATERIALS AND EQUIPMENT

Participants

A total of approximately 700 participants will be recruited for the present study, an elderly group from the age of 65 ($N = 350$) and a control group ranging from 20 to 30 years ($N = 350$). All participants will have normal or corrected to normal vision, since they will have to read a written survey. Besides, participants will be oriented in time and space at the time of test administration because they will have to answer questions about present and past behaviors. They will be asked which date and day of the week it is as well as the place where they will be completing the questionnaire (place, floor, city, and country). These basic questions will allow us to exclude those possible participants who might be confused or do not have a normal cognitive functioning. Participants with hearing deficits will be excluded because the nature of this study is to study music-related behaviors. Participants who currently live in nursing homes or other care facilities will be excluded as well.

The recruitment of participants will take place in Banja Luka (Bosnia and Herzegovina), Brno (Czechia), Meath (Ireland), Bremen and Munich (Germany), and London (UK). The groups will be matched in terms of gender, and similar sample sizes and age distributions will be collected in the different cities.

Design

This study is a transversal, observational study with both descriptive and analytic purposes.

Survey

The survey is divided into three different sections consisting of the Participant's profile, Musical profile, and Well-being questionnaire, which assess sociodemographic aspects, musical activities, and emotional well-being, respectively, through questions and standardized questionnaires. The materials described below were selected for their relevance to the research questions.

The survey will be administered to the elderly group in paper-pencil format and using an online survey platform for the control group. Prior to the administration of the survey, participants will receive an information sheet stating the study procedure and will sign an informed consent.

Participant's Profile

At the beginning of the survey, participants will complete a questionnaire regarding their sociodemographic background, leisure activities, and social support, as well as a short personality screening.

Sociodemographic Background

Information about gender, age, marital status, living situation, education, occupation, financial situation, and religiousness will be enquired.

Leisure Activities

Twelve different leisure activities will be presented and participants will have to answer how often do they participate in these activities by rating the frequency on a 7-point scale from (1) *never* to (7) *daily*. Following responses are possible: listening to or playing music, internet use, watching tv or films, doing crosswords or playing board games, reading, playing sports, walking, doing art activities, such as painting or drawing, gardening, going out for meals or coffee, going out with friends or family, and going to social clubs. This information will allow controlling for other leisure activities influencing well-being.

Social Support

The 12-item Multidimensional Scale of Perceived Social Support (Zimet et al., 1988) will be used to assess social support. This scale comprises three subscales each assessing a different source of support: (a) family, (b) friends, and (c) significant other. There are 12 statements that participants have to rate using a 7-point scale ranging from (1) *very strongly disagree* to (7) *very strongly agree*. The scale and subscales are scored through calculating the mean as according to the manual. Previous studies have shown that a greater social support is positively correlated to well-being

(Siedlecki et al., 2014). Therefore, this factor should be taken into account as a possible confounding variable in the present study. The 12-item Multidimensional Scale of Perceived Social Support shows good internal reliability as well as moderate construct validity ($\alpha = 0.88$; Zimet et al., 1988).

Personality Screening

The Big Five Personality Inventory-10 (BFI-10) (Rammstedt and John, 2007), which measures the Big-Five dimensions of personality, will be administered. The inventory has two items per each of the five personality dimensions: Extraversion, Agreeableness, Conscientiousness, Emotional Stability, and Openness to Experience. Every item uses the same stem, "I see myself as: ...". Participants have to rate the items on a 5-point scale ranging from (1) *disagree strongly* to (7) *agree strongly*. The scales are scored according to the manual. The reliability of the BFI-10 is adequate (Overall mean: $r = 0.75$, Extraversion: $r = 0.83$, Agreeableness: $r = 0.68$, Conscientiousness: $r = 0.43$, Neuroticism: $r = 0.36$, Openness: $r = 0.45$; Rammstedt and John, 2007).

In total, the Participant's profile will take approximately 15 min to complete.

Musical Profile

Participants will answer questions about their education in music, the importance of music for them, habits related to musical activities, style preferences in music, and roles and functions of music in their everyday life.

Mastery

Participants will be asked whether they received special music-related education (such as playing an instrument, singing, and dancing classes).

Importance of Music

The first question assesses the subjective importance of music in the participant's everyday life. Participants will be asked about how important music is for them with answers ranging from (1) *not at all* to (7) *extremely important*.

Habits

To assess the individual habits with regard to musical activities, participants will be asked about the frequency of doing different musical activities (listening to music, singing, playing an instrument, and dancing). A 7-point scale will be presented where they will have to rate how often do they do these activities from (1) *never* to (7) *daily*. Afterward, participants will be asked to estimate how long they listen to music on a normal day (in hours). In addition, information about which devices are used for listening to music will be collected. The possible responses for this last question are: radio, television, stereo, computer, portable device (i.e., walkman, discman, MP3 player), smartphone, concerts, car and public place.

Preferences

Nine different styles of music will be presented, these include: classical music; religious music; country, folk; jazz, swing, blues; disco, electronic; punk, rock, metal; hip-hop, rap; pop; and

reggae, ska. Participants will be asked to judge their level of enjoyment [7-point scale from (1) *dislike extremely* to (7) *enjoy extremely*] with regard to these musical styles. In order to examine how content people feel about the accessibility of their preferred music, a 7-point scale from (1) *not at all* to (7) *very much* will be used.

Roles and Functions of Music

In order to gain insight on the roles and functions music plays in everyday life, participants will be asked to report the frequency of listening to music due to certain reasons. This question contains 24 reasons to listen to music (e.g., for entertainment, to stir up energy), which can be rated from (1) *very seldom* to (5) *very often*. This scale, which originates from a study conducted by Laukka (2007), assesses listening strategies, intending to cover main psychological functions of music listening (i.e., emotional functions, identity, belonging, and agency; e.g., Ruud, 1997).

The Barcelona Music Reward Questionnaire will be used to measure musical reward experience (Mas-Herrero et al., 2013). The Barcelona Music Reward Questionnaire decomposes music reward into five factors: Musical Seeking, Emotion Evocation, Mood Regulation, Social Reward, and Sensory-Motor. Musical reward experience is measured using 20 statements that participants are asked to rate from (1) *completely disagree* to (5) *completely agree*. Scoring is done according to the manual, resulting in a score for each facet and a score for the global sensitivity for music reward. The Barcelona Music Reward Questionnaire also presents good reliability estimates ($r = 0.92$; Mas-Herrero et al., 2013).

To assess different emotions experienced with music, participants will report the frequency of different emotions they felt in response to listening to music, playing an instrument, singing and dancing. Participants are asked to rate 7 different emotions (e.g., happy, nostalgic, moved) on a scale from (1) *never* to (7) *always*. The original question was reported in Laukka (2007), and consisted of 45 emotions. In order to shorten the time of administration for the present study, these were summarized into more basic emotions, which are: happy, nostalgic, anxious, moved, bored, frustrated, sad, lonely, thrills or chills, disappointed, tense, angry, spiritual and relaxed.

The musical profile will take approximately 20 min.

Well-Being Questionnaire

At the beginning of this part of the questionnaire, participants will be asked whether they experience hearing problems, to control for this possible influence on musical experiences. In order to assess emotional well-being, different validated scales will be administered measuring emotional state, health status, and quality of life.

Emotional State

The Positive and Negative Affect Scale, which is comprised by two mood scales: the Positive Affect Scale and the Negative Affect Scale (Watson et al., 1988), will be administered. Ten descriptors are used to define each scale and participants will be asked to respond whether they have felt these emotions in the last week using a 5-point scale that ranges from (1) *very slightly or not at*

all to (5) *extremely*. The Positive and Negative Affect Scale has strong reported validity with measures as general distress and dysfunction, depression, and state anxiety. It was also successfully validated for the elderly population (PA: $\alpha = 0.87$, NA: $\alpha = 0.89$; Kercher, 1992).

Emotion Regulation

To assess emotion regulation problems, the short form of the Difficulties in Emotion Regulation Scale (Kaufmann et al., 2015) will be administered, which consists of 18 items, with answers ranging from (1) *almost never* to (5) *almost always*. The items are grouped into six scales: Strategies, Non-acceptance, Impulse, Goals, Awareness, and Clarity. The short form of the Difficulties in Emotion Regulation Scale will be scored according to the manual and maintains good psychometric properties ($\alpha \geq 0.70$; Kaufmann et al., 2015).

Resilience

The 10-item Connor–Davidson Resilience Scale (Connor and Davidson, 2003) will be administered in order to assess resilience, by focusing on personal resources or qualities deemed appropriate for positive adaptation to adversity. The scale is comprised of 10 items, each rated on a 5-point scale from 0 to 4, with higher scores reflecting greater resilience. The psychometric properties are good ($\alpha = 0.85$; Campbell-Sills and Stein, 2007).

Health Status

Selected subscales from the Short Form Health Survey (Ware and Sherbourne, 1992) will be used to measure emotional role functioning, emotional well-being, and general health. The Short Form Health Survey also showed good reliability and construct validity in terms of distinguishing between groups with expected health differences ($\alpha \geq 0.85$; $r \geq 0.75$; Brazier et al., 1992).

Subjective Well-Being

The Satisfaction with Life Scale (Diener et al., 1985) will be administered to screen subjective well-being. It consists of five items asking about satisfaction with one's life on a 7-point scale from (1) *strongly disagree* to (7) *strongly agree*.

The Well-being questionnaire will take around 30 min to complete, resulting in a total administration time of 65 min.

STEPWISE PROCEDURES

Ethics

The study is approved by the ethical board of the University of Barcelona. Each participant will receive information about the aim and procedures of the study and will provide consent for participation.

Translations of the Survey

Since participants will be recruited from different countries, translated versions of the survey are needed for those countries in which English is not the official language. Therefore, the survey has been translated to Czech, German, and Serbo-Croat-Bosnian following a forward and back translation process

for questions and standardized questionnaires that were not available in those languages. First, a native speaker of the target language with a proficiency level in English translated the original version of the survey in English to the target language (forward translation). Following the guidelines of the WHO for instrument translation, the local researcher of the study provided instructions to the translator, aiming for conceptual rather than literal translations. Second, an expert panel reviewed the forward translation, identifying and discussing inadequate expressions until reaching consensus (reconciliation). The expert panel was composed by three psychologists that were native speakers of the target language with a proficiency level in English and the local researcher. Third, an English native speaker with a proficiency level in the target language translated the survey from the target language to English (back translation). As in the forward translation, the local researcher provided guidelines for a conceptual translation. At the end of the process, the expert panel reviewed and compared both translations. If disagreements occurred in one word or expression, it was discussed until reaching a consensus (harmonization).

Pilot Testing and Feasibility of the Survey

After translating the survey to the target languages, 45 pilots were conducted. The local researchers asked for feedback regarding the clarity and length of the survey, the presentation of the questions and statements, and the display of the items.

Recruitment of Participants

For each recruitment location, a list of social clubs, civil societies, and neighborhood associations will be created with information about the nature of the club or society. Local researchers will contact the person in charge to introduce themselves, explain the purpose of the study and ask if there are members above 65 years old. If so, researchers will ask for permission to recruit participants and administer the survey in their venue. When visiting the clubs and societies, local researchers will explain the aim of the study to participants and ask for the individual consent of each participant. However, elderly participants from these clubs will be individuals relatively independent and active, and thus may not be a representative sample of the elderly people in general. To compensate this bias, local adult day-care centers will be approached following the same procedure as with social clubs and societies to target those individuals who are less active or involved in the community.

Regarding the control group, local researchers will post announcements about the study in different social networks, web pages and online forums. To target the control population (from 20 to 30 years old), local researchers will take into account the nature and content of these online sources. For both groups, the elderly and the control group, the aim of the study will not be disclosed in advance to prevent from attracting participants with a special interest in music.

Administration of the Survey

A guideline for local researchers will be used to explain the nature of the study to participants, and to provide instructions and guidance for each of the questions of the survey should the

participant required assistance. The elderly group will complete the survey in a quiet environment in either single- or small group-sessions whereas the control group will answer using an online platform for surveys.

PROPOSED ANALYSIS

The data collected in different recruitment locations will be managed by each local researcher, who will be responsible for gathering the data in a template database. Statistical analysis will be performed using the Statistical Package for the Social Sciences (IBM SPSS Statistics 23, Armonk, NY, USA). Since the data will be collected in different countries, we will test for measurement invariance before performing any analysis.

For the first aim of the study, which is to investigate the habits and preferences related to music in the elderly in a cross-cultural European sample, descriptive analyses will be performed for variables obtained from the music profile of the survey. These variables are the (1) importance of music, (2) frequency of different musical activities, (3) length and devices for listening to music, (4) music style preferences and level of enjoyment, (5) access to the preferred music, (6) reward and (7) emotions associated to music. These descriptive analyses will be performed first separately per each country (Bosnia and Herzegovina, Czechia, Germany, Ireland, and UK). Exploratory analyses comparing the results of the elderly participants and the control group of each country will be performed using independent *t*-tests. We will also apply descriptive statistics with all the data gathered from the different countries and the results of the overall elderly sample will be compared to the control group to test for differences between age groups.

The same type of analysis, descriptive and exploratory, will be applied to the second aim, which is to study the role of music in everyday life, by selecting the responses obtained from the question about the reasons for engaging in different musical activities. Analysis will be performed first separately for each country and comparing with the control group and then with all the subsamples together. Two-tailed significance tests will be used.

With regard to the third aim, exploring the associations between participation in musical activities (listening to music, singing, playing and instrument and dancing) and emotional well-being, a hierarchical multiple regression with two steps will be applied. Our hypothesis is that participants with higher frequency of participation in musical activities will report higher emotional well-being regardless of the country of origin of the participants.

On the first step of the regression, a model will be created with the information of the Participant's profile and Well-being questionnaire of the survey. The sociodemographic data will be the independent predictor variables, namely: (1) age, (2) marital status, (3) living situation, (4) education, (5) occupation, (6) religiousness, (7) participation in leisure activities, (8) perceived social support, and (9) personality. For the dependent variables, the Satisfaction With Life Scale is the one which is mostly hypothesized to be affected. However, we will use other scales

measuring wellbeing (the Positive and Negative Affect Scale, the Difficulties in Emotion Regulation Scale, the Connor–Davidson Scale, and the selected questions from the Short Form Health Survey) as there is no clear evidence on which practicing music would affect most. We will test them all with the appropriate family-wise error corrections. The variable country will be introduced as a fixed effect. With this step, we expect to have a clearly significant model since previous research supports the idea that some demographic aspects (e.g., marital status, or education or socioeconomic status) have a positive influence on the emotional well-being of individuals (Kahneman and Deaton, 2010). That is why these variables will be included into the model as control variables. Based on previous research on different factors on well-being (Laukka, 2007; Boarini et al., 2012; Oguz et al., 2013), we estimated our background factors to explain $r^2 = 0.35$ of the variance of different well-being measures.

On the second step, the variable about frequency of musical activities will be introduced in the model. With this step, we will test whether the inclusion of participation in musical activities improves the model of well-being. With our estimated sample size (350 elderly participants), we can expect to have >0.95 power with our independent measure adding at least 0.02 to the R^2 of the whole model, after correcting for multiple testing (Bonferroni) for our five different well-being measures. The same procedure will be applied with the participants in the control group and the possible differences between groups will be examined.

ANTICIPATED RESULTS

The first aim of this study is to investigate elderly people's habits and preferences related to music. Based on previous studies, we expect to find that while most elderly people from our sample will be daily listeners, the amount of singers, instrument players or dancers will be lower (Cohen et al., 2002). Moreover, we expect to find a lower frequency of musical activities among elderly people when compared to the control group (Cohen et al., 2002). The main reason for this could be that elderly people do not have as many opportunities to engage in musical activities as younger people do (Cohen et al., 2002; Laukka, 2007). It is thought that music styles and preferences are established during adolescence and early adulthood (Cohen et al., 2002). Indeed previous research indicates that elderly persons have a strong preference for music popular during their youth (Gibbons, 1977). However, there is little recent research investigating this topic, and none of a cross-cultural nature. Nonetheless, we expect to find that both the elderly and control group will show a preference for music popular in their youth.

The second aim of the study is to examine the role music plays in elderly people's everyday life. We anticipate that the importance of music in elderly persons will be evident in its utilization for purposes such as maintaining well-being, entertainment, self-identity, or socializing (Hays and Minichiello, 2005a,b). Furthermore, we expect to find some differences between the elderly and the control group. For example, in the role of music for mood management or emotional regulation, being this more prevalent in younger persons, or functions

such as self-conceptual processing related to autobiographical memories being more salient in the elderly (Lonsdale and North, 2011). Nevertheless, we expect music to have the same importance in both groups, independent of age, mental competence and region (Cohen et al., 2002), and for it to be either more important or more frequent than other leisure activities for a large amount of participants (Lonsdale and North, 2011).

Based on previous research, we expect to have a significant model of background personal variables, such as religiousness, marital status, gender, social support, and personality influencing emotional well-being (Lee and Ishii-Kuntz, 1987; Kahneman and Deaton, 2010). This model will help us to control for possible confounding variables when assessing the relationship between music and well-being. Introducing the frequency of musical activities in this model, we expect that a higher involvement in musical activities will be related to a better emotional well-being, in both, the elderly and the younger group. We expect to find that participation in musical activities, and in particular group activities, will increase emotional well-being, largely as a result of an increase in positive affect, social contact, shared experience, and evocation of positive memories (Laukka, 2007; MacDonald, 2013).

No differences between participants with and without professional musical background are expected (Hays and Minichiello, 2005a). However, due to increased musical education in recent years, we anticipate that formal musical education may be greater in the control group. In accordance with previous literature, we expect that there will be no significant cultural differences in the role of music in emotional well-being (Argstatter, 2016). Studies about neuroscience of music have provided explanation for a similar global experience of music, identifying that music activates the reward, emotion, and arousal regions of the brain (Blood and Zatorre, 2001). This could explain why an increase in positive affect and positive emotion related to musical activities appears to transcend both culture and region.

Despite every effort to control for confounding variables, this study will have a number of limitations. Our strategy for recruitment will lead to a convenient sample of social clubs and churches will likely lead to an over-representation of active elderly persons. Although, we intend on preliminarily screening for participants with dementia, we do not have the resources to carry out full cognitive, neurological or psychiatric screening on participants involved in this study. However, as participants will all be active in the community it is likely that the effects of such conditions will be minimal.

The use of a cross-cultural sample is a limitation, in that individual cultural factors may influence our findings. Indeed, research does suggest that the function of music varies between cultures (Cross, 2001). However, it is essential to note that the focus of this research has to date been between vastly different cultures, such as western and south east Asian nations (Cross, 2001; Kim et al., 2010). In contrast, our study is confined to European countries with more similar musical traditions. Further, our investigation into musical habits and preferences will control for cultural differences in frequency and type of musical engagement.

The use of a number of short self-report questionnaires, may also lead to a several limitations. Although self-report questionnaires have been accused of being especially prone to participant bias in individual studies (Adams et al., 1999; Donaldson and Grant-Vallone, 2002), Chan's (2009) chapter reviewing the literature finds that, although inflation of the observed correlation is a possibility in self-report literature it is not a necessity. While more comprehensive questionnaires often have superior psychometric properties, we have chosen to use shorter questionnaires, in order to prevent extra burden on the participant. This is especially important for an elderly population who may experience fatigue related decline at the end of long testing. For this reason, we have also declined a more in depth examination of the role of other leisure activities on emotional well-being in the elderly. However, this is one possible avenue for future research in the area. A limitation when collecting data in different countries is that we might need to exclude some of the scales or items in the analyses if they do not converge when testing for measurement invariance. Moreover, the use of different methods for data collection represents a limitation in this study when comparing groups. Finally, the use of a cross-sectional, correlational design will limit our ability to make causal interpretations from our research.

We expect our project to contribute to knowledge about music preferences and daily habits in elderly people across Europe. Moreover, this knowledge will help to develop a greater understanding of how music relates to emotional well-being in elderly people, and will therefore be useful to better design musical or recreational programs, not only for elderly persons who experience challenges associated with aging, but also for healthy elderly people. Especially during aging, musical activities can help to maintain physical and mental health and cognitive abilities, however, more research needs to be done to understand how the activities are connected with an individual to better design music interventions.

Music therapy has demonstrated its efficacy in a number of circumstances including dementia (Chu et al., 2014), cognitive decline (Mammarella et al., 2007), pain relief (McCaffrey and Freeman, 2003), and depression (Chu et al., 2014). Moreover, music therapy treatment improves global and social functioning in schizophrenia and serious mental disorders, Parkinson's disease and sleep quality (Kamioka et al., 2014). There is a considerable body of research which indicates that music

therapy interventions are most effective when "preferred music" is used, in comparison to relaxing or other music (Sung and Chang, 2005; Mitchell and MacDonald, 2006). A study illustrated the effectiveness of individualized music designed music interventions for two groups of elderly persons with Alzheimer's disease and related disorders (Gerdner, 2000). One received individualized music and classical relaxation music, the other group received the same protocol in reverse order. Reduction in agitation was shown only during and following individualized music compared to classical relaxation music (Gerdner, 2000). However, elderly people might not be comfortable being involved in some music therapy interventions, as depicted the study by Burns et al. (2005). They suggest that people might be interested in music therapy involving music listening, but not music-making.

CONCLUSION

Individual needs have to be taken into account while designing such interventions. Therefore, by using an exploratory method to identify musical preferences and the role of music in the elderly, we hope to contribute to more effective and targeted music therapy interventions, and to further the use of music to enhance emotional well-being in the elderly.

AUTHOR CONTRIBUTIONS

This study was conceived and initially designed by JG-S. All authors contributed equally to the research design and to the preparation of the manuscript.

ACKNOWLEDGMENTS

The study was conducted as part of the Junior Researcher Programme (JRP). We would like to thank Dr. Laura Ferreri, Joanna Sierpowska, and Lucía Vaquero for their helpful comments. We also want to thank the native speakers and the expert panel who participated in the translation of the survey. This article was supported by the Open Access Publishing Fund of the University of Vienna.

REFERENCES

- Adams, A. S., Soumerai, S. B., Lomas, J., and Ross-Degnan, D. (1999). Evidence of self-report bias in assessing adherence to guidelines. *Int. J. Qual. Health Care.* 11, 187–192. doi: 10.1093/intqhc/11.3.187
- Adams, K. B., Leibbrandt, S., and Moon, H. (2011). A critical review of the literature on social and leisure activity and wellbeing in later life. *Ageing Soc.* 31, 683–712. doi: 10.1017/S0144686X10001091
- Alexopoulos, G. S. (2005). Depression in the elderly. *Lancet* 365, 1961–1970. doi: 10.1016/S0140-6736(05)66665-2
- Argstatter, H. (2016). Perception of basic emotions in music: culture-specific or multicultural? *Psychol. Music* 44, 674–690. doi: 10.1177/0305735615589214
- Blood, A. J., and Zatorre, R. J. (2001). Intensely pleasurable responses to music correlate with activity in brain regions implicated in reward and emotion. *Proc. Natl. Acad. Sci. U.S.A.* 98, 11818–11823. doi: 10.1073/pnas.191355898
- Boarini, R., Comola, M., Smith, C., Manchin, R., and De, F. (2012). What makes for a better life? The determinants of subjective well-being in OECD countries - evidence from the Gallup World Poll. *OECD Stat. Work. Pap.* 3, 1–41. doi: 10.1787/5k9b91tjm937-en
- Boer, D., and Fischer, R. (2010). Towards a holistic model of functions of music listening across cultures: a culturally decentred qualitative approach. *Psychol. Music* 40, 179–200. doi: 10.1177/0305735610381885
- Brazier, J. E., Harper, E., Jones, N. M., O'Cathain, A., Thomas, K. J., Usherwood, T., et al. (1992). Validating the SF-36 health survey questionnaire: new outcome measure for primary care. *BMJ* 305, 160–164. doi: 10.1136/bmj.305.6846.160
- Burns, D. S., Sledge, R. B., Fuller, L. A., Daggy, J. K., and Monahan, P. O. (2005). Cancer patients' interest and preferences for

- music therapy. *J. Music Ther.* 42, 185–199. doi: 10.1093/jmt/42.3.185
- Campbell-Sills, L., and Stein, M. B. (2007). Psychometric analysis and refinement of the Connor-Davidson resilience scale (CD-RISC): validation of a 10-item measure of resilience. *J. Trauma. Stress* 20, 1019–1028. doi: 10.1002/jts.20271
- Chan, D. (2009). “So why ask me? Are self-report data really that bad,” in *Statistical and Methodological Myths and Urban Legends: Doctrine, Verity, and Fable in the Organizational and Social Sciences*, eds C. E. Lance and R. J. Vandenberg (New York, NY: Routledge), 309–336.
- Chang, P. J., Wray, L., and Lin, Y. (2014). Social relationships, leisure activity, and health in older adults. *Health Psychol.* 33, 516–523. doi: 10.1037/hea0000051
- Chu, H., Yang, C. Y., Lin, Y., Ou, K. L., Lee, T. Y., O'Brien, A. P., et al. (2014). The impact of group music therapy on depression and cognition in elderly persons with dementia: a randomized controlled study. *Biol. Res. Nurs.* 16, 209–217. doi: 10.1177/1099800413485410
- Cohen, A., Bailey, B., and Nilsson, T. (2002). The importance of music to seniors. *Psychomusicology* 8, 89–102. doi: 10.1037/h0094049
- Connor, K. M., and Davidson, J. R. (2003). Development of a new resilience scale: the Connor-Davidson resilience scale (CD-RISC). *Depress. Anxiety* 18, 76–82. doi: 10.1002/da.10113
- Croom, A. M. (2012). Music, neuroscience, and the psychology of well-being: a précis. *Front. Psychol.* 2:393. doi: 10.3389/fpsyg.2011.00393
- Cross, I. (2001). Music, cognition, culture, and evolution. *Ann. N. Y. Acad. Sci.* 930, 28–42. doi: 10.1111/j.1749-6632.2001.tb05723.x
- DeNora, T. (2000). *Music in Everyday Life*. Cambridge: Cambridge University Press.
- Diener, E., Emmons, R. A., Larsen, R. J., and Griffin, S. (1985). The satisfaction with life scale. *J. Pers. Assess.* 49, 71–75. doi: 10.1207/s15327752jpa4901_13
- Diener, E., Gohm, C. L., Suh, E., and Oishi, S. (2000). Similarity of the relations between marital status and subjective well-being across cultures. *J. Cross Cult. Psychol.* 31, 419–436. doi: 10.1177/0022022100031004001
- Dodge, R., Daly, A. P., Huyton, J., and Sanders, L. D. (2012). The challenge of defining wellbeing. *Int. J. Wellbeing* 2, 222–235. doi: 10.5502/ijw.v2i3.4
- Donaldson, S. I., and Grant-Vallone, E. J. (2002). Understanding self-report bias in organizational behavior research. *J. Bus. Psychol.* 17, 245–260. doi: 10.1023/A:1019637632584
- Gerdner, L. A. (2000). Effects of individualized versus classical “relaxation” music on the frequency of agitation in elderly persons with Alzheimer’s disease and related disorders. *Int. Psychogeriatr.* 12, 49–65. doi: 10.1017/S1041610200006190
- Gibbons, A. C. (1977). Popular music preferences of elderly people. *J. Music Ther.* 14, 180–189. doi: 10.1093/JMT/14.4.180
- Golden, J., Conroy, R. M., Bruce, I., Denihan, A., Greene, E., Kirby, M., et al. (2009). Loneliness, social support networks, mood and wellbeing in community-dwelling elderly. *Int. J. Geriatr. Psychiatry* 24, 694–700. doi: 10.1002/gps.2181
- Hays, T., and Minichiello, V. (2005a). The meaning of music in the lives of older people: a qualitative study. *Psychol. Music* 33, 437–451. doi: 10.1177/0305735605056160
- Hays, T., and Minichiello, V. (2005b). The contribution of music to quality of life in older people: an Australian qualitative study. *Ageing Soc.* 25, 261–278. doi: 10.1017/S0144686X04002946
- Herzog, A., Franks, M. M., Markus, H. R., and Holmberg, D. (1998). Activities and well-being in older age: effects of self-concept and educational attainment. *Psychol. Aging* 13, 179–185. doi: 10.1037/0882-7974.13.2.179
- Juslin, P. N., and Västfjäll, D. (2008). Emotional responses to music: the need to consider underlying mechanisms. *Behav. Brain Sci.* 31, 559–621. doi: 10.1017/S0140525X08005293
- Kahneman, D., and Deaton, A. (2010). High income improves evaluation of life but not emotional well-being. *Proc. Natl. Acad. Sci. U.S.A.* 107, 16489–16493. doi: 10.1073/pnas.1011492107
- Kamioka, H., Tsutani, K., Yamada, M., Park, H., Okuizumi, H., Tsuruoka, K., et al. (2014). Effectiveness of music therapy: a summary of systematic reviews based on randomized controlled trials of music interventions. *Patient Prefer. Adherence* 8, 727–754. doi: 10.2147/PPA.S61340
- Kaufmann, E. A., Xia, M., Fosco, G., Yaptangco, M., Skidmore, C. R., and Crowell, S. E. (2015). The difficulties in emotion regulation scale short form (DERS-SF): validation and replication in adolescent and adult samples. *J. Psychopath. Behav. Assess.* 38, 443–455. doi: 10.1007/s10862-015-9529-3
- Kercher, K. (1992). Assessing subjective well-being in the old-old the PANAS as a measure of orthogonal dimensions of positive and negative affect. *Res. Aging* 14, 131–168. doi: 10.1177/0164027592142001
- Kim, Y. E., Schmidt, E. M., Migneco, R., Morton, B. G., Richardson, P., Scott, J., et al. (2010). “Music emotion recognition: a state of the art review,” in *Proceedings of the 2010 International Society for Music Information Retrieval Conference*, Utrecht, 255–266.
- Koelsch, S. (2014). Brain correlates of music-evoked emotions. *Nat. Rev. Neurosci.* 15, 170–180. doi: 10.1038/nrn3666
- Kuykendall, L., Tay, L., and Ng, V. (2015). Leisure engagement and subjective well-being: a meta-analysis. *Psychol. Bull.* 141, 364–403. doi: 10.1037/a0038508
- Laukka, P. (2007). Uses of music and psychological well-being among the elderly. *J. Happiness Stud.* 8, 215–241. doi: 10.1007/s10902-006-9024-3
- Lee, G. R., and Ishii-Kuntz, M. (1987). Social interaction, loneliness, and emotional well-being among the elderly. *Res. Aging* 9, 459–482. doi: 10.1177/0164027587094001
- Lonsdale, A. J., and North, A. C. (2011). Why do we listen to music? A uses and gratifications analysis. *Br. J. Psychol.* 102, 108–134. doi: 10.1348/000712610X506831
- Luanagh, C. O., and Lawlor, B. A. (2008). Loneliness and the health of older people. *Int. J. Geriatr. Psychiatry* 23, 1213–1221. doi: 10.1002/gps.2054
- MacDonald, R. A. R. (2013). Music, health, and well-being: a review. *Int. J. Qual. Stud. Health Well Being* 8, 20635. doi: 10.3402/qhw.v8i0.20635
- Mammarella, N., Fairfield, B., and Cornoldi, C. (2007). Does music enhance cognitive performance in healthy older adults? The Vivaldi effect. *Ageing Clin. Exp. Res.* 19, 394–399. doi: 10.1007/BF03324720
- Mas-Herrero, E., Marco-Pallares, J., Lorenzo-Seva, U., Zatorre, R. J., and Rodriguez-Fornells, A. (2013). Individual differences in music reward experiences. *Music Percept.* 31, 118–138. doi: 10.1525/MP.2013.31.2.118
- McCaffrey, R., and Freeman, E. (2003). Effect of music on chronic osteoarthritis pain in older people. *J. Adv. Nurs.* 44, 517–524. doi: 10.1046/j.0309-2402.2003.02835.x
- Mitchell, L. A., and MacDonald, R. A. R. (2006). An experimental investigation of the effects of preferred and relaxing music listening on pain perception. *J. Music Ther.* 43, 295–316. doi: 10.1093/jmt/43.4.295
- Oguz, S., Merad, S., and Snape, D. (2013). *Measuring National Well-Being: What Matters Most to Personal Well-Being?*. Newport: Office for National Statistics, 1–59.
- Rammstedt, B., and John, O. P. (2007). Measuring personality in one minute or less: a 10-item short version of the Big Five Inventory in English and German. *J. Res. Pers.* 41, 207–212. doi: 10.1016/j.jrp.2006.02.001
- Rilling, J. K., Gutman, D. A., Zeh, T. R., Pagnoni, G., Berns, G. S., and Kilts, C. D. (2002). A neural basis for social cooperation. *Neuron* 35, 395–405. doi: 10.1016/S0896-6273(02)00755-9
- Ruud, E. (1997). Music and identity. *Nord. J. Music Ther.* 6, 3–13. doi: 10.1080/08098139709477889
- Ryff, C. D., and Keyes, C. L. M. (1995). The structure of psychological well-being revisited. *J. Pers. Soc. Psychol.* 69, 719–727. doi: 10.1037/0022-3514.69.4.719
- Saarikallio, S. (2011). Music as emotional self-regulation throughout adulthood. *Psychol. Music* 39, 307–327. doi: 10.1177/0305735610374894
- Saarikallio, S. H. (2008). Music in mood regulation: initial scale development. *Music. Sci.* 12, 291–309. doi: 10.1177/102986490801200206
- Schäfer, T., Tipandjan, A., and Sedlmeier, P. (2012). The functions of music and their relationship to music preference in India and Germany. *Int. J. Psychol.* 47, 370–380. doi: 10.1080/00207594.2012.688133
- Siedlecki, K. L., Salthouse, T. A., Oishi, S., and Jeswani, S. (2014). The relationship between social support and subjective well-being across age. *Soc. Indic. Res.* 117, 561–576. doi: 10.1007/s11205-013-0361-4
- Silverstein, M., and Parker, M. G. (2002). Leisure activities and quality of life among the oldest old in Sweden. *Res. Aging* 24, 528–547. doi: 10.1177/0164027502245003
- Sung, H. C., and Chang, A. M. (2005). Use of preferred music to decrease agitated behaviours in older people with dementia: a review of the literature. *J. Clin. Nurs.* 14, 1133–1140. doi: 10.1111/j.1365-2702.2005.01218.x

- Vergheze, J., Lipton, R. B., Katz, M. J., Hall, C. B., Derby, C. A., Kuslansky, G., et al. (2003). Leisure activities and the risk of dementia in the elderly. *N. Engl. J. Med.* 348, 2508–2516. doi: 10.1056/NEJMoa022252
- Ware, J. E., and Sherbourne, C. D. (1992). The MOS 36-item short-form health survey (SF-36): I. Conceptual framework and item selection. *Med. Care* 30, 473–483.
- Watson, D., Clark, L. A., and Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: the PANAS scales. *J. Pers. Soc. Psychol.* 54, 1063–1070. doi: 10.1037/0022-3514.54.6.1063
- Zimet, G. D., Dahlem, N. W., Zimet, S. G., and Farley, G. K. (1988). The multidimensional scale of perceived social support. *J. Pers. Assess.* 52, 30–41. doi: 10.1207/s15327752jpa5201_2

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2017 Grau-Sánchez, Foley, Hlavová, Muukkonen, Ojinaga-Alfageme, Radukic, Spindler and Hundevad. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



‘Talkin’ ‘Bout My Generation’: Using a Mixed-Methods Approach to Explore Changes in Adolescent Well-Being across Several European Countries

Alina Cosma^{1*}, Jelisaveta Belić², Ondřej Blecha³, Friederike Fenski⁴, Man Y. Lo⁵, Filip Murár⁵, Darija Petrovic⁶ and Maria T. Stella⁷

¹ Child and Adolescent Health Research Unit, School of Medicine, University of St Andrews, St Andrews, United Kingdom, ² Department of Psychology, Faculty of Social and Behavioural Sciences, Leiden University, Leiden, Netherlands, ³ Department of Psychology, Faculty of Arts, Masaryk University, Brno, Czechia, ⁴ Department of Psychology, Faculty of Education and Psychology, Free University of Berlin, Berlin, Germany, ⁵ Division of Psychology and Language Sciences, Faculty of Brain Sciences, University College London, London, United Kingdom, ⁶ Department of Psychology, Faculty of Philosophy, University of Novi Sad, Novi Sad, Serbia, ⁷ Department of Psychology and Cognitive Science, University of Trento, Trento, Italy

OPEN ACCESS

Edited by:

Pietro Cipresso,
Istituto Auxologico Italiano (IRCCS),
Italy

Reviewed by:

Fanli Jia,
Seton Hall University, United States
Carol Van Hulle,
University of Wisconsin–Madison,
United States

*Correspondence:

Alina Cosma
apc8@st-andrews.ac.uk

Specialty section:

This article was submitted to
Quantitative Psychology
and Measurement,
a section of the journal
Frontiers in Psychology

Received: 07 November 2016

Accepted: 25 April 2017

Published: 18 May 2017

Citation:

Cosma A, Belić J, Blecha O,
Fenski F, Lo MY, Murár F, Petrovic D
and Stella MT (2017) ‘Talkin’ ‘Bout
My Generation’: Using
a Mixed-Methods Approach
to Explore Changes in Adolescent
Well-Being across Several European
Countries. *Front. Psychol.* 8:758.
doi: 10.3389/fpsyg.2017.00758

The promotion of positive mental health is a becoming priority worldwide. Despite all the efforts invested in preventive and curative work, it is estimated that one in four persons will experience a mental health condition at some point in their lives. Even more worrying is the fact that up to a half of all mental health problems have their onset before the age of 14. Recent statistics (national and international surveys, meta-analyses, international reports) point out to the fact that child and adolescent mental health problems are on the rise. The present study will try to corroborate these results and further explore their meaning, by employing a sequential mixed methods research design (quantitative–qualitative). The quantitative part will analyze time trends using Health Behaviors in School-aged Children data (four survey cycles: 2002, 2006, 2010, 2014) on mental well-being from four European countries (the Czechia, Germany, Italy, and United Kingdom). The qualitative part will rely on focus groups to explore the perspectives of 13- and 15-year-old boys and girls on gender differences and on the changes in adolescent mental well-being over time, as well as measures through which these issues could be addressed. Thematic analysis will be employed to analyze qualitative data. The results of this study could make a major contribution to our understanding of the current trends in adolescent mental well-being, as well as the ways in which existing data could be linked to international and national health policies.

Keywords: adolescence, mental well-being, time trends, mixed methods study, HBSC

INTRODUCTION

The period of adolescence is fundamental for three reasons. Firstly, a healthy adolescence allows for the acquisition of certain developmental tasks, for example emotional and cognitive abilities; secondly, it can be viewed as the period for laying down health behaviors which determine the future health of such individuals; finally, adolescents will be parents of future generations so a

Abbreviations: EU, European Union; HBSC, Health Behaviours in School-aged Children; UK, United Kingdom.

healthy cohort of adolescence can lead to a future of healthy parents and children (Patton et al., 2016). All these premises are highly inter-related with mental health. For the purpose of this paper, the WHO conceptualization of mental health will be employed: “a state of well-being in which the individual realized his or her own abilities, can cope with normal stresses of life, can work productively and fruitfully, and is able to make a contribution to his or her community” (WHO, 2001).

There is a growing body of evidence showing that adolescence appears to be when many psychiatric disorders begin (Kessler et al., 2005). Mental health problems may influence a child’s learning and academic performance in school (Stoep et al., 2003) and have been linked to psychiatric disorders throughout adolescence and adulthood (Costello et al., 2003). Despite this, one in ten young people aged five to sixteen appear to experience a diagnosable mental health problem (Green et al., 2004), reflecting a need to pay greater attention to improving child and adolescent mental health.

Not only are a significant proportion of young people experiencing mental health problems, but, according to data from the HBSC study, it appears that mental well-being of adolescents also appears to be changing across cohorts. The HBSC collects cross-sectional data every four years on 11-, 13-, and 15-year-olds across more than 40 European and North American countries, measuring health and well-being, social environments, and health behaviors. Its advantage, therefore, is that it allows cross-cultural comparisons in data and a look at time trends in adolescent mental well-being. For example, using HBSC data from 35 European and North American countries, Ottová-Jordan et al. (2015) analyzed time trends (from 1994 to 2010) in mental well-being and found that in seven countries (Croatia, Greece, FYR Macedonia, Portugal, Slovenia, Spain, and Switzerland), there was a steady decline across cohorts; in five countries (Flemish Belgium, Denmark, Finland, Greenland, and Norway) a linear increase was found; in four countries (Austria, Canada, Czechia, and Scotland) a U-shaped trend was identified; in six countries (England, Estonia, Lithuania, Poland, Slovakia, and Sweden), an inverted U-shaped trend and unstable patterns in the remaining countries.

In a review that included studies from 12 countries, Bor et al. (2014) found that mental health symptoms in cohorts of children and toddlers generally improve or do not change. On the other hand, mental health trends in adolescents seem to be changing, and seemingly in a negative direction. There were no changes in terms of externalizing problems but evidence exists to suggest that internalizing problems are on the rise in cohorts of adolescents, particularly in adolescent girls. Parent-reported emotional problems in adolescents have also shown increases (Tick et al., 2007; Collishaw, 2015). Sweeting et al. (2010) collected data on a cohort of 15-year-olds in 1987 and compared it to a cohort in 2006, controlled for age, school year, and geographical location (West of Scotland). An increase in self-reported ‘psychological distress’ as measured on the General Health Questionnaire was found; this increase was significant for adolescent boys and girls but higher in girls. On the other hand, emotional problems in children appear to have decreased or not changed (Tick et al., 2007; Hölling et al.,

2014; Matijasevich et al., 2014). Moreover, some studies used a qualitative approach to explore adolescents’ perspectives on well-being, yet it is difficult to find either cross-cultural qualitative studies in which adolescents are asked to interpret observed declining trends in well-being. A recent Navarro et al. (2015) suggests family relationships, peer to peer relationships, and school-related aspects to be the main topics adolescents refer to interpret reduced well-being. In a study on Spanish adolescents (Casas et al., 2012) physical safety, physical exercise, as well as life changes are emerging themes.

However, trends showing increasing mental health problems in adolescence may reflect either a genuine prevalence of emotional problems or increased reporting by adolescence, better recognition and reporting by parents and teachers, and better diagnoses techniques. Evidence suggesting that there is a genuine change comes from cross-cohort comparisons studies, which show increased self-reported depression and anxiety symptoms in adolescents since the 1980s in countries such as Greece, Germany, New Zealand Scotland, and England (Sweeting et al., 2009; Fleming et al., 2014; Collishaw, 2015). Similar studies in lower-income countries are lacking and would add to the amount of supporting evidence. Additionally, Collishaw (2015) reviews agreeing evidence from numerous study methods, informants and trends suggesting changes in some mental health problems, but not others, which would suggest genuine changes.

What new challenges are young people facing today that appear to be contributing to the deteriorating trends seen in adolescent mental health? There is yet to be a clear explanation; however, some possible explanations may include changes in individual vulnerability, socioeconomic and cultural factors, family life, and extrafamilial psychosocial factors (Collishaw, 2015). For example, there is evidence that the current generation of girls are experiencing puberty earlier than previous generations and this may coincide with a greater risk of developing depression (Bor et al., 2014). Moreover, social problems such as youth unemployment and economic recession may contribute to an increased risk of substance misuse and mental disorders in young people (Sawyer et al., 2012). Other studies have also found links between poverty and increased problems in adolescent mental health (Bradley and Corwyn, 2002; Costello et al., 2003; Reiss, 2013). Increasing rates of single parenting and other changes in the family environment may be contributory factors too (Bor et al., 2014). The increasing relevance that social media have in lives of young people may affect their mental well-being as well (Sawyer et al., 2012). Others suggest that lifestyle changes, for example increased sugar consumption, have accelerated rates of depression (Haroon et al., 2012).

Generational differences in adolescents are further complemented by gender differences where girls appear more likely to experience mental health issues than boys (Nolen-Hoeksema and Girgus, 1994; Salk et al., 2016). Possible explanations include differences in global self-esteem (Nolen-Hoeksema and Girgus, 1994), social media use (Monro and Huon, 2005), academic pressure (Byrne et al., 2007), or earlier sexualization, which may be related to worse self-esteem and depressed mood (Hatch, 2011). Many international and national policies, such as The European Mental Health Action

Plan 2013–2020¹, The European child and adolescent health strategy 2015–2020², the Child and Adolescent Mental Health in Europe (CAMHEE) (Braddick et al., 2009), and the European Framework for Action on Mental Health and Wellbeing (Dahlgren and Whitehead, 2006) have been proposed to improve mental health and well-being in young people. However, for such targets to be fulfilled, understanding mental health in young people is essential to assuring that aims are met.

Thus, in this study, we aim to further assess the evidence that changes in cohorts of adolescents' mental well-being are occurring and to gain an insight into the understanding of today's adolescents on the reasons as to why adolescent mental well-being may be decreasing over time. To answer these research questions, this study will employ a mixed methods design (quantitative–qualitative). The quantitative part will aim to identify time trends in adolescent mental well-being across different European countries from 2002 to 2014. To do this, the quantitative part of the present study will rely on running secondary analysis of time trends using HBSC data from the United Kingdom (England, Wales, and Scotland), the Czechia, Germany, and Italy. Specific hypotheses which will be tested through the quantitative part will be the following:

Hypothesis 1: *Based on previous findings, we predict that adolescent mental well-being indicators are decreasing across the cohorts in each of the examined countries.*

Hypothesis 2: *We expect to find a stable gender difference in mental well-being indicators with girls showing worse outcomes than boys.*

Hypothesis 3: *We expect to observe an age effect, the older cohort (15-year olds) will show worse outcomes than the younger cohort.*

The qualitative part will explore young people's perspective on the current trends in adolescent mental well-being (more specifically, the deteriorating mental well-being overtime). To explore these perceptions, series of focus groups with 13- and 15-year-olds will be run in the Czechia, Germany, England, Italy, and Serbia.

To summarize, the present mixed methods study will aim to answer the following two main research questions: *What are possible explanations for the observed trends indicating deteriorating adolescent mental well-being? And why are there gender differences in mental health trends?*

Current policies aiming to improve childhood and adolescent psychological health involve trying to identify indicators of poor psychological well-being at an earlier stage which may increase the success of treatment, increasing accessibility to mental health services for children and adolescents and reducing the stigma surrounding the topic of mental health. Ultimately, we hope that

our findings will be able to impact public health policies around child and adolescent mental well-being, either by supporting the implementation of certain policies or revealing other areas of improvement.

MATERIALS AND METHODS

Mixed Methods Design

In the present study, a mixed methods approach will be employed, with a sequential quantitative–qualitative design (Tashakkori and Teddlie, 2003) (See **Figure 1**). A secondary analysis of quantitative data collected in the HBSC study³ (Inchley et al., 2016) will be performed to assess recent time trends in adolescent mental well-being across six European countries (England, Wales, Scotland, Czechia, Germany, and Italy). Qualitative data will be collected in focus groups

³Open access is available the 2001/02, 2005/06, and 2009/10 HBSC datasets through the HBSC Data Management Center, based at the Norwegian Centre for Research Data (NSD). Because the datasets are restricted for the use of HBSC member country teams for a period of three years from survey completion, the 2013/2014 HBSC dataset is only scheduled for open access release by June 2018. However, aggregated datasets from the 2013/14 survey are available through the WHO Europe's Health for All database (<https://gateway.euro.who.int/en/data-sources/hbsc/>). Further data is available for external use by agreement with the HBSC International Coordinator and the Principal Investigators.

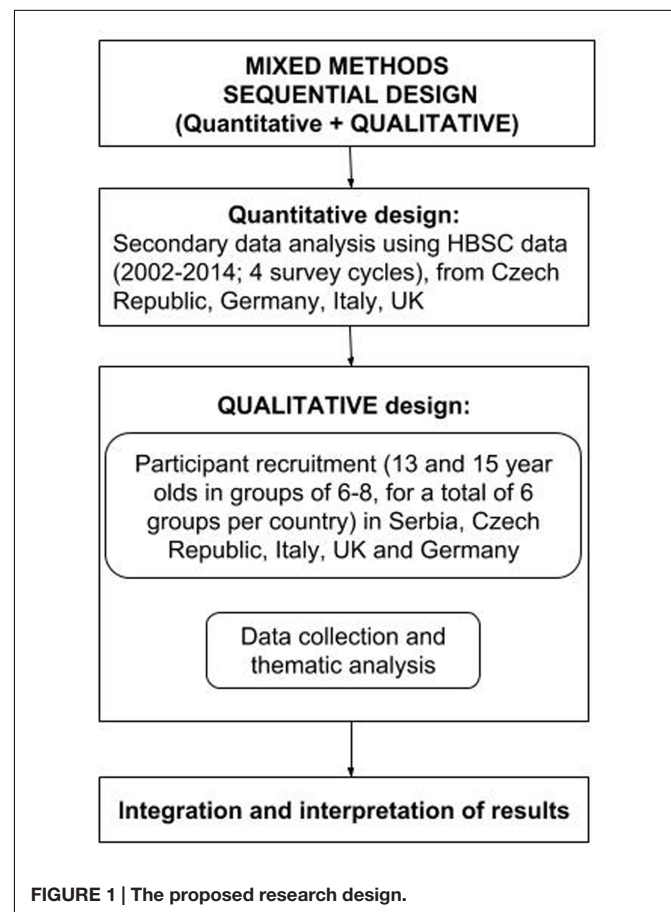


FIGURE 1 | The proposed research design.

¹"The European Mental Health Action Plan 2013–2020 – WHO/Europe" 2015. 24 Sep. 2016. <http://www.euro.who.int/__data/assets/pdf_file/0020/280604/WHO-Europe-Mental-Health-Acion-Plan-2013-2020.pdf>

²"Investing in children: the European child and adolescent health..." 2014. 24 Sep. 2016. <<http://www.euro.who.int/en/health-topics/Life-stages/child-and-adolescent-health/policy/investing-in-children-the-european-child-and-adolescent-health-strategy-20152020>>

with adolescent girls and boys to explore their perspective on the phenomena underlying changes in mental well-being and the challenges they experience. A sequential explanatory quantitative–qualitative design is needed since using quantitative data alone accurately shows emerging changes, but leaves us with limited practical guidelines relating to mental health policies. Therefore, applying a qualitative approach allows directly affected parties, in this case- adolescents, to become our main resource for themes which need to be addressed to understand observed changes in their mental well-being and to allow remedial interventions to be recommended.

Psychosomatic Complaints

To analyze trends in mental well-being, the Psychosomatic Complaints measure of the HBSC will be used. This item consists of a series of questions where participants indicate the frequency with which they had experienced the following eight health complaints over the last six months; “headache”, “stomach ache”, “backache”, “feeling low”, “irritability or bad temper”, “feeling nervous”, “difficulties in getting to sleep” and “feeling dizzy”, (0 = “Rarely or never”, 1 = “About every month”, 2 = “About every week”, 3 = “More than once a week”, 4 = “About every day”). Responses across all eight complaints will be summed to generate a single score between 0 and 32, with higher values reflecting a greater psychosomatic complaint burden. This scale has undergone extensive qualitative and quantitative validation and shows good test-retest reliability and unidimensionality (Haugland et al., 2001). Reporting psychosomatic complaints is an important indicator for measuring subjective well-being, as it reflects personal experience related to negative life events in the social context of school, peers and family (Due et al., 2005; Hjern et al., 2008; Ottová-Jordan et al., 2015). They may also indicate a more serious underlying health problem (Ihlebaek et al., 2002). As opposed to some other well-being-related HBSC items, subjective health complaints are a sensitive measure (Ravens-Sieberer et al., 2009; Eriksson and Sellström, 2010), showing both individual-level (Ottová-Jordan et al., 2015) and international variation (Torsheim et al., 2006). Since somatic symptoms identified in late childhood and adolescence tend to persist into adulthood (Brattberg, 2004; Steinhausen and Winkler Metzke, 2007) and they may be predictive of somatization and anxiety symptoms in early adulthood (Kinnunen et al., 2010), psychosomatic symptoms in adolescence are an important indicator of mental health.

Ethics Statement

For the secondary quantitative analysis, every HBSC participating country included in the present study (Czechia, Germany, England, Italy, Scotland, and Wales) has been given ethical clearance by national and regional governing bodies. For the qualitative part, ethical clearance will be obtained at national level by every research team member (Czechia, Germany, England, Italy, Serbia).

Stepwise Procedure

We will conduct a mixed methods study with a sequential design which consists of the quantitative analysis followed by

a qualitative investigation. Using secondary data analysis of already-collected survey data, the quantitative part will assess and describe changes in mental well-being of adolescents nowadays. It will be carried out by using the HBSC data from United Kingdom, Italy, Germany, and Czechia. The qualitative section of the present study will involve conducting focus groups with adolescents. This approach will allow for an in-depth exploration of possible explanations for the overall deteriorating trend of adolescent mental well-being. Tashakkori and Teddlie’s (2003) mixed methods sequential procedure will serve as a model for the methodological design for our study. One of the main assumptions of this procedure is using a quantitative approach to test theories, followed by a qualitative method that involves a detailed exploration with a few individuals. For example, Hodgkin (2008) used a sequential quantitative–QUALITATIVE approach to study women’s social capital. First, the author conducted a survey to identify different social capital profiles in the population. In a second step, Hodgkin (2008) used in-depth interviews to illuminate the stories behind these profiles with a few participants. The author argues, that the sequential approach allows to give a more powerful voice to groups that are not heard otherwise, in our case it will be the young people themselves.

Quantitative Method

Data from four rounds of the international HBSC study will be used (2002, 2006, 2010, 2014). The HBSC study is a cross-sectional study of adolescent health carried out every four years across several European countries. At the moment, the HBSC network includes research teams from 45 countries across Europe and North America. Data collection is based on a standardized research protocol which specifies sampling methods and questionnaire content across all participating countries. For each survey round, countries collect a nationally representative sample of 11-, 13-, and 15-year-olds, with the timing of fieldwork arranged to achieve mean ages of 11.5, 13.5, and 15.5. Participants from each country were recruited via stratified random cluster sampling, with whole school classes as the sampling unit. Adolescents completed questionnaires in classroom settings, and could leave any question blank that they did not want to answer. Questionnaires were translated from English into respective national languages with back-translation checks. Appropriate ethical consent was gained in each participating country, with schools and adolescents giving active informed consent. For the current study, data will be collected from the Czechia, England, Germany, Italy, Scotland, and Wales, focusing on psychosomatic complaints. Responses across all eight complaints of the Psychosomatic Complaints item will be summed to generate a single score between 0 and 32 (with higher values reflecting a greater psychosomatic complaint burden), which will be used in a linear regression analysis.

Qualitative Method

A qualitative approach in psychology is used mainly when a topic area needs to be further explored, for example when the investigated topic is complex, to empower individuals, to develop theories, and when mainstream quantitative measures simply do not fit the problem and do not provide us with meaningful

information (Maxwell, 2009). In this study, by exploring the adolescents' perspective on this issue, young people are identified as the most important and concerned party and their voices are empowered, making the qualitative approach appropriate for our aims. The qualitative data collection will be done through the use of focus groups. This method was selected by the research team due to its explorative nature (Sim and Snell, 1996). It enables the analysis of the attitudes, opinions and insights of adolescents on the research questions, to better understand the observed trends and give directions for future research.

Participants will be asked different questions regarding adolescence and mental health (please refer to Supplementary Material) to understand the adolescents' perspective on mental well-being trends and the observed generational and gender differences in the trends. The focus groups will also provide us with some representation of what being an adolescent today consists of, including its positive and negative attributes.

Previous HBSC results showed a significant decrease in adolescent mental well-being indicators in both older age groups (13, and especially for the 15- years-olds). These two groups were chosen in the light of this information to be the participants of the focus groups. Their insights will allow us to improve our understanding of the challenges they are facing. The changes in mental well-being observed in 11- year- olds were largely negligible in the past research; therefore, this age group will not be included in the qualitative data collection. Participants in the qualitative part of the research will be 13- and 15- years-old schoolchildren in five participating countries – England, Germany, Czechia, Italy, and Serbia. A convenience sample from each country will be employed. The choice of the countries has been limited to these in particular due to the nature of the Junior Researcher Programme which represents the basic platform for the current study. The Junior Researcher Programme context in which the current study had its onset played the curtail role in the choice of the countries. The aim is to recruit a minimum of 36 adolescents per country. In order to achieve this, six focus groups with six to eight participants per country will be conducted. There will be three different focus group arrangements in both age groups: the single gender focus groups and mixed gender group. The reasoning behind this decision is based on the peculiarity of communication between boys and girls at this age, which may lead them to different answers while discussing the same topic in same-gender or mixed-gender setting.

The focus groups will be conducted in accordance with the focus group questions. This will be developed based on the latest theories and findings about adolescent mental well-being. Questions will originally be developed in English. In next phase the focus group guide will be translated to national languages (German, Italian, Czech, and Serbian) and back-translated to English. Prior to starting the data collection phase, a pilot study of the focus group guide will be conducted in every country. Through this, we aim to assess the overall flow of the questions, wording, question comprehension etc., before the finalization of the questions and the beginning of the field work. Before conducting the focus groups in each country, the recordings of the pilot interviews will be analyzed by the research team in order to ensure that the moderator has not influenced the participants'

in any way, such as asking question in a leading way. In addition, the questions developed for the discussion have been tailored to guarantee that the participants give their opinion on determined sub-topics of our interest, without having their views influenced by the moderator.

Once the participants have been recruited, the meetings will be arranged. Neutral locations are strongly preferred to avoid either positive or negative associations to the site or building. Focus group sessions will last approximately 60 to 90 minutes. At the beginning of every session, the moderator will provide a clear explanation of the purpose of the focus group. Once the purpose of the discussion is clear, participants will be introduced to the basic rules of conduct during the focus group discussions. Participants will be provided with an informed consent form, stating that they agree to take part in research. Within the consent, they will also be asked to agree with an audio-recording of the discussion.

Throughout the session, the moderator will initiate the discussion points and facilitate the exchange of ideas between participants. Depending on the group's dynamics, moderators might need to promote debate, seek clearance, probe for details, or move the debate forward if the conversation loses its focus. Special attention will be given to ensure that each participant is given equal opportunity to speak. The moderators will be expected to avoid demonstration of personal opinions or any kind of preferences that could influence the participants' train of thoughts. Participants will be informed of their valued position as experts on the matter at hand to motivate and empower them. They will also be informed of how unique this chance is to work collaboratively with the researchers and how they will be contributing to the greater good of their generation, and of those to come.

Proposed Analyses

Analysis of Questionnaire Data

The analysis of questionnaire data will follow previous procedures used in other publications which used HBSC data (e.g., Whitehead et al., 2017). More specifically, questionnaire data will be stratified into subpopulations by country, and by gender and age within each country. Dataset weights, available with the raw HBSC dataset, will be applied in order to achieve national representativeness of each country at each time point. Linear regression analyses will be conducted using SPSS v.23 complex samples toolkit, which allows shared variance within sampling to be accounted for. The linear and quadratic effects of survey year on psychosomatic complaints will be evaluated for each subpopulation using general linear modeling.

Analysis of Focus Group Data

Qualitative data will be processed using thematic analysis, as defined by Braun and Clarke (2006). Initially, every author will read and re-read transcripts of recordings of the focus groups that they conducted in order to identify codes in the data, i.e., the basic features of the raw data that carry a meaning pertaining to the research questions (Braun and Clarke, 2006). Codes will be associated with quotes capturing the verbatim expression of the interviewed individuals. Analysis at this stage will be

performed in the original languages, to minimize alteration or loss of meaning due to bulk translations. Once no more codes can be identified, the respective authors will sort the codes into semantically related categories, or themes. A set of candidate themes will then be repeatedly revised; initial themes may be split, merged or abandoned altogether while new themes may be formed. The aim will be to reach internal homogeneity and external heterogeneity of themes. As themes are being formed, they will be organized into a thematic map – a visual representation of the identified themes and their relationships – which may again be refined to achieve a map that validly and accurately reflects information contained in the data. In the next phase, codes, quotes and themes will be translated into English. Finally, these themes will be compared and collated to obtain a single thematic map capturing the insights from all the conducted focus groups. Similarities and differences between individual countries will be noted and the most representative quotes for each theme identified.

Anticipated Results

The present study will contribute to the understanding of time trends in adolescent mental well-being across four European countries and offer possible explanations of their change. In contrast to previous studies, this research project will be based on an innovative sequential mixed methods design. This approach will allow for moving beyond the theoretical and statistical description of these time trends, as it will attempt to connect these results to young people's interpretation of them. We expect that through this novel design (quantitative–qualitative), the findings could have a direct impact on future research, health policies, or interventions targeting adolescent mental well-being.

The quantitative design will allow us to identify the magnitude of change in adolescent mental well-being across the investigated countries. In line with previous studies, we firstly expect to find an overall decreasing trend of adolescent mental well-being between 2002 and 2014 across the investigated countries (for a review, see Bor et al., 2014). Furthermore, the countries are likely to differ in the characteristics of their respective time trends as shown in previous studies of mental well-being trends (Ottová-Jordan et al., 2015). Given that mental well-being also decreases with age, we also assume that, across all age groups included in the analyses, the 15-year-olds will show the greatest decline. This trend will be significantly different from both the 13-year-olds and the 11-year-olds (see, e.g., Ottová-Jordan et al., 2015). Finally, in line with results from the international HBSC report (Inchley et al., 2016), we expect to find gender differences, namely girls reporting more frequent psychosomatic complaints than boys across the investigated countries.

While the quantitative analysis will allow us to identify the extent of change in adolescent mental well-being across the investigated countries, the qualitative section of our study will give us the opportunity to directly ask young people about their opinion on the observed trends. Instead of reviewing possible explanations for them, our design will allow us to discuss actual challenges adolescents face with adolescents themselves. Thus, we expect the focus groups to be an environment that gives adolescents a voice to discuss perceptions, ideas, opinions, and

thoughts about topics that are affecting their mental well-being. The aim of this process is to identify the factors underlying the observed deterioration in adolescent mental well-being in recent cohorts. Possible themes adolescents across participating countries may name are school pressure, the role of social media, or the role of family and friends (West and Sweeting, 2003; Collishaw, 2015; Patton et al., 2016). It is expected that instead of only offering possible explanations for the observed trends, adolescents may also provide possible solutions for tackling current adolescent mental well-being difficulties.

Several challenges in running the present study should be mentioned. Because focus groups in five countries will be conducted with five different official languages (Czech, German, Italian, Serbian, and English), it may be difficult to guarantee uniformity of all focus group guides regardless of the language they are written in. To ensure that questions with the same meaning will be asked across all countries, focus group guides originally developed in the English language will be translated to national languages and back-translated again to English. This process will be repeated until all focus group guides contain the same questions. Another challenge could lie in getting access to adolescents in different European countries in order to run the focus groups. Distinct EU Member States have different rules, regulations and guidelines regarding the participation of children in research projects, particularly concerning ethics approval and informed consent. Whilst some countries require only the parents' informed consent (e.g., Czechia), other countries demand additional approval from respective supervisory school authority (e.g., Germany – consent from the federal state government). To ensure a smooth realization of the focus groups, researchers will make themselves familiar with the requirements of each respective country so that necessary documents can be prepared at an early stage. After successfully running the focus groups, an additional challenge will lie in the transcription of the data as well as the translation and retranslation of the transcripts to ensure that the data has been translated correctly.

Using a mixed methods approach allows a deeper understanding of the underlying processes regarding time trends in adolescent well-being. However, there are several limitations associated with the proposed study design. The main weakness of the design is the large amount of time involved in data collection (Creswell, 2009). Because HBSC survey data is being used, data collection is limited to the qualitative part only. However, the process of conducting the focus groups, the transcription of the audio-recorded focus groups, the translation and retranslation of the transcripts, and the analysis of the transcripts are very time-consuming. On the other hand, acquiring an in-depth insight into young people's views on trends of mental well-being undoubtedly outweighs this limitation. Another frequent limitation of the qualitative designs is a lack of validity (Willig, 2008). It is critical that our focus group guide includes every important question to gain understanding of young people's interpretation of time trends in adolescent mental well-being. To increase validity, pilot focus groups will be conducted where adolescents could provide feedback so that changes to our focus group guide can be made if necessary.

Many national and international policies have been proposed to improve mental well-being in young people, often co-founded under the EU Public Health Programs in recent years (Joint Action on Mental Health and Well-being, 2015). Most of these projects were implemented in schools, the elective setting for the prevention and/or promotion of the mental health of children and adolescents (e.g., CAMHEE, SEYLE, SCMHE, SUPREME). Each of the countries involved in our research has a national campaign for youth development and health. In the Czechia, The National Reference Centre of Programmes for Health Promotion and Disease Prevention supports programs focusing on stress coping, on children in crisis, and protection of children against violence (Jané-llopis et al., 2008). “MindMatters” is a program implemented in Germany that promotes mental health in primary and secondary schools (Braddick et al., 2009). The focus in Italy lies in the prevention of risk-taking behaviors through early identification of psychosocial stress, especially in young people, including interventions based on peer education and life skills education (Jané-llopis et al., 2008). The Strategy of Youth Development and Health in the Republic of Serbia 2007–2012 includes aims and activities to improve the quality, efficiency and accessibility of healthcare as well as to find new approaches for improving young people’s health (Expert Group on Youth Development and Health of the Ministry of Health of Serbia, 2006). In the United Kingdom, a Children and Young People’s Mental Health and Wellbeing Taskforce was set up in September 2014 to consider how to make it easier for children, young people, parents and carers to access help and support when they need it and to improve the help that is offered (Department of Health, 2015).

One of the first and most important priorities of the European Child and Adolescent Health Strategy 2015–2020 was to make adolescents’ lives more visible. We hope that our findings will add to this aim. Furthermore, we hope that they will highlight priority areas for action that can be used to inform the development and implementation of intervention and prevention programs. Overall, we anticipate that our results will contribute to the

comprehension of how current adolescents feel and how they perceive themselves in their environments, as well as to connect these insights to European and national public health policies to counteract the continuous trend of deteriorating adolescent mental well-being around Europe.

AUTHOR CONTRIBUTIONS

This paper was created through collaborative activity and substantial intellectual contribution of all the authors listed above. AC proposed and developed the research idea, wrote parts of the manuscript, and coordinated the writing process. All the other authors wrote parts of the manuscript, provided important feedback when revising the manuscript. The publishing approval was given by all the authors.

FUNDING

Open Access fees were supported by the University of St Andrews.

ACKNOWLEDGMENTS

This paper was made possible due to the virtue of Junior Researcher Programme (<http://jrp.psycholars.org/>). We thank all the members of the Programme for their dedication, engagement, and assistance through each step of the project.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fpsyg.2017.00758/full#supplementary-material>

REFERENCES

- Bor, W., Dean, A. J., Najman, J., and Hayatbakhsh, R. (2014). Are child and adolescent mental health problems increasing in the 21st century? a systematic review. *Aust. N. Z. J. Psychiatry* 48, 606–616. doi: 10.1177/0004867414533834
- Braddick, F., Carral, V., Jenkins, R., and Jané-llopis, E. (2009). *Child and Adolescent Mental Health in Europe?: Infrastructures, Policy and Programmes*. Luxembourg: European Communities. doi: 10.1177/0004867414533834
- Bradley, R. H., and Corwyn, R. F. (2002). Socioeconomic status and child development. *Ann. Rev. Psychol.* 53, 371–399. doi: 10.1146/annurev.psych.53.100901.135233
- Brattberg, G. (2004). Do pain problems in young school children persist into early adulthood? a 13-year follow-up. *Eur. J. Pain* 8, 187–199. doi: 10.1016/j.ejpain.2003.08.001
- Braun, V., and Clarke, V. (2006). Using thematic analysis in psychology. *Qual. Res. Psychol.* 3, 77–101. doi: 10.1191/1478088706qp0630a
- Byrne, D. G., Davenport, S. C., and Mazanov, J. (2007). Profiles of adolescent stress: the development of the Adolescent Stress Questionnaire (ASQ). *J. Adolesc.* 30, 393–416. doi: 10.1191/1478088706qp0630a
- Casas, F., Bello, A., González, M., and Aligué, M. (2012). Personal well-being among Spanish adolescents. *J. Soc. Res. Policy* 3, 19. doi: 10.1016/j.adolescence.2006.04.004
- Collishaw, S. (2015). Annual research review: secular trends in child and adolescent mental health. *J. Child Psychol. Psychiatry Allied Discip.* 56, 370–393. doi: 10.1111/jcpp.12372
- Costello, E. J., Mustillo, S., Erkanli, A., Keeler, G., and Angold, A. (2003). Prevalence and development of psychiatric disorders in childhood and adolescence. *Arch. Gen. Psychiatry* 60, 837–844. doi: 10.1001/archpsyc.60.8.837
- Creswell, J. (2009). Research design. *J. Chem. Inf. Model.* 53, 1689–1699. doi: 10.1017/CBO9781107415324.004
- Dahlgren, G., and Whitehead, M. (2006). *European Strategies for Tackling Social Inequities in Health: Levelling up Part 2. Studies on Social and Economic Determinants of Population Health*, Vol. 3. Copenhagen: WHO Regional Office for Europe, 1–105.
- Department of Health (2015). *Future in Mind*. https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/414024/Childrens_Mental_Health.pdf
- Due, P., Holstein, B. E., Lynch, J., Diderichsen, F., Gabhain, S. N., Scheidt, P., et al. (2005). Bullying and symptoms among school-aged children: international

- comparative cross sectional study in 28 countries. *Eur. J. Public Health* 15, 128–132. doi: 10.1093/eurpub/cki105
- Eriksson, U., and Sellström, E. (2010). School demands and subjective health complaints among swedish schoolchildren: a multilevel study. *Scand. J. Public Health* 38, 344–350. doi: 10.1177/1403494810364683
- Expert Group on Youth Development and Health of the Ministry of Health of Serbia (2006). *Strategy for Youth Development and Health in the Republic of Serbia*. Belgrade: Ministry of Health of the Republic of Serbia.
- Fleming, T. M., Clark, T., Denny, S., Bullen, P., Crengle, S., Peiris-John, R., et al. (2014). Stability and change in the mental health of New Zealand secondary school students 2007–2012: results from the national adolescent health surveys. *Aust. N. Z. J. Psychiatry* 48, 472–480. doi: 10.1177/0004867413514489
- Green, H., McGinnity, A., Meltzer, H., and Ford, T. (2004). *Mental Health of Children and Young People in Great Britain*. Basingstoke: Palgrave Macmillan, doi: 10.1037/e557702010-001
- Haroon, E., Raison, C. L., and Miller, A. H. (2012). Psychoneuroimmunology meets neuropsychopharmacology: translational implications of the impact of inflammation on behavior. *Neuropsychopharmacology* 37, 137–162. doi: 10.1038/npp.2011.205
- Hatch, L. (2011). The american psychological association task force on the sexualization of girls: a review, update and commentary. *Sex. Addict. Compulsivity* 18, 195–211. doi: 10.1080/10720162.2011.613326
- Haugland, S., Wold, B., Stevenson, J., Aaroe, L. E., and Woynarowska, B. (2001). Subjective health complaints in adolescence - a cross-national comparison of prevalence and dimensionality. *Eur. J. Public Health* 11, 4–10. doi: 10.1093/eurpub/11.1.4
- Hjern, A., Alfvén, G., and Östberg, V. (2008). School stressors, psychological complaints and psychosomatic pain. *Acta Paediatrica* 97, 112–117. doi: 10.1111/j.1651-2227.2007.00585.x
- Hodgkin, S. (2008). Telling it all: a story of women's social capital using a mixed methods approach. *J. Mix. Methods Res.* 2, 296–316. doi: 10.1177/1558689808321641
- Hölling, H., Schlack, R., Petermann, F., Ravens-Sieberer, U., and Mauz, E. (2014). Psychische auffälligkeiten und psychosoziale beeinträchtigungen bei kindern und jugendlichen im alter von 3 bis 17 jahren in deutschland - prävalenz und zeitliche trends zu 2 erhebungszeitpunkten (2003–2006 und 2009–2012): ergebnisse der KiGGS-studie - erste folgebefragung (KiGGS Welle 1). *Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz* 57, 807–819. doi: 10.1007/s00103-014-1979-3
- Ihlebaek, C., Eriksen, H. R., and Ursin, H. (2002). Prevalence of Subjective Health Complaints (SHC) in norway. *Scand. J. Public Health* 30, 20–29. doi: 10.1177/14034948020300010701
- Inchley, J., Currie, D., Young, T., Samdal, O., and Torsheim, T. (2016). *Growing Up Unequal Gender and Socioeconomic Differences in Young People's Health and Well-Being: Health Behaviour in School-Aged Children Study, International Report From the 2013/2014 Survey*. Geneva: World Health Organization.
- Jané-Ilopis, E., Braddick, F., Bakker, H., Boeing, D., Calzone, C., Di Cesare, G., et al. (2008). *Mental Health in Youth and Education. Consensus Paper*. Luxembourg: European Communities.
- Joint Action on Mental Health and Well-being (2015). *Mental Health in all Policies: Situation Analysis and Recommendation for Action*. Lisbon: European Commission.
- Kessler, R. C., Berglund, P., Demler, O., Jin, R., Merikangas, K. R., and Walters, E. E. (2005). Lifetime prevalence and age-of-onset distributions of Dsm-Iv disorders in the national comorbidity survey replication. *Arch. Gen. Psychiatry* 62, 593–602. doi: 10.1001/archpsyc.62.6.593
- Kinnunen, P., Laukkanen, E., and Kylmä, J. (2010). Associations between psychosomatic symptoms in adolescence and mental health symptoms in early adulthood. *Int. J. Nurs. Pract.* 16, 43–50. doi: 10.1111/j.1440-172X.2009.01782.x
- Matijasevich, A., Murray, E., Stein, A., Anselmi, L., Menezes, A. M., Santos, I. S., et al. (2014). Increase in child behavior problems among urban brazilian 4-year olds: 1993 and 2004 pelotas birth cohorts. *J. Child Psychol. Psychiatry Allied Discip.* 55, 1125–1134. doi: 10.1111/jcpp.12236
- Maxwell, J. A. (2009). *Designing a Qualitative Study. The SAGE Handbook of Applied Social Research Methods*. Thousand Oaks, CA: Sage, 214–253. doi: 10.1111/1467-9299.00177
- Monro, F., and Huon, G. (2005). Media-portrayed idealized images, body shame, and appearance anxiety. *Int. J. Eat. Disord.* 38, 85–90. doi: 10.1002/eat.20153
- Navarro, D., Montserrat, C., Malo, S., González, M., Casas, F., and Crous, G. (2015). Subjective well-being: what do adolescents say? *Child Family Soc. Work* 22, 175–184. doi: 10.1111/cfs.12215
- Nolen-Hoeksema, S., and Girgus, J. S. (1994). The emergence of gender differences in depression during adolescence. *Psychol. Bull.* 115, 424–443. doi: 10.1037/0033-2909.115.3.424
- Ottová-Jordan, V., Smith, O. R., Gobina, I., Mazur, J., Augustine, L., Cavallo, F., et al. (2015). Trends in multiple recurrent health complaints in 15-year-olds in 35 countries in europe, north america and israel from 1994 to 2010. *Eur. J. Public Health* 25, 24–27. doi: 10.1093/eurpub/ckv015
- Patton, G. C., Sawyer, S. M., Santelli, J. S., Ross, D. A., Afifi, R., Allen, N. B., et al. (2016). Our future: a lancet commission on adolescent health and wellbeing. *Lancet* 387, 2423–2478. doi: 10.1016/S0140-6736(16)00579-1
- Ravens-Sieberer, U., Torsheim, T., Hetland, J., Vollebergh, W., Cavallo, F., Jericek, H., et al. (2009). Subjective health, symptom load and quality of life of children and adolescents in europe. *Int. J. Public Health* 54, 151–159. doi: 10.1007/s00038-009-5406-8
- Reiss, F. (2013). *Socioeconomic Inequalities and Mental Health Problems in Children and Adolescents: A Systematic Review. Social Science & Medicine* 90. Amsterdam: Elsevier, 24–31. doi: 10.1016/j.socscimed.2013.04.026
- Salk, R. H., Petersen, J. L., Abramson, L. Y., and Hyde, J. S. (2016). The contemporary face of gender differences and similarities in depression throughout adolescence: development and chronicity. *J. Affect. Disord.* 205, 28–35. doi: 10.1016/j.jad.2016.03.071
- Sawyer, S. M., Afifi, R. A., Bearinger, L. H., Blakemore, S. J., Dick, B., Ezech, A. C., et al. (2012). Adolescence: a foundation for future health. *Lancet* 379, 1630–1640. doi: 10.1016/S0140-6736(12)60072-5
- Sim, J., and Snell, J. (1996). Focus groups in physiotherapy evaluation and research. *Physiotherapy* 82, 189–198. doi: 10.1016/S0031-9406(05)66923-4
- Steinhausen, H. C., and Winkler Metzke, C. (2007). Continuity of functional-somatic symptoms from late childhood to young adulthood in a community sample. *J. Child Psychol. Psychiatry Allied Discip.* 48, 508–513. doi: 10.1111/j.1469-7610.2006.01721.x
- Stoep, A. V., Weiss, N. S., Kuo, E. S., Cheney, D., and Cohen, P. (2003). What proportion of failure to complete secondary school in the us population is attributable to adolescent psychiatric disorder? *J. Behav. Health Serv. Res.* 30, 119–124. doi: 10.1007/BF02287817
- Sweeting, H., West, P., Young, R., and Der, G. (2010). Can we explain increases in young people's psychological distress over time? *Soc. Sci. Med.* 71, 1819–1830. doi: 10.1016/j.socscimed.2010.08.012
- Sweeting, H., Young, R., and West, P. (2009). GHQ increases among scottish 15 year olds 1987–2006. *Soc. Psychiatry Psychiatr. Epidemiol.* 44, 579–586. doi: 10.1007/s00127-008-0462-6
- Tashakkori, A., and Teddlie, C. (2003). *Handbook of Mixed Methods in Social & Behavioral Research. Book*. Thousand Oaks, CA: SAGE Publications.
- Tick, N. T., Van Der Ende, J., and Verhulst, F. C. (2007). Twenty-year trends in emotional and behavioral problems in dutch children in a changing society. *Acta Psychiatr. Scand.* 116, 473–482. doi: 10.1111/j.1600-0447.2007.01068.x
- Torsheim, T., Ravens-Sieberer, U., Hetland, J., Väiimaa, R., Danielson, M., and Overpeck, M. (2006). Cross-national variation of gender differences in adolescent subjective health in europe and north america. *Soc. Sci. Med.* 62, 815–827. doi: 10.1016/j.socscimed.2005.06.047
- West, P., and Sweeting, H. (2003). Fifteen, female and stressed: changing patterns of psychological distress over time. *J. Child Psychol. Psychiatry Allied Discip.* 44, 399–411. doi: 10.1111/1469-7610.00130
- Whitehead, R., Berg, C., Cosma, A., Gobina, I., Keane, E., Neville, F., et al. (2017). Trends in adolescent overweight perception and its association with psychosomatic health 2002–2014: evidence from 33 countries. *J. Adolesc. Health* 60, 204–211. doi: 10.1016/j.jadohealth.2016.09.029
- WHO (2001). *Strengthening Mental Health Promotion*. Geneva: World Health Organization.

Willig, C. (2008). Introducing qualitative research in psychology adventures in therapy and method. *Perspect. Clin. Res.* 4, 192. doi: 10.4103/2229-3485.115389

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2017 Cosma, Belić, Blecha, Fenski, Lo, Murár, Petrovic and Stella. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



A Study Protocol for Testing the Effectiveness of User-Generated Content in Reducing Excessive Consumption

Atar Herziger^{1*}, Amel Benzerga², Jana Berkessel³, Niken L. Dinartika⁴, Matija Franklin⁵, Kamilla K. Steinnes⁶ and Felicia Sundström⁷

¹ Cologne Graduate School in Management, Economics and Social Sciences, University of Cologne, Cologne, Germany, ² Division of Psychology and Language Sciences, University College London, London, United Kingdom, ³ Department of Psychology, University of Cologne, Cologne, Germany, ⁴ Department of Psychology, Maastricht University, Maastricht, Netherlands, ⁵ Department of Psychology, University of Cambridge, Cambridge, United Kingdom, ⁶ Department of Psychology, University of Oslo, Oslo, Norway, ⁷ Department of Psychology, Uppsala University, Uppsala, Sweden

OPEN ACCESS

Edited by:

Pietro Cipresso,
Istituto Auxologico Italiano (IRCCS),
Italy

Reviewed by:

Eyal Peer,
Bar-Ilan University, Israel
Elisa Pedrolì,
Istituto Auxologico Italiano (IRCCS),
Italy

*Correspondence:

Atar Herziger
herziger@wiso.uni-koeln.de

Specialty section:

This article was submitted to
Quantitative Psychology and
Measurement,
a section of the journal
Frontiers in Psychology

Received: 06 November 2016

Accepted: 26 May 2017

Published: 09 June 2017

Citation:

Herziger A, Benzerga A, Berkessel J, Dinartika NL, Franklin M, Steinnes KK and Sundström F (2017) A Study Protocol for Testing the Effectiveness of User-Generated Content in Reducing Excessive Consumption. *Front. Psychol.* 8:972. doi: 10.3389/fpsyg.2017.00972

Excessive consumption is on the rise, as is apparent in growing financial debt and global greenhouse gas emissions. Voluntary simplicity, a lifestyle choice of reduced consumption and sustainable consumer behavior, provides a potential solution for excessive consumers. However, voluntary simplicity is unpopular, difficult to adopt, and under researched. The outlined research project will test a method of promoting voluntary simplicity via user-generated content, thus mimicking an existing social media trend (Minimalism) in an empirical research design. The project will test (a) whether the Minimalism trend could benefit consumers interested in reducing their consumption, and (b) whether self-transcendence (i.e., biospheric) and self-enhancement (i.e., egoistic and hedonic) values and goals have a similar impact in promoting voluntary simplicity. A one-week intervention program will test the efficacy of watching user-generated voluntary simplicity videos in reducing non-essential consumption. Each of the two intervention conditions will present participants with similar tutorial videos on consumption reduction (e.g., decluttering, donating), while priming the relevant values and goals (self-transcendence or self-enhancement). These interventions will be compared to a control condition, involving no user-generated content. Participants will undergo baseline and post-intervention evaluations of: voluntary simplicity attitudes and behaviors, buying and shopping behaviors, values and goals in reducing consumption, and life satisfaction. Experience sampling will monitor affective state during the intervention. We provide a detailed stepwise procedure, materials, and equipment necessary for executing this intervention. The outlined research design is expected to contribute to the limited literature on voluntary simplicity, online behavioral change interventions, and the use of social marketing principles in consumer interventions.

Keywords: voluntary simplicity, ethical consumption, intervention studies, user-generated content, excessive consumption, consumer values, self-enhancement, self-transcendence

EXCESSIVE CONSUMPTION

Consumption becomes excessive when it exhausts the consumer's financial or mental resources, or when it creates extreme environmental consequences, thus negatively affecting individual and societal well-being (Sheth et al., 2011). During the last decade, US consumption has risen dramatically; in 2007, consumption accounted for 72% of the American GDP. One factor which may contribute to excessive consumption is Materialism. Materialistic individuals consider material goods as an important part of their identity (Belk, 1984). They perceive material possessions as central to their status and happiness (Richins and Dawson, 1992). In a recent meta-analysis examining materialism and personal well-being, a significant negative relationship was found between the two (Dittmar et al., 2014). In an effort to reduce excessive consumption and increase consumer well-being, this project proposes a method to engage consumers in a reduced consumption lifestyle, namely, voluntary simplicity.

VOLUNTARY SIMPLICITY

A frequently proposed solution to excessive consumption and its negative affects is voluntary simplicity (Etzioni, 1998; Alexander and Ussher, 2012); the consumer movement of adopting a simplistic lifestyle. Voluntary simplicity involves a consumer's independent decision to reduce and replace non-essential products and services with non-material life elements, in an attempt to increase life satisfaction and meaning (Etzioni, 1998; Huneke, 2005). Non-essential items and services can be defined as things that are irrelevant for the achievement of one's life purpose (McGouran and Prothero, 2016). For example, voluntary simplicity behaviors include: composting, recycling, decluttering (Huneke, 2005), self-sufficiency and use of pro-environmental transportation (Alexander and Ussher, 2012). While this lifestyle is promising in its potential to reduce consumption and increase well-being, is it difficult to adopt (McGouran and Prothero, 2016).

Minimalism

The scholarly concept of voluntary simplicity has recently been adapted into popular culture, under the term *Minimalism*. Minimalism often encompasses user-generated content, such as videos, texts, and images that promote consumption reduction. Minimalism content is largely shared via YouTube, Facebook, Twitter and similar online social networking services. Additionally, there are numerous blogs, online forums and websites devoted to Minimalism¹. In these online sources, viewers can find information about Minimalism, tips on how to adopt this lifestyle, tutorials, and personal narratives of successful minimalists.

The Minimalism phenomenon, to a lesser degree than voluntary simplicity, seems to utilize an aesthetic approach.

¹See for example; <https://www.youtube.com/watch?v=AhJc8wtQbxQ> (YouTube video), https://www.reddit.com/r/minimalism/comments/53l2m3/extreme_minimalists_post_your_storystuff_here/ (online forum), <http://time.com/3738202/minimalism-clutter-too-much-stuff/> (news article), <http://www.theminimalists.com/minimalism/> (website), <http://minimalismfilm.com/> (documentary).

Minimalism can be studied as an aesthetically pleasing trend, relating to both fashion and web design. For example, narratives of fashion bloggers dedicated to a Minimalistic style revealed a strong emphasis on simplicity, elegance, sophistication and cleanliness, which were achieved in particular by choice of coloring and scarce use of patterns (Karg, 2015). It is yet early to determine whether the Minimalism phenomenon is a passing trend or a new wave of conscious consumerism, but Minimalism may have the power to engage consumers in sustainable behaviors.

SELF-TRANSCENDENCE AND SELF-ENHANCEMENT VALUES AND GOALS

One way through which consumer interventions could reduce materialism and promote sustainable consumption, is through the activation of relevant values and goals (Kasser, 2016). In other words, Minimalism videos may not only serve the purpose of directly increasing voluntary simplicity behaviors through tutorials; these videos are also a channel through which relevant ecological values could be activated and strengthened. Thus, user-generated content on Minimalism could promote personal norms of sustainable consumption, creating longer-term impact on consumer behavior.

Previous research has consistently shown that a biospheric value orientation is positively associated with ecological attitudes and beliefs, and is thus conducive to sustainable behavior (e.g., Stern, 2000; de Groot and Steg, 2008). A biospheric value orientation is a type of self-transcendence value, which emphasizes the environment and the biosphere over one's own personal benefit (Stern et al., 1993).

Self-enhancement values, on the other hand, promote one's own benefit over that of others or the environment, e.g., egoistic and hedonic values (Schwartz, 1992). Egoistic and hedonic values are negatively associated with ecological attitudes and sustainable behaviors (Steg et al., 2014). Moreover, these values are closely clustered with materialism, and thus could promote excessive consumption (Burroughs and Rindfleisch, 2002; Kasser, 2016).

Self-Transcendence and Self-Enhancement in Minimalism Videos

Although, previous literature suggests that self-enhancement values have a negative impact on sustainable behavior, user-generated content on Minimalism seems to be focused exactly on these values. Exploratory research of the Minimalism phenomenon found that engagement in Minimalism is associated with a personal goal to reduce stress (hedonism) and save money (egoism), but not with environmental concern (Hertziger, unpublished manuscript). Thus, we argue that self-enhancement goals might be utilized to promote sustainable consumer behavior. However, it remains unclear whether self-enhancement goals could be as effective as self-transcendence goals in promoting sustainable behavior.

OBJECTIVES

The project will test whether user-generated content promoting either self-transcendence or self-enhancement values and goals could be effective in reducing excessive consumption, as well as increasing well-being. The effectiveness of the interventions will be tested by measuring their effects on life-satisfaction, intent to adopt voluntary simplicity, and purchasing intent and behavior. The project's main research questions are:

- (1) Could user-generated content promoting voluntary simplicity reduce excessive consumption?
- (2) Are self-enhancement goals as effective in promoting sustainable behavior as self-transcendence goals, in user-generated Minimalism content?

We are also interested in whether exposure to user-generated content promoting voluntary simplicity affects one's well-being. Finally, the relationship between one's intent to reduce consumption and one's consumption behavior might generate meaningful insights.

DESIGN

A weeklong intervention program will test the efficacy of watching voluntary simplicity videos in reducing non-essential consumption. The first condition will promote voluntary simplicity while priming self-transcendence values and goals, and the second condition will promote voluntary simplicity while priming self-enhancement values and goals. These two interventions will be compared with a control condition which only measures participant's affective state during the intervention week. All participants will undergo baseline and post-intervention evaluations of values and goals in consumption reduction, voluntary simplicity attitudes and behaviors, spending behaviors, and life satisfaction directly before and directly following the intervention, respectively. A one-month follow up measurement will also be taken. The difference between measurements before and after the intervention will serve as our dependent variables.

INTERVENTION

A key component in this study is the video stimuli employed. Videos have proven to be more effective and persuasive than other modalities of communication (Mohammadi et al., 2013). To maintain a high level of reliability and validity, the research team will create the intervention videos in-house. To create valid video stimuli several measures will be taken:

Video checklists

Video checklists describe the key concepts and behaviors to be addressed in the video stimuli. These checklists detail the following stimuli requirements, per condition:

- (1) Content: information conveyed (stable across intervention conditions)

- (2) Condition priming: what values and goals will be primed in the opening and closing statements of the videos, as well as the visual stimuli (self-transcendence vs. self-enhancement)
- (3) Style: the presenter's tone and expression (stable across intervention conditions).

Due to internal reliability concerns, one presenter will appear in all videos, for both intervention conditions. The presenter will be a female in her twenties, similarly to many Minimalism vloggers. Appendix A presents an example of a video checklist.

The research team will contact several social and consumer psychologists to review these checklists and approve that they (a) prime self-transcendence and self-enhancement values and goals, per condition, and (b) promote consumption reduction, prior to video creation.

Video Pre-Testing

The videos will be pretested in a convenience sample of international psychology students at the advanced undergraduate and graduate level. We aim to test the face validity of the videos. Thus, the pre-test population will consist of young consumers with a basic scientific knowledge on motivation, values, and goals. The pre-test sample will therefore be slightly different from the population addressed in the main study. Participants will watch subsets of videos from both intervention conditions and rate the face validity of the videos (e.g., "Did the presenter promote biospheric values?", "Did the presenter promote egoistic values?"). They will also note any manipulation concerns or potential confounds in the video stimuli of both conditions. Additionally, participants will rate how entertained, focused and interested they were when watching the videos. The latter measures will be used to estimate main-study participants' motivation in watching the daily video stimuli during the weeklong intervention. If a video receives a low score on either validity or entertainment value during the pre-test, the research team will revise the videos based on pre-test comments, and re-test them on a similar sample.

Mobile Application

A mobile application will be developed and used in the study for stimuli presentation, data collection, communication, and once the study is complete, a personalized progress report. Video stimuli will be embedded into Qualtrics questionnaires, which will be linked in the mobile application. This will allow participants to easily fill-out the questionnaires before and after viewing the video stimuli. Time spent viewing the videos will be recorded by a "timing" function in Qualtrics.

The mobile application push-notification system will remind participants to view the video and fill in a questionnaire on a daily basis throughout the weeklong intervention. Thus, participants will receive one push-notification a day, until the intervention is complete. The mobile application will also allow both the participants and research team to communicate via email. This will be useful if any problems or questions arise. Finally, users of the mobile application will receive a personalized progress report once the study is completed.

PROCEDURE

At baseline, all participants will download a free mobile application. Participants will complete an intake survey on Qualtrics, through a link in the mobile application. Following intake, participants will be randomly allocated to one of the three experimental conditions, and complete a battery of questionnaires measuring dependent and independent variables (see t_0 in **Figure 1**). After baseline measurements, the intervention will commence. In the two intervention conditions, participants will watch one video daily for 6 days. Both intervention and control condition participants will complete a daily mood measurement (see t_1 through t_6 in **Figure 1**). On the last day of the intervention, participants will complete questionnaires measuring the dependent and independent variables a second time (see t_7 in **Figure 1**). One month after intervention completion, the dependent variables will be measured again (see t_{end} in **Figure 1**).

METHODS

Ethics Process

Due to the study's intervention design, careful consideration was put into ethical procedures executed by the research team. Ethical approval has been granted for this study. In addition to standard procedures, such as voluntary participation and dropout, several processes are in place to reduce potential harm to participants' well-being.

Firstly, participants who struggle with compulsive buying, hoarding, or other addictive behaviors are regarded as sensitive populations, and will be asked through a consent form not to take part in the study. Individuals diagnosed with Compulsive Buying Disorder (CBD) repeatedly engage in excessive spending behaviors and cognitions due to irrepressible impulses, which result in anxiety or impairment (Black, 2007). CBD diagnosed individuals will be ineligible to participate in the study because of two main reasons; (a) the topic and nature of the study could potentially cause harm or distress to the individual and (b) the study's purpose, to explore potential solutions to reduce excessive consumption, will be directed toward the general population of consumers and not those suffering from a clinical buying disorder. Thus, an in-take survey will employ a validated self-report clinical screener for CBD (Faber and O'Guinn, 1992). Participants scoring over the clinical threshold will be excluded from the study; they will be notified of their ineligibility, via email, within two days of their in-take questionnaire submission. They will also be sent additional information on professional help networks for compulsive buyers². A diagnosis will not be provided to participants due to the research team's lack of clinical training and the absence of differential diagnosis.

Similarly, participants will be asked to report any undesirable changes in mood and behavior during the intervention by contacting the research team. Participants reporting an increase in buying behavior or severe decline in well-being, will be

interviewed by e-mail, and dismissed from the study. Full reporting of these instances, should there be any, will be included in the results section of disseminated work emerging from this project.

After the intervention is completed, including a one-month follow-up, the research team will conduct initial analysis on the effectiveness of the employed interventions. Specifically, conducted analysis will test whether either one or two of the employed interventions prove to be beneficial in reducing excessive consumption and increasing well-being, in comparison to the control condition. If one or both interventions prove to be effective, they will be offered to participants who did not experience them while participating in the study. Thus, all participants will be afforded the opportunity to take part in an effective intervention for consumption reduction, should the research team identify one. Participants will also receive debriefing of the study and will be exposed to all three conditions employed in the study design.

Participant Outreach

Sample Size

Two hundred and fifty English-speaking subjects will be recruited for this study globally, via online methods. The number of participants needed is based on an a priori power analysis using G*Power software (*F*-test, ANOVA: Repeated measures, within-between-interaction; $1 - \beta = .80$, $\alpha = .05$, Cohen's $f = .10$), which revealed that the study would require a total sample size of $N = 198$ (Faul et al., 2007). An additional 52 participants have been added to this estimated sample size to account for an assumed 20% dropout rate.

Participant Criteria

Participants will sign a consent form electronically stating that; (a) they are at least 18 years of age or of legal age in their country of residence, (b) they have not been clinically diagnosed with a disorder related to impulse control or addictions, (c) they express both a desire and difficulty to reduce their non-essential consumption, (d) they have not previously followed the social media trend of consumer Minimalism.

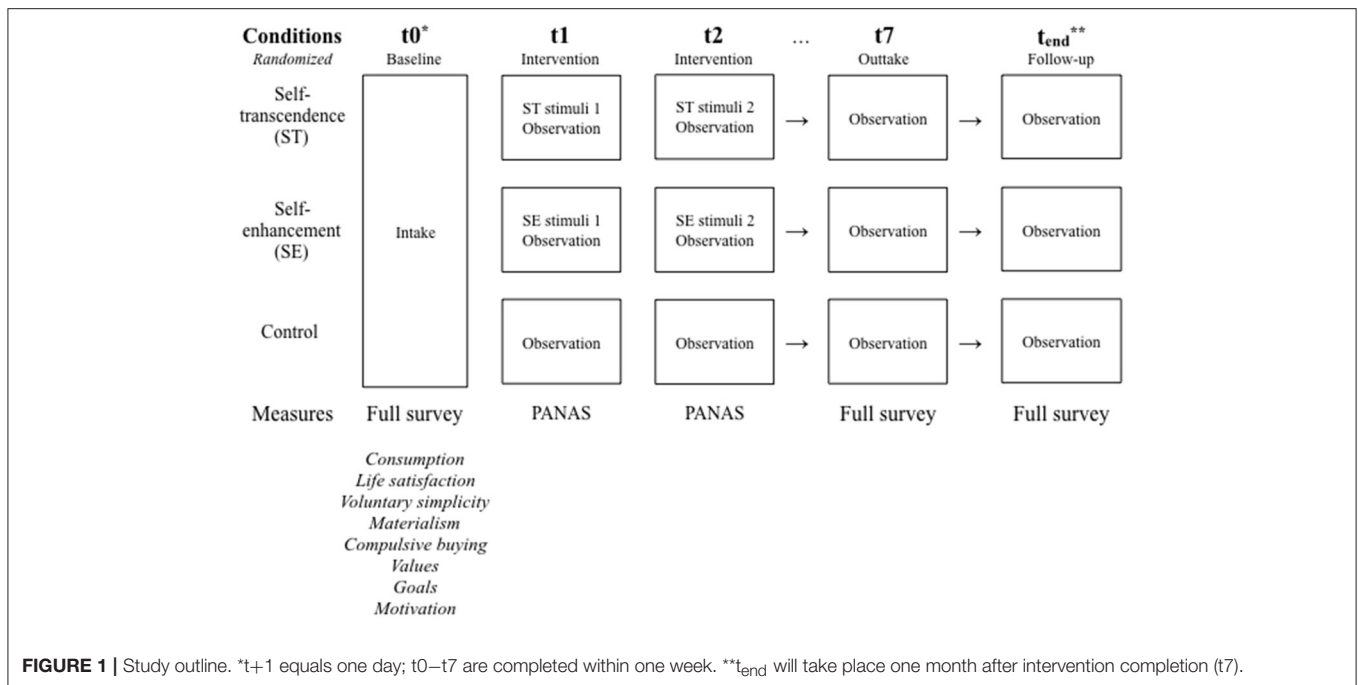
Recruitment Methods

Participants will be recruited through convenience sampling via two main avenues; (a) social media posts in consumption related forums and sites as well as Facebook pages and groups, and (b) campus posters circulated around the universities of the research team. The main appeal to participants in all avenues will follow a "question and proposal" format, i.e., "Do you want to decrease your buying? Sign-up to try out an experimental treatment!" Each poster and online recruitment posting will feature a QR code that participants can scan with their smartphones to download the intervention mobile application. The consent form will be embedded in the in-take questionnaire.

Participant Commitment Through Feedback

Efforts to promote continued participation will be aimed at both extrinsic and intrinsic motivation, with a primary emphasis on the latter. Two intrinsic, motivational aspects will be addressed

²Addiction Resources: <http://www.nhs.uk/Livewell/addiction/Pages/addiction-whatitis.aspx>



in the recruitment call; (a) participants will be offered to try treatments that may help them with their reported, personal goal of reducing non-essential material consumption and (b) participants will receive feedback on their personal progress following the intervention.

To further motivate participants to remain in the study throughout the weeklong intervention, a lottery prize of 100 euros will be raffled to one participant who completed the full intervention. The winning participant will be able to receive this prize via either bank transfer or a donation made on their behalf to a charity of their choice.

Surveys and Intervention

After baseline measurement is completed, participants in each intervention group will be requested to watch one video per day for six days, and answer a short questionnaire following each video. The videos will be between three and five minutes long and consist of a young, female presenter, talking to the camera as if talking to an audience.

The main content of the videos in both intervention conditions will be identical. However, the opening and closing sections of each video will be filmed separately, and emphasize either self-transcendence or self-enhancement values and goals, per condition. For example, in the self-transcendence condition the presenter will say, “We are here because we want to *make the world a better place*. By being *environmentally friendly*, *conscience* and *ethical*, we can *make a positive impact on the world*.” Conversely, in the self-enhancement condition, the presenter will say, “We are here because we want to *make our lives better*. By being *less stressed*, *more in control*, and *spending wisely*, we can *make a positive impact on our own lives*.” Moreover, visual stimuli integrated into the videos will prime participants of either

self-transcendence, biospheric values (e.g., gardens, lakes, plants) or self-enhancement, hedonic and egoistic values (e.g., friends chatting and drinking coffee together). In the control condition, there will be no intervention, but participants will be pinged by the mobile application to answer a survey once per day for six days. Follow-up measurements will be conducted via web-site one month after the last intervention day for all three groups.

MATERIALS AND EQUIPMENT

Smartphone Application

The smartphone application will contain four key features:

- (1) Questionnaires: the mobile application will link participants to Qualtrics questionnaires, including video stimuli. These will be presented to the participants at in-take, during the daily intervention, and at post-intervention assessment;
- (2) Notifications: a push-notification system will remind participants to view the stimuli and complete the questionnaires every day, while the intervention is in progress;
- (3) Communication with the research team: an email link will allow participants to contact the research team directly;
- (4) Personalized progress reports: participants will receive a link to a report of their post-intervention progress once the study is complete.

Measures

Demographic Questionnaire

Participants will be asked to provide demographic information, such as gender, native language, education level, age, ethnicity, and country of residence.

Intake Screeners

Assessment of Compulsive Buying Behavior

The Clinical Screener (CBCS; Faber and O'Guinn, 1992) is one of the most validated and replicable self-report scales measuring compulsive buying behavior, having been found to correctly classify 88% of compulsive buyers (Faber and O'Guinn, 1992). The CBCS contains seven statements that respondents are required to rate their agreement with (1 = Strongly disagree to 5 = Strongly agree) (e.g., "If I have any money left at the end of the pay period, I just have to spend it") or their frequency of experiencing given behaviors/feelings (1 = very often to 5 = never) (e.g., "Bought things even though I could not afford them") on five-point Likert scales. A total score is calculated, and a threshold measurement determines whether an individual is considered a compulsive buyer.

Faber and O'Guinn's (1992) compulsive buying scale has, however, been criticized for only addressing the impulse control dimension of compulsive buying behavior and not the obsessive-compulsive dimension, as well as not being accommodating to buyers that have higher incomes (Ridgway et al., 2008). To compensate for these limitations, the Compulsive-Buying Index (CBI) developed by Ridgway et al. (2008) will be used in addition to the CBCS, thus presenting a more comprehensive measurement of compulsive buying behavior.

The Compulsive-Buying Index (CBI) is a well-validated measure of compulsive buying behavior (Ridgway et al., 2008). The CBI contains six items measuring two dimensions of compulsive buying; three items for obsessive-compulsive buying (e.g., "My closet has unopened shopping bags in it") and three items for impulsive buying (e.g., "I buy things I don't need"). Responses are given on a seven-point Likert scale either indicating agreement with statements (1 = strongly disagree to 7 = strongly agree) or frequency of experiencing given behaviors/feelings (1 = never to 7 = very often). A total score is calculated, and a threshold measurement determines whether an individual is considered a compulsive buyer.

Assessment of Voluntary Simplicity

Lifestyles. The Voluntary Simplicity Lifestyles scale used in this study (VSL; Nepomuceno and Laroche, 2015) is an adapted version of Iwata's original voluntary simplicity scale (Iwata, 1997), a validated self-reported measure of the adoption of the voluntary simplicity consumer lifestyle. This adapted scale is made up of nine statements describing a number of behaviors that are consistent with a voluntary simplistic lifestyle, such as the scale item "I fully adhere to a simple lifestyle and only buy necessities." Participants are required to rate their agreement with these statements on a five-point Likert scale (1 = definitely disagree to 5 = definitely agree), indicating the degree to which they engage in the suggested simplistic lifestyle choices.

Involvement. The Personal Involvement Inventory (PII; Zaichkowsky, 1994) is considered a validated context-free measure of one's motivation to be involved with certain concepts, behaviors or products. This study tailored the PII to measure consumers' voluntary simplicity involvement. Involvement is measured on a self-report semantic scale made up of 10

items. The scale can be divided into two subscales: the affective subscale (e.g., interesting, appealing) and the cognitive subscale (e.g., important, relevant). Participants will be asked to judge voluntary simplicity across all cognitive and affective adjectives. Participants rate each item along a bipolar adjective scale (e.g., "interesting" or "boring"), judging how close they deem voluntary simplicity to either opposing adjective. Responses are given on a seven-point scale and six items require reverse scoring. A total involvement score is calculated where higher scores represent higher involvement.

Assessment of Materialism

The Materialism Scale (MS; Richins and Dawson, 1992) is a validated self-report scale measuring materialism value. The 18-item scale measures three dimensions of materialism; success, centrality and happiness, which describe different motivations for acquiring material possessions. Six items measure success, indicating the perception that possessions are indicators of life-success (e.g., "I admire people who own expensive homes, cars, and clothes"). Seven items measure centrality, representing the general importance of acquisition and possession (e.g., "I enjoy spending money on things that aren't practical"). The last five items measure happiness, representing the perception that possession is necessary for happiness (e.g., "My life would be better if I owned certain things I don't have"). The MS is measured on a five-point Likert scale (1 = strongly disagree, 5 = strongly agree) and eight items require reverse coding. An overall scale score is calculated from the three dimensions, a higher score suggesting higher materialism than a lower score.

Assessment of Egoistic, Altruistic, and Biospheric Values Orientations

The Value Instrument is an adapted, validated measure of egoistic, altruistic and biospheric value orientations (de Groot and Steg, 2008). This scale is an adapted form of the Schwartz value scale (Schwartz, 1992), and consists of 13 items: five for egoistic (e.g., wealth), four for altruistic (e.g., equality), and four for biospheric (e.g., unity with nature) value orientations. Internal reliabilities for these sub-scales range from .73 to .86. As in the Schwartz value scale, participants rate the importance of the 13 items "as a guiding principle in their lives" on a nine point Likert-scale (-1 = opposed to my values, 0 = not important to 7 = extremely important). In the instructions segment, respondents are requested to vary the scores they provide each value and only rate very few of them as extremely important.

Assessment of Satisfaction with Life

The Satisfaction with Life Scale (SWLS; Diener et al., 1985) is a validated measure of subjective well-being, characterized by favorable psychometric properties, such as convergent validity with other similar measures, cross-cultural validity, and temporal stability (Pavot and Diener, 2008). The five-item self-report scale measures global life satisfaction by scoring participant agreement to a statement such as "I am satisfied with my life" on a seven point Likert-scale (1 = strongly disagree to 7 = strongly agree). A total score can range between five (low satisfaction) and 35

(high satisfaction) suggesting the degree of life satisfaction of the participant.

Assessment of Motivation

Participants will be asked to rate their motivation source and strength for taking part in the study. Motivation sources will match the aforementioned value orientations (e.g., egoistic and hedonic: “improve my well-being,” “save my money”; altruistic: “improve the lives of others around me,” “donate more to others”; and biospheric: “improve the future of the environment,” “reduce my carbon footprint”). Participants will rate their agreement with these motivations similarly to the value instrument (−1 = opposed to my motivation, 0 = not motivating to 7 = extremely motivating).

Assessment of Consumption Related Attitudes and Behaviors

Shopping intention. Shopping intention will be measured by adapting the three items of the eight-point semantic scale of behavioral buying intention (Baker and Churchill, 1977). The original items were intended to measure participants’ intention to buy a specific product. In this study, the items will be adapted to suit the context of the non-essential product buying intention. Participants will be asked to think of a type of product they would normally buy, which is non-essential (e.g., clothing, sweet snacks, coffee). Then, they will be asked about their purchase intent for this product, for example: “Would you buy this non-essential product within the next seven days?” In response to the question, participants should indicate their intention on an eight-point semantic scale of “yes—definitely” to “no—definitely not.”

Buying frequency and behavior. Two questionnaires (see Appendix B) measuring shopping frequency (Section A, seven items) and average weekly non-essential expenditures (Section B, two items) will be administered to participants. The order of the items within both sections will be randomized. Responses for five of the total nine items of the two questionnaires will be given on a six-point Likert scale, ranging from zero to five (e.g., “How likely are you to shop for something you need in the coming week?”) that will be averaged within each section. The remaining four items will require an open response by the participant (e.g., “How many times do you typically go shopping in a week?”).

Intervention Questionnaire

Short Mood Scale (PANAS).

The Short Form Positive and Negative Affect Schedule (I-PANAS-SF; Thompson, 2007) is a short version of the widely-used measure of the variation in affect, the Positive and Negative Affect Schedule (PANAS; Watson et al., 1988). PANAS is a well-validated measure of affect with strong psychometric properties (Watson et al., 1988; Crawford and Henry, 2004), demonstrated cross-culturally (e.g., Joiner et al., 1997) and within both community and clinical samples (Leue and Beauducel, 2011). The I-PANAS-SF is a 10-item self-report scale that is divided into two subscales, Negative Affect (NA; afraid, ashamed, hostile, nervous, and upset) and Positive Affect (PA; active, alert, attentive, determined, and inspired). Participants are required to judge the extent to which they experience the NA and PA adjectives on a seven-point Likert scale (0 = never to 7 = always)

in response to the statement: “Thinking about yourself and how you normally feel, to what extent do you generally feel.” This phrasing will be revised to address the current moment.

Post-Intervention Questionnaire

The effectiveness of the intervention will be measured by the difference of participants’ pre- and post-intervention measurements, controlling for covariates. Therefore, all intake questionnaires will be repeated in the post-intervention phase. Along with these repeated scales, several additional measures will be administered.

Two additional questions will measure participants’ attitude toward the videos. The first item will measure whether participants would recommend the video to friends with a similar goal (i.e., “Would you recommend these videos to a friend who also wanted to reduce their non-essential consumption?”). The second item will measure participants’ intention to continue watching similar videos online (i.e., “Would you watch similar videos online?”). Participants would be required to indicate their answer on a nine-point scale (0 = not at all to 8 = definitely) for both questions. Lastly, an open-ended question will allow participants to share what they thought was interesting, difficult, or helpful during the one week of intervention.

PROPOSED ANALYSIS

To test the hypotheses, statistical software including Excel, SPSS, and R will be utilized.

Consumption Reduction—Behavior

To examine whether user-generated content will reduce excessive consumption, a multiple regression analysis will be calculated, using condition as an independent variable to predict excessive consumption post-intervention, taking the baseline measurement into account as a covariate, in addition to other controlled variables (e.g., base-line values, demographics). As can be seen in **Figure 2**, we expect a main effect of point of time for all conditions. In other words, we expect to see consumption behavior decreased post-intervention. Additionally, we expect to see an interaction between point of time and condition, showing a stronger reduction of excessive consumption in the experimental groups following the user-generated content intervention in comparison to the control group.

The strength of the effect is expected to be higher between t_0 and t_7 than between t_0 and t_{end} . In other words, we expect a stronger effect during the intervention compared to the post-intervention phase. Condition will be contrast coded to facilitate the interpretation of the interaction (Cohen et al., 2013). This analysis will also be used to examine which aspects of user-generated content affect excessive consumption, comparing self-transcendence to self-enhancement conditions. Although, there is no clear expectation on which content will have a stronger effect, this will be examined through exploratory analyses. Participants’ in-take reported motivations, values, involvement, and personal goals to reduce consumption will be considered as control variables in these analyses.

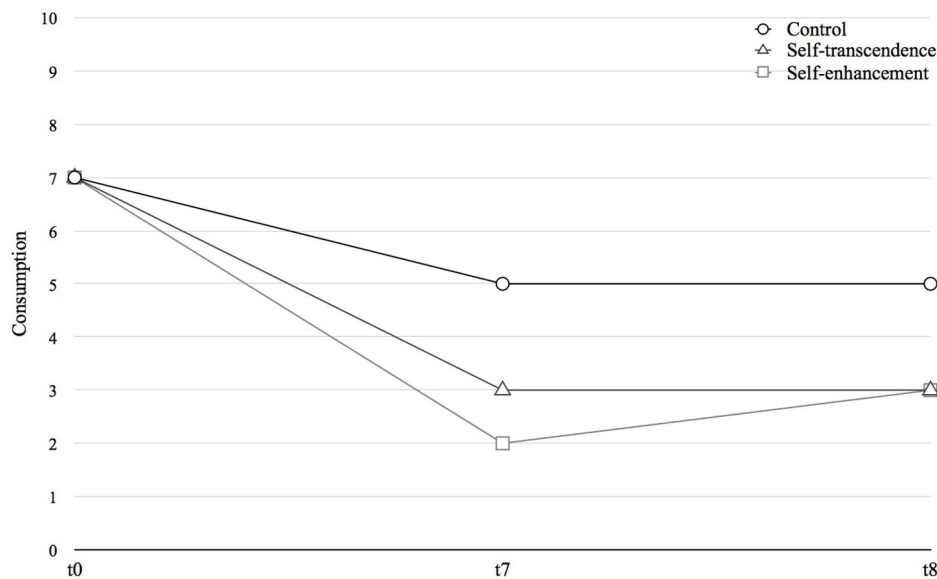


FIGURE 2 | Proposed pattern of results showing reduced consumption within experimental groups.

Consumption Reduction-Intent

A similar regression analysis will be calculated exploring the interventions' effects on reported intent to reduce consumption. Consumption reduction intent is expected to increase, especially for participants taking part in the interventions, showing both a main effect of point of time during the experiment, and an interaction effect between point of time and condition.

Furthermore, a correlational analysis, examining the link between the intent to reduce consumption and reported behavior, will be performed. Pearson's correlation coefficient will be calculated between reported behavior and reported intent, controlling for reported commitment to the study. We expect to find a negative link between consumption reduction-intent and consumption behavior.

Effects of Reduced Consumption

Using another regression, we will further analyze questionnaire data collected at the beginning, during the intervention, as well as during the post-questionnaire, to investigate, whether reduced consumption affects life-satisfaction. We expect it to increase within both experimental groups, negatively correlating with excessive consumption.

Changes in Values and Motivations

Using a repeated-measures ANOVA, with *post-hoc* contrasts, we will also compare in-take measurements of participants' values and motivations to the two post-intervention measurements. Thus, we will test whether exposure to the intervention stimuli affected consumer values and motivations directly. We expect participants in the self-transcendence condition to show an increase in biospheric values and motivations at t7, while participants in the self-enhancement condition will show an increase in egoistic values and egoistic and hedonic motivations

at t7. Control condition participants are not expected to show a significant change in either values or motivations during the intervention. This measurement will serve as a manipulation-check. We will also explore whether the manipulation persists at t_{end}, one month following the intervention.

ANTICIPATED RESULTS

We expect this project to provide insight into the effectiveness of user-generated content in reducing excessive consumption. We hope the project will shed light on specific aspects of user-generated content that help reduce excessive consumption and increase well-being. We predict to find data supporting our hypotheses and we therefore anticipate reduced excessive consumption within the intervention groups, as well as increased life-satisfaction.

Nevertheless, this research project entails some methodological challenges. One project challenge is the limited intervention time. A one-week intervention may not be sufficient time to produce significant impact on consumer intent and behavior. However, we chose this short time-period to minimize respondent's time-burden; prolonged intervention periods could increase potential issues in participant commitment and dropout rates.

Additionally, our choice of a control group is limiting. Control groups in intervention studies could be designed in several ways, and they should provide a comparison measurement to that of the intervention. Our findings will be limited in that they do not control for other, non-video interventions, such as written-text or monetary-incentive interventions. By pinging participants daily, and asking them to answer a mood scale, our outlined control condition does control for a placebo effect and a priming effect. Since our research question is inspired by

a phenomenon currently occurring in the field (i.e., YouTube Minimalism videos) our control condition allows us to test whether these videos are more beneficial for consumers than having no engagement in a consumption-reduction intervention. Thus, the control condition chosen for this study is not optimal, but it does produce the highest ecological validity. If feasible, both monetarily and logistically, we recommend authors utilizing this protocol include additional control conditions to their design.

Due to the exploratory nature of the study, it is possible, that neither intervention significantly impacts buying intent, behavior, affective state, or life satisfaction. Nevertheless, null results will be highly informative. As evident in the number of Minimalism videos on YouTube, and number of subscribers to this content, it would be expected that this user-generated content is effective in promoting voluntary simplicity. If the outlined study finds that even in controlled settings there is no support for a significant influence of this content on behavior, our results could improve the public's understanding of user-generated content. The results—whether supporting the hypothesis or not—will raise further scientific questions and influence our handling of user-generated content.

In sum, the proposed project aims to contribute to the limited literature on voluntary simplicity and the use of user-generated content in consumer interventions. Specifically, this study might add a literary contribution to consumer values and sustainable behavior, specifically by examining its applicability in promoting voluntary simplicity and reducing excessive consumption through media primed

with self-transcendence or self-enhancement messages. Applied implications of the proposed study will be directed to social marketers, social media experts, and proponents of voluntary simplicity.

ETHICS STATEMENT

This study has received ethical approval from the Cambridge Psychology Research Ethics Committee.

AUTHOR CONTRIBUTIONS

This study was conceived by AH. All of the authors (AH, AB, JB, ND, ME, KS, and FS) contributed to the research design, method, analysis plan, and potential implications discussion. All authors approved the final manuscript.

ACKNOWLEDGMENTS

The authors sincerely thank the Junior Researcher Programme and Universidad Francisco de Vitoria for supporting the realisation of this research.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fpsyg.2017.00972/full#supplementary-material>

REFERENCES

- Alexander, S., and Ussher, S. (2012). The voluntary simplicity movement: A multinational survey analysis in theoretical context. *J. Consum. Culture* 12, 66–86. doi: 10.1177/1469540512444019
- Baker, M. J., and Churchill, G. A. (1977). The impact of physically attractive models on advertising evaluations. *J. Market. Res.* 14, 538–555. doi: 10.2307/3151194
- Belk, R. W. (1984). “Three scales to measure constructs related to materialism: Reliability, validity, and relationships to measures of happiness.” in *NA—Advances in Consumer Research*, Vol. 11, ed T. C. Kinnear (Provo, UT: Association for Consumer Research), 291–297.
- Black, D. (2007). A review of compulsive buying disorder. *World Psychiatry* 6, 14–18.
- Burroughs, J. E., and Rindfleisch, A. (2002). Materialism and well-being: a conflicting values perspective. *J. Consum. Res.* 29, 348–370. doi: 10.1086/344429
- Cohen, J., Cohen, P., West, S. G., and Aiken, L. S. (2013). *Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences*. Mahwah, NJ: Routledge.
- Crawford, J. R., and Henry, J. D. (2004). The positive and negative affect schedule (PANAS): Construct validity, measurement properties and normative data in a large non-clinical sample. *Br. J. Clin. Psychol.* 43, 245–265. doi: 10.1348/0144665031752934
- de Groot, J. I. M., and Steg, L. (2008). Value orientations to explain beliefs related to environmental significant behavior. *Environ. Behav.* 40, 330–354. doi: 10.1177/0013916506297831
- Diener, E., Emmons, R., Larsen, J., and Griffin, S. (1985). The satisfaction with life scale. *J. Pers. Assess.* 49, 71–75. doi: 10.1207/s15327752jpa4901_13
- Dittmar, H., Bond, R., Hurst, M., and Kasser, T. (2014). The relationship between materialism and personal well-being: a meta-analysis. *J. Pers. Soc. Psychol.* 107, 879–924. doi: 10.1037/a0037409
- Etzioni, A. (1998). Voluntary simplicity: characterization, select psychological implications, and societal consequences. *J. Econ. Psychol.* 19, 619–643.
- Faber, R. J., and O’Guinn, T. C. (1992). A clinical screener for compulsive buying. *J. Consum. Res.* 19, 459–469.
- Faul, F., Erdfelder, E., Lang, A. G., and Buchner, A. (2007). G*Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behav. Res. Methods* 39, 175–191. doi: 10.3758/BF03193146
- Huneke, M. E. (2005). The face of the un-consumer: an empirical examination of the practice of voluntary simplicity in the United States. *Psychol. Market.* 22, 527–550. doi: 10.1002/mar.20072
- Iwata, O. (1997). Attitudinal and behavioral correlates of voluntary simplicity lifestyles. *Soc. Behav. Personal.* 25, 233–240. doi: 10.2224/sbp.1997.25.3.233
- Joiner, T. E., Sandin, B., Chorot, P., Lostao, L., Marquina, G., Sandin, B., et al. (1997). Development and factor analytic validation of the SPANAS among women in Spain: (more) cross-cultural convergence in the structure of mood. *J. Personal. Assess.* 68, 600–615. doi: 10.1207/s15327752jpa6803_8
- Karg, C. (2015). *New Fashion Minimalism in an Affluent Society: A Paradigm Shift?* Thesis, University of Borås, Borås.
- Kasser, T. (2016). Materialistic values and goals. *Ann. Rev. Psychol.* 67, 489–514. doi: 10.1146/annurev-psych-122414-033344
- Leue, A., and Beauducel, A. (2011). The PANAS structure revisited: on the validity of a bifactor model in community and forensic samples. *Psychol. Assess.* 23, 215–225. doi: 10.1037/a0021400
- McGouran, C., and Prothero, A. (2016). Enacted voluntary simplicity – Exploring the consequences of requesting consumers to intentionally consume less. *Eur. J. Market.* 50, 189–212. doi: 10.1108/EJM-09-2013-0521
- Mohammadi, G., Park, S., Sagae, K., Vinciarelli, A., and Morency, L. P. (2013). “Who is persuasive?: The role of perceived personality and communication modality in social multimedia” in *Proceedings of the 15th ACM on International Conference on Multimodal Interaction* (Sydney, NSW: ACM), 19–26. doi: 10.1145/2522848.2522857

- Nepomuceno, M. V., and Laroche, M. (2015). The impact of materialism and anti-consumption lifestyles on personal debt and account balances. *J. Bus. Res.* 68, 654–664. doi: 10.1016/j.jbusres.2014.08.006
- Pavot, W., and Diener, E. (2008). The satisfaction with life scale and the emerging construct of life satisfaction. *J. Positive Psychol.* 3, 137–152. doi: 10.1080/17439760701756946
- Richins, M. L., and Dawson, S. (1992). A consumer values orientation for materialism and its measurement: scale development and validation. *J. Consum. Res.* 19, 303–316. doi: 10.1086/209304
- Ridgway, N. M., Kukar-Kinney, M., and Monroe, K. B. (2008). An expanded conceptualization and a new measure of compulsive buying. *J. Consum. Res.* 35, 622–639. doi: 10.1086/591108
- Schwartz, S. H. (1992). “Universals in the content and structure of values: theoretical advances and empirical tests in 20 countries,” in *Advances in Experimental Social Psychology*, Vol. 25, ed M. Zanna (San Diego, CA: Academic Press), 1–65.
- Sheth, J. N., Sethia, N. K., and Srinivas, S. (2011). Mindful consumption: a customer-centric approach to sustainability. *J. Acad. Market. Sci.* 39, 21–39. doi: 10.1007/s11747-010-0216-3
- Steg, L., Perlaviciute, G., van der Werff, E., and Lurvink, J. (2014). The significance of hedonic values for environmentally relevant attitudes, preferences, and actions. *Environ. Behav.* 46, 163–192. doi: 10.1177/0013916512454730
- Stern, P. C. (2000). Toward a coherent theory of environmentally significant behavior. *J. Soc. Issues* 56, 407–424. doi: 10.1111/0022-4537.00175
- Stern, P. C., Dietz, T., and Kalof, L. (1993). Value orientations, gender, and environmental concern. *Environ. Behav.* 25, 322–348. doi: 10.1177/0013916593255002
- Thompson, E. R. (2007). Development and validation of an internationally reliable short-form of the positive and negative affect schedule (PANAS). *J. Cross Cult. Psychol.* 38, 227–242. doi: 10.1177/0022022106297301
- Watson, D., Clark, L. A., and Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: the PANAS scales. *J. Personal. Soc. Psychol.* 54, 1063–1070. doi: 10.1037/0022-3514.54.6.1063
- Zaichkowsky, J. L. (1994). The personal involvement inventory: Reduction, revision, and application to advertising. *J. Advert.* 23, 59–70. doi: 10.1080/00913367.1943.10673459

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The reviewer EP and handling Editor declared their shared affiliation, and the handling Editor states that the process nevertheless met the standards of a fair and objective review.

Copyright © 2017 Herziger, Benzerga, Berkessel, Dinartika, Franklin, Steinnes and Sundström. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Eating for Two? Protocol of an Exploratory Survey and Experimental Study on Social Norms and Norm-Based Messages Influencing European Pregnant and Non-pregnant Women's Eating Behavior

Kirsten E. Bevelander^{1*}, Katharina Herte², Catherine Kakoulakis³, Inés Sanguino⁴, Anna-Lena Tebbe⁵ and Markus R. Tünte^{6*}

OPEN ACCESS

Edited by:

Rocio Del Pino,
BioCruces Health Research Institute,
Spain

Reviewed by:

Konrad Schnabel,
International Psychoanalytic University
Berlin, Germany
Pietro De Carli,
Università degli Studi di Padova, Italy

*Correspondence:

Kirsten E. Bevelander
K.Bevelander@bsi.ru.nl
Markus R. Tünte
markusromantuente@gmail.com

Specialty section:

This article was submitted to
Quantitative Psychology
and Measurement,
a section of the journal
Frontiers in Psychology

Received: 30 October 2017

Accepted: 16 April 2018

Published: 08 May 2018

Citation:

Bevelander KE, Herte K, Kakoulakis C, Sanguino I, Tebbe A-L and Tünte MR (2018) Eating for Two? Protocol of an Exploratory Survey and Experimental Study on Social Norms and Norm-Based Messages Influencing European Pregnant and Non-pregnant Women's Eating Behavior. *Front. Psychol.* 9:658. doi: 10.3389/fpsyg.2018.00658

¹ Communication Science, Behavioural Science Institute, Radboud University, Nijmegen, Netherlands, ² Department of Social, Health and Organisational Psychology, Utrecht University, Utrecht, Netherlands, ³ Department of Psychology, University of Nicosia, Nicosia, Cyprus, ⁴ Department of Psychology, King's College London, London, United Kingdom, ⁵ Department of Psychology, University of Mannheim, Mannheim, Germany, ⁶ Faculty of Psychology, University of Vienna, Vienna, Austria

The social context is an important factor underlying unhealthy eating behavior and the development of inappropriate weight gain. Evidence is accumulating that powerful social influences can also be used as a tool to impact people's eating behavior in a positive manner. Social norm-based messages have potential to steer people in making healthier food choices. The research field on nutritional social norms is still emerging and more research is needed to gain insights into why some people adhere to social norms whereas others do not. There are indications stemming from empirical studies on social eating behavior that this may be due to ingratiation purposes and uncertainty reduction. That is, people match their eating behavior to that of the norm set by their eating companion(s) in order to blend in and be part of the group. In this project, we explore nutritional social norms among pregnant women. This population is particularly interesting because they are often subject to unsolicited advice and experience social pressure from their environment. In addition, their pregnancy affects their body composition, eating pattern, and psychosocial status. Pregnancy provides an important window of opportunity to impact health of pregnant women and their child. Nevertheless, the field of nutritional social norms among pregnant women is understudied and more knowledge is needed on whether pregnant women use guidelines from their social environment for their own eating behavior. In this project we aim to fill this research gap by means of an exploratory survey (Study 1) assessing information about social expectations, (mis)perceived social norms and the role of different reference groups such as other pregnant women, family, and friends. In addition, we conduct an online experiment (Study 2) testing to what extent pregnant women are susceptible to social norm-based messages compared to

non-pregnant women. Moreover, possible moderators are explored which might impact women's susceptibility to social norms as well as cultural aspects that co-determine which social norms and guidelines exist. The project's findings could help design effective intervention messages in promoting healthy eating behavior specifically targeted to European pregnant women.

Keywords: social norms, descriptive and injunctive messages, pregnant women, social norm messages, snacking behavior, sugar-sweetened beverages, behavioral nutrition

INTRODUCTION

There is substantial literature examining the general impact of the social environment on behavior (Cialdini et al., 1991), and evidence is accumulating that social influences on eating behavior are powerful as well (Christakis and Fowler, 2007; Salvy et al., 2012; Cruwys et al., 2015). Several methodological approaches have been used to investigate social influence on eating. Besides correlational studies, observational social eating studies have shown that people conform their eating behavior to social norms set by others, and that people converge upon an eating norm when eating together. For example, people who ate with a 'confederate' instructed to eat a certain type or amount of food were found to imitate or adjust their food choices and intake to that of their instructed eating companion (Salvy et al., 2012; Cruwys et al., 2015; Higgs, 2015). To date, the social context is increasingly recognized as an important factor underlying the development of inappropriate weight gain (Christakis and Fowler, 2007). Even though it is completely normal to gain weight for pregnant women, up to 40% gains more weight than recommended by health guidelines while this can have lifelong detrimental effects on both mother and child (Thangaratinam and Jolly, 2010). Surprisingly, the field of nutritional social norms among pregnant women is understudied (Gardner et al., 2012; Hutchinson et al., 2017). This population is of interest in particular, because pregnancy is a crucial period in life at which women may be more or less susceptible to social norms and dietary change than usual (Campbell et al., 2011). That is, pregnant women often experience that they receive unsolicited advice from everyone who knows or sees they are pregnant, and they have to deal with many implicit as well as explicit rules they are subject to by their pregnancy (Root and Browner, 2001; Graham et al., 2013). Moreover, they may feel 'allowed' to consume more unhealthy foods as their body composition changes and weight gain is more socially acceptable, despite being aware that their unborn child benefits from healthy eating (Campbell et al., 2011). Thus, pregnancy can provide women a reason or an excuse to change their diets for better or worse. The overarching goal of this project is to gain more insight into whether and how social norms play a role in pregnant women's eating behavior.

To our knowledge, literature is limited to few correlational studies examining social norms among pregnant women. Health norms during pregnancy in general were brought to attention in the 1970s in a study on smoking, drinking, eating, and physical activity (Barić and MacArthur, 1977). However, only two recent studies have further investigated the influence of the

social environment on dietary intentions and self-reported food intake (Gardner et al., 2012; Hutchinson et al., 2017). A study by Gardner et al. (2012) investigated whether *subjective norms* (i.e., anticipated social approval from the social environment to eat healthy) predicted healthy eating intentions among pregnant women in the United Kingdom. They found positive correlations between approval of eating behavior and family and health care expectations, although there was no direct association between social approval and healthy eating intentions. In a similar manner, a study by Hutchinson et al. (2017) among Australian pregnant women found that endorsement of healthy eating by others was unrelated to dietary intake. The authors speculated that pregnant women may differ from the general population in terms of susceptibility to social influences, because they are more concerned with changing their entire health behavior (e.g., alcohol consumption and smoking) for the benefit of their baby's health regardless of others' views. Additionally, they noted that more knowledge is needed on whether social influence depends on which reference group conveys the norm (e.g., partner, mother, family, pregnant, and non-pregnant friends) as well as whether conveyed norms need to be related to pregnancy. Importantly, both studies concluded that further research is warranted to fully understand the influence of social norms on pregnant women's eating behavior given that food intake usually takes place in social contexts (Gardner et al., 2012; Hutchinson et al., 2017).

To sum up, social norms may determine to a large extent what 'normal' as well as 'acceptable' eating behavior is for pregnant women. In this project, we examine the influence of social norms on pregnant women's eating behavior in a systematic approach. First, we aim to generate and refine hypotheses to give direction to future research (Study 1). Based on previous literature, nutritional social norms, and possible moderators are explored by means of an online survey. We then investigate pregnant women's susceptibility to social norm-based messages in a commonly used experimental research design (Study 2), based on existing literature and the determinants that emerge from Study 1.

Theoretical Framework Study 1: Exploring Pregnant Women's Susceptibility to Social Eating Behavior Social Norm Perceptions

The first aim of Study 1 is to examine the perception of social norms among pregnant women. Perceptions and beliefs about the 'normal' eating behavior of others influence people's

own behavior. For example, a person's perception about what influential others or the majority of peers do (i.e., perceived *descriptive* peer norm) may be an important factor in determining what people choose to eat and drink. Nonetheless, perceived peer norms do not always match the actual norm, which can lead to a 'false consensus' effect. That is, people direct their behavior to a false and *misperceived* norm. Ample studies in normative misperceptions have shown that people often overestimate the unhealthy behaviors peers do (e.g., alcohol, tobacco and substance use, and risky sexual behavior) (Haug et al., 2011; Bertholet et al., 2013; Lewis et al., 2014) whereas healthy behaviors tend to be underestimated (e.g., condom use and seat belt use) (Lewis et al., 2014; Litt et al., 2014). Likewise, studies on consumption behavior have found that people overestimated peers' unhealthy food and drink intake but underestimated their fruit and vegetable consumption (Lally et al., 2011). As directing one's behavior toward a false norm can be harmful and requires correction, gaining insight into misperceived norms is an important topic of investigation. We expect to find similar tendencies among pregnant women and therefore (H1) hypothesize that pregnant women tend to overestimate unhealthy consumption while underestimating healthy consumption norms.

Susceptibility to Social Norms

The second aim of Study 1 is to explore factors that may influence how susceptible pregnant women are to social norms. An important topic of investigation is which reference group women refer to concerning their own eating behavior since the beginning of their pregnancy. It has been shown that perceived shared group membership (i.e., others being similar to the self) plays a role in the degree to which people conform their behavior to others (Pachucki et al., 2011; Robinson et al., 2014b). For example, researchers have found that when an out-group member (dissimilar to the self) provides a healthy eating norm, reactance is triggered by which people eat more unhealthily and vice versa (Oyserman et al., 2007; Berger and Rand, 2008; Stok et al., 2012). More knowledge is needed on whether (and when) pregnant women regard other pregnant women, or their family and friends as their in-group. One could argue that exposure to pregnancy-related norms occurs only during 9 months, which is a relatively brief period compared to general normative influences that are accumulated over a lifetime (Hutchinson et al., 2017). Therefore, Study 1 explores which reference group affects pregnant women's eating behavior (e.g., when sharing a meal with non-pregnant others) utilizing a primarily qualitative approach. It also addresses whether similarity feelings depend on the duration and visibility of pregnancy which can make women feel more or less similar to particular reference groups.

In addition to insights on shared group membership, research has shown that conforming to social norms depends on the social bond between people and that this seems to be motivated by the desire to affiliate and reduce uncertainty (Cruwys et al., 2015). For example, a cold and distant social interaction during dinner led women to direct their eating behavior more toward that of a cold acting eating companion than when the social interaction was warm and friendly, which is believed to reflect an ingratiation

attempt (Hermans et al., 2009). Other studies have focused on the role of empathy, self-esteem, body-esteem, and sociotropy to scrutinize which underlying mechanisms are at play in social modeling behavior (Robinson et al., 2011; Exline et al., 2012; Hirata et al., 2015). For example, in an experimental study asking female dyads to complete a problem solving task together while having access to food, the degree of matching food intake was associated with their empathy and self-esteem (Robinson et al., 2011). That is, women with lower self-esteem were found to match their intake more than those with higher self-esteem. In a study of Exline et al. (2012), sociotropy predicted people's attempts to match their companion's eating with their own to make their companion feel comfortable. In addition, it predicted more personal distress related to social pressure by eating more. Although findings are mixed, the general pattern indicated that people tend to conform their consumption behavior more when they feel uncertain, want to please others or fear to be socially excluded (Cruwys et al., 2015).

With regard to pregnant women, research has shown that they are confronted with changes in their psychosocial status (e.g., anxiety, stress, depression, and self-esteem) (Hickey et al., 1995) that can affect their well-being and weight gain during pregnancy (DiPietro et al., 2003). Based on social eating literature (Cruwys et al., 2015), it is plausible that these factors also play a role in pregnant women's reactions to social normative information. For example, research has shown that self-esteem plays a role in the perception and need of feeling socially accepted (Baumeister and Leary, 1995). That is, people with higher self-esteem tend to worry less about how they are seen by others and conform less to other's behavior (Leary and Baumeister, 2000). Interestingly, pregnancy can be seen as a time when being large is socially acceptable (Campbell et al., 2011) and therefore can give a sense of confidence. In addition, it has been suggested that pregnant women who are preoccupied with the health of their child may feel less uncertain because they are more likely to strictly follow dietary guidelines (Hutchinson et al., 2017). In turn, this might make them less susceptible for specific food norms from their social environment (Hutchinson et al., 2017). In contrast, pregnant women were also found to experience less self-confidence because they feel less physically attractive, subject to public's opinions and limited in their activities (Campbell et al., 2011; Graham et al., 2013). Given that this project is the first that investigates underlying mechanisms of social norm influences among pregnant women, we explore whether and how factors linked to ingratiation and uncertainty reduction (e.g., self-esteem, need to belong and perceived social support or sabotage) play a role in social eating behavior among pregnant women. We (H2) postulate that factors related to affiliation purposes and uncertainty reduction also affect the susceptibility to social norms in pregnant women.

Further, and although not linked directly to social eating behavior, studies have found other factors influencing pregnant women's health behaviors in general such as mindful eating, anxiety, self-regulation, and impulsivity (Rofé et al., 1993; Hickey et al., 1995; Hutchinson et al., 2017). These factors may moderate the degree of pregnant women's responsiveness to social norms. For example, their self-control may be increased

when they are determined to eat healthy for the benefit of their baby resulting in disregard for influences from their social environment. Given that more knowledge is necessary on whether psychological determinants affect pregnant women's susceptibility to social influences, we also explore the role of above mentioned psychological factors outside the ingratiation theme in a qualitative manner.

In conclusion, a deeper understanding of social, personal, and psychological factors underlying pregnant woman's eating behavior is needed. By means of an exploratory survey, Study 1 assesses information about (mis)perceived social norms and social expectations, and the role of different reference groups such as other pregnant women, family, and friends (Campbell et al., 2011). Moreover, we aim to identify underlying mechanisms (e.g., misperceived expectations or sensitivity to social sanctions) that might explain their behavior and whether health considerations may decrease social susceptibility (Hutchinson et al., 2017).

Theoretical Framework Study 2: Experimental Study of Social Norm-Based Messages on Food Choice Social Norm-Based Messages

Next to the empirical social eating literature (Cruwys et al., 2015), people have been found to adhere to norms in situations where individuals were merely exposed to *written* information about what other people did. Typical experimental studies on written norms expose people to information about people's eating behavior and through social norm messages (e.g., by exposure to a poster message in an evaluation task or to information about what prior participants had eating during a task or test) (Robinson et al., 2013, 2014b; Stok et al., 2014). Social norm-based messages have succeeded in changing intentions and behaviors unrelated to health (e.g., pro-environmental behavior) (Goldstein et al., 2008; van der Linden, 2015), but evidence is accumulating that it could be applied in the eating domain as well (Stok et al., 2012; Robinson et al., 2014b; Higgs, 2015; Robinson, 2015). In general, norm-based messages suggesting that others ate large portions of food were associated with increased food intake, and vice versa. In addition, information about food choice norms was found to influence the consumption of unhealthy snack food as well as healthy snacks such as fruit and vegetables (Robinson et al., 2014b; Robinson, 2015). These findings suggest that so-called 'social norm-based' messages have potential to steer people in making healthy food choices.

Research on social norm-based messages related to eating behavior has focused mainly on two types of messages, namely *descriptive* and *injunctive* messages (Burger et al., 2010; Mollen et al., 2013; Robinson et al., 2013, 2014b; Stok et al., 2014; Robinson, 2015). Descriptive messages provide general information and describe what is the 'normal' consumption of (the majority of) others, whereas injunctive messages proscribe what is socially approved of and is found to be the 'appropriate' consumption according to others (i.e., how others want you to behave). There is a limited number of studies testing both descriptive and injunctive norm-based messages (versus a health message or no-message control condition). These studies have

shown mixed findings. For example, a correlational study on fruit intake found that compared to a control group, individuals reported having taken more fruits after being exposed to a descriptive norm but not to an injunctive norm message (Stok et al., 2014). Another study testing a healthy descriptive and injunctive norm and an unhealthy descriptive norm also found that compared to a control group, a healthy descriptive message led to more healthy choices whereas a healthy injunctive norm did not (Mollen et al., 2013). None of the norm messages affected unhealthy food choices in this study (Mollen et al., 2013), whereas another study testing a descriptive norm on junk food intake found opposite results (Robinson et al., 2013). A descriptive norm message reduced junk food intake compared to a control message; however, it did not reduce intake any more than a health message (Robinson et al., 2013). A study that compared a descriptive healthy norm with a healthy message did find a significant effect for the norm over a health message (Robinson et al., 2014a). Overall, descriptive norm messages seemed to have the biggest impact on healthy food intake (Burger et al., 2010; Mollen et al., 2013; Robinson et al., 2013; Stok et al., 2014). As an explanation, it has been argued that a descriptive type of message does not threaten people's sense of freedom compared to an injunctive norm (Stok et al., 2014). That is, telling people explicitly that they should (not) do something (i.e., injunctive norm), may lead to a dismissal of the message or might even provoke an opposite response ('boomerang effect' or 'reactance') (Knowles and Linn, 2004; Brehm and Brehm, 2013). Remarkably, none of the norm-based message studies have investigated whether psychosocial determinants influence people's susceptibility.

Study 2 advances knowledge in the emerging field of research on social norm-based messages on eating behavior by focusing on two parts. In part A, we investigate general differences between pregnant and non-pregnant women after exposure to norm-based messages. We (H3) speculate that a descriptive norm message has a positive effect on pregnant as well as non-pregnant women's healthy food choice compared to a control condition. Given that pregnant women are subject to unsolicited advice, rules and regulations from society (Root and Browner, 2001), we (H4) hypothesize that exposure to an injunctive norm-based message causes a reactance effect on food choice (meaning that women will choose more unhealthy foods) compared to a descriptive norm message and a control condition. We expect this effect to be stronger for pregnant than non-pregnant women. Part B explores moderating variables of social norm-based messages on food choice in the pregnant and non-pregnant women samples separately, based on the outcomes of Study 1 and social eating literature. Similar to Study 1, we explore which factors play a role in the susceptibility to social norms.

Societal Relevance

Lifestyle and dietary habits of pregnant women have lifelong effects on themselves and their child's weight and health (Birdsall et al., 2009). Interventions and educational activities generally aim to inform pregnant women about the harm of smoking, alcohol consumption, and drug use (Johnson et al., 1987; Lumley et al., 2009; Stade et al., 2009; Nilsen, 2010). Regarding nutrition,

most advice is focused on preventing women from suffering deficiencies in micronutrient requirements (e.g., vitamins and folic acid) or eating high risk foods (e.g., unpasteurized milk and soft cheese or raw fish and meat) causing listerial infection or toxoplasmosis that can affect fetal and child development (Ray and Laskin, 1999; Willers et al., 2007; Guelinckx et al., 2008; Janakiraman, 2008; Paquet et al., 2013). An important topic that has received less attention but could nonetheless be fruitful in terms of intervention is the prevention of excess weight gain during pregnancy (Cogswell et al., 1999; Strychar et al., 2000). It is increasingly recognized that inappropriate weight gain during pregnancy has persisting effects on child adiposity, cognitive development, blood pressure, and atopic disease as well as post-partum weight retention among mothers (Birdsall et al., 2009; Graham et al., 2013). Despite the short and long term neonatal and maternal benefits of an appropriate diet during pregnancy, 20–40% of pregnant woman in Europe gain more weight than recommended by health guidelines (Thangaratinam and Jolly, 2010). As explained above, social norms can motivate and direct a person's behavior, because they are linked to social sanctions and rewards for (non)conformity, and social expectations. It is suggested that when one wants to influence behavior permanently, social norms need to be changed first (Barić and MacArthur, 1977). This project is the first to investigate whether social norms can be used as an effective method to impact dietary intake of pregnant women for the benefit of both mother and child.

STUDY 1 – ONLINE SURVEY: EXPLORING PREGNANT WOMEN'S SUSCEPTIBILITY TO SOCIAL EATING BEHAVIOR

Stepwise Procedures

Translation of Materials

For Study 1 (and 2) the authors translate all materials from English to their countries' official language (i.e., forward translation). An English native speaker with a proficiency level in the target language translates the survey from the target language to English (i.e., back translation). The translations are reviewed and compared, discussing disagreements until consensus is reached. In both studies, pilot tests are conducted to receive feedback regarding the clarity and length of the materials and the presentation of the questionnaire measurements.

Design and Participants

The general aim of Study 1 is to explore (mis)perceived social norms and expectations of different reference groups together with potential underlying mechanisms that influence pregnant women's (un)healthy snacking and drinking behavior. We use an exploratory mixed method approach (i.e., qualitative as well as quantitative) to collect information via an anonymous online survey with open-ended and closed questions. Results will be used to extract moderator variables for Study 2.

The study will take place between March and May 2018. Women between 18 and 40 years old with uncomplicated singleton pregnancies are eligible to participate in the study. Given that the study aim is primarily exploratory, a conservative *a priori* power analysis (G*Power 3.1.9.2) for the regression analysis was performed accounting for seven predictors and a moderate effect size (two-tailed, $f^2 = 0.01$; power 0.95, $\alpha = 0.05$). This resulted in a total sample size of at least $N = 132$ participants. Taken into account an attrition rate of 15–20%, we aim to recruit at least 25 participants per country.

Participants are recruited by a purposive sampling method in six European countries (i.e., Netherlands, Germany, United Kingdom, Spain, Austria, and Cyprus/Greece) (Etikan et al., 2016). Advertisements are placed at typical locations where pregnant women come (e.g., midwife practices, medical offices, pregnancy yoga, or swimming classes) and on online platforms (e.g., at parenting and pregnancy forum, specific Facebook groups, and online newsletters). The advertisements fully explain the aim of the study and invite pregnant women to participate in the online study. Before starting the questionnaire, they can read additional information about the study aim, procedure, and context to ensure transparency and allow for proper consideration of participation. The participants are also informed about the anonymity and confidentiality of their answers and the right to withdraw from the study at any stage.

Materials and Equipment

Measures

Participants fill out open and closed-ended online questionnaires using online survey software (Qualtrics), covering demographic and pregnancy-related information, social norms, snacking and drinking behavior and psychological factors. All survey measurements are existing validated questionnaires, translated into the appropriate language. Some questionnaire items are tailored to identify specific behaviors among pregnant women. All measurements are described in detail in the following paragraphs.

Demographics

Self-reported demographics are assessed by asking for participant's age, gender, nationality, height and weight before pregnancy, weight gain, level of education and socio-economic status. To ensure that cultural differences in the study are accurate, participants are asked since when they have lived in their current place of residence. Participant's Body Mass Index (BMI) before pregnancy will be calculated using the standard formula $\text{weight [kg]} / \text{height}^2 \text{ [m]}$.

Pregnancy related measurements

Pregnancy related items that potentially affect women's diet or susceptibility to social norms are assessed, such as duration of pregnancy, singleton or twin pregnancy, parity and experience with miscarriages. In addition, pregnancy ailments are administered on a six-point scale ranging from 'Never' (1) – 'Always' (6), such as nausea, stomach acid, loss or increase of appetite, constipation, tiredness, etc. (Hutchinson et al., 2017).

Eating behavior

Although pregnancy can have influence on a women's entire diet, this project focuses on snack consumption and beverage intake, specifically. An increased consumption of palatable high-sugary and -fat snack foods (e.g., cakes, biscuits, and crisps) and sugar-sweetened beverages (SSB's) has been found to contribute to inappropriate weight gain (Olafsdottir et al., 2005; Crozier et al., 2009; Graham et al., 2013). Meeting recommendations related to sugar intake during pregnancy reduces complications and supports appropriate weight gain, optimal fetal growth, and childhood development resulting in an improved health for both mother and her newborn (Birdsall et al., 2009).

Self-reported fruit consumption. It is assessed by asking participants to report their fruit consumption of the past 2 days. A list of 26 commonly consumed fruits (for every country involved) is provided. Participants indicate the type and amount of fruit they had consumed (in handfuls for small fruits such as raspberries and in pieces for larger fruits such as apples). In addition, three 'other' options are provided enabling participants to add fruits to the list. In line with previous research, consumption is calculated by computing the total amount of portions of fruit consumed (Verkooijen et al., 2015). For example, two or three pieces of smaller fruits (e.g., prunes) equal one portion normal-sized fruits (e.g., apple) whereas parts of large fruits (e.g., melon) count as one portion. Average daily consumption is calculated by dividing the number by 2 days.

Self-reported snack consumption. Similar to the fruit assessment procedure, a list with 13 unhealthy snacks is presented including small or large cookies, sweets, chocolates, warm snacks, etc. The total number of unhealthy snacks will be calculated in the same way as fruit consumption and divided by 2 days (Verkooijen et al., 2015).

Beverage consumption. Beverages are assessed by asking the number of glasses (equaling cans, bottles, and packages of 220 ml) participants drank during the past 2 days. Administered drinks are (sparkling) water, dairy products, sugar-sweetened soft drinks, artificially sweetened (i.e., diet) soft drinks, fruit-flavored drinks (i.e., lemonade), fruit juice, energy drinks and (sweetened) tea and coffee. Response categories range from 'zero glasses per day' (0) to 'five glasses per day' (5) (Smit et al., 2016). Drinks are classified as sugar-sweetened and energy-dense 'unhealthy' or low sugary and low-energy dense 'healthy' (Briefel et al., 2009).

Normative information and influence of the social environment

To develop and refine hypotheses for future studies, participants' social norm expectations and (mis)perceptions as well as the role of different reference groups such as other pregnant women, family, and friends that convey the social norms are explored. In addition, the impact of the social environment is assessed by asking about the women's social surroundings.

Actual norms. The actual norms are calculated by the daily average number of servings of fruits and snacks, and glasses of healthy and unhealthy beverages consumed.

Identification with norm referent group. To assess influential individuals in pregnant women's social surrounding, participants rate the extent to which they identify with different reference groups such as family, partner, and friends (Stok et al., 2012) (e.g., *I feel a strong connection to other pregnant women*), on a six-point scale ranging from 'not at all' (1) to 'very much' (6).

Perceived descriptive norms. Participants are asked to estimate how many servings of fruit, snacks and glasses of SSB's they think other pregnant women generally eat and drink per day. Additionally, the same question is asked for the reference group they indicate to be most important (e.g., family, partner, or friends).

Misperceived norms. Misperceived norms are calculated by subtracting the actual mean number of fruits, snacks, and drinks from the perceived actual mean number (for pregnant and other reference group, separately). A lower score indicates that participants underestimated their consumption, a higher score indicates an overestimation of the norm respectively.

Perceived injunctive norms. Injunctive norms are operationalized by calculating how many servings of fruit, snacks, and drinks participants think other pregnant and non-pregnant women approve of and think they should consume. The same question is asked for the reference group they indicate to be most important.

Social expectations. In addition to the perceived injunctive norm, participants are asked whether they think that it is generally expected by other pregnant women that a pregnant woman should modify her diet (Barić and MacArthur, 1977). This is also assessed for their most important reference group.

Social support and sabotage. Social support from friends and family for eating healthy is measured by selected items of the Friend and Family Support for Healthy Eating Habits scale (Sallis et al., 1987). Participants are asked 'How often...' they feel that friends or family support or sabotage their healthy eating by answering options on a six-point scale ranging from 'Never' (1) - 'Always' (6). Example questions are '...does your family encourage you to eat healthy foods?' and '...do friends eat unhealthy foods in front of you?' Participants are given examples of healthy (e.g., low-fat and low-sugar foods illustrated with example products) and unhealthy foods before starting the questionnaire. Originally, the scale contains items about low-fat and low-sugar foods separately, but we combined them into one 'unhealthy' item to simplify the question and shorten the list of items.

Ingratiation and uncertainty reduction

The extent to which pregnant women are susceptible to social normative influences may be determined by factors related to ingratiation purposes and uncertainty reduction (Cruwys et al., 2015). The following measures are assessed to explore whether these factors play a role among pregnant women as well.

Self-esteem. Social and appearance self-esteem is measured by two subscales of the State Self-Esteem Scale (SSES) (Heatherton and Polivy, 1991). The SSES is a 20-item questionnaire measuring both positive and negative thoughts about oneself related to

social and appearance self-esteem. Items are framed as ‘*How often... participants feel worthy or not about themselves [e.g., ‘How often are you feeling unattractive?’ (appearance) or ‘...are you concerned about the impression that you make?’ (social)] with answering categories ranging from ‘Never’ (1) to ‘Always’ (6).*

Fear of negative evaluation. The fear and distress of being evaluated unfavorably by others in a social situation is measured by the 12-item Brief Fear of Negative Evaluation Scale (BFNE) (Leary, 1983). Items are prefaced with ‘*How often... participants feel distress or social anxiety with answering categories ranging from ‘Never’ (1) to ‘Always’ (6) (e.g., ‘How often are you afraid of other people noticing your shortcomings?’ or ‘...do you worry that you will say or do the wrong things?’).*

Health benefits. Pregnant women’s susceptibility to social norms may be influenced by how preoccupied women are with their own and their baby’s health (Hutchinson et al., 2017). The Nutrition Benefit scale is adapted from previous research (Tiedje et al., 1992) and participants rate how strongly they agree with statement such as ‘*If I don’t eat healthy, there could be something wrong with my baby*’ and ‘*Good nutrition during pregnancy will prevent me from gaining a lot of weight*’ on a six-point scale ranging from ‘*Not at all*’ (1) to ‘*Very much*’ (6). In addition, we use the Figure Rating Scale which depicts nine silhouettes of female adult body figures ranging from very thin (1) to obese (9) (Stunkard et al., 1983). Participants indicate their figure before pregnancy and their desired figure 6 months after pregnancy. The difference score between the desired and actual figure provides an indication of women’s preoccupation about their weight and appearance.

Social desirability bias

Although this research is related to social desirable behavior, we also want to take a possible social desirability bias into account. This refers to the tendency of participants to answer in a particular way that will be viewed favorably by others. Therefore, the short Marlowe-Crowne-Social Desirability Scale (SDS) is used to measure socially desirable responses (Reynolds, 1982; Loo and Thorpe, 2000). It consists of 10 items prefaced by ‘*How often... (e.g., ‘...do you like to gossip?’ and ‘...are you willing to admit it when you make a mistake?’)* with answering categories ranging from ‘*Never*’ (1) – ‘*Always*’ (6). The SDS can detect a form of over-reporting ‘good behavior’ or under-reporting ‘bad’ or undesirable behavior.

Open-ended survey items

To collect more in-depth information about social normative behavior and pregnant women’s susceptibility to social norms, open-ended questions are administered throughout the survey (Root and Browner, 2001). Apart from filling in demographic and pregnancy related information, women are asked whether, how and why pregnancy changed their life in general and in relation to their diet, specifically. In addition, we explore which people (i.e., reference group) may have an important influence on pregnant women’s diet and choices (e.g., ‘*Who in addition to your physician or nurse do you turn to for information or advice during your pregnancy?’*). Next, other factors that have been raised by scarce literature on pregnant women’s health behaviors are presented

(i.e., mindful eating, anxiety, self-control, and impulsivity) (Rofé et al., 1993; Springer et al., 1994; Hickey et al., 1995; Hutchinson et al., 2017). Participants are asked to elaborate on whether and why (not) they think that they influence their own susceptibility to social norms. The factors are clarified before they are presented to the participants. Finally, the participants are invited to think of additional factors concerning the role of social norms and susceptibility on their eating behavior.

Proposed Analysis and Anticipated Results

Aim of the analyses are generating insights into (mis)perceived social norms, reference groups and underlying mechanisms of susceptibility to social norms among pregnant women. Results from these exploratory analyses serve to define and test specific hypotheses for Study 2 in an experimental setting.

The methodological principles of the qualitative data part are based on the grounded theory approach, which is a systematic procedure for building new theory (Corbin and Strauss, 2008). Qualitative data stems from responses to the open-ended survey items. The responses are coded by multiple researchers (i.e., the authors) to minimize coding bias and increase reliability of the coding procedure. Through the coding process, researchers create meaningful labels for participants answers. The procedure allows to make sense of data in a flexible and iterative process in which the analysis procedure goes back and forth until no new coding categories occur and ‘theoretical saturation’ is reached. English transcriptions are analyzed using thematic content analyses (Braun and Clarke, 2006). Data are analyzed (MAXQDA software package) conducting (1) open coding followed by (2) axial coding (i.e., making salient and subcategories) and (3) selective coding (i.e., integrating categories into theoretical concepts) (Corbin and Strauss, 2008; Boeije, 2010). All core categories and identified relationships are documented in a coding dictionary. (1) After familiarizing with the data, the coders generate initial codes and place them into top-level ‘labels’ or ‘themes’ (e.g., home environment, psychosocial factors). Coders compare their individual codes and discuss them until consensus is reached. (2) Axial coding is used to create main and subcategories of the codes. This is done by identifying relationships between and across the themes while examining recurring phenomena, actions, and interactions. The coders discuss their categorizations and after reaching consensus, the identified themes can be further divided into second-level categories. (3) The final step is to integrate the core categories into theory. The themes and subcategories that emerge can provide insight for future directions of research (Corbin and Strauss, 2008; Boeije, 2010). For example, they may support or counter our assumption that ingratiation plays a role in being susceptible for social influence, or point to a different factor that emerged from the open-ended items which needs to be further investigated. In addition, it is possible that pregnancy related items (e.g., the duration of pregnancy and visibility/body figure) appear to play an important role in being susceptible to social norms. The top-level categories and their determinants are also compared between countries to investigate cross-cultural differences.

For the quantitative data analyses, scale reliability and construct validity is assessed by factor analysis according to standard procedures. To test whether participants overestimate the unhealthy and underestimate the healthy consumption norm (H1), the misperceived norms are calculated by subtracting the actual mean number of fruits, snacks and drinks from the perceived actual mean number. Next to exploring whether factors play a role in the susceptibility to social norms in a qualitative manner, we will test whether norm (mis)perceptions can be predicted by factors related to uncertainty reduction or affiliation purposes (e.g., measures 5.1–5.3) using regression analyses (H2). We will only include data in the analyses from participants who completed the survey and did not withdraw their assent. Categorical variables are (dummy) coded as whole numbers. Statistical analysis is performed using SPSS Statistics 23 and R (2013). Statistical significance is considered at the $p < 0.05$ level.

STUDY 2 – ONLINE EXPERIMENTAL STUDY OF SOCIAL NORM-BASED MESSAGES ON FOOD CHOICE: ‘MEMORY AND PLANNING PERFORMANCE OF PREGNANT WOMEN’

Stepwise Procedures

Design and Participants

The first part of Study 2 (Part 2A) has a 2 (pregnant vs. non-pregnant women) \times 3 (poster condition) between-subjects design. In a poster memory task, pregnant- and non-pregnant women are exposed to either a descriptive or an injunctive norm-based message or a message unrelated to eating or social norms (i.e., control condition). The memory task is followed by a planning task in which participants make food choices (i.e., dependent variable). Participants are assigned randomly to one of the experimental conditions. The order at which participants are exposed to the posters in the memory task is not randomized, because the first posters are used as memory practice trials. In order to avoid effects of order or sequences, the food choice pictures in the planning task are presented randomly and the snacks and drinks picture blocks are counterbalanced. Part 2B examines social normative predictors and moderating variables among pregnant and non-pregnant sample separately.

Study 2 takes place between May and September 2018 and has the same inclusion criteria as Study 1. We check whether participants participated in Study 1 and if they did, they will be excluded from Study 2 due to the fact that they are already informed about the general objective of our study. Recruitment of pregnant participants will also follow the same procedure as in Study 1, while non-pregnant females will be recruited by convenience sampling at similar locations (e.g., at yoga classes or Internet fora). Before participating in the online study, participants will provide informed consent.

Power calculations (G*Power 3.1.9.2.) are based on the design of Study Part 2B, requiring more participants than Study Part 2A by conducting the analyses for two samples separately

(i.e., pregnant and non-pregnant women). To detect a medium to large effect size using multiple linear regression ($f^2 = 0.15$; power 0.95, $p = 0.05$) and estimated with seven potential predictors (e.g., hunger, liking of experiment, pregnancy duration, poster condition, self-esteem, health benefit, and fear of negative evaluation), approximately 74 participants are needed. Taking into account an attrition rate of 15–20%, this results in the recruitment of 90 pregnant and non-pregnant women ($N = 180$). Consequently, we aim to recruit at least 15 pregnant and 15 non-pregnant women per country.

Cover Story and Experiment

Participants are delivered a cover story to conceal the actual aim of the study. Participants are told that they are participating in a study called ‘Memory and Planning Performance of Pregnant and Non-Pregnant Women.’ The experimental study consists of three parts. After providing assent and filling in demographic information, participants complete a memory task. In the memory task participants are asked to memorize specific details of posters (e.g., pictures, colors, and text). One at a time, participants are exposed to four posters, three of which are bogus posters to conceal the actual aim of the study. The real stimulus displays a norm-based message (participants in the control group are exposed to a poster unrelated to social norms or food). After each poster, participants have to describe what they remember and answer specific questions about the poster. This procedure ensures that participants pay attention to the stimulus material and their recall of the message is used as a manipulation check.

The second part of the study is the planning task in which participants have to plan ahead their daily activities (*‘Plan your day tomorrow’*) by choosing pictures displaying themes such as clothing, activities and food and drinks they will wear, perform, and consume on the next day, respectively. Thereby, we aim to assess food choice after exposure to the poster message while concealing the actual aim of the study. Next, participants are asked to memorize the pictures of part one again (used as additional manipulation check). In the final part of the study, they answer a questionnaire concerning moderator variables selected from the results of Study 1, and mixed with bogus items. Participants are asked to guess the aim of the study before they are debriefed, and they will have the opportunity to withdraw before ending the study.

Materials and Equipment

Stimulus Material

During the memory task, one of the four posters displays a social-norm based message with either a descriptive or injunctive norm. The descriptive norm provides information about what the majority of others consume and the injunctive norm about what most other people think (non) pregnant women should consume. The control condition involves a message unrelated to social norms, health and food.

Measures

Food choice

The dependent variable consists of the participants’ selected food choice items and quantity. During the planning task,

a combination of 32 pictures of healthy (e.g., apples, snack tomatoes, and dried fruit) and unhealthy snacks (e.g., cookies, sausages, and savory pastry) in small or large portions are shown. Likewise, there is a combination of 20 pictures of healthy (e.g., water and tea) and unhealthy (e.g., energy drink and chocolate milk) drinks, presented in 1 or 3 glasses. Food choice is calculated using kilocalories (kcal) for unhealthy and healthy foods and drinks separately.

Demographics, Pregnancy Related, and Moderator Variables

Administered demographic and pregnancy related measures follow the procedure described for Study 1. Moderator variables are selected based on the outcomes of Study 1 (as presented in the Method section of Study 1).

Manipulation Check

Recall, perception and credibility of normative and non-normative information are assessed in the experimental and control conditions (Stok et al., 2012). Recall of the message is assessed by checking whether participants memorized the norm message. Perception of the norm is checked by asking whether they thought that the percentage of other people referred to in the poster was low or high, and credibility of the normative statement is assessed by asking whether they found the norm to be credible [answering on a six-point scale ranging from 'Not at all' (1) to 'Very much' (6)]. In addition, the software program Qualtrics measures how long participants watched the posters (in seconds) to check whether participants actually paid attention to the poster.

Control Variables

Control variables that have been shown to affect food choice or the social norm manipulation are assessed, such as hunger, time of day, liking of the poster, etc. (Bevelander et al., 2012). These are measured at the end of the experiment to conceal the real aim of the study. Participants' subjective hunger state and liking of the poster are assessed by a slider on a Visual Analog Scale (VAS) ranging from 0 to 100 (e.g., 'Not hungry at all' to 'Very hungry'). It is also registered when participants engage in the experiment, because time of the day can affect participants' food choice (i.e., afternoons are more commonly snack times than mornings) (Cross et al., 1994).

Proposed Analysis and Anticipated Results

Data is screened to identify and remove outlying values as well as participants that terminated their participation early or who guessed the study's aim. Scale reliability and construct validity is assessed by factor analysis according to standard procedures. Randomization checks are performed testing for differences between the poster conditions on demographic variables and potential control variables (e.g., age and hunger) by use of one-factor analysis of variance (ANOVA). Randomization is successful when there are no significant differences between conditions. In case there is a significant difference, the variable is added as a control variable in the main analyses. In a similar manner, manipulation checks are performed to check whether all participants were exposed to the poster and the cover story

was delivered successfully. Participants who do not recall the poster message are excluded from the analysis. If there is a difference between poster conditions on credibility, this variable is added as a control variable in the main analysis. Next, Spearman's rank and Pearson's correlations are calculated for the demographic variables, control variables and the outcome measures (un)healthy food choice (in kcal) to determine which variables have to be included in the main model as covariates.

For Study Part 2A, multivariate analysis of (co)variance (MAN(C)OVA) is performed by testing a main effect of poster conditions on food choice and an interaction with pregnant and non-pregnant women. Pairwise comparisons are carried out with Bonferroni correction to determine significant differences between the experimental conditions. These analyses provide more insights into our hypotheses (H3&H4) that pregnant women exposed to an injunctive norm-based message would have a higher unhealthy intake than pregnant women exposed to a descriptive norm or control condition, and that this effect is stronger in non-pregnant women. We also expect that a descriptive norm message has a greater effect on both pregnant and non-pregnant women's consumption behavior compared to a control condition. The analyses of Study Part 2B further scrutinize the moderating role of determinants relating to affiliation purposes and uncertainty reduction within the pregnant and non-pregnant sample separately. No hypotheses are formed because this is the first study that explores the role of these potential moderating factors. Statistical analysis is performed using SPSS Statistics 23 with a significance level of $p < 0.05$.

LIMITATIONS

Exploratory studies go along with a number of limitations. First, the online studies use non-probability sampling methods (i.e., purposive as well as convenience sampling), possibly resulting in a sampling bias (Etikan et al., 2016). For example, by an over-representation of pregnant women who are concerned about their own and child's health (i.e., women who participate in yoga lessons or search for information on online parenting forums) or an under-presentation of women who are digital illiterate or live in a remote area without appropriate Internet access. To strive for a heterogeneous and representative sample of pregnant women, the studies will be advertised at locations where pregnant women go (e.g., midwives or medical offices and diverse Internet fora). Further, this project is exploratory and the goal is to generate and refine hypotheses which will be particularly valuable for the directions of future research. Second, this project administered self-reported recall and predictions of food intake and choice only. Future research would profit a great deal from including real eating behavior using, for example, daily reports in a diary study or an experimental setup in a natural environment. Third, there may be socio-cultural differences between some of the European countries that affect women's susceptibility to social norms and their eating behavior. For example, it is possible that social norms and expectations are more stringent across Southern than Northern European areas (Heinrichs et al., 2006; Gelfand et al., 2011). That is, in some countries there

may be a greater focus on close family relationships and the community that could magnify the effect on normative social behavior (e.g., Spain and Cyprus) (Campos et al., 2008) whereas this may be less pronounced in other individualistic countries such as United Kingdom and Netherlands (Gelfand et al., 2011). In addition, some literature on eating patterns suggest differences across North-European, Western and Mediterranean cultures (Wolff and Wolff, 1995; Cuco et al., 2006; Northstone et al., 2008; Crozier et al., 2009) whereas other research shows that there is a progressive narrowing of differences in dietary patterns across Northern and Southern European countries (Naska et al., 2005). Therefore, we will explore whether there are socio-cultural differences between countries. If we find differences in Study 1, we take appropriate measures by, for example, dummy coding Northern and Southern European countries in our analyses. In addition, we will adjust the sample size of Study 2 to avoid the experiment from being underpowered by performing new power analyses. This will improve the generalizability of our findings across the different countries. Finally, the use of self-reported data often results in dealing with social desirable answers and participation bias. To prevent this, the aim of the study is explicitly stated and the anonymous data handling is stressed in Study 1. In addition, the cover story used in Study 2 limits social desirable answers.

CONCLUSION

This project aims to fill a research gap by broadening the existing scope of research into social norms. Combining exploratory research with research methodology used previously in social norm research among pregnant and non-pregnant populations enables comparing pregnant with non-pregnant women. This contributes to more knowledge about how women perceive guidelines from their social environment, which underlying mechanisms play a role and whether social norms can be used to stimulate healthy eating. Moreover, cultural aspects that co-determine which social norms and guidelines exist are taken into account. Social norm campaigns have shown to successfully change misperceived norms (DeJong et al., 2006; LaBrie et al., 2008) and promote healthier eating (Robinson et al., 2014a; Stok et al., 2014) by removing uncertainty about how to behave. The project's findings will help to develop and

REFERENCES

- Barić, L., and MacArthur, C. (1977). Health norms in pregnancy. *Br. J. Prev. Soc. Med.* 31, 30–38.
- Baumeister, R. F., and Leary, M. R. (1995). The need to belong: desire for interpersonal attachments as a fundamental human motivation. *Psychol. Bull.* 117, 497–529. doi: 10.1037/0033-2909.117.3.497
- Berger, J., and Rand, L. (2008). Shifting signals to help health: using identity signaling to reduce risky health behaviors. *J. Consum. Res.* 35, 509–518. doi: 10.1086/587632
- Bertholet, N., Faouzi, M., Studer, J., Daeppen, J.-B., and Gmel, G. (2013). Perception of tobacco, cannabis, and alcohol use of others is associated with one's own use. *Addict. Sci. Clin. Pract.* 8:15. doi: 10.1186/1940-0640-8-15

design effective messages for interventions in promoting healthy eating behavior specifically targeted to pregnant women among different European societies.

ETHICS STATEMENT

Both studies were conducted according to the principles laid down in the Declaration of Helsinki (World Medical Association, 2013). The studies were approved by the Ethics Committee of the Faculty of Social Sciences (ECSW-2017-012) of the Radboud University, Nijmegen, Netherlands. The study procedure, materials, and consent forms were reviewed in accordance to ethical guidelines 'Code of Ethics for the Social and Behavioural Sciences' and ethical standards in Dutch as well as European context [EU; General Data Protection Regulation (GDPR)]. Both studies involved online anonymous voluntary data collection procedures among adults without any expected adverse events. Each participant provided active consent for participation after having received information about the aim and procedures of the study. Although Study 2 used a cover story to recruit participants, it had a debriefing at the end of the online experiment. All participants were also given the opportunity to ask questions to the researchers and withdraw their assent during and at the end of the studies.

AUTHOR CONTRIBUTIONS

KB conceived and initially designed and wrote the protocol. KH, CK, IS, A-LT, and MT contributed equally to the research design, wrote parts of the manuscript, and provided important critical feedback when revising the manuscript. All co-authors are listed in alphabetical order. All authors approved the final version of the manuscript.

FUNDING

This protocol article was funded by an Open Access Publishing Fund of the Faculty of Social Sciences, Radboud University Nijmegen, Netherlands, and the Open Access Publishing Fund of the University of Vienna, Austria. The project was conducted as part of the Junior Researcher Programme (JRP).

- Bevelander, K. E., Anschutz, D. J., and Engels, R. C. (2012). The effect of a fictitious peer on young children's choice of familiar v. unfamiliar low- and high-energy-dense foods. *Br. J. Nutr.* 108, 1126–1133. doi: 10.1017/S0007114511006374
- Birdsall, K. M., Vyas, S., Khazaezadeh, N., and Oteng-Ntim, E. (2009). Maternal obesity: a review of interventions. *Int. J. Clin. Pract.* 63, 494–507. doi: 10.1111/j.1742-1241.2008.01910.x
- Boeije, H. (2010). *Analysis in Qualitative Research*. Thousand Oaks, CA: Sage.
- Braun, V., and Clarke, V. (2006). Using thematic analysis in psychology. *Qual. Res. Psychol.* 3, 77–101. doi: 10.1191/1478088706qp0630a
- Brehm, S. S., and Brehm, J. W. (2013). *Psychological Reactance: A Theory of Freedom and Control*. Cambridge, MA: Academic Press.
- Briefel, R. R., Wilson, A., and Gleason, P. M. (2009). Consumption of low-nutrient, energy-dense foods and beverages at school, home, and

- other locations among school lunch participants and nonparticipants. *J. Am. Diet. Assoc.* 109(Suppl. 2), S79–S90. doi: 10.1016/j.jada.2008.10.064
- Burger, J. M., Bell, H., Harvey, K., Johnson, J., Stewart, C., and Dorian, K. (2010). Nutritious or delicious? The effect of descriptive norm information on food choice. *J. Soc. Clin. Psychol.* 29, 228–242. doi: 10.1521/jscp.2010.29.2.228
- Campbell, F., Johnson, M., Messina, J., Guillaume, L., and Goyder, E. (2011). Behavioural interventions for weight management in pregnancy: a systematic review of quantitative and qualitative data. *BMC Public Health* 11:491. doi: 10.1186/1471-2458-11-491
- Campos, B., Schetter, C. D., Abdou, C. M., Hobel, C. J., Glynn, L. M., and Sandman, C. A. (2008). Familialism, social support, and stress: positive implications for pregnant Latinas. *Cultur. Divers. Ethnic Minor. Psychol.* 14, 155–162. doi: 10.1037/1099-9809.14.2.155
- Christakis, N. A., and Fowler, J. H. (2007). The spread of obesity in a large social network over 32 years. *New Engl. J. Med.* 357, 370–379. doi: 10.1056/NEJMsa066082
- Cialdini, R. B., Kallgren, C. A., and Reno, R. R. (1991). “A focus theory of normative conduct: a theoretical refinement and reevaluation of the role of norms in human behavior,” in *Advances in Experimental Social Psychology*, ed. M. P. Zanna (Cambridge, MA: Academic Press), 201–234.
- Cogswell, M. E., Scanlon, K. S., Fein, S. B., and Schieve, L. A. (1999). Medically advised, mother’s personal target, and actual weight gain during pregnancy. *Obstet. Gynecol.* 94, 616–622. doi: 10.1016/S0029-7844(99)00375-0
- Corbin, J., and Strauss, A. (2008). *Basics of Qualitative Research: Techniques and Procedures for Developing Grounded Theory*. Thousand Oaks, CA: Sage Publications, Inc. doi: 10.4135/9781452230153
- Cross, A. T., Babicz, D., and Cushman, L. F. (1994). Snacking patterns among 1,800 adults and children. *J. Am. Diet. Assoc.* 94, 1398–1403. doi: 10.1016/0002-8223(94)92542-9
- Crozier, S. R., Robinson, S. M., Godfrey, K. M., Cooper, C., and Inskip, H. M. (2009). Women’s dietary patterns change little from before to during pregnancy. *J. Nutr.* 139, 1956–1963. doi: 10.3945/jn.109.109579
- Cruwys, T., Bevelander, K. E., and Hermans, R. C. (2015). Social modeling of eating: a review of when and why social influence affects food intake and choice. *Appetite* 86, 3–18. doi: 10.1016/j.appet.2014.08.035
- Cuco, G., Fernandez-Ballart, J., Sala, J., Viladrich, C., Iranzo, R., Vila, J., et al. (2006). Dietary patterns and associated lifestyles in preconception, pregnancy and postpartum. *Eur. J. Clin. Nutr.* 60, 364–371. doi: 10.1038/sj.ejcn.1602324
- DeJong, W., Schneider, S. K., Gomberg Towvim, L., Murphy, M. J., Doerr, E. E., Simonsen, N. R., et al. (2006). A multisite randomized trial of social norms marketing campaigns to reduce college student drinking. *Subst. Abuse.* 67, 868–879. doi: 10.15288/jsa.2006.67.868
- DiPietro, J. A., Millet, S., Costigan, K. A., Gurewitsch, E., and Caulfield, L. E. (2003). Psychosocial influences on weight gain attitudes and behaviors during pregnancy. *J. Am. Diet. Assoc.* 103, 1314–1319. doi: 10.1016/S0002-8223(03)01070-8
- Etikan, I., Musa, S. A., and Alkassim, R. S. (2016). Comparison of convenience sampling and purposive sampling. *Am. J. Theor. Appl. Stat.* 5, 1–4. doi: 10.11648/j.ajtas.20160501.11
- Exline, J. J., Zell, A. L., Bratslavsky, E., Hamilton, M., and Swenson, A. (2012). People-Pleasing through eating: sociotropy predicts greater eating in response to perceived social pressure. *J. Soc. Clin. Psychol.* 31, 169–193. doi: 10.1521/jscp.2012.31.2.169
- Gardner, B., Croker, H., Barr, S., Briley, A., Poston, L., and Wardle, J. (2012). Psychological predictors of dietary intentions in pregnancy. *J. Hum. Nutr. Diet.* 25, 345–353. doi: 10.1111/j.1365-277X.2012.01239.x
- Gelfand, M. J., Raver, J. L., Nishii, L., Leslie, L. M., Lun, J., Lim, B. C., et al. (2011). Differences between tight and loose cultures: a 33-nation study. *Science* 332, 1100–1104. doi: 10.1126/science.1197754
- Goldstein, N. J., Cialdini, R. B., and Griskevicius, V. (2008). A room with a viewpoint: using social norms to motivate environmental conservation in hotels. *J. Consum. Res.* 35, 472–482. doi: 10.1086/586910
- Graham, J. E., Mayan, M., McCargar, L. J., and Bell, R. C. (2013). Making compromises: a qualitative study of sugar consumption behaviors during pregnancy. *J. Nutr. Educ. Behav.* 45, 578–585. doi: 10.1016/j.jneb.2013.03.011
- Guelinckx, I., Devlieger, R., Beckers, K., and Vansant, G. (2008). Maternal obesity: pregnancy complications, gestational weight gain and nutrition. *Obes. Rev.* 9, 140–150. doi: 10.1111/j.1467-789X.2007.00464.x
- Haug, S., Ulbricht, S., Hanke, M., Meyer, C., and John, U. (2011). Overestimation of drinking norms and its association with alcohol consumption in apprentices. *Alcohol Alcohol.* 46, 204–209. doi: 10.1093/alcalc/agg103
- Heatherton, T. F., and Polivy, J. (1991). Development and validation of a scale for measuring state self-esteem. *J. Pers. Soc. Psychol.* 60, 895–910. doi: 10.1037/0022-3514.60.6.895
- Heinrichs, N., Rapee, R. M., Alden, L. A., Bögels, S., Hofmann, S. G., Ja Oh, K., et al. (2006). Cultural differences in perceived social norms and social anxiety. *Behav. Res. Ther.* 44, 1187–1197. doi: 10.1016/j.brat.2005.09.006
- Hermans, R. C., Engels, R. C., Larsen, J. K., and Herman, C. P. (2009). Modeling of palatable food intake. The influence of quality of social interaction. *Appetite* 52, 801–804. doi: 10.1016/j.appet.2009.03.008
- Hickey, C. A., Cliver, S. P., Goldenberg, R. L., McNeal, S. F., and Hoffman, H. J. (1995). Relationship of psychosocial status to low prenatal weight gain among nonobese black and white women delivering at term. *Obstet. Gynecol.* 86, 177–183. doi: 10.1016/0029-7844(95)00161-J
- Higgs, S. (2015). Social norms and their influence on eating behaviours. *Appetite* 86, 38–44. doi: 10.1016/j.appet.2014.10.021
- Hirata, E., Lodder, G. M., Kühnen, U., Lippke, S., and Hermans, R. C. (2015). The possible antecedents and consequences of matching of food intake: examining the role of trait self-esteem and interpersonal closeness. *Front. Psychol.* 6:1920. doi: 10.3389/fpsyg.2015.01920
- Hutchinson, A., Charters, M., Prichard, I., Fletcher, C., and Wilson, C. (2017). Understanding maternal dietary choices during pregnancy: the role of social norms and mindful eating. *Appetite* 112, 227–234. doi: 10.1016/j.appet.2017.02.004
- Janakiraman, V. (2008). Listeriosis in pregnancy: diagnosis, treatment, and prevention. *Rev. Obstet. Gynecol.* 1, 179–185.
- Johnson, S. F., McCarter, R. J., and Ferenczy, C. (1987). Changes in alcohol, cigarette, and recreational drug use during pregnancy: implications for intervention. *Am. J. Epidemiol.* 126, 695–702. doi: 10.1093/oxfordjournals.aje.a114709
- Knowles, E. S., and Linn, J. A. (2004). *Resistance and Persuasion*. Mahwah, NJ: Lawrence Erlbaum Associates.
- LaBrie, J. W., Hummer, J. F., Neighbors, C., and Pedersen, E. R. (2008). Live interactive group-specific normative feedback reduces misperceptions and drinking in college students: a randomized cluster trial. *Psychol. Addict. Behav.* 22, 141–148. doi: 10.1037/0893-164X.22.1.141
- Lally, P., Bartle, N., and Wardle, J. (2011). Social norms and diet in adolescents. *Appetite* 57, 623–627. doi: 10.1016/j.appet.2011.07.015
- Leary, M. R. (1983). A brief version of the fear of negative evaluation scale. *Pers. Soc. Psychol. Bull.* 9, 371–375. doi: 10.1177/0146167283093007
- Leary, M. R., and Baumeister, R. F. (2000). “The nature and function of self-esteem: sociometer theory,” in *Advances in Experimental Social Psychology*, ed. M. P. Zanna (Cambridge, MA: Academic Press), 1–62.
- Lewis, M. A., Litt, D. M., Cronce, J. M., Blayney, J. A., and Gilmore, A. K. (2014). Underestimating protection and overestimating risk: examining descriptive normative perceptions and their association with drinking and sexual behaviors. *J. Sex Res.* 51, 86–96. doi: 10.1080/00224499.2012.710664
- Litt, D. M., Lewis, M. A., Linckenbach, J. W., Lande, G., and Neighbors, C. (2014). Normative misperceptions of peer seat belt use among high school students and their relationship to personal seat belt use. *Traffic Inj. Prev.* 15, 748–752. doi: 10.1080/15389588.2013.868892
- Loo, R., and Thorpe, K. (2000). Confirmatory factor analyses of the full and short versions of the Marlowe-crowne social desirability scale. *J. Soc. Psychol.* 140, 628–635. doi: 10.1080/00224540009600503
- Lumley, J., Chamberlain, C., Dowswell, T., Oliver, S., Oakley, L., and Watson, L. (2009). Interventions for promoting smoking cessation during pregnancy. *Cochrane Database Syst. Rev.* 8:CD001055. doi: 10.1002/14651858.CD001055.pub3
- Mollen, S., Rimal, R. N., Ruiters, R. A., and Kok, G. (2013). Healthy and unhealthy social norms and food selection. Findings from a field-experiment. *Appetite* 65, 83–89. doi: 10.1016/j.appet.2013.01.020
- Naska, A., Fouskakis, D., Oikonomou, E., Almeida, M. D., Berg, M. A., Gedrich, K., et al. (2005). Dietary patterns and their socio-demographic determinants in 10

- European countries: data from the DAFNE databank. *Eur. J. Clin. Nutr.* 60, 181–190. doi: 10.1038/sj.ejcn.1602284
- Nilsen, P. (2010). Brief alcohol intervention—where to from here? Challenges remain for research and practice. *Addiction* 105, 954–959. doi: 10.1111/j.1360-0443.2009.02779.x
- Northstone, K., Emmett, P. M., and Rogers, I. (2008). Dietary patterns in pregnancy and associations with nutrient intakes. *Br. J. Nutr.* 99, 406–415. doi: 10.1017/S0007114507803977
- Olafsdottir, A. S., Skuladottir, G. V., Thorsdottir, I., Hauksson, A., and Steingrimsdottir, L. (2005). Maternal diet in early and late pregnancy in relation to weight gain. *Int. J. Obes.* 30, 492–499. doi: 10.1038/sj.ijo.0803184
- Oyserman, D., Fryberg, S. A., and Yoder, N. (2007). Identity-based motivation and health. *J. Pers. Soc. Psychol.* 93, 1011–1027. doi: 10.1037/0022-3514.93.6.1011
- Pachucki, M. A., Jacques, P. F., and Christakis, N. A. (2011). Social network concordance in food choice among spouses, friends, and siblings. *Am. J. Public Health* 101, 2170–2177. doi: 10.2105/AJPH.2011.300282
- Paquet, C., Yudin, M. H., Yudin, M. H., Allen, V. M., Bouchard, C., Boucher, M., et al. (2013). Toxoplasmosis in pregnancy: prevention, screening, and treatment. *J. Obstet. Gynaecol. Can.* 35, 78–79. doi: 10.1016/S1701-2163(15)31053-7
- Ray, J. G., and Laskin, C. A. (1999). Folic acid and homocyst(e)ine metabolic defects and the risk of placental abruption, pre-eclampsia and spontaneous pregnancy loss: a systematic review. *Placenta* 20, 519–529. doi: 10.1053/plac.1999.0417
- Reynolds, W. M. (1982). Development of reliable and valid short forms of the marlowe-crowne social desirability scale. *J. Clin. Psychol.* 38, 119–125. doi: 10.1002/1097-4679(198201)38:1<119::AID-JCLP2270380118>3.0.CO;2-I
- Robinson, E. (2015). Perceived social norms and eating behaviour: an evaluation of studies and future directions. *Physiol. Behav.* 152(Pt B), 397–401. doi: 10.1016/j.physbeh.2015.06.010
- Robinson, E., Fleming, A., and Higgs, S. (2014a). Prompting healthier eating: testing the use of health and social norm based messages. *Health Psychol.* 33, 1057–1064. doi: 10.1037/a0034213
- Robinson, E., Harris, E., Thomas, J., Aveyard, P., and Higgs, S. (2013). Reducing high calorie snack food in young adults: a role for social norms and health based messages. *Int. J. Behav. Nutr. Phys. Act.* 10, 1–8. doi: 10.1186/1479-5868-10-73
- Robinson, E., Thomas, J., Aveyard, P., and Higgs, S. (2014b). What everyone else is eating: a systematic review and meta-analysis of the effect of informational eating norms on eating behavior. *J. Acad. Nutr. Diet.* 114, 414–429. doi: 10.1016/j.jand.2013.11.009
- Robinson, E., Tobias, T., Shaw, L., Freeman, E., and Higgs, S. (2011). Social matching of food intake and the need for social acceptance. *Appetite* 56, 747–752. doi: 10.1016/j.appet.2011.03.001
- Rofé, Y., Littner, M. B., and Lewin, I. (1993). Emotional experiences during the three trimesters of pregnancy. *J. Clin. Psychol.* 49, 3–12. doi: 10.1002/1097-4679(199301)49:1<3::AID-JCLP2270490102>3.0.CO;2-A
- Root, R., and Browner, C. (2001). Practices of the pregnant self: compliance with and resistance to prenatal norms. *Cult. Med. Psychiatry* 25, 195–223. doi: 10.1023/A:1010665726205
- Sallis, J. F., Grossman, R. M., Pinski, R. B., Patterson, T. L., and Nader, P. R. (1987). The development of scales to measure social support for diet and exercise behaviors. *Prev. Med.* 16, 825–836. doi: 10.1016/0091-7435(87)90022-3
- Salvy, S. J., de la Haye, K., Bowker, J. C., and Hermans, R. C. (2012). Influence of peers and friends on children's and adolescents' eating and activity behaviors. *Physiol. Behav.* 106, 369–378. doi: 10.1016/j.physbeh.2012.03.022
- Smit, C. R., de Leeuw, R. N. H., Bevelander, K. E., Burk, W. J., and Buijzen, M. (2016). A social network-based intervention stimulating peer influence on children's self-reported water consumption: a randomized control trial. *Appetite* 103(Suppl. C), 294–301. doi: 10.1016/j.appet.2016.04.011
- Springer, N. S., Bogue, E.-L., Arnold, M., Yankou, D., and Oakley, D. (1994). Nutrition locus of control and dietary behaviors of pregnant women. *Appl. Nurs. Res.* 7, 28–31. doi: 10.1016/0897-1897(94)90017-5
- Stade, B. C., Bailey, C., Dzendoletas, D., Sgro, M., Dowswell, T., and Bennett, D. (2009). *Psychological and/or Educational Interventions for Reducing Alcohol Consumption in Pregnant Women and Women Planning Pregnancy*. London: The Cochrane Library.
- Stok, F. M., de Ridder, D. T., de Vet, E., and de Wit, J. B. (2012). Minority talks: the influence of descriptive social norms on fruit intake. *Psychol. Health* 27, 956–970. doi: 10.1080/08870446.2011.635303
- Stok, F. M., de Ridder, D. T., de Vet, E., and de Wit, J. B. (2014). Don't tell me what I should do, but what others do: the influence of descriptive and injunctive peer norms on fruit consumption in adolescents. *Br. J. Health Psychol.* 19, 52–64. doi: 10.1111/bjhp.12030
- Strychar, I. M., Chabot, C., Champagne, F., and Ghadirian, P. (2000). Psychosocial and lifestyle factors associated with insufficient and excessive maternal weight gain during pregnancy. *J. Acad. Nutr. Diet.* 100, 353–356. doi: 10.1016/S0002-8223(00)00107-3
- Stunkard, A. J., Sorensen, T., and Schulzinger, F. (1983). Use of the danish adoption register for the study of obesity and thinness. *Assoc. Res. Nerv. Ment. Dis.* 60, 115–120.
- Thangaratinam, S., and Jolly, K. (2010). Obesity in pregnancy: a review of reviews on the effectiveness of interventions. *BJOG* 117, 1309–1312. doi: 10.1111/j.1471-0528.2010.02670.x
- Tiedje, L. B., Kingry, M. J., and Stommel, M. (1992). Patient attitudes concerning health behaviors during pregnancy: initial development of a questionnaire. *Health Educ. Q.* 19, 481–493. doi: 10.1177/109019819201900411
- van der Linden, S. (2015). Exploring beliefs about bottled water and intentions to reduce consumption. *Environ. Behav.* 47, 526–550. doi: 10.1177/001391651513515239
- Verkooijen, K. T., Stok, F. M., and Mollen, S. (2015). The power of regression to the mean: a social norm study revisited. *Eur. J. Soc. Psychol.* 45, 417–425. doi: 10.1002/ejsp.2111
- Willers, S. M., Devereux, G., Craig, L. C. A., McNeill, G., Wijga, A. H., Abou El-Magd, W., et al. (2007). Maternal food consumption during pregnancy and asthma, respiratory and atopic symptoms in 5-year-old children. *Thorax* 62, 773–779. doi: 10.1136/thx.2006.074187
- Wolff, C. B., and Wolff, H. K. (1995). Maternal eating patterns and birth weight of Mexican American infants. *Nutr. Health* 10, 121–134. doi: 10.1177/026010609501000203
- World Medical Association (2013). World medical association declaration of Helsinki: ethical principles for medical research involving human subjects. *JAMA* 310, 2191–2194. doi: 10.1001/jama.2013.281053

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2018 Bevelander, Herte, Kakoulakis, Sanguino, Tebbe and Tünte. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Parental Decision-Making on Childhood Vaccination

Kaja Damjanović^{1*}, Johanna Graeber², Sandra Ilić¹, Wing Y. Lam³, Žan Lep^{4*}, Sara Morales⁵, Tero Pulkkinen⁶ and Loes Vingerhoets⁷

¹ Laboratory for Experimental Psychology, Department of Psychology, Faculty of Philosophy, University of Belgrade, Belgrade, Serbia, ² Department of Psychology, Faculty of Philosophy, Christian-Albrechts-University Kiel, Kiel, Germany, ³ Faculty of Social Sciences, School of Psychology, University of Kent, Canterbury, United Kingdom, ⁴ Department of Psychology, Faculty of Arts, University of Ljubljana, Ljubljana, Slovenia, ⁵ Faculty of Psychology, University of Basque Country, Bilbao, Spain, ⁶ Department of Psychology and Logopedics, Faculty of Medicine, University of Helsinki, Helsinki, Finland, ⁷ Department of Psychology, Faculty of Psychology and Neuroscience, University of Maastricht, Maastricht, Netherlands

OPEN ACCESS

Edited by:

Pietro Cipresso,
Istituto Auxologico Italiano (IRCCS),
Italy

Reviewed by:

Nicola Luigi Bragazzi,
Università di Genova, Italy
Gabriel José Corrêa Mograbi,
Universidade Federal de Mato
Grosso, Brazil

*Correspondence:

Kaja Damjanović
kdamnjan@f.bg.ac.rs
Žan Lep
zan.lep@empirik.si

Specialty section:

This article was submitted to
Quantitative Psychology and
Measurement,
a section of the journal
Frontiers in Psychology

Received: 31 October 2017

Accepted: 26 April 2018

Published: 13 June 2018

Citation:

Damjanović K, Graeber J, Ilić S,
Lam WY, Lep Ž, Morales S,
Pulkkinen T and Vingerhoets L (2018)
Parental Decision-Making on
Childhood Vaccination.
Front. Psychol. 9:735.
doi: 10.3389/fpsyg.2018.00735

A growing number of parents delay vaccinations or are deciding not to vaccinate their children altogether. This increases the risk of contracting vaccine-preventable diseases and disrupting herd immunity, and also impairs the trust in the capacities of health care systems to protect people. Vaccine hesitancy is related to a range of both psychological and demographic determinants, such as attitudes toward vaccinations, social norms, and trust in science. Our aim is to understand those determinants in parents, because they are a special group in this issue—they act as proxy decision makers for their children, who are unable to decide for themselves. The fact that deciding to vaccinate is a socially forced choice that concerns a child's health makes vaccine-related decisions highly important and involving for parents. This high involvement might lead to parents overemphasizing the potential side effects that they know to be vaccine-related, and by amplifying those, parents are more focused on the potential outcomes of vaccine-related decisions, which can yield specific pattern of the outcome bias. We propose two related studies to investigate factors which promote vaccine hesitancy, protective factors that determine parental vaccination decisions, and outcome bias in parental vaccination intentions. We will explore demographic and psychological factors, and test parental involvement related to vaccine hesitancy using an online battery in a correlation panel design study. The second study is an experimental study, in which we will investigate the moderating role of parents' high involvement in the specific domain of vaccination decision making. We expect that higher involvement among parents, compared to non-parents, will shape the pattern of the proneness to outcome bias. The studies will be conducted across eight countries in Europe and Asia (Finland, Germany, Hong Kong, the Netherlands, Serbia, Slovenia, Spain, and the United Kingdom), rendering findings that will aid with understanding the underlying mechanisms of vaccine hesitancy and paving the way for developing interventions custom-made for parents.

Keywords: vaccine, involvement, vaccine hesitancy, immunization, health decisions, decision-making, parents, outcome bias

INTRODUCTION

One of the greatest public health challenges today concerns suboptimal vaccine uptake rates. In 2017, measles affected 21,315 people and caused 35 deaths, according to WHO's press release from 19 February 2018. "The surge in measles cases in 2017 included large outbreaks in 15 of the 53 countries in the (European) region. The highest numbers of affected people were reported in Romania (5,562), Italy (5,006), and Ukraine (4,767)" (World Health Organization, 2018). Greece, Germany, Serbia, the United Kingdom, Spain, Bulgaria, and France also experienced large outbreaks (World Health Organization, 2018). This is a result of suboptimal vaccine rates (World Health Organization, 2017a): in many areas, the coverage rates of common vaccines have decreased below 95% that is postulated as the minimum to herd immunity, the effective halting of the spread of measles and other vaccine-preventable diseases.

While some children cannot be vaccinated for medical reasons and in some areas vaccines are not readily available, a growing number of children are not vaccinated or are vaccinated late largely due to their parents' conscious decision (Pearce et al., 2008). The resistance to be vaccinated or to delay vaccinations despite having available vaccinating services, has been dubbed vaccine hesitancy (Luthy et al., 2009; Gowda and Dempsey, 2013; World Health Organization, 2014). Numerous interventions have been introduced to combat vaccine hesitancy, but many are lacking in success (Sadaf et al., 2013; Dubé et al., 2014; Pluviano et al., 2017). To better combat vaccine hesitancy and optimize interventions, factors associated with parents' decisions on vaccination need to be identified and investigated.

As such, vaccine hesitancy is a multi-layered phenomenon, related, amongst others, to various factors of social and psychological kind. Several studies examining various populations and different vaccines have found that vaccine hesitancy is related, amongst other, to prior beliefs about vaccinations (Smailbegovic et al., 2003; Dubé et al., 2014), perceived benefits of vaccines (Myers and Goodwin, 2011), attitudes toward vaccines (Pareek and Pattison, 2000; Mohd Azizi et al., 2017), whether the child has been previously vaccinated (Pareek and Pattison, 2000), previous experiences with vaccinations (Boes et al., 2017), socioeconomic status (Smith et al., 2004), number of children (Gust et al., 2005), and marital status (Smith et al., 2004). Despite the wide range of findings, the results of several systematic reviews suggest that there are still factors to be identified and further explored (Mills et al., 2005; Larson et al., 2014; Cobos et al., 2015).

The studies about vaccine hesitancy have often focused on parents, who are the key propagators of vaccine hesitancy and consumers of anti-vaccine influences, while the children are the key victims. For parents, vaccinating their children could mean that parents have to witness their child's discomfort and have to face potential potential side effects. At the same time, not vaccinating may lead to contracting vaccine-preventable diseases, potential prosecution in certain countries, enrolment refusal in some schools, disrupting herd immunity, etc. Parents may also face social pressure, such as pressure from health-care professionals (Evans et al., 2001), or other kinds of

social pressure, e.g., to be experts on vaccination mechanism and therefore reliable and informed decision makers. In the contemporary context, parents are prompted to take an active role in their children's healthcare (Pyke-Grimm et al., 1999), which places heavy burden on the parent (Wagenaar et al., 1988).

This is especially true in the realm of intensive parenting; one of the most dominant parenting styles today (Arendell, 2000; Smyth and Craig, 2017). The term was coined by Hays (1996) to describe parenting style closely linked to the pressure felt by parents, mostly women, because of their responsibility for all childcare related tasks, children's outcomes (intellectual, social, emotional, and health-related), and their need to protect the child from any harm or disease. These needs, although very common for majority of parents, in highly individualistic societies outturn as the less communal worldviews, and research shows that intensive (salutogenic) parenting was an important rationale for refusing vaccines, as salutogenic parents have higher sense of advocacy and feel more capable of taking care for the children without expert intervention or vaccines (Reich, 2014; Ward et al., 2017).

On discussions about vaccinating their children, parents emphasize the purpose and safety of vaccination rather than the procedure itself (Salmon et al., 2005; Miton and Mercier, 2015). This parental decision is often accompanied by limited knowledge (Downs et al., 2008; Zingg and Siegrist, 2012), threatening campaigns (Ruiter et al., 2014; Stronach, 2015), societal norms (de Visser et al., 2011; Oraby et al., 2014), and official consent (Leask et al., 2011). Vaccine-hesitant parents thus differ from non-parents in their perception regarding the dangers of vaccines, risk of side effects, and protective benefits. Similarly, the perceived danger of vaccines is associated with the reluctance to vaccinate (Wilson et al., 2008), and it has been suggested that this can play an important role in parents' actual decision on mandatory childhood vaccination (Sporton and Francis, 2001).

This high-stake parental position regarding vaccinations is further complicated by the characteristics of the decision itself. Decisions, among themselves, differ depending on whom we are deciding for: ourselves or someone else (Zikmund-Fisher et al., 2006). People also use different strategies when deciding about other people compared to deciding about inanimate objects (Goldstein and Weber, 1995). Additionally, the importance of a decision differs according to its domain (see Meta-Decision-Making Model; Payne et al., 1993). Especially in health-related decisions, the importance of the decision skews decision-making processes and related phenomena, such as proneness to risky decisions (Wang, 1996a; Fagley and Miller, 1997; Kühberger, 1998; Hanoch et al., 2006; Markiewicz and Weber, 2013; Gummerum et al., 2014; Zimmerman et al., 2014; Damjanović and Gvozdenović, 2016), susceptibility to cognitive biases (McNeil et al., 1982; Wang, 1996b; Tanner et al., 2008) and effort of strategies (Edwards et al., 2001; Almashat et al., 2008).

Health-related decisions also differ according to the extent of their importance (Thompson, 2007), which affects the level of involvement decision makers put into a decision (Solomon et al., 2006). Different decisions range on a continuum from

fairly routine to those that require extensive thought and have a high level of involvement (Solomon et al., 2006). The level of involvement in the same decision can differ between people (Arora and McHorney, 2000). However, some decisions (e.g., health-related decisions) are generally assumed to be important for the great majority of people (Solomon et al., 2006). Parental decision about a child’s health is a special and extreme case of health-related decision (Zikmund-Fisher et al., 2006). It is also highly involving in terms of affect and expectation (Wroe et al., 2004). There is evidence suggesting differences, not only between parents and non-parents (Donovan and Jalleh, 2000), but also parents with children of different ages (Henrikson et al., 2017) regarding intentions to vaccinate and seek information on vaccination. Decomposing involvement and its influence on decision-making processes can help with undermining vaccine hesitancy through minimizing cognitive obstacles to reasoning, which stem from high involvement.

Parental decision on child vaccination is a specific case of health-related decision (Zikmund-Fisher et al., 2006) that is highly involving in terms of affect and expectation (Wroe et al., 2004). When discussing vaccination and immunization, the emphasis is on its purposefulness, potential side effects, and efficacy of vaccination (Salmon et al., 2005; Miton and Mercier, 2015). When normative, but also descriptive theory of decision-making is applied, decisions on vaccination can be analyzed using the decision matrix in which states of nature and alternative decisions are crossed to make cells with different outcomes (see **Table 1**). It can be stipulated that, when deciding on vaccination, people tend to place major weights on the outcomes, that is on the subjective perception of the outcomes. In other words, as well as some other health-related decisions, decisions on vaccination have an inherent feature of a stronger focus toward the outcome (Gellin et al., 2000; Freed et al., 2004, 2010; Gust et al., 2004; Miton and Mercier, 2015). To parents, the important issue while deciding is the outcome of this decision (e.g., well-being and health of their child) (Goldenberg, 2016). Evaluating prior decisions based on their outcomes is a tendency labeled as the outcome bias (Baron and Hershey, 1988). For instance, parents might overemphasize the immediate vaccine side effects, such as rashes or swelling, and use these side effects as justification to avoid vaccinating their child (Callender, 2016). In line with this, parents might judge the quality of the potential decision to vaccinate their child based on the consequences of this decision met previously by them or by the sources they are in contact with. Therefore, this decision is specific due to its explicit orientation toward the outcome (Gellin et al., 2000; Freed et al., 2004, 2010; Gust et al., 2004; Salmon et al., 2005).

Understanding how both psychological and social factors relate to vaccine hesitancy is important for developing effective interventions. Therefore, the aim of the proposed research is to detect factors associated with and affecting the decisions of parents’ regarding vaccination. To do that, we will conduct two separate but related studies. In study 1, socio-demographic and psychological variables will be tested for their connection with differences among parents when it comes to making

vaccine-related decisions for their children. In study 2, the role of involvement in decisions regarding childhood vaccinations will be explored in more details. Specifically, we will study whether involvement will moderate the susceptibility to outcome bias with an experimental design.

STUDY 1 – CORRELATES OF INTENTION TO VACCINATE

Introduction

In study 1, we aim to explore the demographic and psychological factors that influence parents’ vaccine hesitancy. As previously stated, vaccine hesitancy is related to a large range of attitudes, most notably to lower rates of compliance, which lead to drops in vaccination rates (Bloom et al., 2014). Our choice of correlates is in line with the framework of vaccine decision factors proposed by Gowda and Dempsey (2013), and it is important to acknowledge their interrelatedness. Since vaccination intention and hesitancy are multi-layered phenomena, chosen measures are narrowed to broadly cover the three following aspects: parent-specific factors (demographics, knowledge etc.), vaccine-specific factors (perceived vaccine safety and efficacy etc.), and external factors (values, norms, policies, requirements etc.).

Trust Toward Authorities

In the abundance of both affirmative and diminishing information on vaccination, the full picture is seldom easily available and individuals have a hard time forming their own opinions on the topic. Thus, argumentation must rely on evidence, which is accepted largely based on trust (Miton and Mercier, 2015). Trust in relevant actors (such as health professionals, pharmaceutical companies, law makers etc.) in the debate as well as general trust in science play an important role in vaccination decisions (Bedford, 2014; Jolley and Douglas, 2014; Camargo and Grant, 2015). However, this can be challenging as research repeatedly shows that some actors and science as a whole receive a low level of public trust (Lewandowsky and Oberauer, 2016). This can be caused due to high informational pluralism, rendering their argumentation on the topic irrelevant to the public.

Feelings of mistrust could also be part of a general feeling of unease about the complexity of modern society that forces us to rely on others to manage some parts of our lives (Hobson-West, 2007). According to Collins (2009), a general mistrust in science and scientists has enabled a paralyzing form of skepticism and

TABLE 1 | Decision matrix: vaccination case.

		STATES OF NATURE		
		Vaccines are efficient and risk-free	Vaccines are efficient and low-risk	Vaccines are extremely dangerous
Decisions	To vaccinate	Status quo	Status quo	High risk
	Not to vaccinate	High risk	High risk	Status quo

scientific populism that denies the role of science and prompts anti-vaccination decisions.

Parents who have a positive view of the government are more likely to support vaccine policies, and perceive them as beneficial rather than restrictive of their personal freedom (Miton and Mercier, 2015; Highland, 2016). We expect parents who find official pro-vaccination authorities trustworthy to be less vaccine hesitant, show a higher intention to vaccinate, and a higher experience of freedom in the decision.

Previous studies have found that people with higher level of distrust toward authorities are more reluctant to rely on official sources of information (Freed et al., 2011). For that reason, we expect the relationship between level of trust and intention to vaccinate to be moderated by the type of sources consulted to make the decision. We expect level of trust toward authorities to predict the proportion of official sources used for the decision.

Finally, some studies have found that belief in conspiracy theories can predict distrust toward authorities (Darwin et al., 2011; Swami et al., 2011). We expect to replicate this result in the case of vaccine related conspiracy theories.

Perceived Consensus and Norms

Many vaccination decisions are influenced by parents' perception of others when making vaccination decisions (Gust et al., 2004; Leask and MacArtney, 2008; Gowda et al., 2012; Gowda and Dempsey, 2013). People in general tend to rely on consensus cues, because consensus, especially combined judgment of multiple experts, typically implies correctness (Van der Linden and Lewandowsky, 2015; Tom, 2017). However, there is a gap between the low level of scientific consensus perceived by the lay-public, and the actual level of consensus regarding the immunization and vaccination. The Gateway Belief Model proposed by Van der Linden et al. (2015) suggests that reducing the difference between people's subjective perception and the actual level of normative agreement among influential referents can lead to small yet important changes in key personal beliefs. Moreover, perceived scientific consensus has been identified as a key determinant in the public's opinion on, in some aspects equivalent, disputable topics (van der Linden et al., 2017). Due to high involvement aspects of vaccine-related decisions, we assume perceived scientific consensus plays a specific role, as it was the strongest predictor of the acceptance of the scientific arguments in other similar social issues (Lewandowsky et al., 2012). We expect parents who perceive stronger scientific consensus on the topic of vaccination to show less vaccine hesitancy and be more likely to vaccinate their children. We will thus test the correlation of perceived consensus, and the perception of risk with the decision on mandatory childhood vaccination (Wilson et al., 2008; Rolfe-Redding et al., 2012). We expect perceived consensus will correlate more strongly with the intention to vaccinate than perceived vaccination risks. Confidence in vaccines and vaccine-related decisions are also influenced by the individual's perception of societal norms and collective values, as well as their metacognitive perceptions about other groups' (e.g., health professionals) beliefs (Kennedy et al., 2011; Gowda and Dempsey, 2013; van der Linden et al., 2017). Numerous findings suggest normative information is

rated as more trustworthy, less resilient to dismissal, and more influential than anecdotal cases (Carrico et al., 2011; Kahan et al., 2011; Lewandowsky et al., 2012; Rolfe-Redding et al., 2012), hence making it more likely to influence decision-making processes.

In situations where social norms are ambiguous, appealing to consensus and unity of norms tends to be more effective in persuading parents to vaccinate (Lewandowsky et al., 2012). Based on findings by Kahan et al. (2009) and Kahan (2010), we predict that the correlation between parents' intention to vaccinate (what we call adherence to the norm) and norms will be moderated by the perception of social consensus of given norms.

Freedom of Choice, Choice Overload, and Values

Choices are usually considered on a continuum from totally uninfluenced to the ones molded by formal and informal social norms. While norms play an important role in vaccination decisions (Gust et al., 2008; Brunson, 2013a,b), how we perceive those norms and how we perceive freedom when making the choice also influence the decision. This subjective perceived freedom is associated with choice overload in decision making processes (Lau et al., 2015). Decisions and decision-making processes can be exhausting and overwhelming, hence decision makers can find it difficult to retain all the necessary information needed to make an informed decision. We expect parents who experience lower levels of perceived freedom to have a higher tendency to adhere to perceived social norms. We also expect those individuals to be more likely to conform to authorities and to other stakeholders (health professionals, government etc.).

According to the cultural cognition of risk (Kahan and Braman, 2006; Kahan et al., 2010), the evaluation of riskiness is in line with values that we share as a culture. The operationalization of values regarding vaccination is a challenge, therefore we decided to use an indirect measure instead. We will measure participants' actively open-minded thinking style, a construct which was found to predict the tendency to acquire information in order to make competent decisions (Haran et al., 2013; Baron et al., 2015). We expect individuals who are more open-minded to be more likely to seek information from both sides of vaccine hesitancy spectrum (more diverse sources), be less affected by social norms, and show less vaccine hesitancy. The diversity in the sources of information will be rated by the proportion of official and informal sources.

Perception of Danger

Threat perception has been widely used to encourage health-related actions such as vaccinations, but messages that increase risk perceptions are less effective than those increasing perceived effectiveness (Ruiter et al., 2001, 2014). However, parents are more likely to vaccinate or intend to vaccinate their children if they perceive the danger of not vaccinating (e.g., perceived vulnerability of their child contracting a certain disease) as high (Seeman et al., 2010; Healy and Pickering, 2011; Jolley and Douglas, 2014; Highland, 2016).

Health-related decisions often comprise of high levels of uncertainty and varying degrees of potential risk. Given that

health professionals and other experts differ vastly in coping with uncertainty and risk taking (Grol et al., 1990), it is not surprising that parents are also under great strain when deciding. As stated before, parents often perceive vaccinating their children more risky than not vaccinating them, and as willingness to take risks is associated with making obligatory medical decisions (Grol et al., 1990), we expect to find the same connection in parents, with those less willing to take risks to be more vaccine hesitant.

Access to Information

Perception of threat is mediated by access to information. Betsch et al. (2010) found that access to anti-vaccine sources of information increases perceived risks of vaccination. Furthermore, based on their risk perception, different parents trust different kinds of vaccine-related messages. Risk-oriented parents tend to favor statistical over anecdotal arguments, but those who are health-oriented tend to prefer the latter (Downs et al., 2008).

Moreover, knowledge has been identified as an important factor in shaping parents' decisions (Zingg and Siegrist, 2012). A higher number of sources of information has been related to a higher perceived level of knowledge in the frame of decision-making about vaccination (Downs et al., 2008; Rachiotis et al., 2010; Healy and Pickering, 2011; Brunson, 2013a), but the sources of their information can sometimes be problematic. Many parents reported seeking additional information, with most preferring to use the internet rather than consulting a doctor, and would use a general search engine instead of an official or medical website (Downs et al., 2008). Some parents were even found to reserve the decision to vaccinate until enough information was available to them (Highland, 2016). We expect participants who think they have enough access to information to be more likely to score on the extreme ends of the vaccine-hesitant spectrum (i.e., either very pro- or anti-vaccine), and those who are more exposed to anecdotal cases to be more vaccine hesitant. Additionally, in line with previous studies (Rachiotis et al., 2010; Jones et al., 2012; Brunson, 2013b) we expect the type of sources consulted (Official vs. Informal) to predict intention of vaccinate. Anecdotal cases (especially personal experiences) are one of the key forms of communication on the topic of vaccination, particularly among vaccine hesitant groups. This was supported by theoretical models such as the Cultural Attraction theory (Miton and Mercier, 2015), or the Fuzzy-Trace theory (Reyna and Brainerd, 1992; Reyna, 2008). Even vaccine concerns endorsed by a small but vocal group of individuals can heighten vaccine hesitancy in the community (Gowda and Dempsey, 2013). With vaccination being counter-intuitive in its nature (injecting antigens in an already healthy organism to remain healthy), anecdotal cases tap into our intuitive cognitive mechanisms, making individuals less likely to vaccinate. We expect parents that have been exposed to anecdotal cases of bad reactions to vaccines to show higher levels of vaccine hesitancy. However, as mentioned before, Risk vs. Health-orientation can moderate the effect of exposure to anecdotal cases (Downs et al., 2008). For that reason, we expect parents with a negative outcome focus to

be more likely to be affected by the exposure to anecdotal cases.

Methods

Design, Sample, and Procedure

Participants will be parents or primary caregivers of children that are of the recommended age to receive vaccinations from their corresponding countries across Europe and Asia (i.e., Finland, Germany, Hong Kong, the Netherlands, Serbia, Slovenia, and Spain). Participants must be (a) over 18 years old, and (b) a parent or care taker of at least one child under the age of 12 years (Salmon et al., 2005; Jones et al., 2012).

The sample size will be a minimum of 222 participants per country, based on a power analysis for an effect size of 0.15, and 20 predictors in a linear multiple regression model. The final sample size will not have a difference greater than 10% between each country. To make sure that there aren't any significant differences between the sample sizes of the different countries, if the sample of any country exceeds 10% of the mean, some participants will be randomly discarded until the criterion is met.

The questionnaire is programmed in JavaScript and will be administered online. To take part in the study, participants will need an electronic device with access to internet. The link with an unbiased invitation letter will be posted on different social media platforms, forums and websites, targeting a wide range of parents from the entire vaccine-hesitancy spectrum. Once the link is opened, participants will first be given a brief introduction of the study. Participants will then read and sign an informed consent, which states they are able to halt and withdraw from the study at any moment without providing a reason, and that they agree for their anonymized data to be analyzed for future publications. Participants will be asked to compose a unique identification code (consisting of their parents' initials, and month of birth), which will be used to identify their data if they decide to withdraw their data after completing the study. After that, participants will be asked to provide their demographic information and complete a battery of vaccination- and decision-making-related questions, with a total of 115 items (the complete battery is in Appendix 1). Participants will be able to leave items unanswered. The study will take ~25 min to complete, but will vary depending on participants' speed of responding. This procedure was ethically approved by the Institutional Review Board of the Faculty of Philosophy at the University of Belgrade.

Materials and Measures

Choice of measures

Materials from online databases (PsycTESTS, DMIDI) were selected based on psychometric quality, fitness for parental context, translation feasibility, and the significance of usage in a variety of international institutions. Due to the complexity of our proposed model, and to avoid dropout due to the length of the final battery, we have adapted certain instruments and developed questions to assess specific constructs.

Additionally, data about each country will be gathered from different national, international and scientific entities, such as the percentage of vaccinated children in a country (World Health Organization, 2017b), number of physicians per 1000

inhabitants (World Health Organization, 2017a), health system quality (GBD 2015 Healthcare Access and Quality Collaborators, 2017) and about the vaccination programme in each of the countries (obtained from the corresponding governmental communications).

For the measurement of different constructs to be homogenous, all continuous variables will be measured with 7-point Likert-type scales. This is because most of the instruments used in the study originally included this type of scale, and the number of points in Likert-type scales do not influence its metric properties (Contractor and Fox, 2011).

Sociodemographic list

The sociodemographic data to be gathered includes participants' country of residence, age, gender, education (based on European Qualifications framework), marital status, number of children, age of children, the adherence to the official vaccination schedule of the child in the corresponding country, and if the participant themselves adhered to said vaccination schedule (Appendix 1). Subjective Socioeconomic Status (sSES) will rely on the level of education, and on participant's self-report regarding the "difficulty of the household to make ends meet" (Eurostat, 2017).

Intention to vaccinate

We will measure parents' willingness to vaccinate their children by their reported intention to vaccinate when presented with the option. To measure this, a single item will be used: "Would you at this time vaccinate your child according to the official vaccination schedule?" A version of this item, with an additional description: "Regardless if you are a parent or not," was previously used by Stojkovic et al. (2017), who adapted it from the scales by Horne et al. (2015) and Opel et al. (2011). Parents will answer this item on a Likert-type scale ranging from "definitely yes" to "definitely not". Participants' response to this question will be the dependent variable for this study.

Vaccination scales

To measure the perceived risk of (not) vaccinating may cause in children and society, we will use the vaccination scale developed by Horne et al. (2015). It is a 5-item scale that measures people's general attitude toward the vaccination (i.e., "Vaccinating healthy children helps to protect others by stopping the spread of diseases," "I plan to vaccinate my children"). This test has proper psychometric features, including a high internal consistency ($\alpha = 0.84$), and good predictive validity for past and future vaccination behavior, it has not been used in Europe so far.

Furthermore, the Vaccine Conspiracy Belief scale (Shapiro et al., 2016) will be added at this part. This scale contains 10 items, which examine the belief in the conspiracy theory that different entities try to hide the risk of vaccines. This shows the deception rather than the general attitude people have (Shapiro et al., 2016). An example of this item is: "The government is trying to cover up the link between vaccines and autism." This scale has a very high level of internal consistency ($\alpha = 0.94$) and is a good predictor of the willingness of parents to vaccinate their children.

Vaccine hesitancy

To identify vaccine hesitant parents, we will use Opel et al.'s (2011) revised version of Parent Attitudes about Childhood Vaccines (PACV). The measure consists of 15 items, which are divided into three sub-domain scales: Safety and efficacy ($\alpha = 0.74$), General attitudes ($\alpha = 0.84$) and Behavior ($\alpha = 0.74$). The measure has high internal and construct validity, and shows a statistically significant linear association between parents' total score on the 15-item PACV and their child's vaccination status (Opel et al., 2011). To increase the consistency of type of response along the scale and increase the sensitivity of the measure, we transformed some of the multiple-choice items to 7-point Likert type scales, in which each of the extremes represent the options that were included in the original scales. Finally, the item number 15 that referred to parents' trust toward their child's doctor, have been moved to the Trust Toward Authorities Scale (see Appendix 1 for detailed description of the battery).

Perceived freedom

The Experience of Freedom Measure (Lau et al., 2015) will be used to measure parents' perceived freedom when making vaccine-related decisions. Participants will rate 4-items (i.e., "I was able to choose what I wanted") on a Likert-type scale. The measure has good internal consistency ($\alpha = 0.82$).

Choice overload

Lau et al.'s (2015) Choice Overload scale will be used to measure the choice overload within the context of decision-making. Participants will be asked rate the extent to which they agree with 3 statements (i.e., "I felt overwhelmed by the decision") on a Likert-type scale. This instrument shows good psychometric properties ($\alpha = 0.73$) relative to the decision on whether to vaccinate their children.

Actively open-minded thinking

The Actively Open-Minded Thinking Beliefs (AOT) scale will be used to measure participants' beliefs on whether actively open-minded thinking is a desirable personal feature. The scale was originally developed by Stanovich and West (2007) and revised by Haran et al. (2013). We will use the revised version due to its shorter length (7 items) and adequacy for the general, adult population. Participants will rate how much they agree with given statements on a Likert type scale (i.e., "People should take into consideration evidence that goes against their beliefs"). The scale has been found to correlate with various measures of reflective thinking and good performance.

Trust toward authorities scale and sources of information

To measure the perceived credibility and the trust people have toward authorities, Jolley and Douglas' (2014) Trust toward Authorities scale will be used. The scale is built up out of items from previous scales. Participants will rate to what extent they trust corporations, national government, healthcare system, scientists, mainstream media, alternative media, social networks and their child's doctor on a Likert-type scale from strongly mistrust to strongly trust. They will also check all the sources they have used when making a decision about vaccination.

Availability of the relevant information

To assess whether parents believe they have enough information to make a solid decision regarding vaccinating their children, we will use one item from the General Health Styles survey by Gust et al. (2005): “I have access to all the information I need to make good decisions about immunization of my children.” Parents will answer on a Likert scale measuring the level of agreement.

Exposure to anecdotal cases

This variable will be measured through a single item adapted from the PACV scale by Opel et al. (2011) in a yes/no question format (i.e., “Have you ever heard of anyone who had a bad reaction to a shot?”).

Involvement in the vaccination decision

To measure personal involvement on the decision to vaccinate their children, we will include an item asking participants to rate their level of involvement. Additionally, we will ask if any other person is involved in the decision, and how many of them have been. For each of the additional people involved, participants will also be asked to indicate who that person is (the other parent of the child, another family member, a friend or any other person), indicate the gender, and rate the level of involvement of said person.

Perceived consensus, norms, and knowledge about vaccination

Additional Likert-type scale items will be used to assess participants’ perceived scientific and social consensus about vaccination (i.e., “Is there a consensus among scientist about the safety of the vaccines?” “Is the vaccination an issue in your country?”), norms (i.e., “What do you think is the percentage of vaccinated children in your country?”), and knowledge (e.g., “Is vaccination mandatory in your country?”). The items are based on the items used by Van der Linden (2011) in research dealing with topics of perceived consensus and norms.

Passive risk-taking

We will measure participants’ tendency of passive risk-taking using the Passive Risk-Taking Scale (Keinan and Bereby-Meyer, 2012). While risk mostly occurs during action, passive risk-taking can influence potential losses due to inaction. This test contains three subscales regarding risks that involve resources, medical issues and ethical issues. It has 25 items in total and uses a Likert-type rating scale in which participants will rate how likely they will act according to the statements (i.e., “Get vaccinated for the flu in the winter”). This scale has a high internal validity and reliability ($\alpha = 0.82$).

Elaboration of potential outcomes

The Elaboration on Potential Outcomes (EPO, Nenkov et al., 2008) measure will be used to assess participants’ tendencies to generate and evaluate possible positive and negative consequences of their behavior, and measure their attitudes toward risk-taking. The instrument consists of 13 items, which are divided into three subscales with high internal consistency: generation/evaluation (e.g., “I try to anticipate as many consequences of my actions as I can”; $\alpha = 0.88$), positive

outcome focus (e.g., “I keep a positive attitude that things always turn out all right.”; $\alpha = 0.87$), and negative outcome focus (“I am often afraid that things might turn out badly”; $\alpha = 0.87$). The measure was also found to have strong factor structure, high test-retest reliability and high predictive validity (construct of EPO is an important determinant of self-regulation).

Translation and the Pilot Studies

Every text in the study will be presented to participants in their own language: Traditional Chinese, Finnish, German, Spanish, Slovenian, Serbian, and Dutch. The questionnaires are constructed and written in English, then two separate translators for each of the seven languages translated (e.g., English to German) the battery. After that, back translations were compared with the original battery. The translators are native to the language they are translating to, and have at least a C1 level in the Common European Framework of Reference for Languages in said language. The process will be repeated until a third independent translator, native in English, considers that the original text and the back translation are equal in meaning.

The first pilot study will be conducted with the aim of testing the questionnaire framework and interaction with participant, and also to test psychometric characteristics of items and to eliminate those invalid. The battery is administered online in an English-speaking area, such as the UK, with a bigger sample that allows enough variability in the responses to the items to explore their functioning (i.e., 100 participants). We will assess item and subscale reliability and validity, and perform a factorial analysis.

Once the translation is prepared, we will conduct the second pilot study. The aim is to identify problems with items that may have appeared through the translation process. For example, items could be potentially ambiguous, unclear or misleading for participants (Ziegler et al., 2015). The battery will be administered to ~15 parents from each country and their responses will serve to adapt the problematic items and improve the battery.

Proposed Analysis

We will use statistical software (Matlab, R) for data analysis. We will analyse samples from each country separately because the differences in terms of language, legal framework in relation to vaccines, demographical characteristics and representativeness make the comparability between the different samples, and thus, their conjugate analysis, difficult (Ember et al., 1998).

For each sample, a factor analysis will be conducted to reduce the number of factors and control for interrelatedness of the variables. Two clusters of variables will be included: demographics (age, education level, number of children, mean age of children), and vaccine-related decision constructs (scores on scales on vaccination, vaccine hesitancy, perceived freedom of the decision, choice overload, actively open-minded thinking, trust toward authorities, availability of relevant information, perceived scientific consensus on vaccination, subjective estimation of percentage of children vaccinated on the country, perceived social consensus on vaccination, risk taking, and elaboration of potential outcomes). We expect the factor structure of each sample to be similar. With the reduced

number of factors, a multivariate linear regression analysis will be conducted and the predictive power will be tested. The aforementioned factors will serve as predictors, and the parental intention to vaccinate (“I plan to vaccinate my child in accordance to the official vaccination schedule of my country”) will serve as criterion. All variables mentioned will also be included in a multivariate analysis of variance to see if there are differences between them depending on the level of adherence (total adherence, not completely total adherence, no adherence at all) to the official vaccination schedule of their children. We will test the proposed model of mediation (**Figure 1**) using regression analysis and compare it to a more complicated model with involvement as an additional factor.

For the open-ended question in which parents will detail their reasons for not absolutely adhering to the official vaccination schedule in their country, we will perform a qualitative analysis. After a pre-analysis, we will categorize answers based on the main reasons provided (e.g., medical reasons, considering vaccination risky etc.), and report the frequencies for each of them.

We will also check, using ANOVA, if there are differences in dependent variable (i.e., intention to vaccinate) according to participants’ gender, marital status, while also considering whether vaccination is mandatory in their country of residence.

To check if the sources of information parents use to make vaccine-related decisions affect their intention to vaccinate, we will use ANOVA and *post-hoc* analysis.

Anticipated Results

The samples of this study will consist of data obtained in eight separate countries. This presents both challenges and possibilities. As some of the factors underlying vaccine hesitancy are context-specific and vary across time and place (Dubé et al., 2014), research in multiple countries is needed to understand vaccine hesitancy more fully on a local level. By analyzing our samples separately, we expect the study to contribute to knowledge on locally relevant factors related to vaccine hesitancy. However, the samples are unlikely to be representative in their respective countries, which limits the reliability of our conclusions on local, but not so much on the general level. Special care will also be taken to make sure that we sample people from all parts of the vaccine-hesitancy spectrum; the call for participants

will be posted in different interest groups online along with general population call.

Given the sensitive nature of the topic, it is possible for the participants who participate in the study to have stronger convictions toward vaccines in one way or another, which might further decrease the representativeness of the vaccine hesitancy continuum—we want to include parents with various degrees of their hesitancy and strength of their convictions in order to avoid bipolarization. To decrease the effect of the topic on the motivation to participate, all the communication (e.g., invites, instructions) will be neutral in terms of referencing to potential harms and benefits of vaccination as well as any moral judgements of the decision.

Due to the wide range of phenomena of the present study, we hypothesize to find the following factors to reinforce the intent to vaccinate one’s child: trust toward authorities, perceived social and scientific consensus, availability of relevant information, along with previously identified demographic characteristics and open-minded thinking. At the same time, we expect to confirm that vaccine hesitancy, perceived freedom, choice overload, the use of informal sources and susceptibility outcome bias serve as reinforces of delaying or omitting mandatory childhood vaccination. All our hypothesized connections can be seen in **Figure 1**.

STUDY 2—OUTCOME, NOT THE DECISION MAKER, MAKES THE CHOICE

Introduction

In order to further explore involvement as a factor, we will set up a second study. When people are making decisions about important issues (i.e., vaccine-related), the involvement in the decision and the aforementioned phenomenon of choice overload may be exacerbated. This is because ‘the costs associated with making the “wrong” choice, or even beliefs that “wrong” choices do indeed exist, are much more prominent, and substantial time and effort would be required for choosers to make truly informed comparisons among alternatives’ (Iyengar and Lepper, 2000). As the complexity of making choices rises, people tend to simplify their decision-making processes by relying on simple heuristics (Wright, 1975; Payne, 1982;

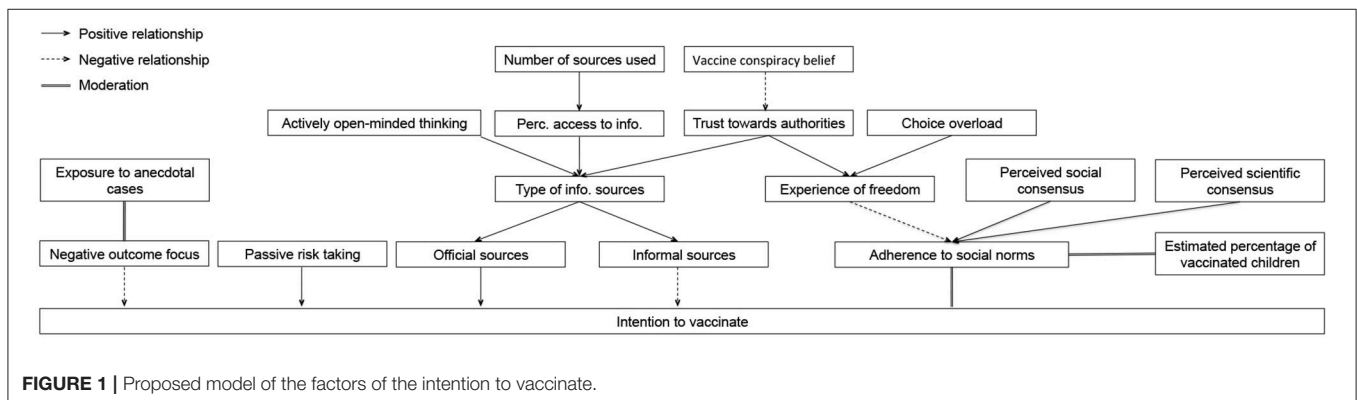


FIGURE 1 | Proposed model of the factors of the intention to vaccinate.

Payne et al., 1988, 1993; Timmermans, 1993). To test whether involvement and choice overload moderate cognitive aspects of vaccination decision-making, we will use one of the empirically well-established cognitive biases as a litmus test, the outcome bias.

Outcome Bias

People make and evaluate their own and other people’s decisions every day, but different timings of these two cognitive processes lead to differences in available information (Baron and Hershey, 1988). All possible outcomes of a decision are mutually exclusive. In the moment of making a decision, the winning outcome and following consequences could not have been known to the decision maker, but these two factors are what the evaluator is familiar with. The outcome of a decision, however, should not be taken into account when evaluating the decision, because this information is irrelevant to the quality of the decision. The systematic tendency to evaluate the quality of a decision based on the outcome is called outcome bias (Baron and Hershey, 1988). People tend to use their knowledge about the outcome in an aforementioned, not logically justified manner (Allison et al., 1996; Gino et al., 2010), and judge the discernment and competence of decision makers based on it (Berg-Cross, 1975; Baron and Hershey, 1988; Lipshitz, 1989; Gino et al., 2010). However, there is no evidence on the relationship between the outcome bias and vaccine hesitancy, or outcome bias’ connectedness to the level of parental involvement in the decision.

In the second study, involvement will be decomposed into three aspects: the situation decision makers are deciding about, decision makers’ role, or who is a protagonist (parents vs. non-parents), and finally, low and high involvement health situations. Because the importance of the decision is related to susceptibility to cognitive biases, we expect that higher involvement among parents, compared to non-parents, will lead to greater proneness to making biased evaluations in health, but not in non-health related dilemmas (experiment 1). In a similar manner, parents will be more susceptible to outcome bias in both high and low involvement situations than non-parents (experiment 2). However, we expect that in dilemmas where the parents are the protagonists, the outcome bias will be stronger in non-parents (experiment 3). The results will help to understand the specific role of parents in vaccine decisions and will also contribute to the research about the relationship between vaccine hesitancy and cognitive biases, which at the current state is often hypothesized about, but lacking in papers.

Methods

Sample

The second study will be conducted in Serbia. The sample will be comprised of parents and caregivers that participated in Study 1, and additionally an equal number of non-parents. Non-parents will be matched with parents who participated in Study 1 in terms of their age, gender and education. To estimate the sample size needed for the second study, special attention was paid to the minimization of Type II error. The analysis results of the test strength show that detecting a statistically significant outcome

bias effect by a bivariate analysis of variance (at level $p < 0.01$) of the effect size of .7, as reported by Baron and Hershey (1988), for a sample of 20 subjects per experimental group amounts to 99.9%.

Design

This study will comprise of three experiments, by which we aim to decompose the parental involvement in a more detailed manner, to test if it moderates susceptibility to outcome bias in parents. Specifically, we will test if parents differ in biased reasoning from non-parents (experiment 1), do those two groups differ when judging about involving decisions (experiment 2), and finally do parents show higher understanding for other parents’ decisions, in comparison with non-parents as judges (experiment 3). The design of all three experiments is mixed, 2×2 with two groups of participants, parents and non-parents.

Experiment 1 (Parenting and biased reasoning): 2 (levels of domain: health and non-health) \times 2 (outcomes).

Experiment 2 (Parenting and involvement): 2 (levels of involvement) \times 2 (outcomes)

Experiment 3 (Parenting and solidarity): 2 (levels of protagonists: parents and non-parents) \times 2 (outcomes).

The independent variables, the domain, level of involvement in the decision, and protagonist will have two levels: health (e.g., Vitamin supplementation) and non-health domain (e.g., free time); low (e.g., should protagonist give a vitamin supplement to a child or not), and high involvement (e.g., should protagonist vaccinate a child or not); and who is making a decision (parent as protagonist or non-parent as protagonist). The outcome variable will have two levels: the positive and the negative outcome. By crossing two levels of each binary independent variable in all three experiments (domain, involvement, decision maker) with the parenthood of participants (parents and non-parents), four experimental situations will be formed: parents in, e.g., high involvement situation, parents in low involvement situation, non-parents in high involvement situation, non-parents in low involvement situation (see **Table 2**). Every situation will be formed with both outcomes (see **Table 3**). Equivalent design will be applied in all three experiments.

Each subject will participate in the procedure in two time slots, separated by 1 week (with opposite outcomes). For each experimental situation, a Latin square design will be used for randomizing the order of presentation of experimental tasks during two experimental sessions.

TABLE 2 | Experimental situations in all three experiments.

	Experiment 1 (domain)		Experiment 2 (involvement)		Experiment 3 (protagonist)	
Factor levels	H	nH	hl	ll	P	nP
Parents	P \times H	P \times nH	P \times hl	P \times ll	P \times P	P \times nP
Non-parents	nP \times H	nP \times nH	nP \times hl	nP \times ll	nP \times P	nP \times nP

H, health-related; nH, non health-related; hl, high involvement; ll, low involvement; P, parent; nP, non parent.

TABLE 3 | Tasks for all three experiments.

	Domain		Involvement		Protagonist	
Factor levels	H	nH	hl	ll	P	nP
Positive outcome (+)	H+	nH+	hl+	ll+	P+	np+
Negative outcome (-)	H-	nH-	hl-	ll-	P-	nP-

H, health-related; nH, non health-related; hl, high involvement; ll, low involvement; P, parent; nP, non parent.

Stimuli and the Third Pilot Study

The stimuli in this study will have the form of an evaluation task, principally used in judging and decision-making research paradigms. The text presented to the participant will consist of a prolog (a description of a situation that contains a dilemma), followed by explicitly stating which option the decision maker (DM; the protagonist of the presented situation) opted for, and the outcome of the decision. Outcomes of DM's decisions will be twofold: positive and negative. Participants' task will be to evaluate the presented decision by rating it on a scale from -3 to +3 [-3—the worst decision the DM could have made, +3—the best decision the protagonist (DM) could have made].

The third pilot study was conducted in Serbia, during December, with the aim to test items. The design was experimental, 2 (outcome) × 2 (involvement). By crossing factors of involvement and outcome, four categories of tasks were formed: high involvement with positive outcome, high involvement with negative outcome, low involvement with positive outcome, low involvement with negative outcome.

Participants were parents with pre-schoolers, younger than seven years old ($N = 49$, 73% female, mean age 34.88). Each participant was presented with 24 pairs of tasks consisting of a prolog, decision maker's decision, and the outcome of the decision in two time slots separated by two weeks. During the second session participants were presented with the same tasks but with outcomes opposite to those from the previous session. Participants' task was to evaluate the decision on a 10-point Likert scale (1 = the worst decision the DM could have made, 10 = the best decision the DM could have made).

Results showed that there was a statistically significant outcome bias detected on a sample as a whole [$F_{(1)} = 283.239$, $p < 0.001$]. Mean evaluation of stimuli with the positive outcome was 6.02 ($S = 0.922$), while the mean evaluation for stimuli with negative outcome was 3.56 ($S = 0.834$). The effect size coefficients were calculated for each pair of tasks. Cohen's *D* coefficients ranged from 0.223 to 1.604, with mean value of 0.957,

Based on these effect sizes, as well as on the analysis of participants' impressions about questions, a number of convenient stimuli will be selected for the Study 2 experiment.

Procedure

After successfully completing Study 1, the participants will be asked to take part in the experiment. The experiment will be conducted at the Faculty of Philosophy of University in Belgrade, in classrooms equipped with computers. The non-parent sample will be collected separately, from the available pool of participants from general population in database of partner

institution. Participants will, again, have a self-generated personal code as their identification. Each subject will be provided with an introductory explanation for the following part of the study and given detailed instructions. Each participant will read a prolog (the situation), the decision the protagonist made and the outcome of the decision will be explicitly stated, and then participants will evaluate that decision. This process will be repeated twice. After finishing the second round, participants will be asked to give consent to store and use the data obtained in the current session, then they will be presented with a thank you note along with a reminder to visit the venue again in a week. They will also be asked to provide their e-mail address to enable us to send them a reminder, and, if they wish, a personal result, in comparison with both samples. After completing the task in the following week, participants will be asked to give consent again, presented with a thank you note, and the authors' email addresses if they have any inquiries regarding the experiment. Debriefing will be provided immediately, with the repeated emphasis that the situations described in the tasks are simulations and that they are not based on real data. The procedure was ethically approved by the Institutional Review Board of the Faculty of Philosophy at the University of Belgrade.

Proposed Analysis

The aim of this study is to investigate the moderating role of involvement in susceptibility to outcome bias.

Like in study 1, different samples will be treated separately. We will be using R and Matlab to conduct our analysis.

There is a possibility that participants might remember their answer from the first session and adapt their answer in the second session accordingly. To control this, we will use a two-way ANOVA. We will test whether the order of presentation of the stimuli, particularly whether the positive or negative outcome presented first, has an influence on the answers obtained during the second session.

To test our hypotheses, we will use a bivariate analysis of variance with repeated measures. As measures of the outcome bias, the difference between mean evaluations of decisions with positive and negative outcomes will be used to form 2 new variables for each of the 3 experiments. These new variables will then be used in the analysis. We will verify the assumption of standard distribution by conducting a Kolmogorov-Smirnov Test. If we cannot assume standard distribution, we will use the Wilcoxon signed rank test instead of an ANOVA. Effect sizes for each pair of tasks will be presented as Cohen's *D*.

JOINT DISCUSSION

With both studies we aim to investigate and come closer to understanding factors which influence vaccine hesitancy as a possible outcome of the vaccine-related decision-making processes, while focusing on testing the role of the involvement, as a potential underlying factor which skews this decision. Results of Study 1 will provide further insights into factors serving as reinforcements of the delay and omission of vaccination but also factors serving as reinforcements of the intention to vaccinate, as well as their interrelatedness. The results will provide insights

into the construct of vaccine hesitancy that is currently lacking different stakeholders combating dropping immunization rates.

Moreover, if the Study 2 yields results consistent with our hypotheses, we will gain valuable proof that in terms of the extent and lengths of the decision-making processes regarding vaccination of their children, parents are indeed a special group, different from people without children, who do not have to face such dilemmas, and are therefore susceptible to different cognitive obstacles to reaching a decision to vaccinate.

With such knowledge it would be possible to draft interventions custom-made for parents aimed toward undermining their vaccine hesitancy by establishing better communication channels and better, more effective formation of relevant, informative and non-patronizing messages, addressing their personal dilemmas and fears with respect and understanding. Parents might not be afraid of vaccine-preventable diseases, but it seems they are afraid of vaccines – and the burden of the immunity of our herd lies on their shoulders.

ETHICS STATEMENT

This study was carried out in accordance with the recommendations of the Institutional Review Board of the

Faculty of Philosophy of the University of Belgrade, who also ethically approved the proposed studies. All subjects gave written informed consent in accordance with the Declaration of Helsinki. The protocol was approved by the Institutional Review Board of the Faculty of Philosophy of the University of Belgrade.

AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct, and intellectual contribution to the work, and approved it for publication.

ACKNOWLEDGMENTS

The project is supported by the Junior Researcher Programme (<http://jrp.psycholars.org/>). We would like to thank the amazing organizing team for their support and dedication.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpsyg.2018.00735/full#supplementary-material>

REFERENCES

- Allison, S. T., Mackie, D. M., and Messick, D. M. (1996). Outcome biases in social perception: implications for dispositional inference, attitude change, stereotyping, and social behavior. *Adv. Exp. Soc. Psychol.* 28, 53–93. doi: 10.1016/S0065-2601(08)60236-1
- Almshat, S., Ayotte, B., Edelstein, B., and Margrett, J. (2008). Framing effect debiasing in medical decision making. *Patient Educ. Couns.* 71, 102–107. doi: 10.1016/j.pec.2007.11.004
- Arendell, T. (2000). Conceiving and investigating motherhood: The decade's scholarship. *J. Marriage Fam.* 62, 1192–1207. doi: 10.1111/j.1741-3737.2000.01192.x
- Arora, N. K., and McHorney, C. A. (2000). Patient Preferences for medical decision making: who really wants to participate? *Med. Care* 38, 335–341. doi: 10.1097/00005650-200003000-00010
- Baron, J., and Hershey, J. C. (1988). Outcome bias in decision evaluation. *J. Pers. Soc. Psychol.* 54, 569–579. doi: 10.1037/0022-3514.54.4.569
- Baron, J., Scott, S., Fincher, K., and Metz, S. E. (2015). Why does the cognitive reflection test (Sometimes) predict utilitarian moral judgment (and other things)? *J. Appl. Res. Mem. Cogn.* 4, 265–84. doi: 10.1016/j.jarmac.2014.09.003
- Bedford, H. (2014). Pro-vaccine messages may be counterproductive among vaccine-hesitant parents. *Evid. Based Med.* 19:219. doi: 10.1136/ebmed-2014-110037
- Berg-Cross, L. G. (1975). Intentionality, degree of damage, and moral judgments. *Child Dev.* 46, 970–974. doi: 10.2307/1128406
- Betsch, C., Renkewitz, F., Betsch, T., and Ulshöfer, C. (2010). The influence of vaccine-critical websites on perceiving vaccination risks. *J. Health Psychol.* 15, 446–455. doi: 10.1177/1359105309353647
- Bloom, B. R., Marcuse, E., and Mnookin, S. (2014). Addressing vaccine hesitancy. *Science* 344, 339. doi: 10.1126/science.1254834
- Boes, L., Boedeker, B., Schmich, P., Wetzstein, M., Wichmann, O., and Remschmidt, C. (2017). Factors associated with parental acceptance of seasonal influenza vaccination for their children—A telephone survey in the adult population in Germany. *Vaccine* 30, 3789–3796. doi: 10.1016/j.vaccine.2017.05.015
- Brunson, E. K. (2013a). How parents make decisions about their children's vaccinations. *Vaccine* 31, 5466–5470. doi: 10.1016/j.vaccine.2013.08.104
- Brunson, E. K. (2013b). The impact of social networks on parents' vaccination decisions. *Pediatrics* 131, 1397–1404. doi: 10.1542/peds.2012-2452
- Callender, D. (2016). Vaccine hesitancy: more than a movement. *Hum. Vaccin. Immunother.* 12, 2464–2468. doi: 10.1080/21645515.2016.1178434
- Camargo, K. Jr., and Grant, R. (2015). Public health, science, and policy debate: being right is not enough. *Am. J. Public Health* 105, 232–235. doi: 10.2105/AJPH.2014.302241
- Carrico, A. R., Vandenbergh, M. P., Stern, P. C., Gardner, G. T., Dietz, T., and Gilligan, J. M. (2011). Energy and climate change: key lessons for implementing the behavioral wedge. *George Washington J. Ener. Environ. Law* 2, 10–24. Available online at: <https://ssrn.com/abstract=1612224>
- Cobos, M. D., Monzón Llamas, L., and Bosch-Capblanch, X. (2015). Exposing concerns about vaccination in low- and middle-income countries: a systematic review. *Int. J. Public Health* 60, 767–780. doi: 10.1007/s00038-015-0715-6
- Collins, H. (2009). We Cannot Live by Scepticism Alone. *Nature* 458:30. doi: 10.1038/458030a
- Contractor, S. H., and Fox, R. J. (2011). An investigation of the relationship between the number of response categories and scale sensitivity. *J. Targe. Meas. Anal. Market.* 19, 23–35. doi: 10.1057/jt.2011.4
- Damjanović, K., and Gvozdenović, V. (2016). Influence of the probability level on the framing effect. *Psihologijske Teme* 25, 405–429. Available online at: <https://hrcak.srce.hr/169520>
- Darwin, H., Neave, N., and Holmes, J. (2011). Belief in conspiracy theories: the role of paranoid belief, paranoid ideation and schizotypy. *Pers. Individ. Diff.* 50, 1289–1293. doi: 10.1016/j.paid.2011.02.027
- de Visser, R., Waites, L., Parikh, C., and Lawrie, A. (2011). The importance of social norms for uptake of catch-up human papillomavirus vaccination in young women. *Sex. Health* 8, 330–337. doi: 10.1071/SH10155
- Donovan, R. J., and Jalleh, G. (2000). Positive versus negative framing of a hypothetical infant immunization: the influence of involvement. *Health Educ. Behav.* 27, 82–95. doi: 10.1177/109019810002700108

- Downs, J. S., de Bruin, W. B., and Fischhoff, B. (2008). Parents' vaccination comprehension and decisions. *Vaccine* 26, 1595–1607. doi: 10.1016/j.vaccine.2008.01.011
- Dubé, E., Gagnon, D., Nickels, E., Jeram, S., and Schuster M. (2014). Mapping vaccine hesitancy-Country-specific characteristics of a global phenomenon. *Vaccine* 32, 6649–6654. doi: 10.1016/j.vaccine.2014.09.039
- Edwards, A., Elwyn, G., Covey, J., Matthews, E., and Pill, R. (2001). Presenting risk information a review of the effects of framing and other manipulations on patient outcomes. *J. Health Commun.* 6, 61–82. doi: 10.1080/10810730150501413
- Ember, C. R., Ember, M., and Peregrine, N. (1998). "Cross-cultural research," in *Handbook of Methods in Cultural Anthropology*, ed H. R. Bernard (Walnut Creek, CA: AltaMira), 647–687.
- Eurostat (2017). *Inability To Make Ends Meet – Eurostat*. Ec.Europa.Eu. Available online at: http://ec.europa.eu/eurostat/en/web/products-datasets/-/ILC_MDES09
- Evans, M., Stoddart, H., Condon, L. J., Freeman, E., Grizzell, M., and Mullen, R. (2001). Parents' perspectives on the MMR immunisation: a focus group study. *Br. J. Gen. Pract.* 51, 904–910. Available online at: <http://bjgp.org/content/bjgp/51/472/904.full.pdf>
- Fagley, N. S., and Miller, P. M. (1997). Framing effects and Arenas of choice: your money or your life? *Org. Behav. Hum. Decis. Proc.* 71, 355–373. doi: 10.1006/obhd.1997.2725
- Freed, G. L., Clark, S. J., Butchart, A. T., Singer, D. C., and Davis, M. M. (2010). Parental vaccine safety concerns in 2009. *Pediatrics* 125, 654–659. doi: 10.1542/peds.2009-1962
- Freed, G. L., Clark, S. J., Butchart, A. T., Singer, D. C., and Davis, M. M. (2011). Sources and perceived credibility of vaccine-safety information for parents. *Pediatrics* 127(Suppl. 1), S107–S112. doi: 10.1542/peds.2010-1722P
- Freed, G. L., Clark, S. J., Hibbs, B. F., and Santoli, J. M. (2004). Parental vaccine safety concerns: the experiences of pediatricians and family physicians. *Am. J. Prevent. Med.* 26, 11–14. doi: 10.1016/j.amepre.2003.09.004
- GBD 2015 Healthcare Access and Quality Collaborators (2017). Healthcare access and quality index based on mortality from causes amenable to personal health care in 195 countries and territories. a novel analysis from the global burden of disease study 2015. *Lancet* 390, 231–266. doi: 10.1016/S0140-6736(17)30818-8
- Gellin, B. G., Maibach, E. W., and Marcuse, E. K. (2000). Do parents understand immunizations? A national telephone survey. *Pediatrics* 106, 1097–1102. doi: 10.1542/peds.106.5.1097
- Gino, F., Shu, L., and Bazerman, M. (2010). Nameless + harmless = blameless: when seemingly irrelevant factors influence judgment of (un)ethical behavior. *Organ. Behav. Hum. Decis. Process.* 111, 93–101. doi: 10.1016/j.obhdp.2009.11.001
- Goldenberg, M. J. (2016). *Public Misunderstanding of Science? Reframing the Problem of Vaccine Hesitancy. Perspectives on Science*. Cambridge, MA: MIT Press.
- Goldstein, W. M., and Weber, E. U. (1995). Content and discontent: indications and implications of domain specificity in preferential decision making. *Psychol. Learn. Motiv.* 32, 83–136.
- Gowda, C., and Dempsey, A. F. (2013). The rise (and Fall?) of parental vaccine hesitancy. *Hum. Vacc. Immunother.* 9, 1755–1762. doi: 10.4161/hv.25085
- Gowda, C., Carlos, R. C., Butchart, A. T., Singer, D. C., Davis, M. M., and Dempsey, A. F. (2012). CHIAS: a standardized measure of parental HPV immunization attitudes and beliefs and its associations with vaccine uptake. *Sex. Transm. Dis.* 39, 475–481. doi: 10.1097/OLQ.0b013e318248a6d5
- Grol, R., Whitfield, M., De Maeseneer, J., and Mokkink, H. (1990). Attitudes to risk taking in medical decision making among British, Dutch and Belgian general practitioners. *Br. J. Gen. Pract.* 40, 134–136.
- Gummerum, M., Hanoch, Y., and Rolison, J. J. (2014). Offenders' risk-taking attitude inside and outside the prison walls *Risk Anal.* 34, 1870–1881. doi: 10.1111/risa.12222
- Gust, D. A., Darling, N., Kennedy, A., and Schwartz, B. (2008). Parents with doubts about vaccines: which vaccines and reasons why. *Pediatrics* 122, 718–725. doi: 10.1542/peds.2007-0538
- Gust, D. A., Kennedy, A., Shui, I., Smith, P. J., Nowak, G., and Pickering, L. K. (2005). Parent attitudes toward immunizations and healthcare providers: the role of information. *Am. J. Prev. Med.* 29, 105–112. doi: 10.1016/j.amepre.2005.04.010
- Gust, D. A., Strine, T. W., Maurice, E., Smith, P., Yusuf, H., Wilkinson, M., et al. (2004). Underimmunization among children: effects of vaccine safety concerns on immunization status. *Pediatrics* 114, 16–21. doi: 10.1542/peds.114.1.e16
- Hanoch, Y., Johnson, J. G., and Wilke, A. (2006). Domain specificity in experimental measures and participant recruitment: an application to risk-taking behavior. *Psychol. Sci.* 17, 300–304. doi: 10.1111/j.1467-9280.2006.01702.x
- Haran, U., Ritov, I., and Mellers, B. A. (2013). The role of actively open-minded thinking in information acquisition, accuracy, and calibration. *Judgm. Decis. Mak.* 8, 188–201. doi: 10.1177/2053168016676705
- Hays, S. (1996). *The Cultural Contradictions of Motherhood*. New Haven, CT: Yale University Press.
- Healy, C. M., and Pickering, L. K. (2011). How to communicate with vaccine-hesitant parents. *Pediatrics* 127 (Suppl. 1), S127–S133. doi: 10.1542/peds.2010-1722S
- Henrikson, N. B., Anderson, M. L., Opel, D. J., Dunn, J., Marcuse, E. K., and Grossman, D. C. (2017). *Longitudinal Trends in Vaccine Hesitancy in a Cohort of Mothers Surveyed in Washington State, 2013–2015*. Public Health Reports. SAGE Publications, Sage, CA, 451–454.
- Highland, J. (2016). *Parental Decision Making and Childhood Vaccination*. CaseWestern Reserve University, 1–30.
- Hobson-West, P. (2007). Trusting Blindly Can Be the Biggest Risk of All: organised resistance to childhood vaccination in the UK. *Sociol. Health Illness* 29, 198–215. doi: 10.1111/j.1467-9566.2007.00544.x
- Horne, Z., Powell, D., Hummel, J. E., and Holyoak, K. J. (2015). Countering antivaccination attitudes. *Proc. Natl. Acad. Sci. U.S.A.* 112, 10321–10324. doi: 10.1073/pnas.1504019112
- Iyengar, S. S., and Lepper, M. R. (2000). When choice is demotivating: can one desire too much of a good thing? *J. Pers. Soc. Psychol.* 79, 995–1006. doi: 10.1037/0022-3514.79.6.995
- Jolley, D., and Douglas, K. M. (2014). The effects of anti-vaccine conspiracy theories on vaccination intentions. *PLoS ONE* 9:e89177. doi: 10.1371/journal.pone.0089177
- Jones, A. M., Omer, S. B., Bednarczyk, R. A., Halsey, N. A., Moulton, L. H., and Salmon, D. A. (2012). Parents' source of vaccine information and impact on vaccine attitudes, beliefs, and nonmedical exemptions. *Adv. Prevent. Med.* 2012, 1–9. doi: 10.1155/2012/932741
- Kahan, D. M. (2010). Fixing the communications failure. *Nature* 463, 296–297. doi: 10.1038/463296a
- Kahan, D. M., and Braman, D. (2006). Cultural cognition and public policy. *Yale Law Policy Rev.* 24, 147–170. Available online at: <http://digitalcommons.law.yale.edu/ylpr/vol24/iss1/5>
- Kahan, D. M., Braman, D., Cohen, G. L., Gastil, J., and Slovic, P. (2010). Who fears the HPV vaccine, who doesn't, and why? an experimental study of the mechanisms of cultural cognition. *Law Hum. Behav.* 34, 501–516. doi: 10.1007/s10979-009-9201-0
- Kahan, D. M., Braman, D., Slovic, P., Gastil, J., and Cohen, G. L. (2009). Cultural cognition of the risks and benefits of nanotechnology. *Nat. Nanotechnol.* 4, 87–90. doi: 10.1038/nnano.2008.341
- Kahan, D. M., Jenkins-Smith, H., and Braman, D. (2011). Cultural cognition of scientific consensus. *J. Risk Res.* 14, 147–174. doi: 10.1080/13669877.2010.511246
- Keinan, R., and Bereby-Meyer, Y. (2012). 'Leaving It to chance' —passive risk taking in everyday life. *Judgm. Decis. Mak.* 7, 705–715. Available online at: <http://journal.sjdm.org/15/15717/jdm15717.pdf>
- Kennedy, A., LaVail, K., Nowak, G., Basket, M., and Landry, S. (2011). Confidence about vaccines in the United States: understanding parents' perceptions. *Health Aff.* 30, 1151–1159. doi: 10.1377/hlthaff.2011.0396
- Kühberger, A. (1998). The influence of framing on risky decisions: a meta-analysis. *Organ. Behav. Hum. Decis. Process.* 75, 23–55. doi: 10.1006/obhd.1998.2781
- Larson, H. J., Jarrett, C., Eckersberger, E., Smith, D. M. D., and Paterson, P. (2014). Understanding vaccine hesitancy around vaccines and vaccination from a global perspective: a systematic review of published literature, 2007–2012. *Vaccine* 32, 2150–2159. doi: 10.1016/j.vaccine.2014.01.081
- Lau, S., Hiemisch, A., and Baumeister, R. F. (2015). The experience of freedom in decisions—questioning philosophical beliefs in favor of psychological determinants. *Conscious. Cogn.* 33, 30–46. doi: 10.1016/j.concog.2014.11.008

- Leask, J. A., Braunack-Mayer, A., and Kerridge, I. (2011). Consent and public engagement in an era of expanded childhood immunisation. *J. Paediatr. Child Health* 47, 603–607. doi: 10.1111/j.1440-1754.2011.02160.x
- Leask, J., and MacArtney, K. (2008). Parental decisions about vaccination: collective values are important. *J. Paediatr. Child Health* 44, 534–535. doi: 10.1111/j.1440-1754.2008.01381.x
- Lewandowsky, S., and Oberauer, K. (2016). Motivated rejection of science. *Curr. Dir. Psychol. Sci.* 25, 217–222. doi: 10.1177/0963721416654436
- Lewandowsky, S., Ecker, U. K. H., Seifert, C. M., Schwarz, N., and Cook, J. (2012). Misinformation and its correction. *Psychol. Sci. Pub. Interest* 13, 106–131. doi: 10.1177/1529100612451018
- Lipshitz, R. (1989). 'Either a Medal or a Corporal': the effects of success and failure on the evaluation of decision making and decision makers. *Organ. Behav. Hum. Decis. Process.* 44, 380–395.
- Luthy, K. E., Beckstrand, R. L., and Peterson, N. E. (2009). Parental hesitation as a factor in delayed childhood immunization. *J. Pediatr. Health Care* 23, 388–393. doi: 10.1016/j.pedhc.2008.09.006
- Markiewicz, Ł., and Weber, E. U. (2013). DOSPERT's gambling risk-taking propensity scale predicts excessive stock trading. *J. Behav. Finance* 14, 65–78. doi: 10.1080/15427560.2013.762000
- McNeil, B. J., Pauker, S. G., Sox, H. C. Jr., and Tversky, A. (1982). On the elicitation of preferences for alternative therapies. *N. Engl. J. Med.* 306, 1259–1262. doi: 10.1056/NEJM198205273062103
- Mills, E. J., Montori, V. M., Ross, C. P., Shea, B., Wilson, K., and Guyatt, G. H. (2005). Systematically reviewing qualitative studies complements survey design: an exploratory study of barriers to paediatric immunisations. *J. Clin. Epidemiol.* 58, 1101–1108. doi: 10.1016/j.jclinepi.2005.01.014
- Mitton, H., and Mercier, H. (2015). Cognitive obstacles to pro-vaccination beliefs. *Trends Cogn. Sci.* 19, 633–636. doi: 10.1016/j.tics.2015.08.007
- Mohd Azizi, F. S., Kew, Y., and Moy, F. M. (2017). Vaccine hesitancy among parents in a multi-ethnic country, Malaysia. *Vaccine* 35, 2955–2961. doi: 10.1016/j.vaccine.2017.04.010
- Myers, L. B., and Goodwin, R. (2011). Determinants of adults' intention to vaccinate against pandemic swine flu. *BMC Public Health* 11:15. doi: 10.1186/1471-2458-11-15
- Nenkov, G. Y., Inman, J. J., and Hulland, J. (2008). Considering the future: the conceptualization. *J. Consu. Res.* 35, 126–141. doi: 10.1086/525504
- Opel, D. J., Taylor, J. A., Mangione-Smith, R., Solomon, C., Zhao, C., Martin, D., et al. (2011). Validity and reliability of a survey to identify vaccine-hesitant parents. *Vaccine* 29, 6598–6605. doi: 10.1016/j.vaccine.2011.06.115
- Oraby, T., Thampi, V., and Bauch, C. T. (2014). The influence of social norms on the dynamics of vaccinating behaviour for paediatric infectious diseases. *Proc. Biol. Sci.* 281:20133172. doi: 10.1098/rspb.2013.3172
- Pareek, M., and Pattison, H. M. (2000). The two-dose Measles, Mumps, and Rubella (MMR) immunisation schedule: factors affecting maternal intention to vaccinate. *Br. J. Gen. Pract.* 50, 969–971.
- Payne, J. W. (1982). Contingent decision behavior. *Psychol. Bull.* 92, 382–402. doi: 10.21236/ADA111655
- Payne, J. W., Bettman, J. R., and Johnson, E. J. (1988). Adaptive strategy selection in decision making. *J. Exp. Psychol. Learn. Mem. Cogn.* 14, 534–552. doi: 10.1037/0278-7393.14.3.534
- Payne, J. W., Bettman, J. R., and Johnson, E. J. (1993). *The Adaptive Decision Maker*.
- Pearce, A., Law, C., Elliman, D., Cole, T. J., and Bedford, H. (2008). Factors associated with uptake of measles, mumps, and rubella vaccine (MMR) and use of single antigen vaccines in a contemporary UK cohort: prospective cohort study. *BMJ* 336, 754–757. doi: 10.1136/bmj.39489.590.671.25
- Pluviano, S., Watt, C., and Sala, S. D. (2017). Misinformation lingers in memory: failure of three pro-vaccination strategies. *PLoS ONE* 12:e0181640. doi: 10.1371/journal.pone.0181640
- Pyke-Grimm, K. A., Degner, L., Small, A., and Mueller, B. (1999). Preferences for participation in treatment decision making and information needs of parents of children with cancer: a pilot study. *J. Pediatr. Oncol. Nurs.* 16, 13–24.
- Rachiotis, G., Mouchtouri, V. A., Kremastinou, J., Gourgoulanis, K., and Hadjichristodoulou, C. (2010). Low Acceptance of Vaccination against the 2009 Pandemic Influenza A (H1N1) among Healthcare Workers in Greece. *Euro Surveill.* 15:19486. doi: 10.2807/ese.15.06.19486-en
- Reich, J. A. (2014). Neoliberal mothering and vaccine refusal: imagined gated communities and the privilege of choice. *Gender Soc.* 28, 679–704. doi: 10.1177/0891243214532711
- Reyna, V. F. (2008). A theory of medical decision making and health: fuzzy trace theory. *Med. Decis. Making* 28, 850–865. doi: 10.1177/0272989X08327066
- Reyna, V. F., and Brainerd, C. J. (1992). "A fuzzy-trace theory of reasoning and remembering: paradoxes, patterns, and parallelism," in *From Learning Processes to Cognitive Processes: Essays in Honor of William K. Estes 2*, eds A. F. Healy, S. M. Kosslyn, and R. M. Shiffrin (Hillsdale, NJ: Erlbaum), 235–259.
- Rolfe-Redding, J. C., Maibach, E. W., Feldman, L., and Leiserowitz, A. (2012). "Republicans and climate change: an audience analysis of predictors for belief and policy preferences," in *62nd Annual International Communication Association Conference* (Phoenix, AZ), 1–48.
- Ruiter, R. A. C., Abraham, C., and Kok, G. (2001). Scary warnings and rational precautions: a review of the psychology of fear appeals. *Psychol. Health* 16, 613–630. doi: 10.1080/08870440108405863
- Ruiter, R. A., Kessels, L. T., Peters, G. J., and Kok, G. (2014). Sixty years of fear appeal research: current state of the evidence. *Int. J. Psychol.* 49, 63–70. doi: 10.1002/ijop.12042
- Sadaf, A., Richards, J. L., Glanz, J., Salmon, D. A., and Omer, S. B. (2013). A systematic review of interventions for reducing parental vaccine refusal and vaccine hesitancy. *Vaccine* 31, 4293–4304. doi: 10.1016/j.vaccine.2013.07.013
- Salmon, D. A., Moulton, L. H., Omer, S. B. M., Patricia deHart, Stokley, S., and Halsey, N. A. (2005). Factors associated with refusal of childhood vaccines among parents of school-aged children: a case-control study. *Arch. Pediatrics Adolesc. Med.* 159, 470–476. doi: 10.1001/archpedi.159.5.470
- Seeman, N., Ing, A., and Rizo, C. (2010). Assessing and responding in real time to online anti-vaccine sentiment during a flu pandemic. *Healthcare Q.* 13, 8–15. doi: 10.12927/hcq.2010.21923
- Shapiro, G. K., Holding, A., Perez, S., Amsel, R., and Rosberger, Z. (2016). Validation of the vaccine conspiracy beliefs scale. *Papillomavirus Res.* 2, 167–172. doi: 10.1016/j.pvr.2016.09.001
- Smailbegovic, M. S., Laing, G. J., and Bedford, H. (2003). Why do parents decide against immunization? the effect of health beliefs and health professionals. *Child Care Health Dev.* 29, 303–311. doi: 10.1046/j.1365-2214.2003.00347.x
- Smith, P. J., Susan Chu, Y., and Lawrence Barker, E. (2004). Children who have received no vaccines: who are they and where do they live? *Pediatrics* 114, 187–195. doi: 10.1542/peds.114.1.187
- Smyth, C., and Craig, L. (2017). Conforming to intensive parenting ideals: willingness, reluctance and social context. *Famil. Relationsh. Soc.* 6, 107–124. doi: 10.1332/204674315X14393034138937
- Solomon, M. G., Bamossy, S., Askegaard, and Hogg, M. K. (2006). *Consumer Behaviour: A European Perspective, 3rd Edn.* New Jersey, NJ: Prentice Hall Inc.
- Sporton, R. K., and Francis, S. A. (2001). Choosing not to immunize: are parents making informed decisions? *Fam. Pract.* 18, 181–188. doi: 10.1093/fampra/18.2.181
- Stanovich, K. E., and West, R. F. (2007). Natural myside bias is independent of cognitive ability. *Think. Reason.* 13, 225–247. doi: 10.1080/13546780600780796
- Stojkovic, A., Lazarevic, J., Anzelm, D., Drljaca, U., Zezelj, I., and Damjanovic, K. (2017). *Psychological Correlates of Resistance to Mandatory Child Vaccination.* University of Zadar Department of Psychology, ZADAR, CROATIA.
- Stronach, A. E. (2015). *Media and Internet Influences on Parental Decision Making Related to Vaccination.* Christchurch: University of Canterbury.
- Swami, V., Coles, R., Stieger, S., Pietschnig, J., Furnham, A., Rehim, S., et al. (2011). Conspiracist ideation in Britain and Austria: evidence of a monological belief system and associations between individual psychological differences and real-world and fictitious conspiracy theories. *Br. J. Psychol.* 102, 443–463. doi: 10.1111/j.2044-8295.2010.02004.x
- Tanner, C., Medin, D. L., and Iliev, R. (2008). Influence of deontological versus consequentialist orientations on act choices and framing effects: when principles are more important than consequences. *Eur. J. Soc. Psychol.* 38, 757–69. doi: 10.1002/ejsp.493

- Thompson, A. G. H. (2007). The meaning of patient involvement and participation in health care consultations: a taxonomy. *Soc. Sci. Med.* 64, 1297–1310. doi: 10.1016/j.socscimed.2006.11.002
- Timmermans, D. (1993). The impact of task complexity on information use in multi-attribute decision making. *J. Behav. Decis. Making* 6, 95–111.
- Tom, J. (2017). Social origins of scientific deviance: examining creationism and global warming skepticism. *Sociol. Perspect.* 61, 341–360. doi: 10.1177/0731121417710459
- Van der Linden, S. (2011). Charitable intent: a moral or social construct? a revised theory of planned behavior model. *Curr. Psychol.* 30, 355–374. doi: 10.1007/s12144-011-9122-1
- van der Linden, S. L., Leiserowitz, A., Rosenthal, S., and Maibach, E. (2017). Inoculating the public against misinformation about climate change. *Global Challenges* 1:1600008. doi: 10.1002/gch2.201600008
- Van der Linden, S. L., Maibach, E., and Leiserowitz, A. (2015). Improving public engagement with climate change. *Perspect. Psychol. Sci.* 10, 758–763. doi: 10.1177/1745691615598516
- Van der Linden, S., and Lewandowsky, S. (2015). How to combat distrust of science: the surprising power of the psychology of consensus. *Sci. Am. Mind*. Available online at: <https://www.scientificamerican.com/article/how-to-combat-distrust-of-science/> (Accessed October 20, 2017).
- Wagenaar, W. A., Keren, G., and Lichtenstein, S. (1988). Islanders and hostages: deep and surface structures of decision problems. *Acta Psychol.* 67, 175–189. doi: 10.1016/0001-6918(88)90012-1
- Wang, X. T. (1996a). Framing effects: dynamics and task domains. organizational behavior and human decision processes. *Organ. Behav. Human Decis. Process.* 68, 145–157.
- Wang, X. T. (1996b). Domain-specific rationality in human choices: violations of utility axioms and social contexts. *Cognition* 60, 31–63.
- Ward, P. R., Attwell, K., Meyer, S. B., Rokkas, P., and Leask, J. (2017). Understanding the perceived logic of care by vaccine-hesitant and vaccine-refusing parents: a qualitative study in Australia. *PLoS ONE* 12:e0185955. doi: 10.1371/journal.pone.0185955
- Wilson, R. S., Arvai, J. L., and Arkes, H. R. (2008). My loss is your loss. sometimes: loss aversion and the effect of motivational biases. *Risk Anal.* 28, 929–938. doi: 10.1111/j.1539-6924.2008.01065.x
- World Health Organization (2014). *Report of the SAGE Working Group on Vaccine Hesitancy*. Available online at: http://www.who.int/immunization/sage/meetings/2014/october/1_Report_WORKING_GROUP_vaccine_hesitancy_final.pdf (Accessed October 28, 2017).
- World Health Organization (2017a). *Density of Physicians (Total Number per 1000 Population, Latest Available Year)*. Available online at: http://www.who.int/gho/health_workforce/physicians_density/en/
- World Health Organization (2017b). *Immunization Covered*. Available online at: <http://www.who.int/mediacentre/factsheets/fs378/en/>
- World Health Organization (2018). *Europe Observes a 4 Fold Increase in Measles Cases in 2017 Compared to a Previous Year*. Available online at: <http://www.euro.who.int/en/media-centre/sections/press-releases/2018/europe-observes-a-4-fold-increase-in-measles-cases-in-2017-compared-to-previous-year>
- Wright, P. (1975). Consumer choice strategies: simplifying vs. optimizing. *J. Market. Res.* 12, 60–67. doi: 10.2307/3150659
- Wroe, A. L., Turner, N., and Salkovskis, P. M. (2004). Understanding and predicting parental decisions about early childhood immunizations. *Health Psychol.* 23, 33–44. doi: 10.1037/0278-6133.23.1.33
- Ziegler, G., Durus, N., Sert, O., and Family, N. (2015). “Analysing ELT in the european arena: multilingual practices,” in *International Perspectives on ELT Classroom Interaction*, eds C. J. Jenks and P. Seedhouse (London, UK: Palgrave), 188–207.
- Zikmund-Fisher, B. J., Sarr, B., Fagerlin, A., and Ubel, P. A. (2006). A matter of perspective: choosing for others differs from choosing for yourself in making treatment decisions. *J. Gener. Int. Med.* 21, 618–22. doi: 10.1111/j.1525-1497.2006.00410.x
- Zimmerman, L., Shalvi, S., and Bereby-Meyer, Y. (2014). Self-reported ethical risk taking tendencies predict actual dishonesty. *Judgm. Decis. Mak.* 9, 58–64. Available online at: <http://journal.sjdm.org/13/131111/jdm131111.pdf>
- Zingg, A., and Siegrist, M. (2012). Measuring people’s knowledge about vaccination: developing a one-dimensional scale. *Vaccine* 30, 3771–3777. doi: 10.1016/j.vaccine.2012.03.01

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2018 Damnjanović, Graeber, Ilić, Lam, Lep, Morales, Pulkkinen and Vingerhoets. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



From Face-to-Face to Facebook: Probing the Effects of Passive Consumption on Interpersonal Attraction

Amy C. Orben^{1*}, Augustin Mutak², Fabian Dablander³, Marlene Hecht⁴, Jakub M. Krawiec⁵, Natálie Valkovičová⁶ and Daina Kosite⁷

¹ Department of Experimental Psychology, University of Oxford, Oxford, United Kingdom, ² Department of Psychology, Faculty of Humanities and Social Sciences, University of Zagreb, Zagreb, Croatia, ³ Department of Psychological Methods, University of Amsterdam, Amsterdam, Netherlands, ⁴ Department of Psychology, Ludwig Maximilian University of Munich, Munich, Germany, ⁵ Department of Psychology, University of Social Sciences and Humanities, Warsaw, Poland, ⁶ Department of Psychology, Faculty of Social Studies, Masaryk University, Brno, Czechia, ⁷ Behaviour and Health Research Unit, University of Cambridge, Cambridge, United Kingdom

OPEN ACCESS

Edited by:

Andrea Cavallo,
Università degli Studi di Torino, Italy

Reviewed by:

Fabrizio Scrima,
Université de Rouen Normandie,
France
Maria Elena Johanna Kempnich,
University of Oxford, United Kingdom

*Correspondence:

Amy C. Orben
amy.orben@psy.ox.ac.uk

Specialty section:

This article was submitted to
Quantitative Psychology and
Measurement,
a section of the journal
Frontiers in Psychology

Received: 01 November 2017

Accepted: 18 June 2018

Published: 10 July 2018

Citation:

Orben AC, Mutak A, Dablander F, Hecht M, Krawiec JM, Valkovičová N and Kosite D (2018) From Face-to-Face to Facebook: Probing the Effects of Passive Consumption on Interpersonal Attraction. *Front. Psychol.* 9:1163. doi: 10.3389/fpsyg.2018.01163

Social media is radically altering the human social landscape. Before the internet era, human interaction consisted chiefly of direct and reciprocal contact, yet with the rise of social media, the passive consumption of other users' information is becoming an increasingly popular pastime. Passive consumption occurs when a user reads the posts of another user without interacting with them in any way. Previous studies suggest that people feel more connected to an artificial person after passively consuming their Facebook posts. This finding could help explain how relationships develop during passive consumption and what motivates this kind of social media use. This protocol proposes two studies that would make both a methodological and a theoretical contribution to the field of social media research. Both studies investigate the influence of passive consumption on changes in interpersonal attraction. The first study tests whether screenshots, which are widely used in present research, can be used as a proxy for real Facebook use. It measures the changes in interpersonal attraction after passive consumption of either a screenshot, an artificial *in situ* profile, or an acquaintance's real Facebook profile. The second study relies on traditional theories of relationship formation and motivation to investigate which variables (perceived intimacy, perceived frequency of posts, perceived variety of post topics, attributional confidence, and homophily) moderate the link between interpersonal attraction before and after passive consumption. The results of the first study provide insights into the generalizability of the effect by using different stimuli, while also providing a valuable investigation into a commonly used method in the research field. The results of the second study supplement researchers' understanding of the pathways linking passive use and interpersonal attraction, giving the field further insight into whether theories about offline relationship formation can be used in an online context. Taken together, this protocol aims to shed light on the intricate relation between passive consumption and interpersonal attraction, and variables moderating this effect.

Keywords: Facebook, passive consumption, homophily, social networking sites, passive use, longitudinal studies, interpersonal attraction

INTRODUCTION

1.32 billion people use Facebook daily, each spending about 50 min on the site (Nowak and Guillermo, 2017). That sums up to 125 years of Facebook use – every day. This trend to use social media, which has spread around the world, shows no signs of slowing down. In the meantime, it has not only changed how we spend our time, it has also introduced many new communicative features to our social life.

While social media use is often treated as a monolithic concept, it supports several different behaviors. Burke et al. (2010) identified two distinct types of social media use: directed communication and passive consumption. Directed communication does not only consist of chatting one-on-one, but also includes tagging in photos, commenting, or writing on profiles. In contrast, passive consumption includes browsing News feeds and scrolling through others' posts and public conversations without direct interaction (Schlosser, 2005).

It is easy to equate social media use to the former activity, interpreting it as a way to socialize directly with others. An interesting trend has, however, emerged over the last years. In 2012, 70% of Facebook users were active users, spending their time posting statuses or chatting with friends. This number decreased to only 52% in 2014 (McGrath, 2015), while the time spent passively consuming markedly increased. Recent research suggests that passive Facebook use (Brandtzæg, 2012) – when users spend most of their time viewing others' content while not creating content of their own (Burke et al., 2011) – is now the norm (Gerson et al., 2017).

Passive Use and Why It Is Interesting

It is, however, still unclear how passive consumption fits into our previous interpersonal social landscape. Trying to understand this, many research projects have focused on whether passive consumption can increase social connection and interpersonal attraction (e.g., Utz, 2015; Orben and Dunbar, 2017). Yet the evidence for whether passive consumption on social media can support relationship development is still sparse and mixed (Frost and Rickwood, 2017). While it is not as intimate as one-on-one conversations, some suggest passive consumption helps maintain broader social networks (Burke et al., 2011). Passive consumption of friends' and acquaintances' posts on the News feed gives users access to a much larger social circle, letting them overcome social and time barriers that make direct contact laborious or impossible (Lewis and West, 2009; Burke et al., 2011). Passive consumption can also positively influence emotional well-being and provide informational and emotional support (Ballantine and Stephenson, 2011; Good et al., 2013).

Yet studies also demonstrate that passive use of social media has potential negative effects. These negative consequences include increased loneliness, undermined affective well-being, raised social anxiety, and envy (Burke et al., 2010; Krasnova et al., 2013; Verduyn et al., 2017). Passive use of social networking sites also promotes weak and low-commitment ties (Lewis and West, 2009), social bonds that are not as intimate as the ones formed via direct interactions. These unique effects make it apparent

that passive consumption is not identical to offline interpersonal interaction.

Nevertheless, increasing proportions of the population are using Facebook passively and it is therefore important to research what mechanisms determine its effect on interpersonal attraction (Steers, 2015). Previous psychological theories – put forth to explain the development of interpersonal attraction in the offline world – can be a helpful guide in that process. These theories suggest a wealth of possible moderators of face-to-face communication which affect interpersonal attraction. By investigating whether these moderators also influence changes in interpersonal attraction during passive consumption, researchers could take substantial steps toward a more comprehensive picture of how passive consumption affects social attraction. Our protocol details two studies that take first steps in this direction.

The Proposed Studies

The two proposed studies aim to make both a methodological and a theoretical contribution to the field of social media research. The first study examines the novel experimental manipulation we intend to use in our subsequent study. In this first study, we test whether previous manipulations used in research are valid proxies of social media use, and whether our manipulation offers an improvement. Instead of having participants passively consume directly on Facebook, previous research has often used screenshots of posts as a proxy. However, the extent to which this substitution is sound has not been assessed before, thus previous results might have been artifacts of the experimental manipulation. We examine this by using different mock-ups of Facebook profiles in the first study, and comparing them to actual passive consumption on the platform. In particular, we want to measure the change in feelings of interpersonal attraction after passive consumption of either a screenshot, an artificial *in situ* profile, or an acquaintance's real Facebook profile.

After exploring the effect of different Facebook mock-ups, we aim to investigate potential variables that might moderate the effects of passive consumption on feelings of interpersonal attraction in our second study. This study will allow us to integrate our work into the existing theoretical literature. Since social media research is a relatively new field, we use traditional psychological theories of relationship formation and motivation to form hypotheses about variables moderating the link between interpersonal attraction before and after passive consumption. We then use these moderators in our models, examining how they operate in a social media context. Doing so will provide clues about what theories might help researchers understand aspects of online sociality and relationship formation over social media in the future. Using a longitudinal design, we furthermore try to detect long-term changes of interpersonal attraction to add up to existing literature focusing mainly on short-term and cross-sectional effects of passive consumption.

Study 1: Comparing the Effect of Different Facebook Profile Mock-Ups

Many previous studies in the field of social media research have used screenshots of real or artificial posts (e.g., Bazarova,

2012; Lin and Utz, 2015; Rains and Brunner, 2015; Orben and Dunbar, 2017) to investigate cross-sectional changes in interpersonal attraction after passive consumption, despite this method not being formally validated. Our protocol aims to examine this widely used methodology using *in situ* Facebook profiles: fake profiles on the actual Facebook platform. We compare the effects of these *in situ* profiles on interpersonal attraction to the effects of screenshots and real Facebook profiles.

We further want to examine how the passive consumption of artificial profiles (screenshots or *in situ* profiles) affects the consumer's interpersonal attraction toward the profile owner in comparison to passive consumption of real profiles. This will, on the one hand, allow us to test whether our *in situ* profile manipulation is valid, while also providing a valuable investigation into a commonly used method in the field.

Perceived physical attractiveness of the profile owner could influence the effects we are studying. For example, physically attractive people are liked more by others and, due to the halo effect, other characteristics like their personality or behavior are also rated more positively (Berscheid and Walster, 1974). The elevation in social attractiveness is higher for users that are perceived as more attractive.

However, we also predict that the effects of different profile mock-ups are dependent on the type of the outcome variable measured. For example, feelings of attraction take longer to develop and it is therefore possible that interpersonal attraction will not change after a single browsing session used in our study. On the other hand, certain processes, such as the formation of interpersonal evaluations, may unfold more quickly and therefore be impacted by the validity of manipulation by different profile mock-ups.

RQ1: Can screenshots or artificial *in situ* profiles be used as a proxy for real Facebook use?

As a similar study has not been previously completed, we do not have any strong predictions. However, as past research has heavily relied on screenshots, we tentatively suggest that there will be no differences in the increase in our dependent variable across all three conditions (screenshots, *in situ* profile and real profile). Thus, our hypotheses are that both screenshots and *in situ* profiles are not inferior to real profiles.

- H1: (a) There will be no difference between post- and pre-test scores in interpersonal attractiveness and interpersonal evaluation for participants who view Facebook posts of an actual Facebook friend's profile compared to participants who view screenshots of posts (Hypothesis 1a).
 (b) There will be no difference between post- and pre-test scores in interpersonal attractiveness and interpersonal evaluation for participants who view Facebook posts of an actual Facebook friend's profile compared to participants who view posts of an artificial profile (Hypothesis 1b).

We further expect an increase in interpersonal attractiveness and interpersonal evaluation for all conditions after passively consuming Facebook posts.

Study 2: Investigating the Change in Interpersonal Attraction After Passive Consumption and the Potential Moderators of This Process

Many previous studies in the field of social media research have focused on short-term cross-sectional changes in interpersonal attraction (e.g., Bazarova, 2012; Lin and Utz, 2015; Rains and Brunner, 2015). We aim to extend this research by investigating potential moderators and by using a longitudinal study design. By setting up three real Facebook profiles that our study participants will add, and by posting on two of them on a regular basis, we aim to mimic real-life Facebook passive consumption in an experimental setting. One profile will not post during the study period, and thus serve as a control condition. This control condition will help us track changes back to the passive consumption of posts: the experimental manipulation.

Based on the previous research of passive consumption we hypothesize the following:

- H2: Feelings of interpersonal attraction toward the Facebook profile owner will increase in the two experimental conditions (updated Facebook profiles), but not in the control condition (Facebook profiles without any new updates).

Generally, disclosure of personal information can help or hinder relationship development (Orben and Dunbar, 2017). This diversity in effects makes it important to examine which variables moderate how disclosure of personal information, either offline or online, affects interpersonal attraction. Several psychological theories have addressed this in offline interaction, examining how relationships are formed through face-to-face self-disclosure. While the theories were conceptualized to explain offline behavior, we believe that they could harbor important insights into how interpersonal attraction evolves after passive consumption.

In particular, theorists have highlighted many different factors surrounding self-disclosures that might facilitate the formation of a social bond. We can re-appropriate these factors and look at them in a social media context to examine the effects of passive consumption. Thereby, we find three theoretical approaches to be most helpful: Social Penetration Theory, Uncertainty Reduction Theory, and Homophily.

Social Penetration Theory highlights the importance of activities which involve intimate informational disclosures about oneself as the driving force of relationship development (Altman and Taylor, 1973). It suggests that relationships become more intimate over time as the members of a dyad intentionally disclose information about themselves. This process is gradual and starts with revealing only the superficial information about oneself. By continuously sharing more and more intimate information, however, individuals involved in dyadic communication develop closer relationships. As people like those who reveal more about themselves, liking increases accordingly (Altman and Taylor, 1973). Hereby, the breadth and depth of the disclosed information are crucial. The dimension is reflected by the intimacy of disclosures, whereas the breadth refers to the quantity of disclosures, reflected by the variety of topics and the frequency

of messages disclosed. According to Social Penetration Theory, disclosure breadth and depth covary, because disclosing on a higher frequency and on a wide variety of topics typically reveals personal information.

Social Penetration Theory emphasizes the need for self-disclosures to be perceived as intimate and to cover a variety of topics for a social relationship to form. Based on this theory, one would predict that higher perceived intimacy and a greater perceived variety of Facebook posts, as well as a greater perceived frequency of posts, would lead to a higher elevation of interpersonal attraction over time.

RQ2: Does the perceived intimacy of posts moderate the change in interpersonal attractiveness after passive consumption?

RQ3: Does the perceived frequency of posts moderate the change in interpersonal attractiveness after passive consumption?

RQ4: Does the perceived variety of posts moderate the change in interpersonal attractiveness after passive consumption?

Uncertainty Reduction Theory (Berger and Calabrese, 1975) contributes another approach to explaining interpersonal attraction. It suggests that a reduction of uncertainty about another person supports the development of interpersonal attraction. To reduce uncertainty, individuals engage in activities that provide information. They can use strategies that are passive (i.e., observing others' behavior), active (i.e., proactive information-seeking effort), or interactive (i.e., direct communication) (Berger and Calabrese, 1975). If a piece of information increases attributional confidence, that is, if people feel like they know each other and can predict each other's behaviors, it should, therefore, increase feelings of connection (Antheunis et al., 2010). The theory, therefore, predicts that the more uncertainty is reduced by passive consumption, the stronger is the change in interpersonal attraction over time. Hereby, passive consumption serves as a form of passive behavior observation reducing uncertainty about the user.

RQ5: Does attributional confidence moderate the change in interpersonal attractiveness after passive consumption?

Additionally, homophily, or the perceived similarity with another person, may elicit liking and develop relationship formation (Carrarini and Mengel, 2016). Shared interests, values, sociodemographic dimensions and other features of similarity impact how attracted a person feels to someone (Montoya et al., 2008). This is related to shared social identity, which makes people who are a part of one's social group more favorable (McPherson et al., 2001). This in-group bias refers to "the systematic tendency to evaluate one's own membership group (the in-group) or its members more favorably than a non-membership group (the out-group) or its members" (Hewstone et al., 2002, p. 576). Thus, if users feel like they share the profile owner's interests or views after browsing through their Facebook posts, they might evaluate these users more favorably compared to the ones who they do not have any information about, and thus express more elevated levels of interpersonal attraction toward the profile owners. In contrast to Social Penetration Theory, which highlights the importance of intimate disclosure of a wide range of topics, and Uncertainty Reduction Theory emphasizing the need to reduce insecurity about the person disclosing

information, the perceived similarity with the person disclosing information may be another crucial element for a relationship to form.

RQ6: Does homophily moderate the change in interpersonal attractiveness after passive consumption?

By examining the outlined moderators, which we derive from traditional psychological theory, we hope to contribute to a greater understanding of the underlying processes of relationship formation and development on social media. We are aware that there might be other variables of interest changing the level of interpersonal attractiveness, augmenting our moderators, and control variables. However, by selecting the most relevant theories in the field, we hope to cover the most important factors influencing interpersonal attractiveness, and to test whether traditional theories of relationship formation are applicable to online relationship formation.

METHODS

Study 1: Examining the Effect in Real Life Participants

Study 1 will take place in a lab. Participants will be recruited via a departmental mailing list, and participants need to be between 18 and 50 years old, proficient English speakers, and Facebook users. If the lab is in a different country from the United Kingdom or United States, we will translate the introduction and conclusion of the study, and we will recruit proficient English speakers so the rest of the study should proceed as normal in English. The participants will be paid a cash reimbursement for their participation.

Procedure

After entering the lab, participants are assigned to one of three conditions: the real Facebook profile, fake profile, or profile screenshots condition. In the study introduction, participants will be led to believe that the study examines their general attitudes toward Facebook. As we use deceptive elements in our study, we will ensure that the participants are debriefed extensively at the end of the study.

At the beginning of the study, participants will be presented with an information sheet and consent form. Participants will be told that they can withdraw from the study at any time. After providing basic demographic information, participants fill out a distractor questionnaire concerning their opinions about privacy, sharing news stories, and promoting hate speech on Facebook, in order to obfuscate the real purpose of the study and avoid biasing effects. Afterwards, participants will be asked to log into their Facebook account. Subsequently, participants will be presented with the top half of a Facebook profile in the artificial and real condition, or with a screenshot of the top half of a profile in the screenshot condition. The target's Facebook profile will show a woman in her early 1920s along with basic demographic information, such as her alma mater, hometown, and employment. The profile will also contain a profile picture and a cover photo. Participants then rate how attracted they are to the person and evaluate the interpersonal characteristics of the

person (for more information on measurements, see the section “Measures”).

Participants then browse through the first 10 posts of the respective Facebook profile (or look at 10 screenshots of Facebook posts). After scrolling through the posts/screenshots, the participants will again complete the measure of interpersonal social attraction, interpersonal evaluation, and the distractor measure. Subsequently, they will complete further measures about homophily, attributional confidence, perceived intimacy, and perceived valence of the posts, as well as physical attractiveness of the profile owner (McCroskey et al., 1975; Clatterbuck, 1979; Watson et al., 1988).

Artificial Facebook profile or screenshot condition

In the artificial Facebook profile condition or the screenshot condition, we use a selection of posts that were pretested to make sure they are perceived as highly appropriate and credible in order to avoid confounding effects. We additionally measured the posts’ perceived valence and intimacy to use these as control variables, i.e., to be able to track back potential changes to these variables if the effect is not merely caused by the change in interpersonal attraction. Generally, we ensure that our posts do not just include written text, but a variety of post types (e.g., photos, photos with text, written text, shared article, and video) in order to make the profile appear as real as possible. In the artificial Facebook profile condition, we combine these posts into a real Facebook profile on the website, whereas in the screenshot condition, we take screenshots of the posts and present them to participants one-by-one.

Real Facebook profile

In the real profile condition, participants will be asked to look through their friend list and choose a person whose profile they have not looked at during the last 2 months, who they feel like they do not know very well, and who they do not have any strong feelings about. Participants that do not have such a person in their friend’s list will be excluded from the study. Participants choosing a close friend might score higher on ratings of interpersonal attraction after passive consumption and thus confound our results. Our instructions and exclusion criteria should therefore diminish potential confounding effects that could arise from previous feelings of connectedness toward the profile owner.

The research assistant will ask the participants who the person they selected is and take notes about the relationship between the participants and the user. The participant is then asked to scroll down the profile looking at the first 10 posts only. As the fake Facebook profile includes 10 publicly visible posts, this will enable us to compare the subsequent measures between the two conditions.

Measures

Interpersonal Attractiveness Questionnaire (IAQ)

We will use the Interpersonal Attractiveness Questionnaire (IAQ) (Montoya and Horton, 2004) to measure participants’ perceived attractiveness of the profile owner. The questionnaire consists of nine items with the response format being a seven-point Likert-type scale, ranging from “I strongly disagree” to “I strongly agree.” By re-analyzing previous data, we found that the

IAQ possesses favorable psychometric properties. A single-factor confirmatory factor analysis model has been fit to the data using the lavaan package (Rosseel, 2012) for R (R Core Team, 2017). Most fit indicators (CFI = 0.98, TLI = 0.97, PCLOSE = 0.08, normed $\chi^2 = 3.31$, SRMR = 0.02) suggest that the model achieves close fit. To assess reliability, tau-congeneric, tau-equivalent, and parallel versions of the model were calculated (Cho, 2016). Comparing the tau-congeneric and the tau-equivalent models using the χ^2 difference test showed that the tau-congeneric model achieves better fit ($\chi^2_{diff} = 84.9$, $df_{diff} = 8$, $p < 0.001$). Reliability was calculated using the procedure for tau-congeneric models proposed by Raykov (1997). This analysis yielded a high reliability of 0.92.

Interpersonal evaluation inventory

To examine how participants evaluate social characteristics of the profile owner, which could evolve more rapidly than the feelings of attraction, we will use the interpersonal evaluation inventory (Kelly et al., 1980). The questionnaire consists of 24 adjectives (e.g., considerate, educated, and honest) and two question-like phrases describing or related to a person (in this case, the profile owner). The participant rates the person on a seven-point Likert-type scale adjusted to each adjective (e.g., “Extremely inconsiderate” to “Extremely considerate”). The scoring direction is determined randomly for each adjective. Previous research indicates that the Inventory is valid with four underlying factors. However, for the purpose of our study we have decided to only use the items from the Likeability subscale. Likeability subscale is the most relevant subscale for this study content wise and with its 14 items it composes the majority of the inventory. Using unifactorial measures will greatly simplify our analyses, with negligible information loss.

Proactive attributional confidence scale (PACS)

We will use the proactive attributional confidence scale (PACS) (Clatterbuck, 1979) to measure attributional confidence. The scale consists of seven items, with the response format being a six-point Likert-type scale ranging from “Not confident at all” to “Very confident.”

Attitude homophily scale (AHS)

The attitude homophily scale (AHS) (McCroskey et al., 1975) will be used to measure homophily. The scale consists of four items. Each item consists of two opposite claims (e.g., “Doesn’t think like me” and “Thinks like me”). These two claims are anchor points of a seven-point Likert-type scale which is used to indicate a participant’s response.

Distractor measures

To conceal the true purpose of the study, which could become evident if the participants are only completing social attraction/evaluation measures at two close timepoints, we will use a distractor measure to lead the participants to believe that the aim of the study is to measure their attitudes on Facebook. To achieve this, we designed a questionnaire with good face validity containing items such as “Facebook makes it easier to keep in touch with other people,” “The way Facebook is set up is a threat to privacy,” “Some Facebook users incite violence

toward minorities using the network.” We believe that attitudes toward Facebook are a good distractor measure because the central activity of the experimental procedure is Facebook use. The content of the items was chosen with caution not to prime the participant to form a certain attitude toward the profile owner (e.g., we did not use any items related to Facebook addiction).

Control variables

As additional measures, we also include perceived intimacy, perceived valence, and perceived physical attractiveness. To assess the perceived intimacy of posts, participants will be asked to rate, how intimate the posts made by each of the profile owners were on a seven-point Likert scale from 1 (not at all intimate) to 7 (very intimate). Participants will be asked to rate the general valence on a seven-point Likert scale from 1 (very negative) to 7 (very positive). Perceived physical attractiveness of the profile owner will be assessed by a single item question “How physically attractive do you find this person?” by rating from 1 (not attractive at all) to 7 (very attractive). Data on perceived intimacy and perceived valence will be collected only after the participant has scrolled through or saw screenshots of 10 Facebook posts.

Analysis

Statistical analysis will be carried out using the R programming language (R Core Team, 2017) and, together with the raw data, will be made publicly available on the Open Science Framework.

Our goal in Study 1 is to test whether artificial Facebook profiles and screenshots can be used as a proxy for real Facebook profiles. If so, we expect that there is no difference in the mean change on the dependent variable across these conditions. To test this, we employ a multiple-group latent difference model (McArdle, 2009, pp. 584–585; Allison, 1990) using the change in the mean interpersonal attractiveness scores and in the mean interpersonal evaluation scores as dependent variable, respectively. Since we cannot experimentally control for the content and type of the post in the real Facebook condition, we use perceived intimacy of posts, perceived valence of posts, and perceived physical attractiveness of the profile owner as control variables.

Study 2: Investigating the Change in Interpersonal Attraction

Participants

We aim to recruit about 400 participants using a convenience sample over social media and through mailing lists. For inclusion in the study, participants need to be between 18 and 50 years old, proficient English speakers in order to understand study instructions and Facebook profile content, and Facebook users. Participants will get a chance to win a raffle prize upon completion of Study 2 for reimbursement. We will endeavor to collect all participants of Study 2 within 2 weeks.

Procedures

First, participants will receive an information sheet and a consent form via email. If they consent to participate in our study,

participants will be asked to indicate their demographic data (age, gender, race, and location) and to log into their Facebook accounts. By following a link, they will be instructed to take a look at a profile of “another research participant” for 1 min. Participants will be able to see only the basic information about the profile owner – name, profile photo, and cover photo. Subsequently, participants will be asked to return to the survey platform to fill in questionnaires which measure interpersonal attraction toward the profile owner and perceived physical attractiveness of the profile owner.

The same procedure as outlined above will be repeated with two other Facebook profiles of “other research participants.” Participants will be instructed to be friend the Facebook profiles and will be informed that within 2 weeks, the three other research participants whose posts they saw will be instructed to accept the friend request. They will be asked to mark the new Facebook friends’ profiles as “see first,” which are an option on Facebook that ensures participants see all of the profiles’ posts in their News feeds, by putting these posts on top of all others. This will ensure that all participants are exposed to the posts, and that we can attribute changes in interpersonal attraction to the passive consumption. Participants will be able to see the new Facebook friends’ profiles and posts; however, they are not allowed to directly interact with them by messaging, commenting, or liking their posts.

The Facebook profiles will be set up prior to the study with the aim to make them seem realistic. To ensure this, we will remove the date our profile was created from the information box on the set-up Facebook profile. All three profiles will have 10 posts in their history. However, only two of the profiles will add status updates throughout the study period, i.e., be the experimental condition profiles. One profile will not add any new posts during the study period, i.e., be the control condition profile.

Participants will be sent a questionnaire after 2 weeks (Wave 2), 4 weeks (Wave 3), and a final questionnaire after 6 weeks (Wave 4) via email. In Wave 2, Wave 3, and Wave 4, they will be asked to fill in the interpersonal attraction measure and measures of control variables that will include, aside from (1) perceived physical attractiveness of the profile owner, (2) perceived intimacy of posts, (3) perceived frequency of posts, (4) perceived variety of post topics, (5) attributional confidence, (6) homophily, and (7) perceived valence of the posts. Next to the questions about perceived intimacy and valence of the posts, there will be the option “this person did not post during the last 2 weeks” that participants can select. At the end of Wave 4, after completion of the questionnaires, participants will be asked whether they remember three profile posts, only two of which were actually posted. This will serve as a manipulation check to test whether participants were exposed to the profile’s posts. Finally, participants will be asked to rate how credible the profile was overall, and will be debriefed about the real aim of the study.

Measures

Study 2 will include the same measures used in Study 1 (described in Section Measures), but not all of the measures

will be used. Specifically, we will use the IAQ, PACS, AHS and measures of perceived intimacy, perceived valence, perceived physical attraction, and perceived frequency of posts. Participants will rate the perceived frequency of posts on a seven-point Likert scale from 1 (posts very rarely) to 7 (posts very often).

Analysis

Our goal in Study 2 is to explore whether theories of offline communication can inform online communication, particularly passive consumption. Concretely, we aim to test whether (1) perceived intimacy of posts, (2) perceived frequency of posts, (3) perceived variety of post topics, (4) attributional confidence, and (5) homophily influence the change in interpersonal attractiveness over time (see RQ2-6).

Our analysis uses structural equation modeling and proceeds as follows. As a prerequisite, we will first test whether the factor structure of interpersonal attractiveness is invariant across time; that is, we will test for strong measurement invariance (Little et al., 2007, pp. 358–359). First, to see whether our intervention, that is, participants passively consuming posts, was successful, we will compare the change in interpersonal attractiveness across the three conditions (see above) using a multiple-group latent growth-curve model (e.g., Newsom, 2015, chapter 7). We will also test whether the change can be adequately captured by a linear function. Second, to test whether the variables derived from theory influence the change in interpersonal attractiveness, we will include them as time-invariant covariates, controlling for perceived physical attractiveness of the profile owner, and perceived valence of the posts.

LIMITATIONS AND IMPACT

As we will implement a complex procedure in a relatively new field, our study has multiple limitations. However, we strongly believe that the insights acquired by running this study will outweigh these limitations. Our work will allow researchers to reassess traditional psychological theory and will provide a stepping stone in the use of more complex longitudinal and realistic methodology in social media research.

General Limitations

Our sampling strategy might be criticized as participants will be recruited through social media or mailing lists and therefore might not be representative of the population. While true, we do not see an issue with this. As our study requires participants to be (frequent) social media users, our sample cannot be fully representative by intent and we can generalize our results only to social media users.

Participants may also guess the purpose of our study or they might know that the artificial profiles are fake. In Study 1, this could mean that using fake profiles is not a solid substitute for simulating real life Facebook use. However, in Study 2, this would be a cause of concern as we cannot generalize attraction developed toward a non-existent person

to attraction developed toward a real person. Therefore, at the end of both studies, we will include a question about what the purpose of the study was, to find out whether any of the participants guessed the real aims of the experiment. In case they did, we will exclude these participants from the study.

Limitations Study 1

One of the main limitations of the design of Study 1 is the lab setting. The lab environment is artificial and may lead to behavior that would not occur in a natural setting. The laboratory will, however, allow us to control that the participants focus solely on the study and do not digress to other websites. The artificial setting of Study 1 is not just limited to the laboratory setting, but also includes the task of scrolling through Facebook profiles themselves. Being asked to look through someone else's profile for a set period of time might provoke a different kind of reaction than the one that is experienced while browsing through profiles voluntarily. Participants might, for example, speculate whether they chose the right person from their list of friends, which could decrease their focus. Therefore, we have to be careful when generalizing our results.

Another limitation is linked to the selection of the real Facebook profile. As the chosen profile owner is already acquainted with the participant, confounding variables due to previous feelings, or certain attitudes toward this particular person might impact our study in an unpredictable manner. We, however, try to control for these variables by telling the participants to choose a person whose profile they have not looked at in the last 2 months, who they feel like they do not know very well, and who they do not have any strong feelings about. We need to use a real person in order to be able to compare real Facebook consumption to simulated Facebook consumption used in social media research, and we therefore need to take these limitations into account.

In the screenshot conditions, participants do not scroll through the posts but instead view them one by one. While unlikely, it could be that scrolling affects the interpersonal attractiveness judgments of the participants. However, as previous studies also did not include scrolling in the screenshot conditions, we decided against it to be as close to previous studies as possible.

Limitations Study 2

While developing the design of longitudinal studies, choosing the optimal lag between time points is critical (Gollob and Reichardt, 1987; Little, 2013). In this respect, there is a potential limitation of our study. First, our proposed time period might be an overestimate, meaning that interpersonal attractiveness might change more rapidly. However, it could also be an underestimate as the change in interpersonal attractiveness might be a long process which cannot be tracked during the span of our study. Overall, choosing the lag is difficult, especially in areas without precise theoretical guidance. We thus acknowledge this limitation fully.

With longitudinal studies, the loss to follow-up is a known threat. While we recognize this risk, we hope to reduce it by running a fairly short longitudinal study (6 weeks). We will also send reminders to fill out the questionnaires in each wave.

Another general limitation is that all study participants befriend three unbeknownst to them, young, Caucasian, female students, and receive their status updates. We can therefore, strictly speaking, generalize our conclusions only to these particular profiles; our stimulus sample size is $n = 3$ (Wells and Windschitl, 1999). This befriending of unknown (fake) profiles is also different from the typical Facebook dynamic, where users typically know persons they add to their friend list, have met them before or have mutual friends. While this might at first appear as a limitation, we will design the profile to be representative of a prototypical Facebook user, therefore our ability to generalize will not be limited severely. Previous studies in this area have used artificial profiles successfully and we believe that our improvement in methodology (conducting the study on Facebook) adds greatly to the existing literature. Furthermore, because we test whether our employed measures are invariant across the fake profile and the real profiles, we can, to some extent, empirically assess this generalizability problem.

Impact

By completing the studies set forth in this proposal, we expect to shed further light on the motivational background and outcomes of passive consumption. The trend toward spending time online passively consuming has developed for several years, yet it has not been extensively researched. To gain further understanding about this process, we aim to utilize traditional social psychological theory. Can classic theories of relationship formation be applied to our online use of social media? If so, they can help us explain the effects in more detail and give us further insight into whether interpersonal relationships can be formed solely by passively consuming someone's online information. By examining variables derived from traditional psychological theory, we can also test what might be moderating effects of passive consumption and whether these moderators are similar to those influencing face-to-face communication. This will allow us to highlight theories that could be utilized by social media researchers in the future.

Our protocol also contributes to the field methodologically. Using a longitudinal design with *in situ* artificial profiles that will be active throughout the duration of study instead of screenshots of posts and profiles, we take a novel and innovative step toward highlighting drivers and outcomes of passive consumption. Furthermore, our study is one of the first to examine whether the artificial passive consumption paradigm (participants looking at a stranger's profile) is a good simulation of actual passive consumption. It combines both real and fake Facebook profiles as well as screenshots of posts that have been used in social media research up until now, to inspect whether we can apply previous studies which used an artificial environment (e.g., Deters

and Mehl, 2013; Orben and Dunbar, 2017) to explain real-life Facebook use. It is not known whether these widely used results are solely an artifact of the research paradigm because comparable validity checks have not been conducted yet. By conducting our research, we can, therefore, possibly increase the validity of methodological approaches in future studies of social networks. Thus, our proposed studies are an important addition to social media research in multiple different ways.

CONCLUSION

With social media constantly growing in popularity and becoming indispensable in the lives of the majority of people in developed countries, it is important to study the drivers and impacts of its use. In this study, we focus on passive consumption on social media as it has overtaken active use in overall popularity over the past years (Gerson et al., 2017). The reasons behind this trend, and its impact on social life, have not been thoroughly researched. Previously, passive use has been described as lacking the reciprocity and depth to aid form social connections. However, there has been evidence that passive use might help us feel more connected which we aim to investigate further. Using two studies, we want to take a critical look at the methodology used in the social media research area, by comparing previously used screenshots of fake profiles to actual fake profiles on Facebook and to real profiles, while also examining possible changes in interpersonal attraction after passive consumption on Facebook, as well as possible factors influencing this change. By conducting this research, we hope to answer important questions regarding social media use and bring the research field closer to understanding these phenomena.

ETHICS STATEMENT

The study materials and consent forms have been developed in accordance with ethical norms and guidelines from all participating institutions. The study received approval by the Central University Research Ethics Committee (CUREC) at the University of Oxford.

AUTHOR CONTRIBUTIONS

AO conceived the idea for this project. AO, AM, FD, MH, JK, NV, and DK have made substantial intellectual contribution to this work. All authors gave the publishing approval.

ACKNOWLEDGMENTS

This project was conducted under the Junior Researcher Program.

REFERENCES

- Allison, P. D. (1990). Change scores as dependent variables in regression analysis. *Soci. Method.* 20, 93–114. doi: 10.2307/271083
- Altman, I., and Taylor, D. A. (1973). *Social Penetration: The Development of Interpersonal Relationships*. New York, NY: Holt, Rinehart and Winston.
- Antheunis, M. L., Valkenburg, P. M., and Peter, J. (2010). Getting acquainted through social network sites: testing a model of online uncertainty reduction and social attraction. *Comput. Hum. Behav.* 26, 100–109. doi: 10.1016/j.chb.2009.07.005
- Ballantine, P. W., and Stephenson, R. J. (2011). Help me, I'm fat! Social support in online weight loss networks. *J. Consum. Behav.* 10, 332–337. doi: 10.1002/cb.374
- Bazarova, N. (2012). Public intimacy: disclosure interpretation and social judgments on facebook. *J. Commun.* 62, 815–832. doi: 10.1111/j.1460-2466.2012.01664.x
- Berger, C. R., and Calabrese, R. J. (1975). Some explorations in initial interaction and beyond: toward a developmental theory of interpersonal communication. *Hum. Commun. Res.* 1, 99–112. doi: 10.1111/j.1468-2958.1975.tb00258.x
- Berscheid, E., and Walster, E. (1974). Physical Attractiveness. *Adv. Exp. Soc. Psychol.* 7, 157–215. doi: 10.1016/S0065-2601(08)60037-4
- Brandtzæg, P. B. (2012). Social networking sites: their users and social implications - a longitudinal study. *J. Comput. Mediat. Commun.* 17, 467–488. doi: 10.1111/j.1083-6101.2012.01580.x
- Burke, M., Kraut, R., and Marlow, C. (2011). "Social capital on Facebook: differentiating uses and users," in *CHI 2011 Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, Vancouver, BC, 571–580. doi: 10.1145/1978942.1979023
- Burke, M., Marlow, C., and Lento, T. (2010). "Social network activity and social well-being," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, New York, NY, 1909–1912. doi: 10.1145/1753326.1753613
- Cho, E. (2016). Making reliability reliable: a systematic approach to reliability coefficients. *Organ. Res. Methods* 19, 651–682. doi: 10.1177/1094428116656239
- Clatterbuck, G. W. (1979). Attributional confidence and uncertainty in initial interaction. *Hum. Commun. Res.* 5, 147–157. doi: 10.1111/j.1468-2958.1979.tb00630.x
- Curranini, S., and Mengel, F. (2016). Identity, homophily and in-group bias. *Eur. Econ. Rev.* 90, 40–55. doi: 10.1016/j.eurocorev.2016.02.015
- Deters, F. G., and Mehl, M. R. (2013). Does posting facebook status updates increase or decrease loneliness? an online social networking experiment. *Soc. Psychol. Pers. Sci.* 4, 579–586. doi: 10.1177/1948550612469233
- Frost, R. L., and Rickwood, D. J. (2017). A systematic review of the mental health outcomes associated with Facebook use. *Comput. Hum. Behav.* 76, 576–600. doi: 10.1016/j.chb.2017.08.001
- Gerson, J., Plagnol, A. C., and Corr, P. J. (2017). Passive and active facebook use measure (PAUM): validation and relationship to the reinforcement sensitivity theory. *Pers. Individ. Diff.* 117, 81–90. doi: 10.1016/j.paid.2017.05.034
- Gollob, H. F., and Reichardt, C. S. (1987). Taking account of time lags in causal models. *Child Dev.* 58, 80–92. doi: 10.2307/1130293
- Good, A., Sambhantham, A., and Panjanj, V. (2013). "Looking back at facebook content and the positive impact upon wellbeing: exploring reminiscing as a tool for self soothing BT - online communities and social computing," in *Proceedings of the 5th International Conference, OCSC 2013, Held as Part of HCI International 2013, Las*, eds A. A. Ozok and P. Zaphiris (Berlin: Springer), 278–286.
- Hewstone, M., Rubin, M., and Willis, H. (2002). Intergroup Bias. *Annu. Rev. Psychol.* 53, 575–604. doi: 10.1146/annurev.psych.53.100901.135109
- Kelly, J. A., Kern, J. M., Kirkley, B. G., Patterson, J. N., and Keane, T. M. (1980). Reactions to assertive versus unassertive behavior: differential effects for males and females and implications for assertiveness training. *Behav. Ther.* 11, 670–682. doi: 10.1016/S0005-7894(80)80066-2
- Krasnova, H., Wenninger, H., Widjaja, T., and Buxmann, P. (2013). *Envy on Facebook: A Hidden Threat to Users' Life Satisfaction?*. Available at: <https://ktwop.com/2013/01/22/facebook-envy-is-a-hidden-threat-to-life-satisfaction/>
- Lewis, J., and West, A. (2009). "Friending": London-based undergraduates' experience of Facebook. *New Med. Soc.* 11, 1209–1229. doi: 10.1177/1461444809342058
- Lin, R., and Utz, S. (2015). The emotional responses of browsing Facebook: happiness, envy, and the role of tie strength. *Comput. Hum. Behav.* 52, 29–38. doi: 10.1016/J.CHB.2015.04.064
- Little, T. D. (2013). *Longitudinal Structural Equation Modeling*. New York, NY: Guilford Press.
- Little, T. D., Preacher, K. J., Selig, J. P., and Card, N. A. (2007). New developments in latent variable panel analyses of longitudinal data. *Int. J. Behav. Dev.* 31, 357–365. doi: 10.1177/0165025407077757
- McArdle, J. J. (2009). Latent variable modeling of differences and changes with longitudinal data. *Annu. Rev. Psychol.* 60, 577–605. doi: 10.1146/annurev.psych.60.110707.163612
- McCroskey, J. C., Richmond, V. P., and Daly, J. A. (1975). The development of a measure of perceived homophily in interpersonal communication. *Hum. Commun. Res.* 1, 323–332. doi: 10.1111/j.1468-2958.1975.tb00281.x
- McGrath, F. (2015). *4 in 10 Facebookers Now Browsing the Site Passively*. Available at: <http://blog.globalwebindex.net/chart-of-the-day/4-in-10-facebookers-now-browsing-the-site-passively/>
- McPherson, M., Smith-Lovin, L., and Cook, J. M. (2001). Birds of a feather: homophily in social networks. *Annu. Rev. Sociol.* 27, 415–444. doi: 10.1146/annurev.soc.27.1.415
- Montoya, R. M., and Horton, R. S. (2004). On the importance of cognitive evaluation as a determinant of interpersonal attraction. *J. Pers. Soc. Psychol.* 86, 696–712. doi: 10.1037/0022-3514.86.5.696
- Montoya, R. M., Horton, R. S., and Kirchner, J. (2008). Is actual similarity necessary for attraction? A meta-analysis of actual and perceived similarity. *J. Soc. Pers. Relationsh.* 25, 889–922. doi: 10.1177/0265407508096700
- Newsom, J. T. (2015). *Longitudinal Structural Equation Modeling: A Comprehensive Introduction*. Abingdon: Routledge.
- Nowak, M., and Guillermo, S. (2017). *Two Billion People Coming Together on Facebook*. Available at: <https://newsroom.fb.com/news/2017/06/two-billion-people-coming-together-on-facebook/>
- Orben, A. C., and Dunbar, R. I. M. (2017). Social media and relationship development: the effect of valence and intimacy of posts. *Comput. Hum. Behav.* 73, 489–498. doi: 10.1016/j.chb.2017.04.006
- Rains, S. A., and Brunner, S. R. (2015). The outcomes of broadcasting self-disclosure using new communication technologies: responses to disclosure vary across one's social network. *Commun. Res.* 45, 1–29. doi: 10.1177/0093650215598836
- Raykov, T. (1997). Estimation of composite reliability for congeneric measures. *Appl. Psychol. Meas.* 21, 173–184. doi: 10.1177/01466216970212006
- R Core Team (2017). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.
- Rossee, Y. (2012). lavaan: an r package for structural equation modeling. *J. Stat. Softw.* 48, 1–36. doi: 10.18637/jss.v048.i02
- Schlosser, A. E. (2005). Posting versus lurking: communicating in a multiple audience context. *J. Consum. Res.* 32, 260–265. doi: 10.1086/432235
- Steers, M.-L. (2015). "It's complicated": facebook's relationship with the need to belong and depression. *Curr. Opin. Psychol.* 9, 22–26. doi: 10.1016/j.copsyc.2015.10.007
- Utz, S. (2015). The function of self-disclosure on social network sites: not only intimate, but also positive and entertaining self-disclosures increase the feeling of connection. *Comput. Hum. Behav.* 45, 1–10. doi: 10.1016/j.chb.2014.11.076
- Verduyn, P., Ybarra, O., Leuven, K., Jonides, J., and Kross, E. (2017). Do social network sites enhance or undermine subjective well-being? a critical review maxime rébois do social network sites enhance or undermine subjective well-being? *Soc. Issues Policy Rev.* 11, 274–302. doi: 10.1111/sipr.12033

- Watson, D., Clark, L. A., and Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: the PANAS scales. *J. Pers. Soc. Psychol.* 54, 1063–1070. doi: 10.1037/0022-3514.54.6.1063
- Wells, G. L., and Windschitl, P. D. (1999). Stimulus sampling and social psychological experimentation. *Pers. Soc. Psychol. Bull.* 25, 1115–1125. doi: 10.1177/01461672992512005

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The reviewer MEJK declared a shared affiliation, with no collaboration, with one of the authors, AO, to the handling Editor.

Copyright © 2018 Orben, Mutak, Dablander, Hecht, Krawiec, Valkovičová and Košič. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Internet Users' Valuation of Enhanced Data Protection on Social Media: Which Aspects of Privacy Are Worth the Most?

Jasmin Mahmoodi^{1*}, Jitka Čurdová², Christoph Henking³, Marvin Kunz⁴, Karla Matic⁵, Peter Mohr⁶ and Maja Vovko⁷

¹ Swiss Center for Affective Sciences, University of Geneva, Geneva, Switzerland, ² Department of Psychology, Masaryk University, Brno, Czechia, ³ Department of Psychological and Behavioural Science, London School of Economics and Political Science, London, United Kingdom, ⁴ Faculty of Social and Behavioral Science, University of Groningen, Groningen, Netherlands, ⁵ Department of Psychology, University of Leuven, Leuven, Belgium, ⁶ Department of Psychology, University of Amsterdam, Amsterdam, Netherlands, ⁷ Department of Psychology, University of Ljubljana, Ljubljana, Slovenia

OPEN ACCESS

Edited by:

Jin Eun Yoo,
Korea National University
of Education, South Korea

Reviewed by:

Pam Briggs,
Northumbria University,
United Kingdom
Meinald T. Thielsch,
Universität Münster, Germany

*Correspondence:

Jasmin Mahmoodi
j.mahmoodi@outlook.com

Specialty section:

This article was submitted to
Quantitative Psychology
and Measurement,
a section of the journal
Frontiers in Psychology

Received: 31 October 2017

Accepted: 31 July 2018

Published: 22 August 2018

Citation:

Mahmoodi J, Čurdová J, Henking C,
Kunz M, Matic K, Mohr P and
Vovko M (2018) Internet Users'
Valuation of Enhanced Data
Protection on Social Media: Which
Aspects of Privacy Are Worth
the Most? *Front. Psychol.* 9:1516.
doi: 10.3389/fpsyg.2018.01516

As the development of the Internet and social media has led to pervasive data collection and usage practices, consumers' privacy concerns have increasingly grown stronger. While previous research has investigated consumer valuation of personal data and privacy, only few studies have investigated valuation of different privacy aspects (e.g., third party sharing). Addressing this research gap in the literature, the present study explores Internet users' valuations of three different privacy aspects on a social networking service (i.e., Facebook), which are commonly captured in privacy policies (i.e., data collection, data control, and third party sharing). A total of 350 participants will be recruited for an experimental online study. The experimental design will consecutively contrast a conventional, free-of-charge version of Facebook with four hypothetical, privacy-enhanced premium versions of the same service. The privacy-enhanced premium versions will offer (1) restricted data collection on side of the company; (2) enhanced data control for users; and (3) no third party sharing, respectively. A fourth premium version offers full protection of all three privacy aspects. Participants' valuation of the privacy aspects captured in the premium versions will be quantified measuring willingness-to-pay. Additionally, a psychological test battery will be employed to examine the psychological mechanisms (e.g., privacy concerns, trust, and risk perceptions) underlying the valuation of privacy. Overall, this study will offer insights into valuation of different privacy aspects, thus providing valuable suggestions for economically sustainable privacy enhancements and alternative business models that are beneficial to consumers, businesses, practitioners, and policymakers, alike.

Keywords: information privacy, privacy concerns, willingness-to-pay, social networking services, Facebook, premium products, privacy dimensions

INTRODUCTION

The advent of the Internet and social media has drastically transformed all aspects of our lives; how we work, consume, and communicate (see also Stewart and Segars, 2002; Paine et al., 2007). While this has had considerable advantages for society overall, the growing influence of the Internet and technologies has always been linked to concerns for privacy and the collection and use of personal

information (e.g., Zuboff, 1988). The threats to individual privacy through these technologies have been repeatedly documented. Over the past years, sensitive personal data were repeatedly unlawfully obtained and mishandled in numerous data breaches. Most recently, sensitive personal information, including credit scores, of almost 150 million people was compromised in the 2017 Equifax data breach (e.g., Zou and Schaub, 2018) and around 87 million Facebook users were impacted by the Cambridge Analytica data scandal in 2018 (e.g., Revell, 2018).

While some consumers are unaware of the data they produce or of the full extent to which their data are mined and analyzed (e.g., Turow et al., 2005), others do not care (Garg et al., 2014). A majority of consumers, however, report concerns about their online privacy (e.g., Phelps et al., 2000; Pew Research Center, 2014), and, yet, most people often trade their personal data for online services and products (Carrascal et al., 2013). For instance, even privacy-concerned individuals join social networking services, such as Facebook, and share large amounts of personal information on these platforms (Acquisti and Gross, 2006).

Several factors play a role in explaining the discrepancy between people's concerns and their online sharing behaviors, such as bounded rationality, cognitive biases and heuristics, or social factors (see Kokolakis, 2017 for a review). One explanation is the so-called privacy calculus, which postulates that people perform a calculus of the costs (i.e., loss of privacy) and benefits (i.e., gain from information disclosure). Their final decisions and behaviors are a result of this calculus and determined by the outcome of this trade-off. When the perceived benefits outweigh the perceived costs, people are likely to disclose information (Culnan and Armstrong, 1999; Dinev and Hart, 2006b). Other factors accounting for this discrepancy are, for instance, that privacy functionalities are often not usable leaving users with little choice or alternatives and making it almost impossible for users to act upon their concerns (Iachello and Hong, 2007; Lipford et al., 2008). Experts call for better data and privacy regulations as well as alternative business models to balance the asymmetric relationship between consumers and business (e.g., Zuckerman, 2014; Tufekci, 2015; Gasser, 2016; New York Times, 2018; Quito, 2018). Understanding Internet users' privacy concerns and valuations is essential to develop strategies that match users' needs and enable them to act in accordance to their concerns.

The present research investigates Internet users' concerns and valuation of privacy in the context of the social networking service Facebook. In the experimental online study, participants will be presented premium versions of Facebook that offer different privacy enhancements (e.g., less data collection, more data control, and no third party sharing) for a monthly fee. Participants will be asked to indicate their willingness-to-pay for these privacy enhancements. In addition, psychological mechanisms underlying these valuations will be examined. In the following, the scientific literature underlying this research will be reviewed and the research hypotheses for this research will be developed. The experimental design and research methods will be outlined and the anticipated results presented and discussed.

THEORETICAL BACKGROUND

Privacy concerns have become one of the most central themes in the digital era, likewise for scholars, consumers, businesses, practitioners, and policy-makers. Acquisti and Gross (2009), for example, demonstrated the threat to individual privacy by inferring identities (i.e., social security numbers) through supposedly "anonymized" data. Other research showed that sensitive personal information, such as sexual orientation, could be inferred from Facebook Likes and facial images (Kosinski et al., 2013; Wang and Kosinski, 2018). Most recently, several data breaches, such as the Cambridge Analytica scandal that compromised personal data of about 87 million Facebook users worldwide (Revell, 2018), have sparked ethical debates on users' online privacy (e.g., Zunger, 2018).

Although not a novel concept, there is no clear consensus on the definition of privacy (Solove, 2006). Privacy is a complex, multidimensional construct that has been studied from different perspectives (Laufer and Wolfe, 1977) and, accordingly, has been operationalized in many different ways (e.g., as an attitude in Buchanan et al., 2007; as a value in Earp et al., 2005; Alashoor et al., 2015; as a behavior in Jensen et al., 2005; or as a right in McCloskey, 1980; Warren and Brandeis, 1890; see also Bélanger and Crossler, 2011 for a review). In order to tackle privacy in a standardized and reliable manner, most contemporary research concerned with online privacy uses the construct of privacy concerns as a proxy to explore information privacy (see Dinev et al., 2009; Smith et al., 2011). Hence, a control-centered definition of information privacy prevails, where privacy is defined as individual ability to control disclosure and use of personal information (Westin, 1968; Altman et al., 1974; Margulis, 1977). Accordingly, privacy concerns can be defined as consumers' perceptions of how the information they provide online will be used (Dinev and Hart, 2006a), and if this use can be regarded as 'fair' (Malhotra et al., 2004). Two widely accepted models of privacy exist that treat privacy concern as a multidimensional construct: The multidimensional instrument developed by Smith et al. (1996) assesses "individuals' concerns about organizational information privacy practices" (p. 167). This instrument has been adapted by Malhotra et al. (2004), making it applicable to the context of online privacy. The Internet User's Information Privacy Concerns (IUIPC) model consists of three dimensions, namely collection, control, and awareness. The dimension collection refers to users' concerns regarding the collection of their personal information. The dimension control refers to users' beliefs to have the right to determine and control how their information are collected, stored, and shared. The dimension awareness refers to users' awareness of data privacy practices of companies (i.e., online service providers).

Despite the importance of privacy in the digital era, people – paradoxically even those holding strong privacy concerns – often trade their personal data for online services and products (Carrascal et al., 2013). For example, Acquisti and Gross (2006) demonstrated that even privacy-concerned individuals join the social networking service Facebook disregarding its privacy policies and revealing large amounts of personal

information. The term “privacy paradox” has been coined to describe this dichotomy between expressed privacy concerns and actual online disclosure and sharing behaviors (Norberg et al., 2007). This paradox is particularly pronounced on social networking platforms, given the seemingly contradictory relationship between information privacy and social networking (i.e., connecting and sharing personal information with an online network; Lipford et al., 2012).

Many researchers have attempted to unravel and explain the privacy paradox (e.g., Barnes, 2006; Pötzsch, 2008; Sundar et al., 2013; Motiwalla and Li, 2016). One explanation defines the privacy paradox in terms of trade-offs between the benefits of using digital products and services and disclosing information online at the cost of a (partial) loss of privacy. These cost-benefit analyses are modeled as privacy calculus (Culnan and Armstrong, 1999), where privacy and personal information are conceptualized in economic terms as commodities (Klopfner and Rubenstein, 1977; Bennett et al., 1995). Willingness-to-pay is a commonly used indicator to quantify consumers’ economic valuation of commodities, such as goods and services (e.g., Casidy and Wymer, 2016; Lee and Heo, 2016). Accordingly, many scholars use willingness-to-pay as an indicator for economic valuations of privacy and information disclosure (e.g., Grossklags and Acquisti, 2007; Beresford et al., 2012; Spiekermann et al., 2012; Acquisti et al., 2013; Schreiner and Hess, 2015). Tsai et al. (2011) demonstrated that, when sufficient privacy information is available, people are willing to pay a premium to be able to purchase from websites that offer greater privacy protection. Studying low-priced products, the authors found that people were willing to pay up to 4% – around US\$0.60 – more for enhanced privacy. Egelman et al. (2009) showed that people are willing to pay up to US\$0.75 for increased privacy when online shopping, particularly when shopping for privacy-sensitive items. Similarly, a quarter of smartphone users were willing to pay a US\$1.50 premium to use a mobile app that made fewer requests to access users’ personal information (Egelman et al., 2013). In a study by Hann et al. (2007) among U.S. Americans, personal information was worth US\$30.49 – US\$44.62. In another study, participants expressed high sensitivity to and concern for privacy, but only half of the participants were actually willing to pay for a change in data protection laws that would give them property rights to their personal data. The economic value placed on these privacy rights averaged around US\$38 (Rose, 2005). Schreiner et al. (2013) tested privacy-enhanced premium versions of Facebook and Google and measured consumers’ propensity to pay for these services. The authors found that the optimal price for Facebook was €1.67/month and the optimal price for Google’s search engine lay between €1.00 and €1.50/month. Even though participants in the study were willing to pay for privacy-enhanced premium version, these valuations are relatively low (see also Bauer et al., 2012). Different explanations can account for the rather low valuations of privacy and data protection. For example, individuals who have not experienced invasion of their information privacy (e.g., through breaches or hacks) do not understand all the possible consequences resulting from information privacy violations and, therefore,

tend to undervalue privacy (Hann et al., 2002). It might also be because many costs associated with the invasion of privacy occur from secondary use of information (Laudon, 1996), of which the consequences are often only experienced *ex post* (Acquisti, 2004). What is more, not all the costs of unprotected personal information are easy to quantify – while some of the consequences are tangible (e.g., identity theft), others are intangible (e.g., revealing personal life history to strangers; Brandimarte et al., 2015). Hence, it seems likely that people value privacy aspects that are tangible and immediate more than others.

In addition to these factors, several psychological characteristics have been identified in explaining consumers’ concerns and valuation of privacy. A large body of the literature shows that cognitive biases and heuristics, such as comparative optimism, overconfidence, or affect bias play an important role (see Kokolakis, 2017 for a review). For example, low privacy valuations are associated with people’s underestimation of one’s own and overestimation of other’s likelihood of experiencing misuse of personal data (Syverson, 2003; Baek et al., 2014), which could translate into low privacy valuations. Valuation of online privacy has also been linked to perceptions of usefulness, risk, and trust toward companies or services (e.g., Malhotra et al., 2004; Milne and Culnan, 2004; Dinev and Hart, 2006a; Garg et al., 2014; Schreiner and Hess, 2015). Prior context-specific disclosure behaviors are additional indicators of consumers’ valuations (Motiwalla et al., 2014). Therefore, it seems that the willingness-to-pay for online privacy is a telling measure, but only if considered in light of its psychological drivers.

While there is no shortage of willingness-to-pay studies trying to quantify the valuation of privacy (see also Acquisti et al., 2013), only very few studies have investigated the perception or valuation of different aspects of privacy. Hann et al. (2002) used conjoint analysis to examine the importance people ascribe to the different privacy concern dimensions of Smith et al. (1996), showing that websites’ secondary use of personal information is perceived as most important, followed by improper access of personal information. An earlier study using consumer ratings yielded similar results showing that consumers were more concerned about improper access and unauthorized secondary use than about data collection and possible errors in their data (Esrock and Ferre, 1999). Another conjoint analysis identified consumer segments based on their differing levels of privacy concerns, highlighting the need for different premium accounts that cater to consumers’ differing privacy preferences (Krasnova et al., 2009).

To our knowledge, no study has so far investigated whether these patterns can be replicated for Malhotra et al.’s (2004) adapted model of privacy concerns and no study has investigated consumers’ valuation of these privacy aspects in the context of social networking services. For example, a study by Schreiner et al. (2013) examined social media users’ willingness-to-pay for information privacy on Facebook, but did not differentiate between the three dimensions of privacy and, therefore, does not provide insights into which aspects of privacy are most valued by users. Additionally, the study

by Schreiner and colleagues was limited in that they excluded non-members of Facebook, which constitutes an interesting consumer segment when it comes to privacy-enhanced premium versions of social networking services, as this segment may be especially interested in joining privacy-enhanced versions of such platforms.

STUDY OBJECTIVES AND RESEARCH HYPOTHESES

Filling these research gaps, the overarching objectives of the present study are twofold: first, the study will explore users' valuation of three different privacy aspects in the context of social networking services and, second, the study will investigate the psychological mechanisms underlying users' overall valuation of privacy.

Investigating the former, three privacy aspects will be studied that are captured in Facebook's Data Policy (Facebook Inc., 2016) as well as in Malhotra et al.'s (2004) multidimensional model of privacy. These three privacy aspects are (1) data collection, (2) data control, and (3) third party use. Accordingly, participants will be offered enhancement of these three privacy aspects within hypothetical premium versions of Facebook. Precisely, these privacy-enhanced premium versions of Facebook will offer (1) restricted data collection on side of the company, (2) enhanced data control for users, and (3) no sharing of users' data with third parties. Willingness-to-pay for the premium versions will be used as a proxy for participants' valuation of these privacy aspects. Expanding on previous studies (e.g., Schreiner et al., 2013), this study's insights will provide a more detailed understanding of users' valuation of different aspects of privacy. It is explored whether Internet users value some aspects of privacy more than others. Though previous research suggests that third-party sharing may be valued most (Esrock and Ferre, 1999; Hann et al., 2002), we argue that it is also possible that companies' restrictions on data collection may be valued more, since if no data are collected, users may be less worried about their data being shared with third parties. At the same time, the prevailing control-centered definition of privacy may invoke stronger valuations of the data control aspect. In light of these contradictory assumptions, for the present research no directional hypotheses can be formulated for the valuation of the three privacy aspects.

Investigating the latter, that is, the psychological mechanisms underlying valuation of privacy on social networking services, the present study will test a theoretical model that is developed and adapted based on proposed models by Malhotra et al. (2004) and Wilson and Valacich (2012). These models propose that privacy concerns increase perceived risk of information disclosure online and, thus, influence people's intentions to protect their data. This relationship is expected to be further moderated by several other psychological and socio-demographic characteristics measured in this study. It is hypothesized that the proposed model will explain the psychological mechanisms underlying valuation of privacy on Facebook (see Section Theoretical Model).

MATERIALS AND METHODS

Participants

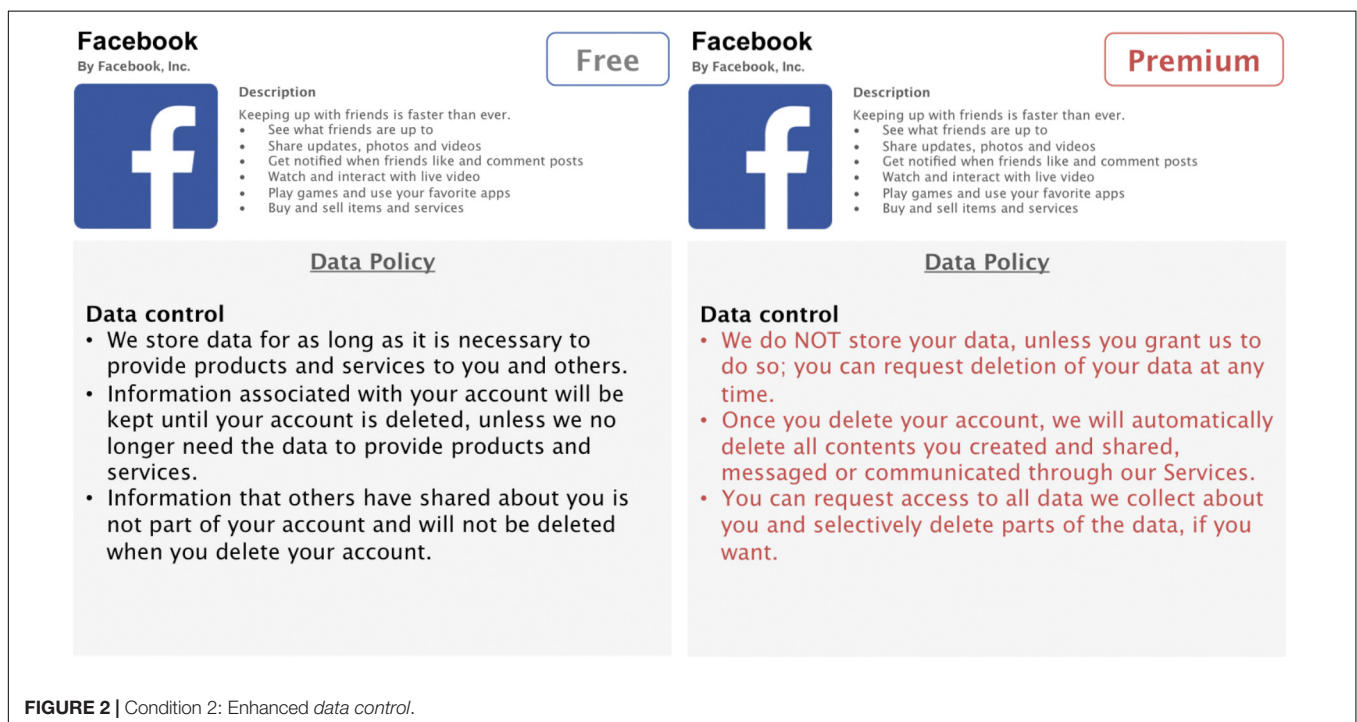
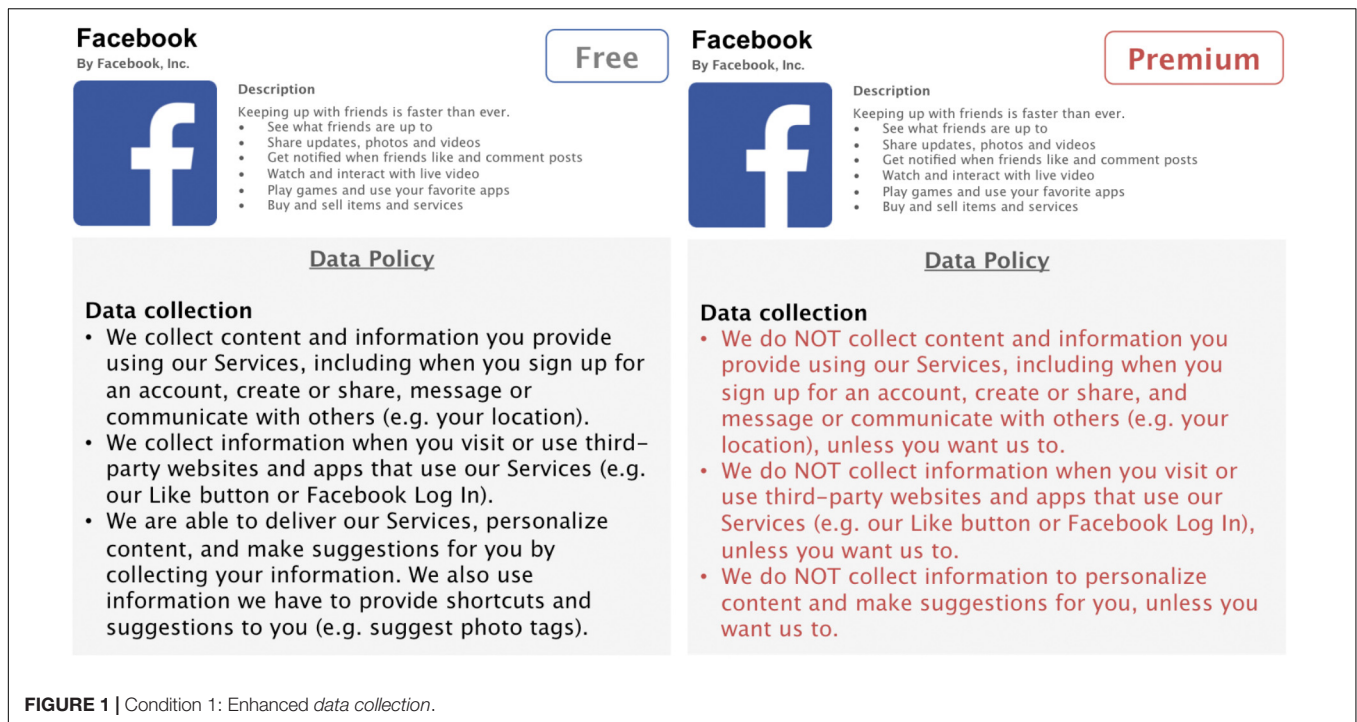
We aim to recruit at least 350 English-speaking adults (i.e., minimum age of 18 years). The estimated sample size is based on Lipovetsky's (2006) estimation that a minimum of 256 participants are needed to set up a price model with the precision of $\epsilon = 0.05$ and to reach value close to 80%. Taking into account potential dropouts and invalid participant responses, we aim to reach sample size of a minimum of 350 participants. Though participant recruitment is restricted to English-speaking adults, we will, unlike previous studies (e.g., Schreiner et al., 2013), recruit participants across different countries¹. As statistics report differing levels of privacy concerns and social media use across countries and cultures (e.g., Eurobarometer, 2016), we hope that our recruitment strategy will enable us to capture a heterogeneous participant sample with respect to the level of concern for and valuation of privacy. Furthermore, we will include both Facebook members and non-members in the sample. Facebook non-members are an important subsample, as this consumer segment could have a particular interest in privacy-enhanced versions of social networking services like Facebook. To ensure these sampling criteria, we will make use of various online channels, such as social networks and specialized study recruitment pages (e.g., findparticipants.com), as well as mailing lists, university platforms, and topic-relevant online forums.

Experimental Design

In the present online study, we will create four hypothetical privacy-enhanced premium versions of Facebook. The privacy enhancements of the premium versions will be based on three privacy aspects that are captured both in the IUIPC model (Malhotra et al., 2004) as well as in Facebook's Data Policy (Facebook Inc., 2016). Three of these premium versions will have one specific privacy aspect enhanced: in the first condition, data collection policies will be less permissible, thus, granting users the option that Facebook collects less data about them; the second condition will offer enhanced data control for users and allows complete or selective deletion of stored data; in the third condition, users will have the option to opt out from having Facebook share their data with third parties, such as advertisers (see **Figures 1–3**). An additional fourth condition will consist all three privacy enhancements in a full-design premium version.

Designing these hypothetical premium versions as realistic as possible, we will rely on Facebook's Data Policy to extract three central privacy aspects, namely data collection, data control, and third party sharing (Facebook Inc., 2016). We will adapt relevant parts of the policy accordingly to match the increased privacy functionalities of our premium versions. The conventional, free version of Facebook used for side-by-side comparisons consists of shortened and simplified, but otherwise unaltered, parts of Facebook's original Data Policy. The premium versions are written in such a fashion to correspond to the original policy

¹Given that many of the participants may not be native English-speakers, participants' English language proficiency will be assessed in a one-item self-report measure.

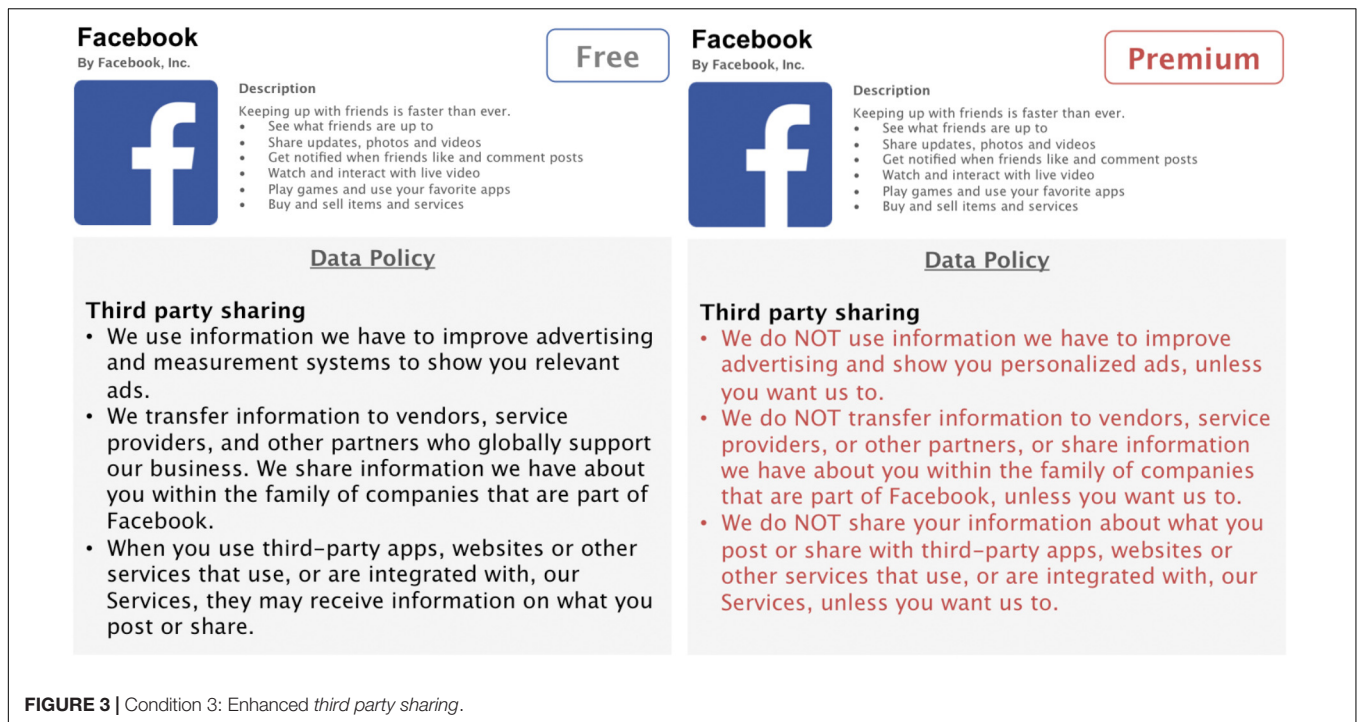


as much as possible, while enhancing specific privacy aspects. To facilitate readability, this information is presented in form of concise and comprehensive bullet points.

Willingness-to-Pay Measure

Quantifying Internet users’ valuation of the different privacy aspects, the van Westendorp’s (1976) Price Sensitivity Meter

model (PSM) will be employed as a willingness-to-pay measure. The PSM is a descriptive statistical procedure labeled the “psychological price” modeling (Lipovetsky et al., 2011). Rather than asking a single price indicator, the PSM allows capturing economic valuation in psychological terms. Furthermore, it ensures comparability of the results with the study by Schreiner et al. (2013). The PSM consists of four questions that ask



participants to balance the value of certain products or services against the price. Precisely, participants will answer the following questions about the four premium versions (as compared to the free version) presented:

- (1) At what price does this product become *too cheap*, that is, so cheap that you would question its quality and not buy it?
- (2) At what price does this product start to seem *cheap* to you, that is, when does it start to seem like a bargain?
- (3) At what price does this product start to seem *expensive* to you?
- (4) At what price does this product become *too expensive*, that you would not consider buying it?

The questions will be presented simultaneously and in the above order below the two versions of Facebook (i.e., conventional, free-of-charge versus hypothetical, privacy-enhanced version of Facebook). Participants will be asked to indicate a monthly price they are willing to pay for the privacy enhancement of each premium version. Combining the answers from the four PSM questions will allow identifying the upper and the lower price limit that participants are willing to pay for privacy. Based on this, the optimal price can be calculated as described in more detail in Section Proposed Analysis.

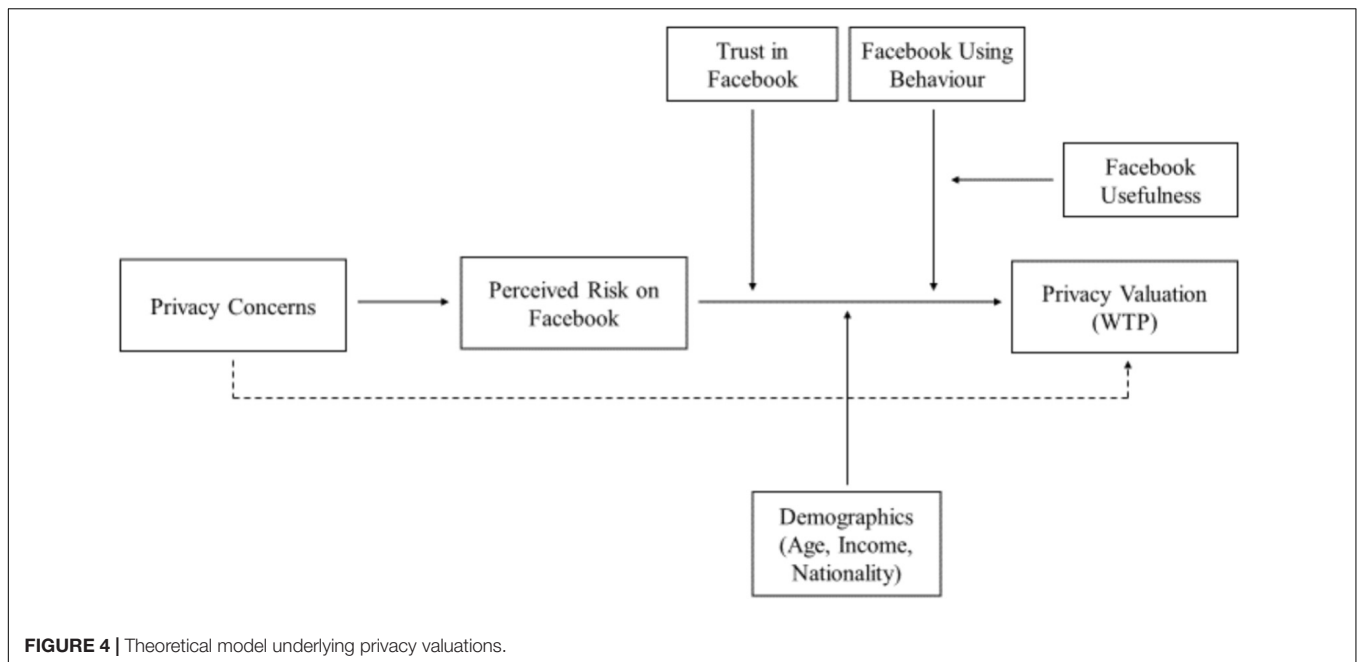
After answering the four PSM questions, a single-item willingness-to-pay measure will be employed to additionally assess the overall willingness-to-pay for the different privacy aspects (“Overall, how much would be willing to pay for this premium version of Facebook?”). This overall valuation measure will be used to validate the results of the PSM and to conduct the multiple comparisons between the three privacy enhancements,

which will allow drawing conclusions about which privacy aspects are valued the most.

Theoretical Model

To unravel the psychological mechanisms underlying privacy valuations on social networking services, a theoretical model will be tested. The present model is developed based on previously suggested models by Wilson and Valacich (2012) and Malhotra et al. (2004). The theoretical model presented here outlines the expected relationships between the psychological variables in predicting Internet users’ privacy valuations on social networking services (see **Figure 4**). The modeled psychological variables are selected based on previous research demonstrating their relevance in the context of information privacy. Where necessary, the psychological measures are adapted to suit the context of Facebook.

The present model proposes that perceived risk on Facebook mediates the relationship between privacy concerns (see also Malhotra et al., 2004) in predicting valuation of privacy, and that this relationship is further moderated by trust in Facebook and its Data Policies (adapted from Milne and Culnan, 2004) as well as by the level of Facebook use (adapted from Jenkins-Guarnieri et al., 2013). More specifically, we propose that high levels of privacy concerns predict high willingness-to-pay for privacy, mediated through increased privacy-related risk perception on Facebook. Additionally, the valuation of privacy is expected to depend on Facebook members’ current Facebook use or non-members’ perceived usefulness of Facebook, respectively (adapted from Rauniar et al., 2014). Among frequent Facebook users, those with greater privacy concerns are expected to express greater willingness-to-pay for



privacy on Facebook than those with lower privacy concerns. Among non-members of Facebook, those with strong privacy concerns and perceptions of Facebook's usefulness are expected to express higher willingness-to-pay for privacy than those non-members who do not perceive Facebook as useful. The rationale behind this is that privacy-concerned people who perceive Facebook as useful but are not member of the network, may abstain due to their privacy concerns, rather than due to lacking benefits from Facebook membership, and may thus be more likely to pay for privacy on Facebook. In addition to these psychological characteristics, socio-demographic information and the psychological characteristics social norms and comparative optimism will also be assessed, as these may have additional explanatory power beyond the primary variables included in the model. The psychological characteristics and socio-demographic information that are expected to explain participants' privacy valuations are explained in more detail in the next section (see Section Psychological Characteristics).

Psychological Characteristics

Privacy Concerns

The IUIPC scale developed by Malhotra et al. (2004) is a widely used measure of privacy concerns consisting of 10 items. The items (e.g., "It usually bothers me when online companies ask me for personal information") assess the three privacy dimensions data collection, data control, and awareness of the company's data practices on a 7-point Likert scale from one (strongly disagree) to seven (strongly agree). All three subscales have a composite reliability score of above 0.70 and have been validated in predicting behavioral intentions and Internet users' reactions to online privacy threats (Malhotra et al., 2004). The relationship between privacy concerns and willingness-to-pay for

privacy on social networking services will be examined. It is hypothesized that high levels of privacy concerns will predict greater willingness-to-pay for privacy directly through perceived risks on Facebook as well as through moderation of further psychological characteristics.

Perceived Risk on Facebook

Along with the IUIPC, Malhotra et al. (2004) used and adapted the risk perception scale validated by Jarvenpaa et al. (1999). As suggested in Malhotra et al. (2004), we adapted the six risk perception items to make them specific to the context of Facebook (e.g., "The risk that personal information submitted to Facebook could be misused is immense"). The scale has a reliability score of Cronbach's $\alpha = 0.70$ and uses a 7-point Likert scale ranging from one (strongly disagree) to seven (strongly agree). We hypothesize perceived risk on Facebook to be the main mediator of the effect of privacy concerns on willingness-to-pay. For participants with high privacy concerns but low risk perceptions on Facebook, however, valuation of privacy is expected to be low.

Perceived Internet Privacy Risk and Personal Internet Interest

Two scales will be used that were developed and validated by Dinev and Hart (2006a) and measure general Internet privacy risk and interest. Perceived Internet privacy risk consists of four items (e.g., "I am concerned that the information I submit on the Internet could be misused"), while personal Internet interest consists of three items (e.g., "The greater my interest to obtain a certain information or service from the Internet, the more I tend to suppress my privacy concern"). The items are assessed on a 5-point Likert scale ranging from one (very low risks/strongly disagree) to five (very high risk/strongly agree). For both scales, Cronbach's alpha indicates reliability above 0.66, which is the recommended cut-off score (Nunnally, 1978). Dinev and Hart (2006a) find

that higher privacy risk perceptions are related to higher levels of privacy concerns and lower willingness to transact personal information on the Internet, and that higher Internet interest is related to higher willingness to transact personal information on the Internet. While perceived risk on Facebook (see Section Perceived Risk on Facebook) is included as the main mediator in the model, the more general perceived Internet privacy risk measure will be tested as potential moderator for non-members of Facebook.

Trust in Facebook

Trust has been described as important foundation for all economic transactions (Ben-Ner and Halldorsson, 2010) and previous research demonstrated that customers' trust in companies and the Internet are important predictors of online disclosure and sharing behaviors (Metzger, 2004). Trust in the social networking service Facebook will be assessed via the trust in privacy notices subscale by Milne and Culnan (2004), defining trust as consumers' willingness to accept a level of risk in the face of incomplete information and as their belief that businesses will adhere to the privacy practices they declare (see Gefen et al., 2003 for a review on the trust literature). The relationship between trust in privacy notices with perceived risk and privacy concerns has been validated in Milne and Culnan (2004). In the present study, this relates to the belief that changes in Facebook's Data Policy can generally be trusted and the scale will be adapted to the context of Facebook. The scale consists of five items (e.g., "I believe that the Facebook privacy statements are truthful"), which are assessed on a 5-point Likert scale ranging from one (strongly disagree) to five (strongly agree). The scale's Cronbach's alpha is 0.82. Trust is hypothesized to moderate the relationship between privacy concerns and willingness-to-pay for privacy. Precisely, to invoke willingness-to-pay for privacy, participants need to generally trust Facebook and trust in Facebook's adherence to the offered privacy enhancements.

Facebook Use

Facebook use will be measured only among participants who, at the time of participation in this study, are members of Facebook. Facebook use will be assessed using the social media use integration scale by Jenkins-Guarnieri et al. (2013). The validated scale consists of 10 items (e.g., "I feel disconnected from friends when I have not logged into Facebook"), which are assessed on a 6-point Likert scale ranging from one (strongly disagree) to six (strongly agree). The scale has a Cronbach's alpha reliability of 0.91 and assesses social integration in and emotional connectedness to Facebook. It is hypothesized that frequent Facebook use will moderate the effect of privacy concerns through risk perceptions on participants' willingness-to-pay. Precisely, frequent Facebook users with strong privacy concerns are assumed to indicate greater willingness-to-pay.

Perceived Usefulness of Facebook

Perceived usefulness of Facebook will be assessed only in participants who, at the time of participation in this study, are non-members of Facebook. The perceived usefulness scale from the revised social media technology acceptance model (TAM) by

Rauniar et al. (2014) will be administered and adapted to the context of Facebook. The scale has been validated by Rauniar and colleagues and consists of five items (e.g., "Using Facebook makes it easier to stay informed with my friends and family"), which are assessed on a 5-point Likert scale ranging from one (strongly disagree) to five (strongly agree). The scale has a composite reliability score of above 0.70. We hypothesize that perceived usefulness of Facebook will moderate the relationship between privacy concerns and willingness-to-pay for non-members of Facebook. Precisely, we expect that when non-members of Facebook with high privacy concerns and risk perceptions still consider the usefulness of Facebook to be high, they could be willing to use a version of Facebook that protects their data and therefore indicate a higher willingness-to-pay.

Socio-Demographic Information

Previous research showed that socio-demographic factors, such as age and gender, influence Internet users' valuation of personal data and privacy (e.g., Krasnova et al., 2009). Therefore, socio-demographic information will be assessed, including gender, age, level of education, employment status, type of work, socioeconomic status, country of residence, and nationality. Socioeconomic status is predicted to have an influence on willingness-to-pay, as economic status (e.g., income) impacts people's overall readiness to pay a certain financial amount for the usage of a service or a product (Onwujekwe et al., 2009). We assume that socio-demographic information will influence the relationship between privacy concerns and willingness-to-pay for privacy on social networking services and control for these influences in our model.

Social Norms

Social norms are a strong predictor of human offline behaviors (Cialdini and Trost, 1998) and have been shown to be a significant antecedent of adopting online behaviors too (Chiasson and Lovato, 2001; Spottswood and Hancock, 2017). We will employ the questionnaire developed by Charng et al. (1988) to assess perceptions of social online norms and adapt the questionnaire to the context of Facebook. The questionnaire was validated for online use and has a reliability of Cronbach's alpha of 0.86 (Choi and Chung, 2013). The five items (e.g., "Many of the people that I know expect me to continuously use Facebook") are assessed on a 7-point Likert scale ranging from one (strongly disagree) to seven (strongly agree). We hypothesize that perceived social norms positively correlate with perceived usefulness of Facebook in non-members and with Facebook use in current Facebook users. Hence, social norms could further moderate the impact of privacy concerns on willingness-to-pay. If confirmed in the analysis, this variable may be included in the theoretical model.

Comparative Optimism

Participants' comparative optimism in the online context will be assessed using the approach by Baek et al. (2014). This approach relies on the indirect method (Harris et al., 2000) to assess participants' likelihood estimation of experiencing a certain event as compared to others experiencing the same

event. In two separate items, participants make judgments about their perceived personal and target group risk (i.e., “How likely are you [target group] to fall victim to improper use of online information?”). Both items will be assessed on a 5-point Likert scale ranging from one (least likely) to five (most likely). It is expected that participants who underestimate their own risk to fall victim to improper use of online information, as compared to others, have lower privacy concerns and risk perceptions, which may result in lower willingness-to-pay for privacy. Similar to social norms, we will test the relevance of this variable for the model.

STEPWISE PROCEDURES

The present experiment will be administered online using the web-based survey tool Qualtrics that allows designing, running, and collecting data through online experiments and surveys. The stepwise procedures of the experiment are as follows: After informed consent is given, participants will first answer a baseline measure that assesses if participants would be willing to pay for the current, free-of-charge version of Facebook. Afterward, participants will be presented a short vignette describing a scenario in which Facebook may consider developing premium versions of their service that would offer enhanced privacy for users in return for a monthly fee. In the first part of the online experiment, four hypothetical, privacy-enhanced premium versions of Facebook are presented consecutively and participants indicate their willingness-to-pay for each of the premium versions using the four questions of the PSM and the additional overall willingness-to-pay item (see Section Willingness-to-Pay Measure). Each privacy-enhanced version of Facebook is contrasted with the conventional, free-of-charge version of Facebook to facilitate comparability and increase participants’ understanding of the enhancements of the premium versions. To control for order effects, the three privacy-enhanced premium versions of Facebook (i.e., data collection, data control, and third party sharing) will be presented in randomized order. The fourth full-design premium version, which combines all three privacy enhancements in one version, will be presented last.

The second part of the study will assess several psychological characteristics (see Section Psychological Characteristics) to test the proposed theoretical model (see Section Theoretical Model) that specifies the psychological mechanisms underlying Internet users’ privacy valuations. The items of each scale will be presented in randomized order. Short control questions will be included in the online survey to ensure participants understand the privacy enhancements in the premium versions and to assess for how useful, credible, and technologically feasible these are rated. Two more general items will control whether participants answer the online study truthfully (e.g., “In general, I answered all of the questions seriously”). Lastly, socio-demographic information will be assessed. Once the survey is completed, participants will be thanked and further debriefed about the topic and purpose of the present study and those interested can read more about privacy and how to protect their online data. Those participants wishing

to enter the prize draw will be invited to follow a link to a separate survey where they can enter their email addresses. This way participants’ anonymity will be preserved and linking survey responses to identifiable information will be avoided.

PROPOSED ANALYSIS

In the first step, a cumulative frequency will be calculated for each of the enhanced privacy aspects captured in the hypothetical premium versions of Facebook (Figure 5).

In a second step, the range of acceptable prices that each participant is willing to pay for the different privacy-enhanced premium versions will be determined. The range of acceptable prices is defined by its endpoints marginal cheapness and marginal expensiveness (van Westendorp, 1976). Marginal cheapness is determined by the point where the cumulative frequencies of “too cheap” prices (reversed) and “cheap” prices intersect (MGP in Figure 6). In contrast, the point of marginal expensiveness is determined by the intersection of the cumulative frequencies of “too expensive” prices (reversed) and “expensive” prices (MEP in Figure 6).

In a third step, we will follow the approach by Lipovetsky (2006) who proposes that the four questions of the PSM and their corresponding cumulative distributions split the price continuum into five price perception intervals. These five price perception intervals are *too cheap*, *bargain*, *acceptable price*, *premium*, and *too expensive*. Thus, instead of the four thresholds of the questions of the PSM (Figure 5), five price ranges will be considered that are defined as discrete states with a continuous price variable and modeled as ordinal logistic regressions.

Following this model, the logistic cumulative probabilities for each price threshold will be determined and the appropriate thresholds for the particular model will be subtracted (i.e., for the acceptable price model the expensive price threshold is subtracted from the cheap price threshold). This procedure leads to smooth regression lines and allows determining the maximum of a specific price perception range. These maxima will be used as a proxy for participants’ willingness-to-pay (WTP in Figure 7).

Ordinal logistic regression models will be applied to test for statistical differences between participants’ willingness-to-pay

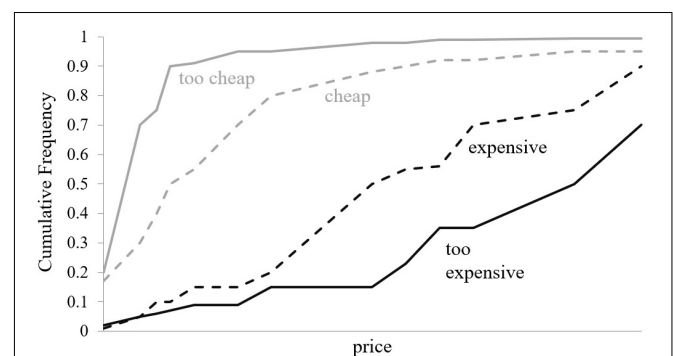


FIGURE 5 | Cumulative frequencies of the questions of the PSM.

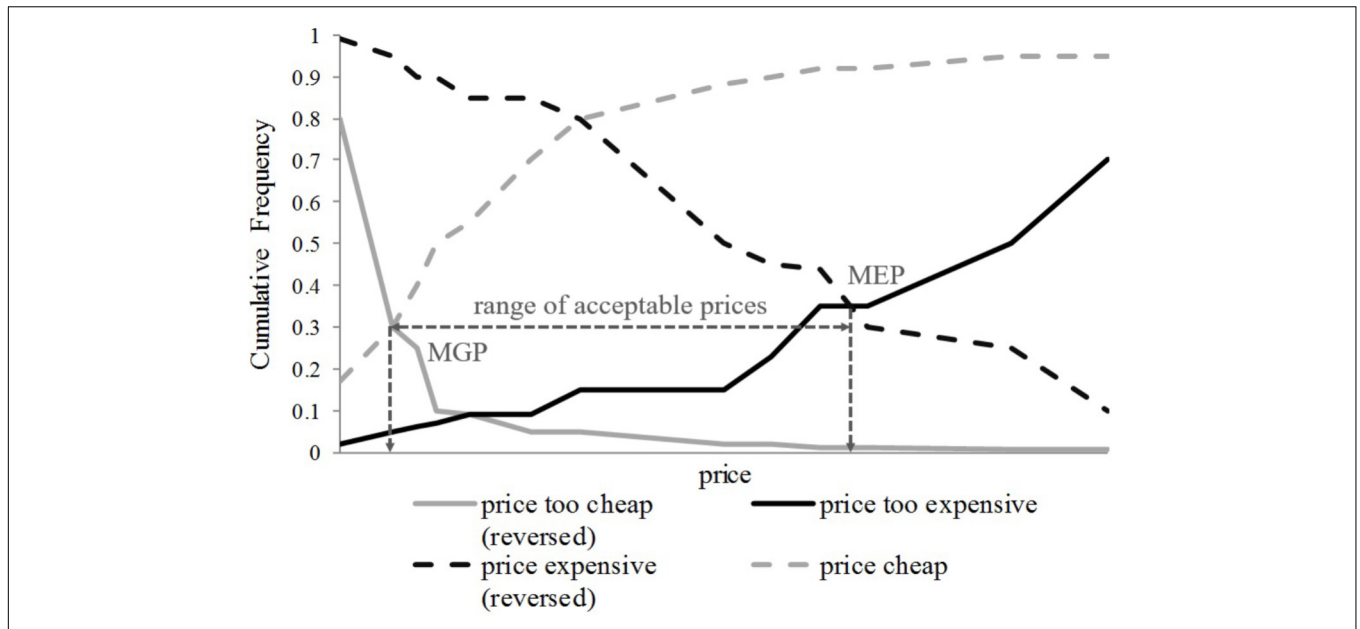


FIGURE 6 | Range of acceptable prices.

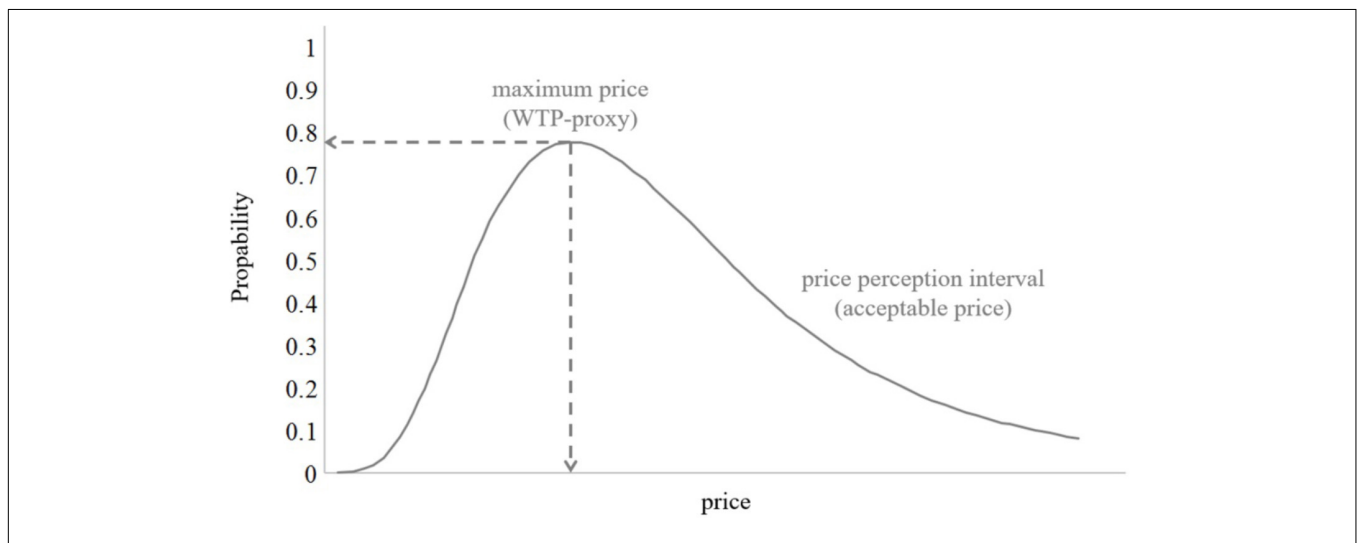


FIGURE 7 | Price sensitivity for the acceptable price perception interval.

for the different privacy aspects captured in the hypothetical premium versions of Facebook. Furthermore, the regression models can be extended to multiple predictors (e.g., privacy concerns and socio-demographic characteristics), since we hypothesize that psychological characteristics influence participants’ propensity to pay for the privacy enhancements. Together with the range of acceptable prices, the proxies will be used to test for intra-individual and inter-individual differences between willingness-to-pay for the four privacy-enhanced premium versions of Facebook. In addition, repeated-measure ANOVAs will be calculated for participants’ willingness-to-pay for the four different premium versions of Facebook, using the participant answers on the overall valuation measure (i.e.,

“Overall, how much would be willing to pay for this premium version of Facebook?”) as dependent variable. Where applicable, *post hoc* tests will be employed to determine the specific group differences. Data analysis will be conducted in R studio (R Core Team, 2017) and the conventional significance level of $\alpha = 0.05$ will apply to all analyses.

With respect to the theoretical model (see Figure 4, Section Theoretical Model), we follow previous approaches (Malhotra et al., 2004; Schreiner and Hess, 2015) and assume linear relationships between the indicated psychological variables (see Section Psychological Characteristics), which will be statistically tested using structural equation modeling to identify the path coefficients. As outcome variable in the tested model,

participants' overall willingness-to-pay for the hypothetical, full-design premium version of Facebook will be used.

ANTICIPATED RESULTS

In the proposed experiment, Internet users' valuation of different privacy aspects will be investigated in the context of social networking services. Four hypothetical, privacy-enhanced premium versions of Facebook will be developed, each offering the enhancement of one specific privacy aspect, namely data collection, data control, and third party sharing. A fourth version incorporates all three privacy enhancements. Valuation of privacy will be quantified using willingness-to-pay. The main aims of the experiment are to identify differences in the valuation of the three privacy aspects as well as to unravel the psychological mechanisms underlying these valuations.

For the purpose of the study, the PSM will be employed to measure willingness-to-pay for the premium versions of Facebook. The PSM allows estimating acceptable price ranges for each of the examined privacy aspects. Ordinal logistic regression as well as ANOVAs and according *post hoc* testing will be employed to investigate within-subject valuations of the three privacy aspects (i.e., data collection, data control, and third party sharing). In a second analysis step, the proposed theoretical model encompassing relevant psychological characteristics will be tested in order to unravel the psychological mechanisms underlying valuations of privacy. We expect overall willingness-to-pay (i.e., for the full-design premium version of Facebook) to be explained by privacy concerns, mediated by the perceived risk on Facebook, as well as by several moderating variables (see Sections Theoretical Model and Psychological Characteristics).

The results from this study will be a valuable contribution to the existing literature on information privacy. Most of the previous research has treated privacy as a one-dimensional construct and, thus, has not addressed consumer valuation of different aspects of privacy. Also, previous studies have largely disregarded non-members of social networking services, who constitute a large subsample that could be attracted to join social networking services, if these offered users enhanced privacy. The findings will, hence, complement several previous studies that examined the privacy paradox and valuation of privacy (e.g., Tsai et al., 2011) by offering a more detailed examination of the valuation of different privacy aspects, while also including non-members of certain services and products in this examination. Moreover, the findings will provide insights into the psychological mechanisms underlying these valuations. In comparison to Schreiner and Hess (2015), for example, who explained willingness-to-pay for privacy-enhanced premium services using the theory of planned behavior, the model proposed in this study emphasizes risk perceptions as a mediator for the effect of privacy concerns on willingness-to-pay for privacy on social networking services. It thereby focuses less on the valuations of the premium version itself, and rather serves to explain the individual differences in online privacy valuations. Furthermore, Schreiner and Hess did not find a link between perceived Internet risk and willingness-to-pay for

privacy-enhanced premium services. We suggest that the use of a general risk perception measure, rather than a Facebook-specific measure, could likely account for the unidentified link between these two related constructs. Therefore, in the present study, we will use a risk perception measure adapted specific to the context of Facebook. Besides the novel scope and the adapted constellation of the psychological factors in our proposed model, the present model also adds a cross-cultural dimension by sampling participants internationally and across cultures. Previous studies often collected data in only one country (e.g., Schreiner et al., 2013) or were predominantly relying on student populations (e.g., Krasnova et al., 2009).

Beyond the scientific contributions, the findings from the present research have considerable practical relevance, particularly in light of recent events such as the Cambridge Analytica Scandal (Revell, 2018) and the data protection laws that came into effect in the European Union in May 2018 (i.e., General Data Protection Regulation [GDPR]; Regulation (EU) 2016/679, 2017). Alternative business models may receive greater attention, as these could balance the asymmetric relationship between consumers and businesses and offer Internet users new privacy functionalities (e.g., Crook, 2018). Identifying which privacy features (e.g., third party sharing) are valued most, direct suggestions for the most important privacy enhancements can be derived. This will allow providing valuable suggestions for economically sustainable privacy enhancements and urgently needed alternative business models that are beneficial to consumers, service providers, and policymakers, alike.

Despite the study's important contributions to the existing scientific literature on information privacy and its practical relevance, there are a number of limitations that need to be addressed. First, as this study relies on a hypothetical scenario, no actual behaviors will be measured. Thus, this study only provides insights into Internet users' valuation of privacy based on hypothetical premium versions of Facebook. Though this study uses willingness-to-pay an indicator to quantify valuation of privacy, it is a rather intentional measure and does not provide a reliable economic value that translates into actual willingness-to-pay in a real-world settings (see intention-action gap; Sheeran and Webb, 2016). Second, as privacy concerns are context-dependent (e.g., Nissenbaum, 2009), the findings from this study are not generalizable to other platforms, but are specific to Facebook. Similarly, other measures assessed in this study, such as privacy concerns or risk perceptions, differ across countries, and culture (Wildavsky and Dake, 1990; Krasnova et al., 2012; Morando et al., 2014; Eurobarometer, 2016). Therefore, we will control for this by employing an international, cross-border sampling strategy. Third, despite our attempts to reach a heterogeneous sample by recruiting internationally and advertising our study on different platforms, our sample strategy may nonetheless be affected by sample bias, such as self-selection bias. Future studies could employ panel-based recruitment in order to reduce self-selection bias. Lastly, the presentation of the privacy policies will likely have an influence on users' willingness-to-pay. Privacy policies are usually far from the brevity and level of user-friendliness offered in this experiment. Future studies could more closely investigate the influence of presentation of

such policies to suggest more user-friendly alternatives and test willingness-to-pay in real-world setting using actual premium versions.

NOMENCLATURES

IUIPC, the 'Internet User's Information Privacy Concerns' is an instrument which measures the perception of acceptability of personal information collection practices; PSM, the 'Price Sensitivity Meter' is a descriptive statistical procedure used for calculating willingness-to-pay developed by van Westendorp; MGP, the point of 'marginal cheapness' is the intersection of the reversed 'too cheap' curve with the 'cheap' curve, defined by van Westendorp in his price sensitivity meter; MEP, the point of 'marginal expensiveness' is the intersection of the reversed 'expensive' curve with the 'too expensive' curve, defined by van Westendorp in his price sensitivity meter; WTP, 'Willingness-to-pay'; TAM, the revised social media 'technology acceptance model' by Rauniar et al. (2014); ANOVA, analysis of variance is a statistical procedure used to analyze the differences among group means in a sample; GDPR, the 'General Data Protection Regulation' is a regulation in European law that came into effect on 25 May 2018, serving to strengthen data protection and privacy for all individuals within the European Union and the European Economic Area.

REFERENCES

- Acquisti, A. (2004). "Privacy in electronic commerce and the economics of immediate gratification," in *Proceedings of the 5th ACM Conference on Electronic Commerce* (New York, NY: ACM), 21–29. doi: 10.1145/988772.988777
- Acquisti, A., and Gross, R. (2006). "Imagined communities: awareness, information sharing, and privacy on the Facebook," in *International workshop on privacy enhancing technologies* (Berlin: Springer), 36–58. doi: 10.1007/11957454_3
- Acquisti, A., and Gross, R. (2009). Predicting social security numbers from public data. *Proc. Natl. Acad. Sci. U.S.A.* 106, 10975–10980. doi: 10.1073/pnas.0904891106
- Acquisti, A., John, L. K., and Loewenstein, G. (2013). What is privacy worth? *J. Legal Stud.* 42, 249–274. doi: 10.1086/671754
- Alashoor, T., Keil M., Liu L., and Smith, J. (2015) "How values shape concerns about privacy for self and others," in *Proceedings of the 2015 International Conference on Information Systems (ICIS2015)*, December 13–16, Bangkok.
- Altman, E. I., Jacquillat, B., and Levasseur, M. (1974). Comparative analysis of risk measures: France and the United States. *J. Fin.* 29, 1495–1511.
- Baek, Y. M., Kim, E. M., and Bae, Y. (2014). My privacy is okay, but theirs is endangered: why comparative optimism matters in online privacy concerns. *Comput. Hum. Behav.* 31, 48–56. doi: 10.1016/j.chb.2013.10.010
- Barnes, S. B. (2006). A privacy paradox: social networking in the United States. *First Monday* 11. doi: 10.5210/fm.v11i9.1394
- Bauer, C., Korunovska, J., and Spiekermann, S. (2012). "On the value of information-what Facebook users are willing to pay," in *Proceedings of the 20th European Conference on Information Systems (ECIS)*, Barcelona, 1–12.
- Bélanger, F., and Crossler, R. E. (2011). Privacy in the digital age: a review of information privacy research in information systems. *MIS Quart.* 35, 1017–1042. doi: 10.2307/41409971
- Ben-Ner, A., and Halldorsson, F. (2010). Trusting and trustworthiness: what are they, how to measure them, and what affects them. *J. Econ. Psychol.* 31, 64–79. doi: 10.1016/j.joep.2009.10.001
- Bennett, C. H., Brassard, G., Crépeau, C., and Maurer, U. M. (1995). Generalized privacy amplification. *IEEE Trans. Inform. Theory* 41, 1915–1923. doi: 10.1109/18.476316

ETHICS STATEMENT

The proposed research was approved by the ethics committee of the Department of Psychology at the University of Geneva.

AUTHOR CONTRIBUTIONS

JM conceived the original idea for the research and provided supervision and guidance throughout. All authors made significant intellectual contributions to the study design and written protocol, were involved in all steps of the process, and approved the final version for publication.

FUNDING

This article was partly funded by the University of Geneva, Switzerland, and the Swiss National Science Foundation (SNSF).

ACKNOWLEDGMENTS

The Junior Researcher Programme (JRP) made this research possible. The authors would like to sincerely thank the JRP for their guidance and support.

- Beresford, A. R., Kübler, D., and Preibusch, S. (2012). Unwillingness to pay for privacy: a field experiment. *Econ. Lett.* 117, 25–27. doi: 10.1016/j.econlet.2012.04.077
- Brandimarte, L., Acquisti, A., and Loewenstein, G. (2015). Privacy and human behavior in the age of information. *Science* 347, 509–514. doi: 10.1126/science.aaa1465
- Buchanan, T., Paine, C., Joinson, N., and Reips, U. (2007). Development of measures of online privacy concern and protection for use on the Internet. *J. Am. Soc. Inform. Sci. Technol.* 58, 157–165. doi: 10.1002/asi.20459
- Carrascal, J. P., Riederer, C., Erramilli, V., Cherubini, M., and de Oliveira, R. (2013). "Your browsing behavior for a big mac: Economics of personal information online," in *Proceedings of the 22nd International Conference on World Wide Web* (New York, NY: Association for Computing Machinery), 189–200. doi: 10.1145/2488388.2488406
- Casidy, R., and Wymer, W. (2016). A risk worth taking: perceived risk as moderator of satisfaction, loyalty, and willingness-to-pay premium price. *J. Retail. Consum. Ser.* 32, 189–197. doi: 10.1016/j.jretconser.2016.06.014
- Chang, H., Piliavin, J. A., and Callero, P. L. (1988). Role-identity and reasoned action in the prediction of blood donation. *Soc. Psychol. Quart.* 51, 303–317. doi: 10.2307/2786758
- Chiasson, M. W., and Lovato, C. Y. (2001). Factors influencing the formation of a user's perceptions and use of a DSS software innovation. *ACM SIGMIS Database* 32, 16–35. doi: 10.1145/506724.506727
- Choi, G., and Chung, H. (2013). Applying the technology acceptance model to social networking sites (SNS): impact of subjective norm and social capital on the acceptance of SNS. *Int. J. Hum. Comput. Interact.* 29, 619–628. doi: 10.1080/10447318.2012.756333
- Cialdini, R. B., and Trost, M. R. (1998). "Social influence: Social norms, conformity, and compliance," in *The Handbook of Social Psychology*, eds D. T. Gilbert, S. T. Fiske, and G. Lindzey (New York, NY: McGraw-Hill), 151–192.
- Crook, J. (2018). Mark Zuckerberg: 'There will Always be a Version of Facebook That is free'. *TechCrunch*. Available at: <https://tcrn.ch/2GNu9ga> [accessed May 31, 2018].

- Culnan, M. J., and Armstrong, P. K. (1999). Information privacy concerns, procedural fairness, and impersonal trust: an empirical investigation. *Organ. Sci.* 10, 104–115. doi: 10.1287/orsc.10.1.104
- Dinev, T., Goo, J., Hu, Q., and Nam, K. (2009). User behaviour towards protective information technologies: the role of national cultural differences. *Informat. Syst. J.* 19, 391–412. doi: 10.1111/j.1365-2575.2007.00289.x
- Dinev, T., and Hart, P. (2006a). Internet privacy concerns and social awareness as determinants of intention to transact. *Int. J. Electron. Comm.* 10, 7–29. doi: 10.2753/JEC1086-4415100201
- Dinev, T., and Hart, P. (2006b). Privacy concerns and levels of information exchange: an empirical investigation of intended e-service use. *E-Serv. J.* 6, 25–59. doi: 10.2979/esj.2006.4.3.25
- Earp, J. B., Antón, A. I., Aiman-Smith, L., and Stufflebeam, W. H. (2005). Examining Internet privacy policies within the context of user privacy values. *IEEE Trans. Eng. Manage.* 52, 227–237. doi: 10.1109/TEM.2005.844927
- Egelman, S., Felt, A. P., and Wagner, D. (2013). “Choice architecture and smartphone privacy: there’s a price for that,” in *The Economics of Information Security and Privacy*, ed. Böhme (Berlin: Springer), 211–236. doi: 10.1007/978-3-642-39498-0_10
- Egelman, S., Tsai, J., Cranor, L. F., and Acquisti, A. (2009). “Timing is everything?: the effects of timing and placement of online privacy indicators,” in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (New York, NY: ACM), 319–328. doi: 10.1145/1518701.1518752
- Esrock, S. L., and Ferre, J. P. (1999). A dichotomy of privacy: personal and professional attitudes of marketers. *Bus. Soc. Rev.* 104, 107–120. doi: 10.1111/0045-3609.00042
- Eurobarometer (2016). *Standard Eurobarometer 86: Media Use in the European Union*. Available at: <http://ec.europa.eu/COMMFrontOffice/publicopinion/index.cfm/Survey/getSurveyDetail/instruments/STANDARD/surveyKy/2137> [accessed May 31, 2018].
- Facebook Inc. (2016). *Data Policy*. Available at: <https://www.facebook.com/about/privacy/your-info> [accessed December 28, 2017].
- Garg, V., Benton, K., and Camp, L. (2014). “The privacy paradox: a Facebook case study,” in: *Proceedings of the 2014 Telecommunications Policy Research Conference (TPRC)*, Arlington, VA. doi: 10.2139/ssrn.2411672
- Gasser, U. (2016). recoding privacy law: reflections on the future relationship among law. *Technol. Privacy Harvard Law Rev. Forum* 130, 61–70.
- Gefen, D., Karahanna, E., and Straub, D. W. (2003). Trust and TAM in online shopping: an integrated model. *MIS Quarterly* 27, 51–90. doi: 10.2307/30036519
- Grossklags, A., and Acquisti, J. (2007). “When 25 cents is too much: An experiment on willingness-to-sell and willingness-to-protect personal information, in *Proceedings of the Sixth Workshop on the Economics of Information Security (WEIS 2007)*, June 6, Pittsburgh, PA, 7–18.
- Hann, I. H., Hui, K. L., Lee, S. Y. T., and Png, I. P. (2007). Overcoming online information privacy concerns: An information-processing theory approach. *J. Manage. Inform. Syst.* 24, 13–42. doi: 10.2753/MIS0742-1222240202
- Hann, I. H., Hui, K. L., Lee, T., and Png, I. (2002). “Online information privacy: measuring the cost-benefit trade-off,” in *Proceedings of the Twenty-Third International Conference on Information Systems ICIS*, Barcelona, 1–10.
- Harris, P., Middleton, W., and Joiner, R. (2000). The typical student as an in-group member: eliminating optimistic bias by reducing social distance. *Eur. J. Soc. Psychol.* 30, 235–253. doi: 10.1002/(SICI)1099-0992(200003/04)30:2<235::AID-EJSP990>3.0.CO;2-G
- Iachello, G., and Hong, J. (2007). End-user privacy in human-computer interaction. *Found. Trends Hum. Comput. Interact.* 1, 1–137. doi: 10.1561/1100000004
- Jarvenpaa, S. L., Tractinsky, N., and Saarinen, L. (1999). Consumer trust in an Internet store: a cross-cultural validation. *J. Comput. Mediat. Commun.* 5:JCMC526.
- Jenkins-Guarnieri, M. A., Wright, S. L., and Johnson, B. (2013). Development and validation of a social media use integration scale. *Psychol. Popular Media Cult.* 2, 38–50. doi: 10.1037/a0030277
- Jensen, C., Potts, C., and Jensen, C. (2005). Privacy practices of Internet users: self-reports versus observed behavior. *Int. J. Hum. Comput. Stud.* 63, 203–227. doi: 10.1016/j.ijhcs.2005.04.019
- Klopfer, P. H., and Rubenstein, D. I. (1977). The concept privacy and its biological basis. *J. Soc. Issues* 33, 52–65. doi: 10.1111/j.1540-4560.1977.tb01882.x
- Kokolakis, S. (2017). Privacy attitudes and privacy behaviour: a review of current research on the privacy paradox phenomenon. *Comput. Sec.* 64, 122–134. doi: 10.1016/j.cose.2015.07.002
- Kosinski, M., Stillwell, D., and Graepel, T. (2013). Private traits and attributes are predictable from digital records of human behavior. *Proc. Natl. Acad. Sci. U.S.A.* 110, 5802–5805. doi: 10.1073/pnas.1218772110
- Krasnova, H., Hildebrand, T., and Günther, O. (2009). “Investigating the value of privacy in online social networks: conjoint analysis,” in *Proceedings of the Thirtieth International Conference on Information Systems, December 2009*, Phoenix, AZ, 173–191.
- Krasnova, H., Veltri, N. F., and Günther, O. (2012). Self-disclosure and privacy calculus on social networking sites: the role of culture. *Bus. Inform. Syst. Eng.* 4, 127–135. doi: 10.1007/s12599-012-0216-6
- Laudon, K. C. (1996). Markets and privacy. *Commun. ACM* 39, 92–104. doi: 10.1145/234215.234476
- Laufer, R. S., and Wolfe, M. (1977). Privacy as a concept and a social issue: a multidimensional developmental theory. *J. Soc. Issues* 33, 22–42. doi: 10.1111/j.1540-4560.1977.tb01880.x
- Lee, C. Y., and Heo, H. (2016). Estimating willingness to pay for renewable energy in South Korea using the contingent valuation method. *Energy Policy* 94, 150–156. doi: 10.1016/j.enpol.2016.03.051
- Lipford, H., Besmer, A., and Watson, J. (2008). “Understanding privacy settings in Facebook with an audience view,” in *Proceedings of the 1st Conference on Usability, Psychology, and Security* (Berkeley: USENIX Association), 1–8.
- Lipford, H. R., Wisniewski, P. J., Lampe, C., Kisselburgh, L., and Caine, K. (2012). “Reconciling privacy with social media,” in *Proceedings of the ACM 2012 Conference on Computer Supported Cooperative Work Companion* (New York, NY: ACM), 19–20. doi: 10.1145/2141512.2141523
- Lipovetsky, S. (2006). Entropy criterion in logistic regression and Shapley value of predictors. *J. Modern Appl. Stat. Methods* 5, 95–106. doi: 10.22237/jmasm/1146456480
- Lipovetsky, S., Magnan, S., and Zanetti-Polzi, A. (2011). Pricing models in marketing research. *Intell. Inform. Manage.* 3, 167–174. doi: 10.4236/iim.2011.35020
- Malhotra, N. K., Kim, S. S., and Agarwal, J. (2004). Internet users’ information privacy concerns (IUIPC): the construct, the scale, and a causal model. *Informat. Sys. Res.* 15, 336–355. doi: 10.1287/isre.1040.0032
- Margulis, S. T. (1977). Conceptions of privacy: current status and next steps. *J. Soc. Issues* 33, 5–21. doi: 10.1111/j.1540-4560.1977.tb01879.x
- McCloskey, H. J. (1980). Privacy and the right to privacy. *Philosophy* 55, 17–38. doi: 10.1017/S0031819100063725
- Metzger, M. J. (2004). Privacy, trust, and disclosure: exploring barriers to electronic commerce. *J. Comput. Mediat. Commun.* 9:942, doi: 10.1111/j.1083-6101.2004.tb00292.x
- Milne, G. R., and Culnan, M. J. (2004). Strategies for reducing online privacy risks: Why consumers read (or don’t read) online privacy notices. *J. Interact. Market.* 18, 15–29. doi: 10.1002/dir.20009
- Morando, F., Lemma, R., and Raiteri, E. (2014). Privacy evaluation: what empirical research on users’ valuation of personal data tells us. *Int. Policy Rev.* 3, 1–11.
- Motiwalla, L. F., and Li, X. B. (2016). Unveiling consumers’ privacy paradox behaviour in an economic exchange. *Int. J. Bus. Inform. Syst.* 23, 307–329. doi: 10.1504/IJBIS.2016.079523
- Motiwalla, L. F., Li, X. B., and Liu, X. (2014). *Privacy Paradox: Does Stated Privacy Concerns Translate into the Valuation of Personal Information?* Kaohsiung: PACIS, 281–296.
- New York Times (2018). *Facebook Is Not the Problem. Lax Privacy Rules Are. [Editorial]*. *New York Times*. Available at: <https://www.nytimes.com/2018/04/01/opinion/facebook-lax-privacy-rules.html> [accessed May 31, 2018].
- Nissenbaum, H. (2009). *Privacy in Context: Technology, Policy, and the Integrity of Social Life*. Redwood City, CA: Stanford University Press.
- Norberg, P. A., Horne, D. R., and Horne, D. A. (2007). The privacy paradox: personal information disclosure intentions versus behaviors. *J. Consum. Affairs* 41, 100–126. doi: 10.1111/j.1745-6606.2006.00070.x

- Nunnally, J. C. (1978). *Psychometric Theory*, 2nd Edn. Hillsdale, NJ: McGraw-Hill.
- Onwujekwe, O., Okereke, E., Onoka, C., Uzochukwu, B., Kirigia, J., and Petu, A. (2009). Willingness to pay for community-based health insurance in Nigeria: do economic status and place of residence matter? *Health Policy Plann.* 25, 155–161. doi: 10.1093/heapol/czp046
- Paine, C., Reips, U. D., Stieger, S., Joinson, A., and Buchanan, T. (2007). Internet users' perceptions of 'privacy concerns' and 'privacy actions'. *Int. J. Hum. Comput. Stud.* 65, 526–536. doi: 10.1016/j.ijhcs.2006.12.001
- Pew Research Center (2014). *Public Perceptions of Privacy and Security in the Post-Snowden Era*. Available at: <http://www.pewinternet.org/2014/11/12/public-privacy-perceptions/> [accessed May 31, 2018].
- Phelps, J., Nowak, G., and Ferrell, E. (2000). Privacy concerns and consumer willingness to provide personal information. *J. Public Policy Market.* 19, 27–41. doi: 10.1509/jppm.19.1.27.16941
- Pötzsch, S. (2008). "Privacy awareness: a means to solve the privacy paradox?" in *The Future of Identity in the Information Society. Privacy and Identity 2008. IFIP Advances in Information and Communication Technology*, Vol. 298, eds V. Matyáš, S. Fischer-Hübner, D. Cvrček, and P. Švenda (Berlin: Springer), 226–236.
- Quito, A. (2018). *The Case for Paying for Google and Facebook: Quartz Media*. Available at: <https://qz.com/1249955/jaron-lanier-at-ted-2018-to-fix-the-internet-we-have-to-start-paying-for-google-and-facebook/> [accessed May 31, 2018].
- R Core Team (2017). *R: A Language and Environment for Statistical Computing*. Vienna: R Core Team.
- Rauniar, R., Rawski, G., Yang, J., and Johnson, B. (2014). Technology acceptance model (TAM) and social media usage: an empirical study on Facebook. *J. Enter. Inform. Manage.* 27, 6–30. doi: 10.1108/JEIM-04-2012-0011
- Regulation (EU) 2016/679 (2017). *Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the Protection of Natural Persons with Regard to the Processing of Personal Data and on the Free Movement of Such Data, and Repealing Directive 95/46/EC (General Data Protection Regulation)*. Available at: <https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=CELEX:32016R0679>
- Revell, T. (2018). *How Facebook Let a Friend Pass My Data to Cambridge Analytica. The New Scientist*. Available at: <https://www.newscientist.com/article/2166435-how-facebook-let-a-friend-pass-my-data-to-cambridge-analytica/> [accessed May 31, 2018].
- Rose, E. A. (2005). "Data users versus data subjects: are consumers willing to pay for property rights to personal information?" in *Proceedings of the 38th Annual Hawaii International Conference on System Sciences* (Washington, D.C.: IEEE Computer Society Press), 1–10. doi: 10.1109/HICSS.2005.184
- Schreiner, M., and Hess, T. (2015). "Why are consumers willing to pay for privacy? An application of the privacy-freemium model to media companies," in *Proceedings of the 23rd European Conference on Information Systems (ECIS)*, Münster, 1–15.
- Schreiner, M., Hess, T., and Fathianpour, F. (2013). "On the willingness to pay for privacy as a freemium model: first empirical evidence," in *Proceedings of the 21st European Conference on Information Systems (ECIS)*, London, 30–36.
- Sheeran, P., and Webb, T. L. (2016). The intention-behavior gap. *Social Pers. Psychol. Compass* 10, 503–518. doi: 10.1111/spc3.12265
- Smith, H. J., Dinev, T., and Xu, H. (2011). Information privacy research: an interdisciplinary review. *MIS Quart.* 35, 989–1016. doi: 10.2307/41409970
- Smith, H. J., Milberg, S. J., and Burke, S. J. (1996). Information privacy: measuring individuals' concerns about organizational practices. *MIS Quart.* 20, 167–196. doi: 10.2307/249477
- Solove, D. J. (2006). A taxonomy of privacy. *Univ. Pennsylv. Law Rev.* 154, 477–560. doi: 10.2307/40041279
- Spiekermann, S., Korunovska, J., and Bauer, C. (2012). "Psychology of ownership and asset defense: why people value their personal information beyond privacy," in *Proceedings of the 33rd International Conference on Information Systems (ICIS2012)*, December 16–19, Orlando, FL. doi: 10.2139/ssrn.2148886
- Spottswood, E. L., and Hancock, J. T. (2017). Should I share that? Prompting social norms that influence privacy behaviors on a social networking site. *J. Comput. Mediat. Commun.* 22, 55–70. doi: 10.1111/jcc4.12182
- Stewart, K. A., and Segars, A. H. (2002). An empirical examination of the concern for information privacy instrument. *Informat. Syst. Res.* 13, 36–49. doi: 10.1287/isre.13.1.36.97
- Sundar, S. S., Kang, H., Wu, M., Go, E., and Zhang, B. (2013). "Unlocking the privacy paradox: do cognitive heuristics hold the key?," in *Proceedings of the CHI'13 Extended Abstracts on Human Factors in Computing Systems* (New York, NY: ACM), 811–816. doi: 10.1145/2468356.2468501
- Syverson, P. (2003). "The paradoxical value of privacy, in *Proceedings of the 2nd Annual Workshop on Economics and Information Security, May 29-30*, University of Maryland, College Park, MD, USA.
- Tsai, J. Y., Egelman, S., Cranor, L., and Acquisti, A. (2011). The effect of online privacy information on purchasing behavior: an experimental study. *Informat. Syst. Res.* 22, 254–268. doi: 10.1287/isre.1090.0260
- Tufekci, Z. (2015). *Mark Zuckerberg, Let Me Pay for Facebook*. New York Times. Available at: <https://www.nytimes.com/2015/06/04/opinion/zeynep-tufekci-mark-zuckerberg-let-me-pay-for-facebook.html> [accessed May 31, 2018].
- Turow, J., Feldman, L., and Meltzer, K. (2005). *Open to Exploitation: American Shoppers Online and Offline*. Philadelphia, PA: A Report from the Annenberg Public Policy Center of the University of Pennsylvania.
- van Westendorp, P. H. (1976). "NSS price sensitivity meter (PSM)—A new approach to study consumer perception of prices," in *Proceedings of the ESOMAR Congress*, Venice, 139–167.
- Wang, Y., and Kosinski, M. (2018). Deep neural networks are more accurate than humans at detecting sexual orientation from facial images. *J. Pers. Soc. Psychol.* 114, 246–257. doi: 10.1037/pspa0000098
- Warren, S. D., and Brandeis, L. D. (1890). The right to privacy. *Harvard Law Rev.* 4, 193–220. doi: 10.2307/1321160
- Westin, A. F. (1968). Privacy and freedom. *Washing. Lee Law Rev.* 25, 166.
- Wildavsky, A., and Dake, K. (1990). Theories of risk perception: who fears what and why? *Daedalus* 119, 41–60.
- Wilson, D., and Valacich, J.S. (2012). "Unpacking the privacy paradox: irrational decision making within the privacy calculus," in *Proceedings of the 33rd International Conference on Information Systems (ICIS2012)*, December 16–19, Florida, USA.
- Zou, Y., and Schaub, F. (2018). "Concern but no action: consumers' reactions to the equifax data breach," in *Proceedings of the Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems* (New York, NY: ACM), LBW506. doi: 10.1145/3170427.3188510
- Zuboff, S. (1988). *In The Age of the Smart Machine: The Future of Work and Power*. New York, NY: Basic books.
- Zuckerman, E. (2014). *The Internet's Original Sin. The Atlantic*. Available at: <https://www.theatlantic.com/technology/archive/2014/08/advertising-is-the-internets-original-sin/376041/> [accessed May 31, 2018].
- Zunger, Y. (2018). *Computer Science Faces an Ethics Crisis. The Cambridge Analytica Scandal Proves It. The Boston Globe*. Available at: <https://www.bostonglobe.com/ideas/2018/03/22/computer-science-faces-ethics-crisis-the-cambridge-analytica-scandal-proves/IzaXx2BsYBtwM4nxzgcP/story.html> [accessed May 31, 2018].

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2018 Mahmoodi, Čurdová, Henking, Kunz, Matić, Mohr and Vovko. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



The Power of Choice: A Study Protocol on How Identity Leadership Fosters Commitment Toward the Organization

Mafalda F. Mascarenhas¹, Felix Dübbers², Magdalena Hoszowska³, Aylin Köseoğlu⁴, Ralitsa Karakasheva⁵, Ayse B. Topal⁶, David Izydorczyk⁷ and Jérémy E. Lemoine^{8,9*}

¹ ISPA Instituto Universitário, Lisbon, Portugal, ² Department of Psychology, Maastricht University, Maastricht, Netherlands, ³ University of Social Sciences and Humanities, Warsaw, Poland, ⁴ Faculty of Social Sciences, Özyeğin University, Istanbul, Turkey, ⁵ Department Neuroscience, Psychology and Behaviour, University of Leicester, Leicester, United Kingdom, ⁶ Faculty of Arts and Social Sciences, Sabancı University, Istanbul, Turkey, ⁷ Faculty of Human Sciences, Psychology Department, University of Cologne, Cologne, Germany, ⁸ ESCP Europe Business School, London, United Kingdom, ⁹ C2S, Laboratory of Psychology: 'Cognition, Health, Socialization', University of Reims Champagne-Ardenne, Reims, France

OPEN ACCESS

Edited by:

Rocio Del Pino,
BioCruces Health Research Institute,
Spain

Reviewed by:

Alejandro Amillano,
University of Deusto, Spain
Charles Jacob,
University of Pennsylvania,
United States

*Correspondence:

Jérémy E. Lemoine
jeremy.lemoine@univ-reims.fr

Specialty section:

This article was submitted to
Quantitative Psychology
and Measurement,
a section of the journal
Frontiers in Psychology

Received: 30 October 2017

Accepted: 20 August 2018

Published: 06 September 2018

Citation:

Mascarenhas MF, Dübbers F,
Hoszowska M, Köseoğlu A,
Karakasheva R, Topal AB,
Izydorczyk D and Lemoine JE (2018)
The Power of Choice: A Study
Protocol on How Identity Leadership
Fosters Commitment Toward
the Organization.
Front. Psychol. 9:1677.
doi: 10.3389/fpsyg.2018.01677

Identity leadership (IL) describes that the effectiveness of a leader will depend upon his capacity to represent a given group, to make the group go forward, to create a group identity, and to make the group matter. An identity leader may increase commitment among his followers by increasing the perception of shared identity and giving more weight in the decision process to his followers. We aim to explore the mechanisms through which a leader who creates a shared group identity can increase organizational commitment. In the first study, we plan to conduct a cross-cultural correlational study where we aim to test if the relationship between IL and organizational commitment is mediated by team identification and mediated-moderated by participation in decision making (PDM) and collective efficacy. In the second study, we aim to explore the direction of the causality between IL and PDM. To test this hypothesis, we will conduct an experimental study in which (1) we will manipulate IL to test its influence on the perception of PDM and (2) we will manipulate PDM to test its influence on the perception of IL. Thus, we will be able to identify the role of IL and the perception of PDM on organizational commitment.

Keywords: identity leadership, organizational commitment, participation in decision making, collective efficacy, team identification, SEM

INTRODUCTION

Leadership research in psychology theorizes about what makes successful leaders attract and bind their followers as well as keeping them committed to their goals. Numerous theories behind leadership have evolved through very different paths (Day et al., 2014). At the very beginning of organizational psychology, it was common sense that someone was either born as a leader, or not, and that there was only one effective leadership style (Day et al., 2014). The task of the born leaders was to tell followers in an effective way what to do (Durrue et al., 2011). More behavioral approaches then alleged the contrary; leaders are made instead of born. It was proposed that there are characteristic traits which make you a good leader and these traits can be defined, measured, and taught so that theoretically, everybody could become a leader (Blake and Mouton, 1979). Overall

those theories exclusively focused on the characteristics of the leader and did not take into account the relationship between a leader and his followers.

More contemporary theories focused on this relationship between the leader and his followers. Authentic leadership theory focuses on leaders who have an honest relationship with their subordinates and are self-aware of their goals and aspirations. There is a focus on the value they give to their subordinates and their input (Gardner et al., 2011). In the charismatic leadership theory, the leader instigates followers by his innate charisma which is attributed to the leader based on the displayed behavior (Conger and Kanungo, 1988). Regarding the relationship between leader and follower, transformational leadership theory (Judge and Piccolo, 2004) proposes that a leader can exert influence by activating and serving higher order needs in his followers. Transformational leaders guide by vision, inspire their followers, and support them in personal growth (Judge and Piccolo, 2004). Similarly, in the leader-member-exchange theory, leaders and members influence each other within a dyadic bond which is built on trust and respect (Graen and Uhl-Bien, 1995). The follower and leader often develop even an emotional relationship and support each other (Graen and Uhl-Bien, 1995).

In all of the more recent leadership theories, the goal was to extend the leader's behavior toward the relationship between the leader and a single member. Indeed, one of the criticisms to the leadership literature is its focus on dyadic relationships (e.g., Yukl, 1999) while ignoring group level processes and the dynamic relationship between leaders and their teams (Hunter et al., 2007). Furthermore, a major part of social interaction has not been included in leadership theories so far: the relationship between a leader and his group (Gardner et al., 2010; Dinh et al., 2014). The theory of identity leadership (IL) is attempting to close that gap (Hogg, 2001) by focusing on the group identity and the group level processes that happen within a group: between the leader and his group members and between the group members themselves. Importantly, IL also differs from previous approaches by acknowledging the fact that leaders often need to create a shared sense of identity for the team to be more effective. As will be discussed in the following section, one aspect of IL (identity entrepreneurship; Haslam et al., 2011) describes that creating and shaping a shared sense of identity increases a leader's effectiveness. Thus, IL provides guidelines on how to transform a group of people with little in common into an effective team with a shared identity, which is often not the case in newly companies.

Identity Leadership

This new approach of leadership has appeared more recently and it perceives leadership as a group process rather than the result of leader characteristics or of a one-to-one relationship (Hogg, 2001). This model is based on the social identity theory (Tajfel and Turner, 1979) which claims that individuals have both an individual and social identity regarding the groups they belong to. Social identity theory has been used to think about processes that happen in organizational settings (Hogg and Terry, 2000). Incorporating social identity theory into the leadership literature allows for considering not only the leader, the follower, or their dyadic relationship, but the whole group and their

relationship to the leader. Leadership comes as a group and a social influence process that happens within a group with a shared identity (Hogg, 2001; Hogg et al., 2012). In the first decade, social identity theory of leadership was more concerned with leader prototypicality. Empirical evidence suggested that leaders who were more prototypical of the group were more supported and more trusted by their followers (Hogg et al., 2012). Later, another model tried to identify other dimensions that enable a leader to create and maintain a social identity within its team: IL (Haslam et al., 2011). The authors defined four dimensions of IL: *identity prototypicality* – refers to being “one of us,” to be an ideal member of the group; *identity advancement* – refers to the leader's vision for the group and his ability to make the group go forward in achieving their goals and improving their situation; *identity entrepreneurship* – the ability to create “a sense of us,” which means that the leader should be able to create a shared identity (as when politicians use “we” instead of “I” and “you”; Steffens and Haslam, 2013), and *identity impresarianship* – the ability to create moments that make the group matter. Reicher et al. (2005) described that for a leader to be efficient, it is not only necessary to create a shared social identity (i.e., a group has to exist for the leader to lead), but it is also necessary to create structures that maintain and promote the shared social identity (i.e., initiating a regular meeting to discuss group related matters and problems).

Moreover, IL leads to a better perceived performance of the leader and lower turnover intentions by followers (Steffens and Haslam, 2013). Followers are also more willing to follow and support the leader (Haslam and Platow, 2001). IL was also found to increase positive feelings among followers such as higher job satisfaction (Cicero et al., 2010). In addition, a meta-analysis by Mathieu and Zajac (1990) found that a leader who initiates structure also increases organizational commitment among his followers. Organizational commitment refers to a psychological relationship an individual develops with an organization (Nascimento et al., 2008), both for emotional reasons and a moral obligation to stay (Meyer et al., 1993), giving individuals a sense of belonging within an organization (Nascimento et al., 2008). Therefore, we expect that IL will significantly predict organizational commitment.

Team Identification

Prior research suggests that IL, especially leader prototypicality, increases team identification (TI; Hogg and van Knippenberg, 2003) a term which refers to a feeling of identification within a group and is often expressed by an individual seeing himself with similar characteristics to other members of the group (Dutton et al., 1994). Furthermore, TI has been found to be highly positively correlated with group commitment (Wann and Pierce, 2003). Also, identifying with a collective is proposed to lead to an increase in organizational commitment (Meyer et al., 2006; Johnson and Yang, 2010). Finally, Zhu et al. (2013) found that TI fully mediated the positive effect of transformational leadership on affective organizational commitment. We suspect that the positive effect of IL on organizational commitment is partly explained by how much the individual identifies with the team and thus that TI partially mediates the relationship between IL and organizational commitment.

Participation in Decision Making

Team identification is probably not the only concept that explains the relationship between IL and organizational commitment. A leader who is “one of us,” who is “doing it for us,” who is “crafting a sense of us” and who is “making us matter” (Haslam et al., 2011), will likely facilitate the willingness of group members to participate in decision making. When there is a shared sense of social identity, the leader might create more opportunities for the members to participate and group members might be more willing to participate. Participation in decision making (PDM) can be conceptualized as a process of decision making that includes various parties (Knoop, 1995; Witt et al., 2000): from a decision made by one person to a combined group decision. PDM allows people to have more control over their work and environment (Witt et al., 2000) and results in both higher job satisfaction (Witt et al., 2000; Scott-Ladd et al., 2006) and greater work commitment (Mathieu and Zajac, 1990; Knoop, 1995; Scott-Ladd et al., 2006). Mathieu and Zajac (1990) found in their meta-analysis that leader communication and participative leadership were strongly correlated with organizational commitment. Thus, we expect that IL predicts PDM and that PDM predicts organizational commitment. We also expect that the effect of IL on organizational commitment is partially explained by perceived PDM.

Collective Efficacy

While we are expecting that the relationship between IL and organizational commitment is partially mediated by PDM, this may not be the case for all workers in all teams. This mediation may be influenced by the level of team efficacy perceived by the team members. Self-efficacy is refers to one’s own belief that he is capable of producing certain effects through his actions (Bandura, 1998). When acting as a group, human agency is complemented by collective agency. Thus, collective efficacy (CE) is about shared beliefs in a group’s collective power and an emerging group-property rather than just the sum of the individual self-efficacy beliefs (Bandura, 1997). We suspect that people who are given the opportunity of applying shared decision making within their group decision profit more greatly when they believe that the CE of their group is high. Hence, we suspect that the positive relationship between PDM and organizational commitment will be moderated by CE. In addition, we suspect that people will strive for more participation if they perceive themselves as being capable of succeeding as a team. Therefore, we propose that CE also moderates the relationship between IL and PDM.

Culture

For different cultures, working as a group has a different value. The Hofstede study (Hofstede, 2001) measures a global orientation toward the individual and its own interests or the collective (individualism/collectivism). Considering that the IL theory is based on social identity and group processes, and that different cultures behave differently within groups (House et al., 2004), we decided to investigate whether the model is generalizable across cultures that vary in terms

of individualism-collectivism values by using two clusters of countries: one composed of individualistic countries and the other composed of collectivistic countries. In a large research study, van Dick et al. (2018) analyzed the generalizability of the IL model across 20 countries. Although those countries varied in terms of individualism-collectivism, the IL model was generalizable among all of them except Nepal. In all countries, IL predicts various outcomes such as job satisfaction, burnout, or organizational citizenship behavior. Furthermore, efficacy beliefs work cross-culturally, in individualistic as well as collectivistic cultures (Earley, 1994). Hence, we hypothesize that the model is generalizable across individualistic and collectivistic cultures.

Aims and Hypotheses

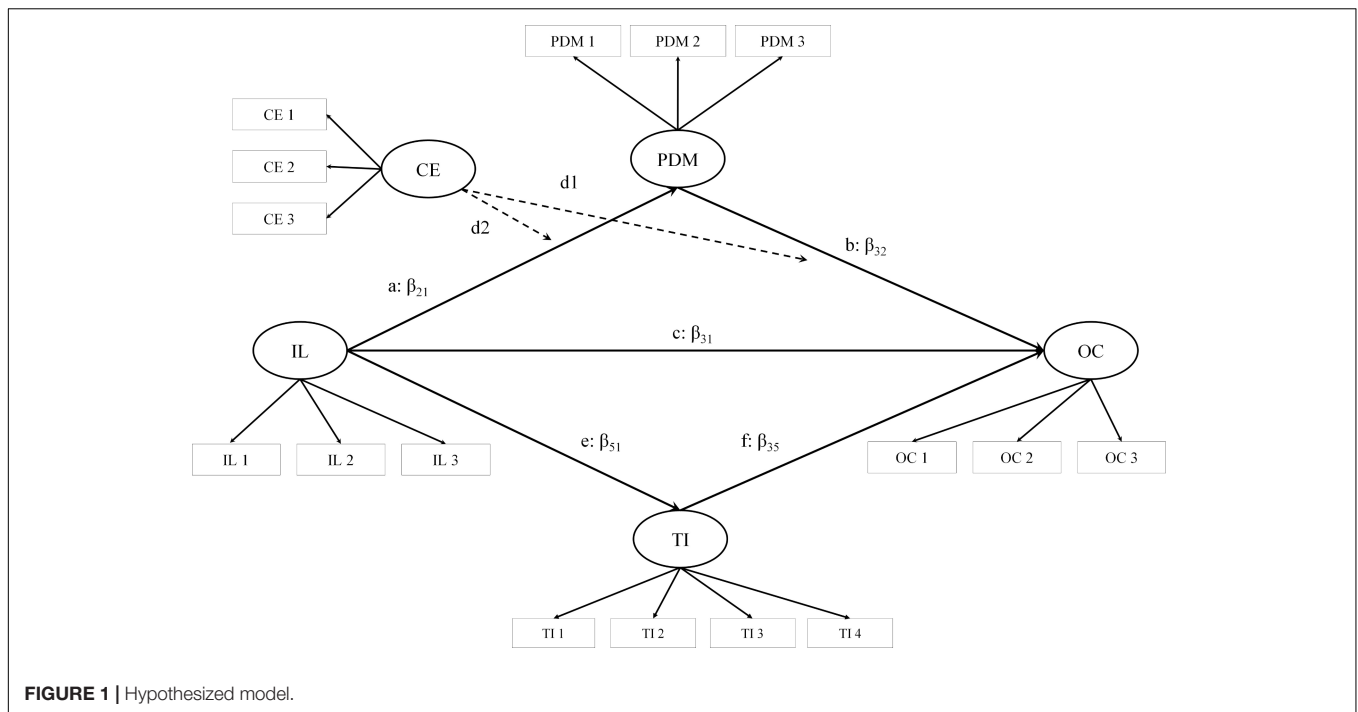
To summarize, IL is a promising new approach in leadership research. Having a leader who creates and fosters a shared identity in the team leads to positive outcomes in the workplace such as higher job satisfaction (Cicero et al., 2010), lower turnover intentions by followers (Steffens and Haslam, 2013), and it may also increase organizational commitment. We propose that one way in which IL leads to increased organizational commitment is by increasing the individual identification with the team and by facilitating the willingness of group members to participate in decision making. These relationships, however, may be moderated and influenced by factors like perceived CE, and the individualism-collectivism values in a country. This leads to the following hypotheses.

In Study 1, we propose a model (see **Figure 1**) in which IL positively influences organizational commitment (OC; H1). We hypothesize that the impact of IL on OC is mediated by TI and PDM. A higher level of IL will lead to both a higher level of TI (H2) and a higher level of PDM (H3). Moreover, we expect that a higher level of TI (H4) and PDM (H5) will increase OC. Furthermore, we hypothesize that the relationship between IL and PDM (H6) as well as PDM and OC (H7) are moderated by CE. Finally, we suppose that this model is generalizable across individualistic and collectivistic cultures (H8). In Study 2, we will use an experimental design and focus on the causal relationship between IL and PDM. We hypothesize that there is a bidirectional causal relationship between IL and PDM, i.e., a leader who creates a shared sense of identity will make his subordinates more willing to participate in decisions (H9) and, in turn, greater PDM will increase the sense of shared identity created by the leader (H10).

STUDY 1

Method Participants

Participants in Study 1 will be recruited from two clusters of countries according to their individualistic-collectivistic orientation. We used the measure of individualism/collectivism defined by the Hofstede study (Hofstede, 2001) which measures a global orientation toward the individual and its own interests or the collective. The scores range from 1 (collectivistic) to 100 (individualistic). In order to construct our two clusters, we used 50 as a cut-off point (Hofstede, 2001). Therefore, the



collectivist cluster is comprised of Bulgaria (30), Portugal (27), and Turkey (37) and the individualist cluster is composed of France (71), Germany (67), Netherlands (80), Poland (60), and United Kingdom (89).

Based on our hypothesized model depicted in **Figure 1**, we will need to estimate 43 parameters (20 error variances, 11 factor loadings, five variances, seven regression path coefficients) which yields an estimated sample size of 430 per country cluster when 10 observations per estimated parameter are used (Bentler and Chou, 1987; Bollen, 1989). Due to the explorative nature of this study, we have no clear idea about the majority of effect sizes in the model. Therefore, we used this rule of thumb estimation of sample size over the more sophisticated Monte Carlo estimates.

There are three inclusion criteria: (1) participants should work in an organization, (2) have a direct supervisor, and (3) be part of a team of at least three people. Participation will be anonymous and voluntary.

Measures

Leadership

Participants will be asked to evaluate how their current supervisor/manager at work scores on each of the four dimensions of the IL by completing the IL Inventory (Steffens et al., 2014). The inventory is a 15-item questionnaire reflecting the four dimensions of the IL theory: identity prototypicality (e.g., “this leader embodies what the group stands for”), identity advancement (e.g., “this leader stands up for the group”), identity entrepreneurship (e.g., “this leader makes people feel as if they are part of the same group”), and identity impresarioship (e.g., “this leader devises activities that bring the group together”). Items will be rated on a seven-point Likert scale (1 = *not at all*;

7 = *completely*). Cronbach’s α varied from 0.88 to 0.92 (Steffens et al., 2014).

Organizational commitment

Participants’ organizational commitment will be measured by the 18-item scale developed by Meyer et al. (1993). The inventory measures three dimensions of organizational commitment: affective, continuance, and normative commitment. These three dimensions reflect the personal desire of respondents to stay in the organization (e.g., “I would be very happy to spend the rest of my career with this organization”), necessity to stay (e.g., “right now, staying with my organization is a matter of necessity as much as desire”), and loyalty to the organization (e.g., “I would feel guilty if I left my organization now”). Respondents rated items on a seven-point Likert scale (1 = *strongly disagree*; 7 = *strongly agree*). Cronbach’s α ranged from 0.77 to 0.85 (Meyer et al., 1993).

Team identification

Respondents will be administered the group identification measure (Doosje et al., 1995) in which four items regarding one’s group identification (e.g., “I identify with the other team members”) will be rated on a seven-point Likert scale (1 = *not at all*; 7 = *extremely*). For the aim of this study, the items will be adapted for an organizational context. The scale has a good reliability, Cronbach $\alpha = 0.83$ (Doosje et al., 1995).

Participation in Decision Making

Participants will be asked to complete a group adapted version of the PDM scale (Witt et al., 2000). It is a six-item inventory which asks respondents to indicate how they and their managers make decisions in various contexts such as work appraisal. Answers will be scored on a five-point scale (1 = *we discuss things in a great*

detail and come to a decision based on consensus regarding the issue; 2 = we discuss things in a great deal and his/her decision is usually adopted; 3 = we discuss things in a great deal and the group decision is usually adopted; 4 = we don't discuss things very much as his/her decisions are usually adopted; and 5 = we don't discuss things very much and the group make most of the decisions). The scale is reported to have good reliability, Cronbach $\alpha = 0.90$ (Witt et al., 2000).

Collective efficacy

Participant will complete the seven-item CE Beliefs scale (Riggs et al., 1994) which measures CE in an organizational setting. On a seven-point Likert scale, respondents would rate items such as ("the team I work with has above average ability"). The scale is reported to have good reliability, Cronbach $\alpha = 0.88$ (Riggs et al., 1994).

Socio-demographic information

Participants will be asked to provide socio-demographic information (sex, age, nationality, and education), as well as information regarding their job: work field, type of employment (e.g., full time or part time), type of contract (temporary or permanent), country in which they are working, team size, and number of years spent working in their team.

Procedure

The materials will be translated into the languages of the targeted countries. The questionnaires (i.e., IL, PDM, OC, CE, and TI) which did not previously exist in the target languages (Bulgarian, Dutch, French, German, Polish, Portuguese, and Turkish) will be translated into the respective languages following the back-translation technique (Brislin, 1970). The IL measure was already validated in Dutch, French, German, and Turkish (van Dick et al., 2018). The adapted versions of OC in Dutch, French, German, Polish, Portuguese, and Turkish were already validated (de Gilder et al., 1997; Vandenberghe et al., 2001; Wasti, 2002; Süß, 2007; Nascimento et al., 2008; Baňka and Wooska, 2015); the Bulgarian version of OC will be translated with the back-translation technique. Furthermore, PDM, TI, and CE will need to be translated in all targeted languages. The factor structure of all translated measures will be assessed.

Data will be collected using an online questionnaire on Qualtrics. Participants will be recruited using the snowball sampling technique: via email, social media, personal contact, and work environment. Participation will be anonymous and voluntary. After giving their informed consent, participants will answer three inclusion criteria questions (e.g., "are you currently working as an employee?"). Those who correspond to our target population will then have to answer the questionnaire including measures of the IL, PDM, CE, organizational commitment, TI, and the socio-demographic information. The presentation order of the measures will be randomized, in order to avoid any order effects.

Planned Analysis

We will report all data exclusions (if any), all manipulations, and all measures in the studies. All analyses will be done with the GNU R software (R Core Team, 2013).

Moderation and mediation effects based on entire sample

To investigate the underlying relationships between IL and OC as well as the potential mediation of TI and PDM and the moderation effect of CE, we will first conduct a structural equation modeling analysis based on the entire sample. The hypothesized model is depicted in **Figure 1**. The latent variables IL, OC, PDM, and CE are defined by three indicators each. These indicators will be generated by item parceling. Parceling is when items are combined (summed or averaged) prior to an analysis and the parcels (instead of the original items) are used as the manifest indicators of latent constructs (Cattell, 1956). Instead of parcels, for TI, we will use the four corresponding items as indicators. As shown in **Figure 1**, we hypothesize that the relationship between IL and OC is partially mediated by TI. Furthermore, PDM partially mediates the relationship between IL and OC. In addition, CE potentially moderates both the relationship between IL and PDM as well as the relationship between PDM and OC. For the analyses, we will use the lavaan package in R (Rosseel, 2012). Before the actual analyses, we will evaluate the univariate normality assumption by examining skewness and kurtosis using the psych package (Revelle, 2017). Absolute values of skewness and kurtosis < 1 implicate univariate normality (Kline, 2011). We also will assess the multivariate normality assumption with Mardia's multivariate test (Mardia, 1970) by using the MVN package (Korkmaz et al., 2014).

Model fit. We will follow the recommendations from Kline (2011) and Schermelleh-Engel et al. (2003) and use several fit indices to interpret the model fit in general. First, we will use Chi-square (χ^2) and its associated p value, χ^2/df . Because χ^2 is sensitive to sample size and the violation of the multivariate normality assumption, we will also include different classes of goodness-of-fit criteria: the root-mean-square error of approximation (RMSEA; Steiger, 1990), the comparative fit index (CFI; Bentler, 1990), the standardized root-mean-square residual (SRMR; Jöreskog and Sörbom, 1989), and the Tucker-Lewis index (TLI; Tucker and Lewis, 1973). As recommended by Chen et al. (2008) as well as Marsh et al. (2004a), we will interpret the global model fit based on the constellation of these indices.

Mediation. To investigate the mediation effect, we will use bias-corrected bootstrapping to estimate confidence intervals for the indirect effect based on the recommendation of MacKinnon et al. (2007) and results of MacKinnon et al. (2004) and Fritz and MacKinnon (2007). This will allow us to test H2–H5 and to study if the relationship between IL is mediated by TI and PDM. In addition, we will use the ratio of indirect effect to total effect (Wen and Fan, 2015) as additional indicator.

Moderation. Although interaction or moderation effects are common in social sciences, estimating such effects in SEM, however, is difficult and not straightforward. A plethora of competing strategies and statistical approaches have been proposed (see, e.g., Jaccard and Wan, 1995; Jöreskog and Yang, 1996; Algina and Moulder, 2001; Marsh et al., 2004b; Little et al., 2006), which are mostly based on the product indicator model from Kenny and Judd (1984). We will use the double-mean-centering approach proposed by Lin et al. (2010) which is

identical or superior to the single-mean-centering (Marsh et al., 2004b) and orthogonalizing strategies (Little et al., 2006) that have been proposed previously. Thus, we can test H6 and H7 and explore if CE moderates the relationship between IL and PDM as well as PDM and OC.

Differences between groups

We will use multiple-group analyses to explore and test H8, i.e., whether differences in the structural parameters across groups of individualistic and collectivistic countries were statistically significant. To test for group invariance, we will compare two nested models with a likelihood ratio test (Bentler and Bonett, 1980; Bollen, 1989; Ryu, 2015). First, we will compare a baseline model wherein no constraints were specified and a second model where all factor loadings were constrained to be invariant between the groups. In the next step, we will compare this model with a model where all path coefficients were constrained to be invariant between the groups. In addition, when there are differences between the unconstrained and constrained models, subsequent likelihood ratio tests will be conducted, where different paths will be constrained and tested against the unconstrained model.

Anticipated Results

A leader who creates a shared sense of identity can lead to more commitment to the organization. We therefore expect to find a positive relationship between IL and OC. Based on previous research studies, we hypothesize that TI mediates this relationship. We further hypothesize that PDM is mediating the relationship between IL and OC, because when there is a shared sense of social identity, the leader might create more opportunities for the members to participate and group members might be more willing to participate. Additionally, CE is supposed to act as a moderator on this mediation. This is proposed to happen in such a way that the positive effect of IL on OC mediated by PDM is higher for participants with higher CE. Finally, we propose this as a cross-cultural model which is generalizable across individualistic and collectivistic cultures.

Depending on which hypotheses are supported by the results, different pieces of advice could be given to developing or established leaders in organizations. When there is a positive relationship between IL and OC, one might argue that leaders and organizations will profit from adopting IL behaviors, since OC leads to beneficial outcomes such as higher performance and employee well-being (Kurtessis et al., 2017). A possible mediation of the effect of IL on OC by TI and PDM would suggest that if a leader is wondering how to best foster organizational commitment in his followers, we would advise them to focus on the following: creating opportunities for followers to be actively involved and able to participate in decision making. Furthermore, a leader might strengthen OC by promoting identification within the group. Also if we find that CE is moderating the mediation of PDM, we would advise leaders that when attempting to increase OC by letting their followers participate in decision making, they should make sure that members of the team perceive their group as competent and effective when dealing with challenges and coming to a decision together.

It is important to identify the potential mediators and moderators of the relationship between IL and work commitment. This would help to better understand the impact of IL, to explore the way it works and to design programs that maximize its impact on organizational commitment and other organizational outcomes.

Anticipated Limitations

There are certain limitations of this study that future research should focus on. First, the clusters of individualistic and collectivistic countries that will be tested include only European countries and Turkey, thus sampling cultures from other continents can be potentially helpful to improve our understanding. Additionally, although the study will provide significant insight into this research question, the correlational nature of the study is problematic for the internal validity of this piece of research. Based on the literature, we assumed that IL of a leader will predict the level of PDM among his followers. However, it is also possible that followers who have a higher level of PDM will be more likely to perceive their leader as an identity leader than those who have a lower level of PDM. The direction of this relationship is neither clear in the literature nor based on our first study. Therefore, the second study will address this issue by using an experimental design which will provide in-depth understanding of the relationship between IL and PDM. By manipulating the degree of IL and PDM, we will aim to establish whether scoring high on IL encourages PDM or PDM shapes one's perception of the leader as creating a shared sense of identity.

STUDY 2

Overview

In Study 1, we studied the correlational relationship of IL, OC, TI, PDM, and CE within two country clusters. To further investigate the direction of the underlying causal processes, in Study 2, we will use an experimental design and focus on the causal relationship between IL and PDM. We hypothesize that there is a bidirectional causal relationship between IL and PDM, i.e., (1) a leader who creates a shared sense of identity will make his subordinates more willing to participate in decisions (H9) and (2) greater PDM will increase the perception of shared identity created by the leader (H10).

Method

Participants

An *a priori* sample size calculation with G-Power 3.1 (Faul et al., 2007) showed that for six conditions of IL manipulation, 36 participants per condition are required to achieve a power of 0.80 with $\alpha = 0.05$ and an expected medium effect size, $f = 0.25$. In addition, for the two conditions of manipulation of the level of PDM, the sample size estimate resulted in 64 participants per group, based on $\alpha = 0.05$, a power of 0.80, and an expected medium effect size, $d = 0.5$. Therefore, this study aims to recruit 344 participants. Participants will be recruited in English speaking countries using the snowball sampling technique: via

email, social media, personal contact, and work environment. There are three inclusion criteria: (1) participants should work in an organization, (2) have a leader, and (3) be part of a team of at least three people. Participation will be anonymous and voluntary.

Materials and Procedure

After giving their informed consent, participants will answer the same three inclusion criteria questions as used in Study 1. Those who correspond to our target population will be randomly assigned to one of eight possible conditions. Six of these conditions are dedicated to manipulate IL, the other two manipulate one's level of participation in the decision making process at work. In regard of IL manipulation, participants will be presented with a short description of the behavior of a manager in the workplace based on the IL model. In every description, each of the four dimensions of IL will be manipulated to be either in the high or in the low version (e.g., "your manager exemplifies/does not exemplify what it means to a member of this group"). Thus, there will be six possible versions: one with all dimensions being in the high version, one with all dimensions being in the low version, and four more in which the hypothetical manager would score high on one dimension but low on the other three dimensions. In the conditions dedicated to manipulating PDM, participants would read a paragraph in which one would be either highly involved in discussions and the decision making process (e.g., "your supervisor listens to each and every one of you and you, all together, come to a decision that everybody agrees to") or barely involved in the decision making process (e.g., "your supervisor comes to most of the decisions, without considering what you have to say"). All manipulations (six IL and two PDM) will be pre-tested in an online survey. After reading the manipulation, participants will complete the two scales measuring IL (Steffens et al., 2014) and PDM (Witt et al., 2000). These two measures are described in section "Study 1." For the six manipulation of IL conditions, participants will answer the PDM first and then the IL (which will be used as a manipulation check). For the manipulation of the two PDM conditions, they will answer the IL first and then the PDM (which will be used as a manipulation check). At the end of the study, participants will have to provide the same socio-demographic information as in Study 1.

Planned Analysis

R (R Core Team, 2013) will be employed to investigate the direction of the relationship between IL and PDM with an online experiment.

Identity leadership

In order to examine how IL, defined as four dimensions by Haslam et al. (2011), affects the PDM processes of group members (H9), a one-way ANOVA will be performed. The manipulation of IL will result in six conditions: the presentation of the leader will be either (1) high on the four dimensions, or (2) low on the four dimensions, or high on one of the dimensions and low on the other – (3) prototypicality dimension is high – rest is low, (4) advancement dimension is high – rest is low,

(5) entrepreneurship dimension is high – rest is low, and (6) impresarioship dimension is high – rest is low. Thus, we will observe the effect of the six different descriptions of the leader on PDM. First, we will test the normality of the distribution of the residuals by analyzing the skewness and kurtosis as well as using residuals vs. fitted and normal QQ plots (Hwu et al., 2002; Field, 2013). Afterward, the Levene test will be employed to assess the homogeneity of variances (Levene, 1960). If the normality of the distribution and homogeneity of variances are confirmed, a one-way ANOVA with planned contrast will be performed to compare any differences between the six groups (Norris, 2008). First, we will compare the high on all dimension condition with the five other conditions (1 vs. 2, 3, 4, 5, 6) and then we will compare the low on all dimension condition with the four other conditions (2 vs. 3, 4, 5, 6). As we do not have specific hypotheses regarding the four conditions in which IL is high on one dimension and low on the three others, we will use *post hoc* tests following the guidelines of Field (2013) to compare the differences between these four groups.

Participation in decision making

A *t*-test will be performed in order to detect if the level of PDM of team members has an effect on the perception of IL (H10). After checking the assumptions (normality, homogeneity of variances), we will compare the level of perceived IL between the low and high PDM conditions using an independent-samples *t*-test. Lastly, we plan to report the 95% confidence interval (Coe, 2002) and effect size using Cohen's *d* (Cohen, 1988).

Anticipated Results

In Study 2, we manipulate the degree of IL practiced by a leader in six conditions. We anticipate that a leader with a high degree of IL in every dimension of IL will make team members want to participate more in decision making compared to a leader with a medium or low degree of IL. In addition, we also anticipate that by manipulating the amount of PDM in a group, the leader will appear more as creating and fostering a shared identity in his team. This would allow us to formulate practical guidelines for increasing the organizational commitment of workers.

Anticipated Limitations

This piece of research will advance our understanding of the relationship between IL and PDM and may contribute to the way managers approach decision making with their employees. Investigating the way in which one's level of PDM shapes the image of their manager is particularly important, as managers can use PDM as a tool to enhance the shared sense of identity amongst the team. There are, however, certain limitations with regard to the generalizability of the results that future research should address. Conducting the experiment through an online platform can potentially influence responses and future research should aim to test the model in more realistic settings. Additionally, a more complex experimental design may want to establish additional relationships by including variables such as TI and CE.

CONCLUSION

Even though IL is still in its infancy, numerous studies have suggested its importance in predicting positive work related outcomes (e.g., Cicero et al., 2010; Hogg et al., 2012). The goal of our research study is to identify how a supervisor who adopts behaviors based on the IL principles can increase commitment in the organization among his followers. Therefore, this study may help to provide practical guidelines for supervisors as a way to increase commitment. Depending on the results, we could advise leaders to focus on promoting identification of their followers within their team and providing their followers with choices and opportunities to participate in decision making. Regarding the latter, we would recommend that leaders improve the perceptions of collective-efficacy held by their followers (e.g., through creating success stories, team encouragement, and the promotion of in-group collaboration (Bandura, 1998; Goddard et al., 2007), with the purpose of moderating the positive relationship toward organizational commitment. The model is not restricted for application in the work environment, but can also be transferred to other non-organizational contexts such as education, sports, politics or NGOs.

ETHICS STATEMENT

This study will be carried out in accordance with the recommendations of “Comité d’Ethique Interne du laboratoire

C2S” with written informed consent from all subjects. All subjects will give written informed consent in accordance with the Declaration of Helsinki. The protocol was approved by the “Comité d’Ethique Interne du laboratoire C2S.”

AUTHOR CONTRIBUTIONS

The initial design of this study was conceptualized by JL who also supervised the project and provided feedback. The rest of the authors also contributed to refining the design upon the start of their team work. MH, AK, and RK were in charge of the method section. MM and FD worked on the introduction section. The planned analysis section was produced by DI and AT.

FUNDING

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

ACKNOWLEDGMENTS

The project was initiated through and supported by the Junior Researcher Programme (JRP). We would like to thank the JRP team for their effort and support.

REFERENCES

- Algina, J., and Moulder, B. C. (2001). A note on estimating the Jöreskog-Yang model for latent variable interaction using LISREL 8.3. *Struct. Equ. Modeling* 8, 40–52. doi: 10.1207/s15328007sem0801_3
- Bandura, A. (1997). *Self-efficacy: The Exercise of Control*. New York, NY: Freeman.
- Bandura, A. (1998). “Personal and collective efficacy in human adaptation and change,” in *Advances in Psychological Science: Personal, Social, and Cultural Aspects*, Vol. 1, eds J. G. Adair, D. Belanger, and K. L. Dion (Hove: Psychology Press). 51–71.
- Bañka, A., and Wooska, A. (2015). Polish version of Meyer and Allen’s organizational commitment scales polska wersja meyera i allen skali przywiązania do organizacji. *Czasopismo Psychologiczne* 8, 65–84.
- Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychol. Bull.* 107, 238–246. doi: 10.1037/0033-2909.107.2.238
- Bentler, P. M., and Bonett, D. G. (1980). Significance tests and goodness of fit in the analysis of covariance structures. *Psychol. Bull.* 88, 588–606. doi: 10.1037/0033-2909.88.3.588
- Bentler, P. M., and Chou, C.-P. (1987). Practical issues in structural modeling. *Sociol. Methods Res.* 16, 78–117. doi: 10.1177/0049124187016001004
- Blake, R. R., and Mouton, J. S. (1979). What’s New With The Grid? *Train. Dev. J.* 32, 3–8. doi: 10.1177/103841117901600412
- Bollen, K. A. (1989). *Structural Equations with Latent Variables*. New York, NY: Wiley. doi: 10.1002/9781118619179
- Brislin, R. W. (1970). Back-translation for cross-cultural research. *J. Cross Cult. Psychol.* 1, 185–216. doi: 10.1177/135910457000100301
- Cattell, R. B. (1956). Validation and intensification of the sixteen personality factor questionnaire. *J. Clin. Psychol.* 12, 205–214. doi: 10.1002/1097-4679(195607)12:3<205::AID-JCLP2270120302>3.0.CO;2-0
- Chen, F., Curran, P. J., Bollen, K. A., Kirby, J., and Paxton, P. (2008). An empirical evaluation of the use of fixed cutoff points in RMSEA test statistic in structural equation models. *Sociol. Methods Res.* 36, 462–494. doi: 10.1177/0049124108314720
- Cicero, L., Pierro, A., and van Knippenberg, D. (2010). Leadership and uncertainty: how role ambiguity affects the relationship between leader group prototypicality and leadership effectiveness. *Br. J. Manag.* 21, 411–421. doi: 10.1111/j.1467-8551.2009.00648.x
- Coe, R. (2002). *It’s the Effect Size, Stupid: What Effect Size is and Why It Is Important*. Paper presented at the Annual Conference of the British Educational Research Association, University of Exeter, Exeter.
- Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences*, 2 Edn. New York, NY: Academic press.
- Conger, J., and Kanungo, R. N. (1988). *Charismatic Leadership: The Elusive Factor in Organizational Effectiveness*. San Francisco, CA: Jossey-Bass.
- Day, D. V., Fleenor, J. W., Atwater, L. E., Sturm, R. E., and McKee, R. A. (2014). Advances in leader and leadership development: a review of 25 years of research and theory. *Leadersh. Q.* 25, 63–82. doi: 10.1016/j.leaqua.2013.11.004
- de Gilder, D., van den Heuvel, H., and Ellemers, N. (1997). Het 3-componenten model van commitment A three component model of organizational commitment. *Gedrag. Organ.* 10, 95–106.
- Dinh, J. E., Lord, R. G., Gardner, W. L., Meuser, J. D., Liden, R. C., and Hu, J. (2014). Leadership theory and research in the new millennium: current theoretical trends and changing perspectives. *Leadersh. Q.* 25, 36–62. doi: 10.1016/j.leaqua.2013.11.005
- Doosje, B., Ellemers, N., and Spears, R. (1995). Perceived intragroup variability as a function of group status and identification. *J. Exp. Soc. Psychol.* 31, 410–436. doi: 10.1006/jesp.1995.1018
- Durue, D. S., Nahrgang, J. D., Wellman, N., and Humphrey, S. E. (2011). Trait and behavior theories-of leadership: an intergration and: meta-analytic test of their relative validity. *Pers. Psychol.* 64, 7–52. doi: 10.1111/j.1744-6570.2010.01201.x
- Dutton, J. E., Dukerich, J. M., and Harquail, C. V. (1994). Organizational images and member identification. *Adm. Sci. Q.* 39, 239–263. doi: 10.2307/2393235
- Earley, P. C. (1994). Self or group? Cultural effects of training on self-efficacy and performance. *Adm. Sci. Q.* 39, 89–117. doi: 10.2307/2393495

- Faul, F., Erdfelder, E., Lang, A.-G., and Buchner, A. (2007). G* Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behav. Res. Methods* 39, 175–191. doi: 10.3758/bf03193146
- Field, A. (2013). *Discovering Statistics Using IBM SPSS Statistics: And Sex and Drugs and Rock 'n' Roll*, 4 Edn. London: Sage.
- Fritz, M. S., and MacKinnon, D. P. (2007). Required sample size to detect the mediated effect. *Psychol. Sci.* 18, 233–239. doi: 10.1111/j.1467-9280.2007.01882.x
- Gardner, W. L., Cogliser, C. C., Davis, K. M., and Dickens, M. P. (2011). Authentic leadership: a review of the literature and research agenda. *Leadersh. Q.* 22, 1120–1145. doi: 10.1016/j.leaqua.2011.09.007
- Gardner, W. L., Lowe, K. B., Moss, T. W., Mahoney, K. T., and Cogliser, C. C. (2010). Scholarly leadership of the study of leadership: a review of the leadership quarterly's second decade, 2000–2009. *Leadersh. Q.* 21, 922–958. doi: 10.1016/j.leaqua.2010.10.003
- Goddard, Y. L., Goddard, R. D., and Tschannen-Moran, M. (2007). A theoretical and empirical investigation of teacher collaboration for school improvement and student achievement in public elementary schools. *Teach. College Record* 109, 877–896.
- Graen, G. B., and Uhl-Bien, M. (1995). Relationship-based approach to leadership: development of leader-member exchange (LMX) theory of leadership over 25 years: applying a multi-level multi-domain perspective. *Leadersh. Q.* 6, 219–247. doi: 10.1016/1048-9843(95)90036-5
- Haslam, A. S., and Platow, M. J. (2001). The link between leadership and followership: how affirming social identity translates vision into action. *Pers. Soc. Psychol. Bull.* 27, 1469–1479. doi: 10.1177/01461672012711008
- Haslam, A. S., Reicher, S. D., and Platow, M. J. (2011). *The New Psychology of Leadership: Identity, Influence and Power*. New York, NY: Psychology Press.
- Hofstede, G. (2001). *Culture's Consequences: Comparing Values, Behaviors, Institutions and Organizations Across Nations*, 2 Edn. Thousand Oaks, CA: Sage.
- Hogg, M. A. (2001). A social identity theory of leadership. *Pers. Soc. Psychol. Rev.* 5, 184–200. doi: 10.1207/s15327957pspr0503_1
- Hogg, M. A., and Terry, D. J. (2000). Social identity and self-categorization processes in organizational contexts. *Acad. Manag. Rev.* 25, 121–140. doi: 10.5465/AMR.2000.2791606
- Hogg, M. A., and van Knippenberg, D. (2003). "Social identity and leadership processes in groups", in *Advances in Experimental Social Psychology*, Vol. 35, ed. M. P. Zanna (San Diego, CA: Academic Press).
- Hogg, M. A., van Knippenberg, D., and Rast, D. E. III (2012). The social identity theory of leadership: theoretical origins, research findings, and conceptual developments. *Eur. Rev. Soc. Psychol.* 23, 258–304. doi: 10.1080/10463283.2012.741134
- House, R. J., Hanges, P. J., Javidan, M., Dorfman, P. W., and Gupta, V. (2004). *Culture, Leadership, and Organizations: The GLOBE Study of 62 Societies*. Thousand Oaks: Sage Publications.
- Hunter, S. T., Bedell-Avers, K. E., and Mumford, M. D. (2007). The typical leadership study: assumptions, implications, and potential remedies. *Leadersh. Q.* 18, 435–446. doi: 10.1016/j.leaqua.2007.07.001
- Hwu, T.-J., Han, C.-P., Rogers, K., and Rogers, K. J. (2002). The combination test for multivariate normality. *J. Stat. Comp. Simul.* 72, 379–390. doi: 10.1080/00949650213536
- Jaccard, J., and Wan, C. K. (1995). Measurement error in the analysis of interaction effects between continuous predictors using multiple regression: multiple indicator and structural equation approaches. *Psychol. Bull.* 117, 348–357. doi: 10.1037//0033-2909.117.2.348
- Johnson, R. E., and Yang, L.-Q. (2010). Commitment and motivation at work: the relevance of employee identity and regulatory focus. *Acad. Manag. Rev.* 35, 226–245. doi: 10.5465/amr.2010.48463332
- Jöreskog, K. G., and Sörbom, D. (1989). *LISREL 7: A Guide to the Program and Applications*. Chicago, IL: SPSS.
- Jöreskog, K. G., and Yang, F. (1996). "Nonlinear structural equation models: The Kenny-Judd model with interaction effects," in *Advanced Structural Equation Modeling: Issues and Techniques*, eds G. A. Marcoulides and R. E. Schumacker (Mahwah, NJ: Lawrence Erlbaum), 57–88.
- Judge, T. A., and Piccolo, R. F. (2004). Transformational and transactional leadership: a meta-analytic test of their relative validity. *J. Appl. Psychol.* 89, 755. doi: 10.1037/0021-9010.89.5.755
- Kenny, D. A., and Judd, C. M. (1984). Estimating the nonlinear and interactive effects of latent variables. *Psychol. Bull.* 96, 201–210. doi: 10.1037//0033-2909.96.1.201
- Kline, R. B. (2011). *Principles and Practice of Structural Equation Modeling*, 3 Edn. New York, NY: Guilford.
- Knoop, R. (1995). Relationships among job involvement, job satisfaction, and organizational commitment for nurses. *J. Psychol.* 129, 643–649. doi: 10.1080/00223980.1995.9914935
- Korkmaz, S., Goksuluk, D., and Zararsiz, G. (2014). MVN: an R package for assessing multivariate normality. *R J.* 6, 151–162.
- Kurtessis, J. N., Eisenberger, R., Ford, M. T., Buffardi, L. C., Stewart, K. A., and Adis, C. S. (2017). Perceived organizational support: a meta-analytic evaluation of organizational support theory. *J. Manag.* 43, 1854–1884. doi: 10.1177/0149206315575554
- Levene, H. (1960). "Robust tests for equality of variances," in *Contributions to Probability and Statistics: Essays in Honor of Harold Hotelling*, ed. I. Olkin (Palo Alto, CA: Stanford University Press), 278–292.
- Lin, G.-C., Wen, Z., Marsh, H. W., and Lin, H.-S. (2010). Structural equation models of latent interactions: clarification of orthogonalizing and double-mean-centering strategies. *Struct. Equ. Modeling* 17, 374–391. doi: 10.1080/10705511.2010.488999
- Little, T. D., Bovaird, J. A., and Widaman, K. F. (2006). On the merits of orthogonalizing powered and product terms: implications for modeling interactions among latent variables. *Struct. Equ. Modeling* 13, 497–519. doi: 10.1207/s15328007sem1304_1
- MacKinnon, D. P., Fairchild, A. J., and Fritz, M. S. (2007). Mediation analysis. *Annu. Rev. Psychol.* 58, 593–614. doi: 10.1146/annurev.psych.58.110405.085542
- MacKinnon, D. P., Lockwood, C. M., and Williams, J. (2004). Confidence limits for the indirect effect: distribution of the product and resampling methods. *Multivariate Behav. Res.* 39, 99–128. doi: 10.1207/s15327906mbr3901_4
- Mardia, K. V. (1970). Measures of multivariate skewness and kurtosis with applications. *Biometrika* 57, 519–530. doi: 10.1093/biomet/57.3.519
- Marsh, H. W., Hau, K.-T., and Wen, Z. (2004a). In search of golden rules: comment on hypothesis-testing approaches to setting cutoff values for fit indexes and dangers in overgeneralizing Hu and Bentler's (1999) findings. *Struct. Equ. Modeling* 11, 320–341. doi: 10.1207/s15328007sem1103_2
- Marsh, H. W., Wen, Z., and Hau, K.-T. (2004b). Structural equation models of latent interactions: evaluation of alternative estimation strategies and indicator construction. *Psychol. Methods* 9, 275–300. doi: 10.1037/1082-989x.9.3.275
- Mathieu, J., and Zajac, D. (1990). A review and meta-analysis of the antecedents, correlates, and consequences of organizational commitment. *Psychol. Bull.* 108, 171–194. doi: 10.1037/0033-2909.108.2.171
- Meyer, J. P., Allen, N. J., and Smith, C. A. (1993). Commitment to organizations and occupations: extension and test of a three-component conceptualization. *J. Appl. Psychol.* 78:538. doi: 10.1037//0021-9010.78.4.538
- Meyer, J. P., Becker, T. E., and Van Dick, R. (2006). Social identities and commitments at work: toward an integrative model. *J. Organ. Behav.* 27, 665–683. doi: 10.1002/job.383
- Nascimento, J. L., Lopes, A., and Salgueiro, M. D. F. (2008). Estudo sobre a validação do "Modelo de Comportamento Organizacional" de Meyer e Allen para o contexto português. *Comportamento Organ. Gestão* 14, 115–133. doi: 10.3900/fpj.04.06.341.p
- Norusis, M. J. (2008). *SPSS Statistics 17.0 Guide to Data Analysis*. Upper Saddle River, NJ: Prentice Hall.
- R Core Team. (2013). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.
- Reicher, S., Haslam, A. S., and Hopkins, N. (2005). Social identity and the dynamics of leadership: leaders and followers as collaborative agents in the transformation of social reality. *Leadersh. Q.* 16, 547–568. doi: 10.1016/j.leaqua.2005.06.007
- Revelle, W. (2017). *Psych: Procedures for Psychological, Psychometric, and Personality Research. R package version 1.7.5*. Available at: <https://CRAN.R-project.org/package=psych>
- Riggs, M. L., Warka, J., Babasa, B., Betancourt, R., and Hooker, S. (1994). Development and validation of self-efficacy and outcome expectancy scales for job-related applications. *Educ. Psychol. Meas.* 54, 793–802. doi: 10.1177/0013164494054003026
- Rosseel, Y. (2012). lavaan: an R package for structural equation modeling. *J. Stat. Softw.* 48, 1–36. doi: 10.18637/jss.v048.i02

- Ryu, E. (2015). Multiple-group analysis approach to testing group difference in indirect effects. *Behav. Res. Methods* 47, 484–493. doi: 10.3758/s13428-014-0485-8
- Schermelleh-Engel, K., Moosbrugger, H., and Müller, H. (2003). Evaluating the fit of structural equation models: tests of significance and descriptive goodness-of-fit measures. *Methods Psychol. Res. Online* 8, 23–74.
- Scott-Ladd, B., Travaglione, A., and Marshall, V. (2006). Causal inferences between participation in decision making, task attributes, work effort, rewards, job satisfaction and commitment. *Leadership Organ. Dev. J.* 27, 399–414. doi: 10.1108/01437730610677990
- Steffens, N. K., and Haslam, A. S. (2013). Power through ‘us’: leaders’ use of we-referencing language predicts election victory. *PLoS One* 8:e77952. doi: 10.1371/journal.pone.0077952
- Steffens, N. K., Haslam, A. S., Reicher, S. D., Platow, M. J., Fransen, K., Yang, J., et al. (2014). Leadership as social identity management: introducing the identity leadership inventory (ILI) to assess and validate a four-dimensional model. *Leadersh. Q.* 25, 1001–1024. doi: 10.1016/j.leaqua.2014.05.002
- Steiger, J. H. (1990). Structural model evaluation and modification: an interval estimation approach. *Multivariate Behav. Res.* 25, 173–180. doi: 10.1207/s15327906mbr2502_4
- Süß, S. (2007). *Die Psychologische Beziehung Zwischen Unternehmen und Freien Mitarbeiter : Eine Empirische Untersuchung des Commitments und der Arbeitsbezogenen Erwartungen von IT-Freelancern*. Hagen: FernUniversität Hagen.
- Tajfel, H., and Turner, J. C. (1979). “An integrative theory of intergroup conflict,” in *The Social Psychology of Intergroup Relations*, eds W. G. Austin and S. Worchel (Monterey, CA: Brooks), 33–47.
- Tucker, L. R., and Lewis, C. (1973). A reliability coefficient for maximum likelihood factor analysis. *Psychometrika* 38, 1–10. doi: 10.1007/bf02291170
- van Dick, R., Lemoine, J. E., Steffens, N. K., Kerschreiter, R., Akfirat, S. A., Avanzi, L., et al. (2018). Identity leadership going global: validation of the identity leadership inventory across 20 countries. *J. Occup. Organ. Psychol.* doi: 10.1111/joop.12223
- Vandenbergh, C., Stinglhamber, F., Bentein, K., and Delhaise, T. (2001). An examination of the cross-cultural validity of a multidimensional model of commitment in Europe. *J. Cross Cult. Psychol.* 32, 322–347. doi: 10.1177/0022022101032003005
- Wann, D. L., and Pierce, S. (2003). Measuring sport team identification and commitment: an empirical comparison of the sport spectator identification scale and the psychological commitment to team scale. *N. Am. J. Psychol.* 5, 365–372.
- Wasti, S. A. (2002). Affective and continuance commitment to the organization: test of an integrated model in the Turkish context. *Int. J. Intercult. Relat.* 26, 525–550. doi: 10.1016/s0147-1767(02)00032-9
- Wen, Z., and Fan, X. (2015). Monotonicity of effect sizes: questioning kappa-squared as mediation effect size measure. *Psychol. Methods* 20, 193–203. doi: 10.1037/met0000040
- Witt, L. A., Andrews, M. C., and Kacmar, K. M. (2000). The role of participation in decision-making in the organizational politics-job satisfaction relationship. *Hum. Relat.* 53, 341–358. doi: 10.1177/0018726700533003
- Yukl, G. (1999). An evaluation of conceptual weaknesses in transformational and charismatic leadership theories. *Leadersh. Q.* 10, 285–305. doi: 10.1016/s1048-9843(99)00013-2
- Zhu, W., Wang, G., Zheng, X., Liu, T., and Miao, Q. (2013). Examining the role of personal identification with the leader in leadership effectiveness: a partial nomological network. *Group Organ. Manag.* 38, 36–67. doi: 10.1177/1059601112456595

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2018 Mascarenhas, Dübbers, Hozowska, Köseoğlu, Karakasheva, Topal, Izydorczyk and Lemoine. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



The Effect of Moral Congruence of Calls to Action and Salient Social Norms on Online Charitable Donations: A Protocol Study

Nikola Erceg¹, Matthias Burghart², Alessia Cottone³, Jessica Lorimer^{4*}, Kiran Manku⁴, Hannah Pütz⁵, Denis Vlašiček¹ and Manou Willems²

¹ Department of Psychology, Faculty of Humanities and Social Sciences, University of Zagreb, Zagreb, Croatia, ² Department of Clinical Psychological Science, Maastricht University, Maastricht, Netherlands, ³ Department of Psychology, University of Leicester, Leicester, United Kingdom, ⁴ Department of Psychiatry, University of Oxford, Oxford, United Kingdom, ⁵ Department of Social Policy and Intervention, University of Oxford, Oxford, United Kingdom

OPEN ACCESS

Edited by:

Pietro Cipresso,
Istituto Auxologico Italiano (IRCCS),
Italy

Reviewed by:

John McAlaney,
Bournemouth University,
United Kingdom
Charles Jacob,
University of Pennsylvania,
United States

*Correspondence:

Jessica Lorimer
jessica.lorimer@psych.ox.ac.uk

Specialty section:

This article was submitted to
Quantitative Psychology
and Measurement,
a section of the journal
Frontiers in Psychology

Received: 31 October 2017

Accepted: 18 September 2018

Published: 26 October 2018

Citation:

Erceg N, Burghart M, Cottone A,
Lorimer J, Manku K, Pütz H,
Vlašiček D and Willems M (2018) The
Effect of Moral Congruence of Calls
to Action and Salient Social Norms on
Online Charitable Donations:
A Protocol Study.
Front. Psychol. 9:1913.
doi: 10.3389/fpsyg.2018.01913

Online advertising is an important tool that can be utilized by charities to elicit attention and funding. A critical examination of advertisement strategies is thus necessary to increase the efficacy of fundraising efforts. Previous studies have shown that individuals' moral views and perceptions of social norms can play important roles in charitable behavior. Thus, the current protocol describes a study to examine whether framing charitable advertisements in line with participants' morality and increasing the salience of descriptive social norms increases subsequent charitable behavior. We describe experimental, online methods, whereby participants are provided with a framed call-to-action and normative information within a custom-developed application or existing survey platform. Furthermore, in the exploratory fashion, we discuss the possibility of collecting participants' Facebook data and predicting moral profiles from this data. If there is an increased rate of donations as a result of moral compatibility and/or increased norm salience, charities can leverage this knowledge to increase the donations by tailoring their campaigns in a more appealing way for their prospective donors. Moreover, if it turns out to be possible to predict one's moral profile from Facebook footprints, charities can use this knowledge to find and target people that are more likely to support their cause. However, this introduces important ethical questions that are discussed within this protocol.

Keywords: charitable behavior, moral foundations, moral identity internalization, social norms, Facebook data

INTRODUCTION

Charities often provide a vital service for marginalized and vulnerable people in society. Given that individual charitable giving now contributes the largest proportion of income for all registered charities in the United Kingdom (The National Council for Voluntary Organisations [NCVO], 2017), it is vital that charities maximize the effectiveness of their campaigns. A promising avenue for increasing the efficiency of charitable fundraising with little or no additional cost are campaigns in the digital sphere, a rapidly growing platform for philanthropy. In various

reports, MacLaughlin (2012, 2014, 2017) has shown that online giving has been steadily growing in past years. For example, between 2013 and 2014, it increased by 8.9% (MacLaughlin, 2014). Furthermore, from 2015 to 2016, online giving increased by 2.8% in the United Kingdom and 7.9% in the United States (MacLaughlin, 2017). The following protocol suggests a method to study whether tailored appeals and salient social norms can be utilized to increase individual charitable behavior. The findings based on this protocol could significantly aid the online marketing strategies of charitable organizations.

The Determinants of Charitable Behavior

In this protocol, charitable behavior (CB) refers to the measurable actions of supporting a charity through donating money or time. To construct an intervention that influences CB, it is first necessary to identify its potential determinants. CB has often been examined in the context of the Theory of Planned Behavior (TPB; Ajzen, 1985, 1991). In the TPB, the immediate causes of any behavior are (1) intentions to perform that behavior and (2) the actual control one has over performing it. In turn, behavioral intentions result from individuals' attitudes, social norms and perceived behavioral control. Indeed, some studies found that these TPB variables are good predictors of CB. For example, van der Linden (2011) found that attitudes, perceived behavioral control, past behavior and moral norms significantly predicted charitable giving intentions. Smith and McSweeney (2007) reported similar results, but identified injunctive norms as an additional predictor of charitable giving intentions. Thus, one way to increase CB is to influence one or more of its immediate causes, such as changing attitudes and referencing a social norm.

Changing Behavior by Changing Attitudes

An intervention aimed at increasing CB may focus on changing people's attitudes toward CB. One promising way of doing this is by using the assumptions of Regulatory Fit Theory (Higgins, 2005). According to Regulatory Fit Theory (RFT), it is possible to increase the effectiveness of a persuasive appeal by framing the arguments of a persuasive message in a way that fits one's psychological characteristics. This could include motivational orientation (Updegraff et al., 2007), personality (Hirsh et al., 2012) or moral characteristics (Feinberg and Willer, 2013). Regulatory fit is hypothesized to shift attitudes through three intertwined mechanisms: by making the recipient 'feel right' during the message reception, by increasing the recipient's strength of engagement with the message, which contributes to processing fluency, and by influencing elaboration likelihood (Cesario et al., 2008).

What psychological characteristics are most relevant in the context of CB? It seems that an individual's morals play a significant role. For example, Smith and McSweeney (2007) and van der Linden (2011) showed that personal moral norms are one of the strongest predictors of CB. Furthermore, Aquino and Reed's (2002) moral identity internalization, or degree to which moral traits are central to one's self-concept, has been shown to influence: (a) type of charitable donations (time versus money), (b) donation intentions, (c) actual donations, and (d) emotions

experienced during donations. Those who feel that morality is central to their self-concept: (a) prefer donating time instead of money, (b) show greater intentions to donate money, (c) are willing to actually donate more money, and (d) experience more positive donation related emotions than those who are lower on moral identity (Reed et al., 2007; Winterich et al., 2012). This evidence makes a strong case for the importance of individual morals in predicting CB.

Therefore, following from the RFT, it can be hypothesized that a person's attitudes would change if a persuasive message was congruent with his/her moral views. One of the most influential and extensive theories that describe individuals' moral systems is the Moral Foundations Theory (MFT; Haidt and Joseph, 2004; Haidt and Graham, 2007; Graham et al., 2011, 2012). MFT postulates that five different, innate, moral foundations provide a "first draft" of moral intuitions; but these intuitions can also be revised through exposure to social context and culture (for a detailed review, see Graham et al., 2012). The five moral foundations proposed by MFT are Care/Harm, Fairness/Cheating, Loyalty/Betrayal, Authority/Subversion, and Sanctity/Degradation (Graham et al., 2012). According to the theory, differences in these five foundations are responsible for differences in morality across individuals and cultures.

The first two foundations are called the *individualizing foundations*, meaning that they emphasize inter-individual relations. Individuals scoring highly on these foundations are primarily sensitive to possible cruelty, unfairness, and inequality when making moral judgments (Graham et al., 2012). In contrast, the last three foundations are called the *binding foundations*, which bind individuals into communities. According to the MFT, those scoring highly on these foundations are primarily sensitive to social community, hierarchical relations, and physical and spiritual purity when making moral judgments. Graham et al. (2009) found that politically liberal individuals primarily endorsed and used individualizing moral foundations when making judgments (i.e., Care/Harm and Fairness/Cheating), whereas conservatives endorsed and used all five foundations more equally. These correlational patterns between morality and political ideology have been shown to be stable across cultures (e.g., Bobbio et al., 2011; Graham et al., 2011; Kosugi et al., 2014). This suggests at least two distinct, universal moral foundations profiles: a liberal and a conservative one.

Recent studies have examined the influence of (in)congruence of messages and individuals' moral foundations on attitudes toward CB, charitable intentions and actual CB. For example, Feinberg and Willer (2013) showed that framing messages about the environment in terms of sanctity, rather than only care, shifted conservatives' attitudes in a pro-environmental direction. Building on this, Wolsko et al. (2016) showed that the attitude change was also accompanied by increased donations to pro-environmental causes. Additionally, the congruence between individuals' moral foundations and both the charity cause and persuasive calls-to-action has been shown to increase donation intentions and donations, but only for those individuals high on moral identity

internalization (Winterich et al., 2012; Nilsson et al., 2016).

Despite this, previous studies have not explicitly contrasted charitable causes and calls-to-action in order to see, for example, whether congruent calls-to-action can have a positive impact on CB even if the charity cause is not in line with one's moral foundations (e.g., conservatives donating to a charity supporting immigration). Therefore, studying attitudes and subsequent behavior change in response to the (in)congruence between persuasive appeals of differing charity causes and individuals' moral foundations is a promising research avenue.

Changing Behavior by Changing Social Norms

In addition to changing attitudes toward CB, a fruitful approach may be to influence perceptions of social norms about CB. Social norms are the perceived rules of a community or group that dictate desired behavior (Kandori, 1992). Although some studies show that norms are the weakest of the TPB predictors (see Armitage and Conner, 2001), others have pointed out the need to distinguish between several types of norms before assessing their contribution to behaviors. Specifically, three distinct types of normative influences have been identified. First, norms can be *injunctive*, representing the information about what most others approve or disapprove. Second, norms can be *descriptive*, conveying information about what most others actually do (Cialdini et al., 1990). Finally, personal injunctive norms, or moral norms can be defined as *individual internalized moral rules* (Smith and McSweeney, 2007; van der Linden, 2011). Thus far, several studies have investigated the role of different types of norms on prosocial behavior, showing mixed results (e.g., Shang and Croson, 2009; van der Linden, 2011).

Several studies showed that moral and injunctive norms significantly predicted charitable intentions, after controlling for attitudes, perceived behavioral control and past behavior (Smith and McSweeney, 2007; van der Linden, 2011). However, in these studies, descriptive norms did not have a significant influence on charitable giving. In contrast, other studies have shown that descriptive norms can be significant determinants of CB and prosocial behavior in general. For example, Shang and Croson (2009) demonstrated that providing individuals with information about the amount of money that others donated, influenced the amounts donated in public radio fundraising, both immediately and in renewals the year after. Moreover, Croson et al. (2009) showed that the effect of providing social information in such a way on donations to public radio is fully mediated by changes in the perception of descriptive social norms. It is possible that these mixed findings are the result of differences in the saliency of norms that were used in those studies.

A norm must be salient to be efficient in changing behavior (Cialdini et al., 1990). This may explain why descriptive norms did not influence CB in some of the previous studies: CB is often performed privately (Smith and McSweeney, 2007). Because CB is often performed privately, individuals may not have an accurate sense of the extent to which other people engage in charitable action. In other words, descriptive norms may be

ineffective in the context of individual CB because they are not salient enough. Therefore, providing explicit information about the behavior of others, and thus making descriptive norms about CB salient, is hypothesized to be a useful approach to changing perceptions about descriptive social norms, and consequently CB itself.

Leveraging Social Networks to Foster Charitable Behavior

The percentage of CB conducted online grows every year (e.g., MacLaughlin, 2017). As such, it would be beneficial for charities to take advantage of this and make their online fundraising campaigns more efficient. For example, if charities could target interested potential donors more precisely and approach them with tailored, congruent calls-to-action, this could significantly improve their fundraising outcomes.

One way to target potential donors more precisely is by using big data produced by social networks. When browsing social networks and engaging in behaviors on those networks, people leave digital footprints. Previous research has shown that these footprints can be predictive of different psychological characteristics. Kosinski and his colleagues have conducted multiple studies exploring the potential uses of digital footprints created on social media sites to identify users' psychological characteristics (e.g., Quercia et al., 2011; Kosinski et al., 2013, 2014; Youyou et al., 2015).¹

Their research shows that it is possible to predict people's personality trait scores (e.g., openness and extraversion) based on their digital footprints – specifically, Facebook *likes*. In fact, these predictions can be as accurate as those made by human judges, such as colleagues, friends or spouses (Youyou et al., 2015). Other researchers have demonstrated that it is possible to use social media data to predict various other characteristics. For example, Conover et al. (2011) demonstrated that people's political alignment can be predicted from their Twitter data with 90.8% accuracy. Furthermore, since an individual's different psychological properties are reflected in his/her digital footprints, it is possible that these footprints could be used to predict the individual's moral foundations. If so, it would be possible for charities to directly target people who are more likely to support their cause and become donors, while avoiding those who are less likely to support that specific cause.

However, it has to be noted that the future of data collection and advertising on Facebook is questionable, both pragmatically and from an ethical standpoint. From a practical point of view, there are two limitations: declining use and more stringent data sharing policies. For example, a Pew Research Center² study completed in 2018 found that Facebook is no longer the most popular online platform among teens, with only half of teens reporting using it. In addition, mostly due to the Cambridge

¹Data were obtained via a Facebook app developed under the *myPersonality* project. Users had the chance to complete a personality questionnaire, an intelligence test and a satisfaction with life questionnaire. Various demographic attributes were collected from users' Facebook profiles and some characteristics were obtained through a survey (Kosinski et al., 2013). Users could also volunteer their social media data – specifically, Facebook *likes* – for research purposes.

²<http://www.pewinternet.org/2018/05/31/teens-social-media-technology-2018/>

Analytica scandal³, Facebook has made its data sharing policies much more stringent. In practice, they currently do not allow new applications to collect users' data for research purposes. Moreover, the scandal echoed extremely negatively among its users, decreasing users' trust toward the company. For example, a survey from the Ponemon Institute (2018) found that between 2017 and 2018, the percentage of people who believed that Facebook was committed to privacy dropped by 52 percentage points. Therefore, users may be reluctant to give away their information, even for scientific purposes.

However, as will be described in the protocol, the collection of Facebook data and prediction of moral views from it is completely optional and constitutes the exploratory part of this protocol. It is perfectly possible to skip this part and follow only the confirmatory part of the protocol. This would significantly simplify the procedure, while still potentially providing theoretically and practically meaningful findings.

AIMS AND HYPOTHESES

The project has five aims:

- (i) to investigate the effect of morally (in)congruent calls-to-action and salient descriptive social norms on CB;
- (ii) to explore whether these effects occur when the charity's cause is not aligned with a person's moral profile;
- (iii) to examine whether the effects of (in)congruent calls-to-action are moderated by participants' moral internalization scores;
- (iv) to investigate whether attitudes toward CB mediate the effect of (in)congruent calls-to-action on CB and whether descriptive social norms mediate the effect of normative information on CB;
- (v) to investigate whether individuals' moral foundations can be predicted based on their Facebook behavior.

In previous studies, calls-to-action congruent with one's moral foundations increased CB (Winterich et al., 2012). Therefore, we hypothesize that morally congruent calls-to-action will have a greater positive impact on CB compared with morally

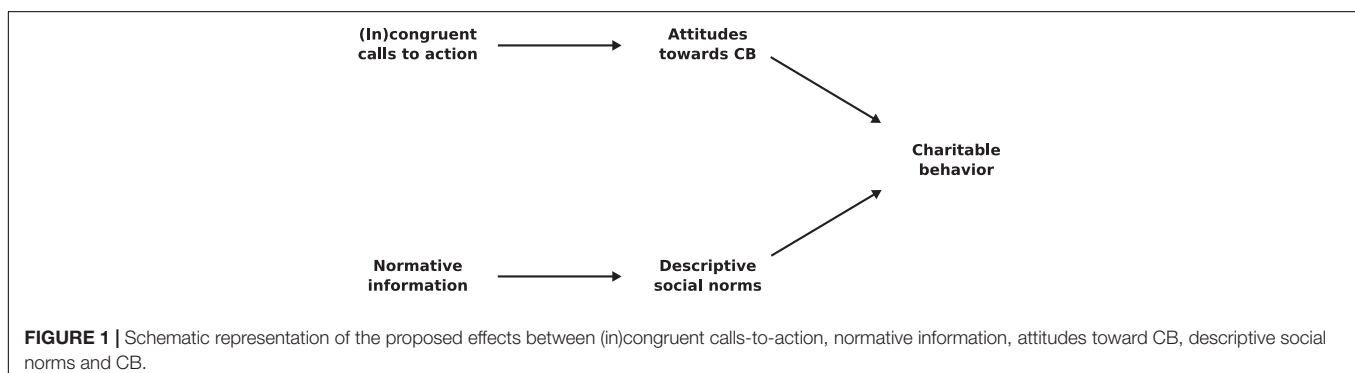
incongruent and neutral calls-to-action, regardless of whether the charity cause itself is in line with one's moral foundations. Furthermore, we hypothesize that providing participants with normative information, thus making the descriptive norm salient, will significantly affect CB. Additionally, we expect that the impact of morally congruent calls-to-action will be greater in individuals with high moral internalization compared to individuals with low moral internalization. Also, in line with the assumptions of TPB (Ajzen, 1985, 1991), we expect that the effects of calls-to-action congruence on CB will be mediated by attitudes toward CB, and that the effect of normative information on CB will be mediated by the perception of descriptive social norms. Finally, although this part of the research would be purely exploratory, we expect to be able to predict some of the participants' moral foundations by his/her Facebook behavior. The proposed effects are shown in the **Figure 1**.

MATERIALS AND EQUIPMENT

Participants

The number of participants needed when conducting the study according to this protocol mostly depends on the desired effect size we would like to be able to detect and the statistical power we would like to achieve. For example, using the multivariate analysis of variance, if we would like to achieve a power of at least $1-\beta = 0.90$ to detect small effects ($f^2 = 0.02$) with the probability of *Type I* error set at $\alpha = 0.05$, we would need a sample size of $N = 439$ as calculated using *G*Power 3* (Faul et al., 2007). However, given that it is expected that researchers will have to omit some participants whose scores do not meet the criteria for individualizing and binding groupings which will be determined from the data [due to calls-to-action not being (in)congruent enough with their moral views], it would be advisable to further increase the number of participants by 20%. Therefore, the ideal final sample would be around $N = 527$. Of course, if one would, for example, like to study the effects of (in)congruent calls-to-action and norms on donations to two different charity causes separately (e.g., those hypothesized to be more in line with liberals' and conservatives' values), one would need to double this number. On the other hand, if one is willing to accept somewhat lower power of $1-\beta = 0.80$, that is usually accepted in psychological research, one would need the sample

³https://en.wikipedia.org/wiki/Facebook/%E2%80%9993Cambridge_Analytica_data_scandal



size of $N = 344$, or around $N = 412$ if one decides to increase it by 20%.

The recruitment of participants can take place online. For example, in order to obtain a mixture of conservative and liberal participants, the link to the survey can be disseminated through different conservatively and liberally oriented Facebook groups (see **Table A1** in **Appendix A**). Other online platforms where it is relatively easy to approach the users of different ideological positions can be also utilized, such as Reddit, Twitter etc.

Facebook App

In this part we will describe the implementation of the survey and data collection using the custom made Facebook app. The Facebook app has several advantages over more traditional online surveys. First, it offers the possibility to customize the procedure and randomize some of its steps, which will be a useful feature for this study. Second, it allows automatic calculation of participants' scores and instant customized feedback for each participant, which acts as a motivator for participation. Third, it allows us to collect some of the Facebook data that will be crucial for our exploratory part of the study (i.e., predicting the moral foundations). Finally, it facilitates the dissemination of the study by allowing the participants to share their scores and invite others to participate. A clickable URL will take participants to a custom-made Facebook app that will host the survey. The survey will include measurements of moral foundations, moral identity internalization, attitudes toward CB, descriptive social norms and CB in addition to relevant (in)congruent calls-to-action for the charities. The charities will be used based on the pre-study results. The wireframe of an app is presented in **Figure A1** in **Appendix A**.

However, as noted in the introduction, it is perfectly possible to conduct the study based on this research protocol without collecting Facebook data or developing the custom Facebook app. In this case one could simply use one of the many existing research platforms to create the survey and disseminate it over different online platforms.

Pre-study

For the pre-study purposes, all the measures except the MFQ were translated into German, Dutch, and Italian (i.e., the calls-to-action, Moral Identity Internalization Scale and the measures of attitudes and social norms) The translated copies of the MFQ were already publicly available on the original authors' website⁴. The translation of the other scales was done using a forward and back translation method, in which a native speaker first translated the original version of the measure from English to the target language (German/Dutch/Italian), trying to keep conceptual rather than literal meaning. Next, a separate bilingual speaker translated the translation back to English. Finally, the group of authors reviewed and compared the translations. Consensus among the authors was reached before continuing with pilot testing.

After translation of the materials, we conducted the pre-study ($N = 50$). There were several goals of this pre-study. First,

we confirmed that there were no ambiguities and confusions regarding the translations and instructions. Second, the pilot study was used to check the appropriateness of the calls-to-action which are the main independent variable. Specifically, we wanted to receive feedback on whether our calls-to-action were perceived as being based on the individualizing and binding moral foundations. To do this, we first briefly familiarized our participants with the MFT and each of the moral foundations. Thereafter, we asked them to estimate on a seven-point scale the degree to which each of the calls-to-action rely on each of the five foundations. As we specifically created our calls-to-action using words from the *Moral Foundations Dictionary* (it can be found online: <http://moralfoundations.org/othermaterials>), we expected that our individualizing calls will be estimated to rely mostly on the Care/Harm and Fairness/Cheating foundations. Accordingly, we expected binding calls to be estimated to rely more heavily on the Loyalty/Betrayal, Authority/Subversion, and Sanctity/Degradation foundations.

Third, we wanted to examine whether different charities were perceived as being more acceptable for liberals, more acceptable for conservatives or universally acceptable. Specifically, we listed brief descriptions of 12 different charities and asked participants to estimate whether they thought they would be more acceptable for liberals or conservatives, scored on a 7-point-scale ($-3 = \text{liberal}$, $3 = \text{conservative}$). The final goal of the pre-study was to test the main dependent variables of donating time and money. Specifically, we wanted to estimate the proportion of participants that were willing to donate any portion of money and time. Participants were asked to imagine that they had won a £/€50 Amazon gift card. Consecutively, they were asked whether they were willing to donate money out of this £/€50 to one of the charities that had previously been described and if they were interested in volunteering for one of these charities.

Generally, results of this pre-study indicated that the estimations of our calls-to-action were in the intended directions. Specifically, the individualizing call-to-action was estimated to be substantially more in line with the Care/Harm ($M = 5.58$; $SD = 1.50$) and Fairness/Cheating ($M = 4.88$; $SD = 1.52$) foundations than was the binding call-to-action (Care/Harm: $M = 3.92$; $SD = 2.11$; Fairness/Cheating: $M = 3.82$; $SD = 1.60$). Conversely, the binding call-to-action was estimated to be substantially more in line with the Loyalty/Betrayal ($M = 4.50$; $SD = 1.64$), Authority/Subversion ($M = 4.90$; $SD = 1.82$), and Sanctity/Degradation ($M = 3.38$; $SD = 2.02$) foundations than the individualizing one (Loyalty/Betrayal: $M = 3.32$; $SD = 1.71$; Authority/Subversion: $M = 2.72$; $SD = 1.75$; Sanctity/Degradation: $M = 2.74$; $SD = 1.70$). Estimations for the neutral call-to-action were between these two estimations for each foundation (**Figure C1** in **Appendix C**).

Furthermore, results indicated that participants mostly perceived the charities as being more appropriate for liberals than conservatives, with several charities perceived as being more or less ideologically neutral (see **Figure C2** in the **Appendix C**). For example, *European association for cancer research*, *Eurochild* and *MAGmine* were perceived to be

⁴<http://moralfoundations.org/questionnaires>

relatively neutral, while the *Group for transcultural relation* was perceived as the most liberal one. These results can be used by any researcher that decides to draw on this protocol for studying CB. Finally, 86% of the participants in the pre-study were willing to donate at least some money out of this £/€50 ($M = 19.84$; $SD = 15.67$), and 78% were interested in volunteering for one of the charities ($M = 4.00$; $SD = 2.21$).

The Independent Variables

Moral Foundations

Participants' moral foundations will be assessed using the latest version of the MFQ (Graham et al., 2008; see <http://www.moralfoundations.org> for more information and to access the questionnaire). The questionnaire comprises 30 items, asking participants about the extent to which different factors are relevant in their moral decision-making. Each of the five foundations is assessed with six items, and the participants' task is to evaluate the relevance of an item or agreement with the item on a six-point scale (0 = not at all relevant/strongly disagree; to 5 = extremely relevant/strongly agree). To obtain the final score on each of the subscales, the evaluations are summed up and consecutively divided by the number of items related to the particular subscale. A score between 0 and 5 will be calculated for each of the moral foundations subscales. The internal consistencies of the subscales were reported to be $\alpha = 0.69$ (Harm), $\alpha = 0.65$ (Fairness), $\alpha = 0.71$ (Ingroup), $\alpha = 0.74$ (Authority) and $\alpha = 0.84$ (Purity) (Graham et al., 2011). The MFQ is open-access and freely accessible for research purposes.

Moral Identity Internalization

Moral identity internalization will be measured using a five-item subscale ("Internalization" subscale) of Aquino and Reed (2002) "Moral Identity Scale". This scale is also open access for researchers. The Internalization scale measures the degree to which moral traits are central to an individual's self-concept. The participants' task is to evaluate every item using a 5-point Likert scale (1 = strongly disagree; to 5 = strongly agree). The internalization scale

has been shown to have good internal consistency with $\alpha = 0.73$.

(In)congruent Calls-to-Action

We designed three types of calls-to-action with the goal of having one that is neutral, one that is congruent with liberals' moral foundations and one that is congruent with conservatives' moral foundations. The assumption is that most participants will be grouped into two categories: those who mostly rely on individualizing foundations, and those who mostly rely on binding foundations. Thus, two types of morally relevant calls-to-action were used. The first emphasized individualizing foundations of Care/Harm and Fairness/Cheating, whereas the second one emphasized binding foundations of Loyalty/Betrayal, Authority/Subversion and Sanctity/Degradation. The final, neutral, call-to-action was constructed without involving any moral foundations. The proposed morally (in)congruent and neutral calls-to-action are presented in **Table 1**. We are presenting hypothetical calls to donate to two of the charities we used in the pre-study, *Eurochild* and *City of Sanctuary*. Incidentally, one is also estimated to be more liberal (i.e., *City of Sanctuary*), while the other appears to be relatively ideologically neutral (i.e., *Eurochild*; **Figure C2** in **Appendix C**).

Normative Information

Normative information will either be included or omitted in addition to the calls-to-action. This normative message will read "In similar studies, the majority of people decided to donate some of the money should they win it," followed by a question about the participants' willingness to donate (see **Table 1**). Although the normative message could be stronger than the one suggested (i.e., "In **this** study, the majority of participants decided to donate"), it would risk being deceitful if it turned out to be inaccurate. The one we are proposing is subtler but non-deceitful, as it is based both on our pre-study results and various previous studies that showed that a substantial number of participants actually decides to donate. For example, in Nilsson et al. (2016) study, 60 and 62% of the participants donated some of the money they earned in the study, while in Reed et al. (2007) study, 90% of the participants decided to donate either time or money.

TABLE 1 | Calls-to-action that are going to be used in the study.

Individualizing call-to-action	Show us your compassion by supporting <i>City of Sanctuary/Eurochild</i> in achieving their goal to <i>care for refugees/protect children's well-being</i> . Helping vulnerable refugees/children and preventing suffering is the right thing to do . By supporting <i>City of Sanctuary/Eurochild</i> you will get the chance to increase fairness and justice and reduce harm all over the world. Show us that you care for others and make a difference now! * In similar studies, the majority of people decided to donate some of the money should they win it. (included only in the social-norm condition).
Binding call-to-action	Show us your integrity by supporting <i>City of Sanctuary/Eurochild</i> in achieving their goal to <i>support refugees/promote children's well-being</i> . Protecting innocent refugees/children and their families is your civil duty . By supporting <i>City of Sanctuary/Eurochild</i> you will get the chance to follow proud religious and national leaders doing their obligation . Show us that you are a responsible community member and join the fight now! * In similar studies, the majority of people decided to donate some of the money should they win it. (included only in the social-norm condition).
Neutral call-to-action	Support <i>City of Sanctuary/Eurochild</i> in achieving their goal to <i>support refugees/promote children's wellbeing</i> . You can help make a real difference to this very important issue. By supporting <i>City of Sanctuary/Eurochild</i> you will play an active role in finding a peaceful outcome. Show us that you are willing to support the cause now! * In similar studies, the majority of people decided to donate some of the money should they win it. (included only in the social-norm condition).

Bolded words represent the individualizing and binding moral frames. Italicized words pertain to the charity dealing with refugees and children, respectively.

Apart from these measurements, participants will also be asked to provide some basic demographic information such as age, gender, education, nationality, ideology, and previous donations.

The Dependent Variables

Attitudes Toward CB

In line with proposals for constructing TPB questionnaires (Ajzen, 2006), attitudes will be assessed with several semantic differential scales. Participants will respond to the following question: "I believe that making a donation to a described charity in terms of money or time would be:" (1) unpleasant – (7) pleasant, useful – useless, satisfying – unsatisfying, favorable – unfavorable, positive – negative, considerate – inconsiderate, pointless – worthwhile, and bad – good. Items will be scored such that higher scores indicate a more positive attitude toward CB. The internal consistency of a similar scale was shown to be $\alpha = 0.93$ (Smith and McSweeney, 2007).

Descriptive Norms

Similarly to several previous studies (e.g., Smith and McSweeney, 2007; Croson et al., 2009; van der Linden, 2011), descriptive norms will be assessed with the following items: "How likely do you think is that people you know would donate to this charity?" (1 – very unlikely; 7 – very likely); "How much do you think an average person doing this survey would donate if they won the gift card?" (1 – nothing; 7 – all of it); "How many of the people you know would donate to this charity?" (1 – none of them; 7 – all of them); "How many of the people doing this survey would donate to this charity?" (1 – none of them; 7 – all of them). In previous studies, the reliability of similar four-item measurements of descriptive norms was found to be $p = 0.76$ (Smith and McSweeney, 2007).

Charitable Behavior

Since CB is not a unidimensional construct but can be classified into three main categories - helping a stranger, giving time and giving money (Charities Aid Foundation [CAF], 2017) – in this study we decided to measure CB in two different ways:

Time donated

The time participants are willing to donate to our charity will be assessed through participants' interest to volunteer for the charity. Upon finishing with the questionnaires, participants will be asked the following question: "How interested are you in donating your time to volunteer for the charity?". The question will be scored on a 7-point Likert scale (1 = not interested, 7 = very interested).

Money donated

As a part of the study, we will be giving three €/£50 Amazon gift cards to three random participants. Thus, each participant has an equal chance to win a €/£50 Amazon gift card. The amount of money participants are willing to donate to our partner charity will be operationalised as the amount they are willing to donate if they were to win the Amazon gift card. Therefore, participants will answer the following question: "If you were to win the €/£50 Amazon gift card, how much money out of the 50€/£ would you be willing to donate to this charity?" If a participant decides

to donate e.g., 20€/£, should he/she actually win the gift card, he/she would receive a 30€/£ Amazon gift card, while 20€/£ would actually be donated to the charity. All the measurements that will be used in this study are presented in the **Appendix B**.

Design and Procedure

The procedures described in this protocol are partly based on several other studies. Specifically, when discussing the (in)congruent calls-to-action, we are following the procedures developed in the Winterich et al. (2012) and Kidwell et al. (2013) studies. Winterich et al. (2012) manipulated the description of the charities to be in line either with individualizing or binding foundations, and measured the intentions to donate a part of an Amazon \$50 gift card, should the participant win it. However, unlike Winterich et al. (2012) who manipulated the charity description, we decided to manipulate the calls-to-action which is more in line with Kidwell et al. (2013) study in which they constructed individualizing and binding appeals for recycling.

Although neither of these two studies referred to attitudes and social norms, we nevertheless decided to address them in the protocol, as they seem to be important determinants of CB. Here, we draw on the methodology used in the Smith and McSweeney (2007) and van der Linden (2011) studies, but especially on Croson et al.'s (2009) study, who presented participants with descriptive information and measured its influence on the perception of social norms. Similarly, we are measuring both social norms and attitudes in order to gain insights not only into whether (in)congruent appeals and normative information influence donations, but also into the mechanisms of that relationship.

In sum, although the current protocol is based on several well-established procedures effectively used in previous studies, by combining them and adding several new features, we believe we managed to create a study protocol capable of yielding rich and comprehensive insights into the relationship between morality, attitudes, norms and CB.

STEPWISE PROCEDURE

This experimental study uses a 3 (morally (in)congruent/neutral calls-to-action) \times 2 (presence/absence of social norm) between-subjects design and can be conducted online through a custom developed app or existing survey platform. The study procedure can be split into eight steps. (1) Firstly, participants receive a brief description of the study and its goals (e.g., "we want to explore the relationship between morality and CB and test the effectiveness of appeals to donate") and are asked to provide informed consent for participation. (2) Secondly, participants complete the MFQ and the Moral Internalization subscale. (3) Thirdly, participants are asked to complete a basic demographic questionnaire. (4) Thereafter, the participants are randomly assigned to one of the two different charity causes and to one of six different experimental conditions. We use random assignment to ensure that samples are similar across conditions in terms of observed and unobserved characteristics. After the experimental exposure, participants are first asked about their attitudes toward charitable

behavior and perception of social norms regarding the CB (5) and are then asked to indicate their willingness to donate (a) time and (b) money to the charity they were presented (6). The sequence of presenting the questions to donate time versus money will vary randomly in order to minimize the potential influence of sequence on both donation measures. Next, the participants are provided with a feedback regarding their scores on the MFQ and moral identity internalization questionnaire (7). Finally, an optional step is asking permission to collect participants' Facebook data. Participants can either accept or deny this request. Regardless of their choice, in the last step, debriefing about the study and contact for further inquiries are provided (8). See **Figure 2** for a visualization of this procedure.

STATISTICAL ANALYSIS

To test for the effects of moral congruence of the calls-to-action, descriptive social norms, and their interaction on donations of time and money, a multivariate analysis of variance (MANOVA) will be conducted. In the same analysis, a test for the moderating effect of moral internalization on the congruence of the calls-to-action will also be conducted. This approach allows researchers to account for potential relationships between our dependent variables, and to test whether the experimental manipulations influence participants on a combination of different types of donations (Field, 2009). Furthermore, in comparison with univariate ANOVA, that requires several individual tests, using MANOVA reduces the probability of making a *Type I* error.

To test whether attitudes toward CB mediate the effects of (in)congruence of the calls-to-action on CB, and whether descriptive social norms mediate the effects of normative information on CB, a mediation analysis will be conducted. This can be done by using the bootstrapping method developed by Preacher and Hayes (Hayes, 2013). This is because previous approaches related to this analysis, such as the causal step approach (Baron and Kenny, 1986), have been criticized for lack of power, the underlying assumptions and inability to directly evaluate potential mediation (MacKinnon et al., 2002).

To determine if people's moral foundations can be predicted from their Facebook data, various machine learning algorithms can be employed⁵. However, as this part of the study is exploratory and primarily concerned with maximizing prediction accuracy, we cannot specify the exact algorithm which will be reported. Various models and approaches can be tried out, choosing the one that exhibits the least amount of generalization error, i.e., the one that performs best on unseen data, which will be determined through cross-validation. Some possible approaches are LASSO (*Least Absolute Shrinkage and Selection Operator*) regression, multiple linear regression on clustered data and

⁵The first step is usually data preprocessing, which also includes trimming down the data set (i.e., removal of non-informative entries; e.g., a Facebook page that is *liked* by less than 150 users; Kosinski et al., 2016). After that, a form of dimensionality reduction is usually applied to the data (Kosinski et al., 2013, 2016). This is done because analyzing high-dimensional data sets can exacerbate issues such as model overfitting or multicollinearity (James et al., 2013). In the end, machine learning algorithms are trained in order to identify or predict users' characteristics.

principal component regression⁶ (James et al., 2013; Kosinski et al., 2016).

The analyses can be conducted using the R language (R Core Team, 2017). For implementing the various machine learning algorithms, we will use existing packages, such as glmnet (Friedman et al., 2010), irlba (Baglama et al., 2017) and topicmodels (Grün and Hornik, 2011).

IMPACT AND LIMITATIONS

Anticipated Results

Based on the theory and the literature reviewed in the introductory part, we can make some educated expectations regarding our results. First, in line with previous research (e.g., Winterich et al., 2012; Kidwell et al., 2013; Nilsson et al., 2016), we expect that our intervention in terms of morally (in)congruent calls-to-action will have a significant impact on CB. Specifically, we expect that morally congruent appeals, regardless of whether they are individualizing or binding, will foster significantly more CB compared to incongruent or neutral appeals.

Regarding the social norms, we expect that making the descriptive norms salient ("*most of the participants donate*") will influence one's subsequent CB. Although some of the previous studies did not find an effect of descriptive norms on CB (e.g., Smith and McSweeney, 2007; van der Linden, 2011), it seems that these studies failed to make the descriptive norms salient enough, which seems to be a prerequisite for them to be effective in influencing behavior (Cialdini et al., 1990). Thus, we hypothesize that participants who receive descriptive norms in their appeal will donate more of their money and time compared to those who do not receive a descriptive norm.

Since we expect both morally (in)congruent calls-to-action and descriptive social norms to have an effect on CB, we hypothesize that participants receiving a morally congruent call-to-action with a descriptive norm will donate the most. Furthermore, based on previous findings regarding the influence of moral identity internalization on CB (e.g., Winterich et al., 2012), we expect to find a significant two-way interaction effect of morally (in)congruent appeals and moral identity internalization on CB. A larger effect of morally (in)congruent appeals on CB is expected for those who are high on moral identity internalization as compared to those who are low on moral identity internalization (see **Figure 3**).

We also expect to find mediating effects of attitudes and descriptive norms on the relationship between (in)congruent calls-to-action and normative messages on the one side and CB on the other. More precisely, we expect that the individualizing/binding calls-to-action will have a positive impact on donation attitudes of those participants who are high on

⁶For example, Youyou et al. (2015) used LASSO regression when judging participants' scores on the Big Five dimensions. LASSO is an approach to regression problems that applies penalties for the number of variables in a model, and which can shrink certain variables' coefficients to zero, thus serving as a variable selection technique (James et al., 2013). Kosinski et al. (2013) first conducted a singular value decomposition on their dataset in order to reduce its dimensionality, and then trained linear and logistic regression models to predict users' characteristics.

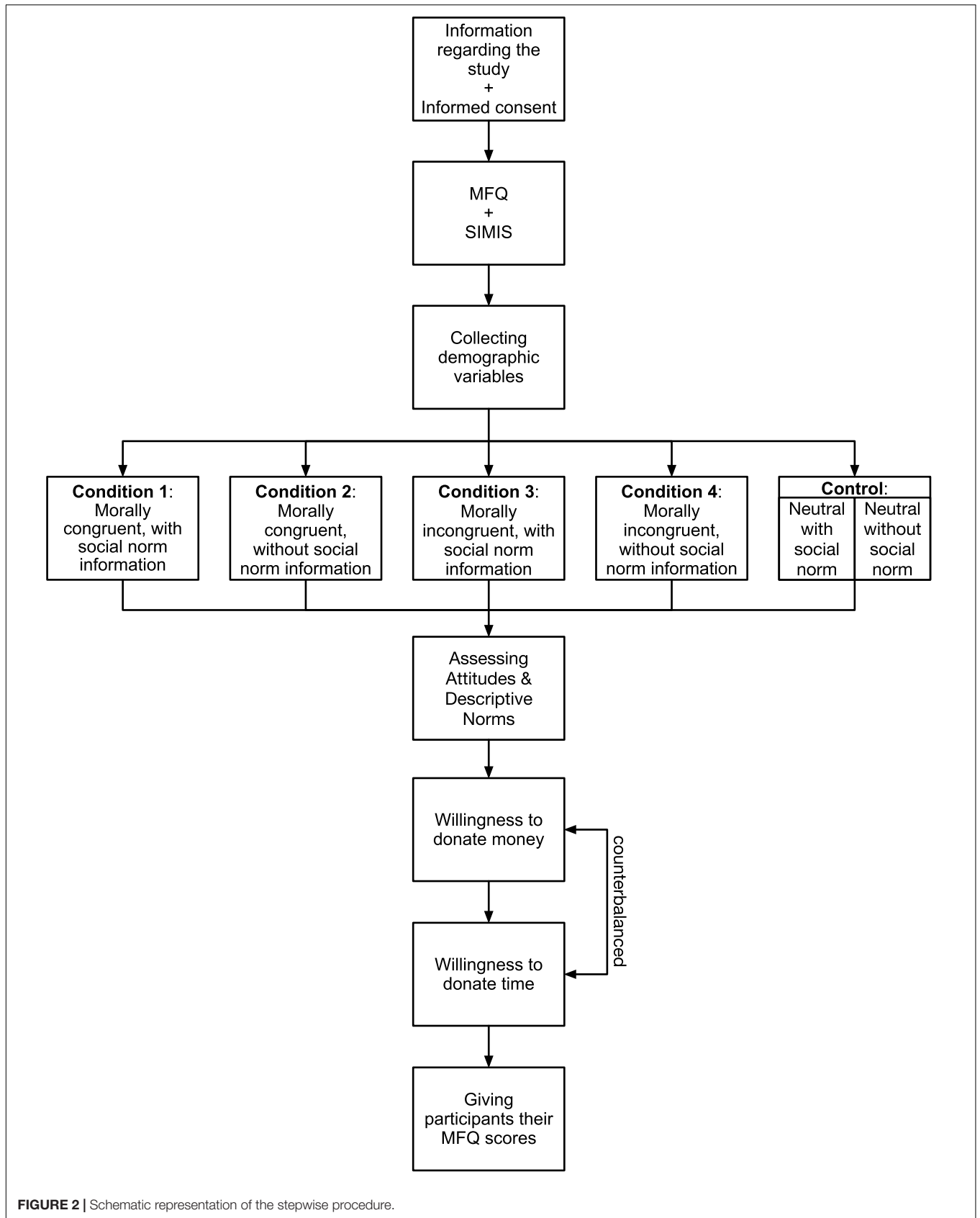


FIGURE 2 | Schematic representation of the stepwise procedure.

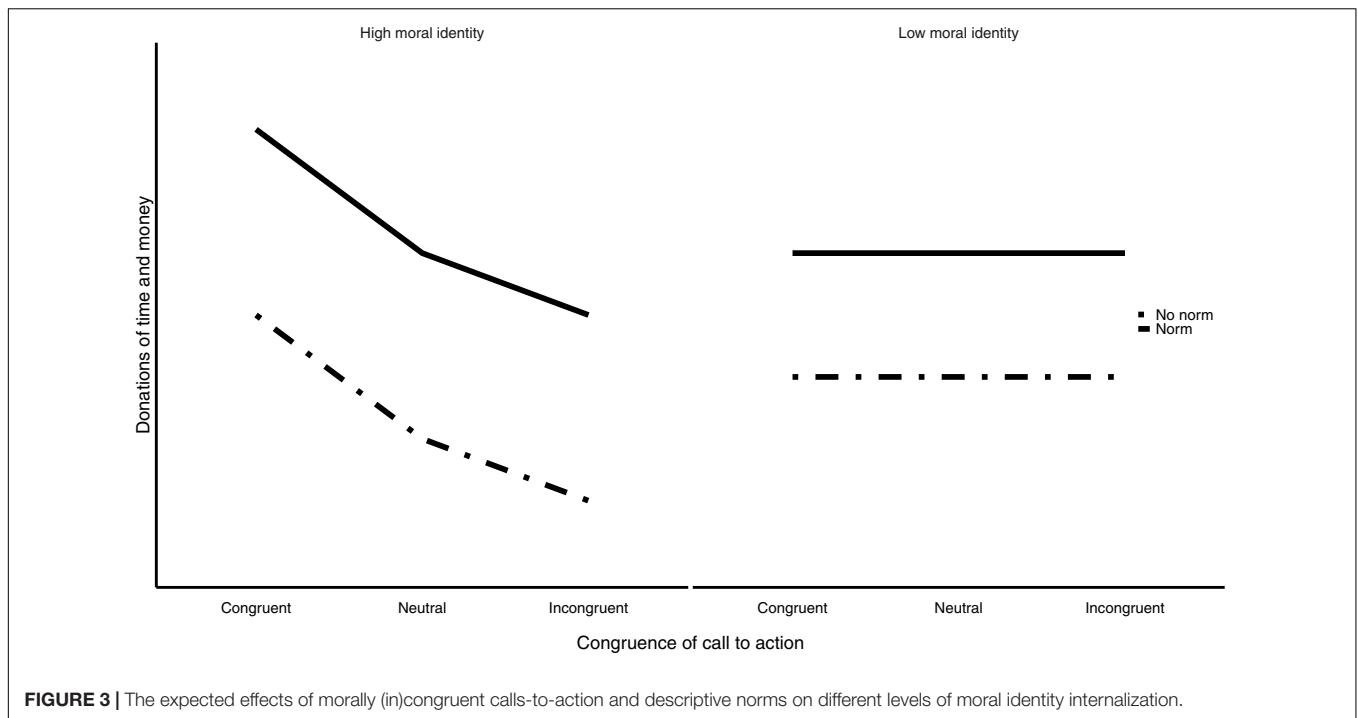


FIGURE 3 | The expected effects of morally (in)congruent calls-to-action and descriptive norms on different levels of moral identity internalization.

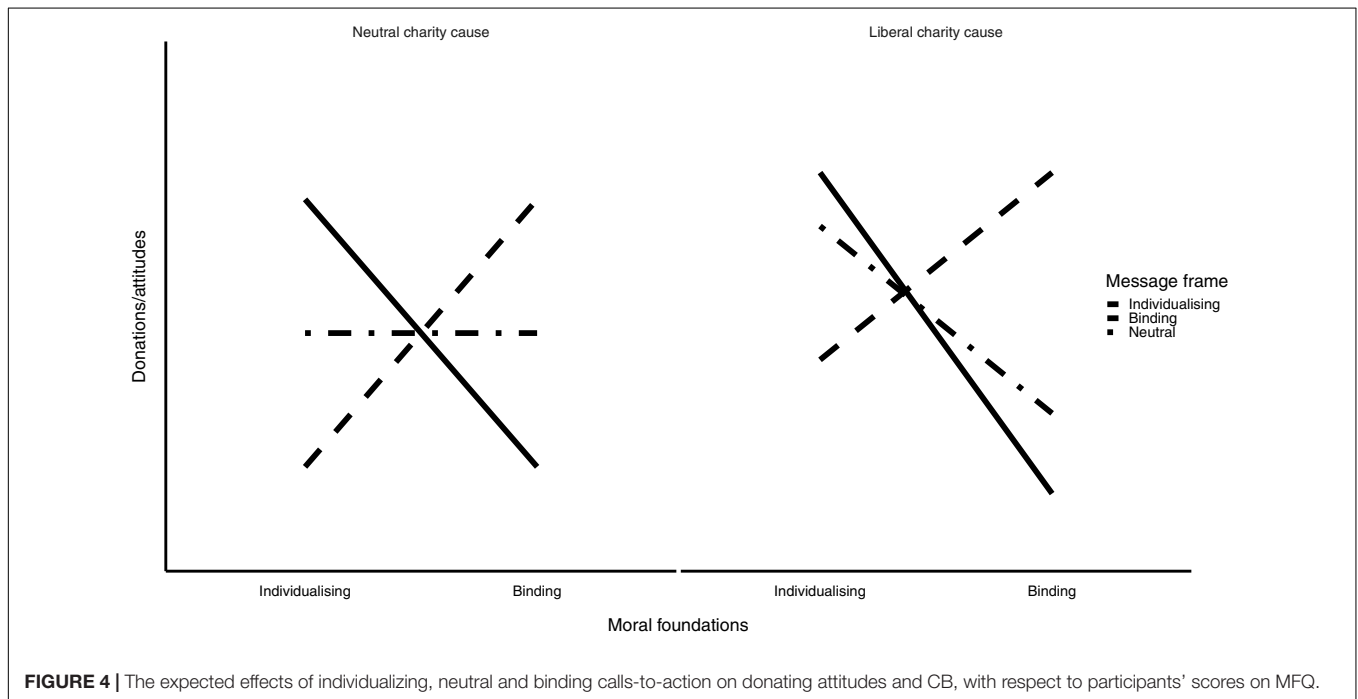


FIGURE 4 | The expected effects of individualizing, neutral and binding calls-to-action on donating attitudes and CB, with respect to participants' scores on MFQ.

individualizing/binding moral foundations, resulting in their willingness to donate more of their time and money. Since our second charity cause (refugee support) is already supposed to be more appealing for those high on individualizing foundations, we do not expect the individualizing calls-to-action to have a major additional impact on participants' donating attitudes or behavior. On the other hand, we do expect that the binding calls-to-action will have a bigger impact, shifting the attitudes and increasing

CB of those higher on binding moral foundations, since they are expected to have relatively negative attitudes toward such charity. See **Figure 4** for a schematic representation of these effects.

Significance of Research

We believe that the current project makes an important contribution both in theoretical and in practical terms. In

theoretical terms, we hope to expand the knowledge about the scope of morally (in)congruent appeals. Rare previous research did examine the effects of the congruence between both the appeals and individual's morality, as well as charity causes and individual's morality. However, within the current project, we plan to explicitly contrast these two. We aim to explore whether the appeals tailored in line with one's morality can enhance one's donations to charitable causes, even in the case of charities for which there is no original affinity. This way, we contribute to the theoretical considerations of the role of morality in CB, but also sketch potential fundraising strategies for charities with different causes. Furthermore, another significant theoretical contribution of this project is the investigation of the role of attitudes in CB change. Specifically, within this project we not only want to show that congruent calls-to-action can have a positive impact on CB, but also to elucidate the mechanism through which this effect operates (i.e., by affecting attitudes toward CB).

Besides the effect of congruent appeals on attitudes and CB, within this project we will investigate the role of descriptive social norms in CB. Specifically, we want to test potential boundary conditions for the effectiveness of descriptive norms, especially related to their saliency. However, here we also want to go a step further and show that, if salient normative information indeed can affect CB, it probably does so through changing one's perceptions about descriptive norms related to CB. Therefore, we believe that the current project contributes to the theoretical knowledge in several important ways.

More importantly, however, we hope that the project will have useful practical implications. Specifically, we hope to be able to provide charities with some simple yet effective tools that they could utilize in subsequent fundraising campaigns. For example, charities will be able to benefit from the often easily attainable knowledge about the approximate ideological positions of their target group by changing the way in which they deliver information about their causes and frame donation pitches. In this way, they could decrease the risk of turning away people who would potentially become their donors, by addressing them the right way. Given the low implementation costs of these interventions, we believe that charities can significantly benefit from them even if the effect sizes are modest.

Looking at the previous literature, we can expect to find small to medium effect sizes of our interventions on donating behavior. For example, Winterich et al. (2012) found small to medium effect sizes for both charitable intentions and donations (Cohen's d between 0.28 and 0.46). Croson et al. (2009) found a medium impact of descriptive social norms on donations to public radio's (Cohen's $d = 0.58$). However, we believe that, from the perspective of charity donations, even modest effects can have big practical importance. Statistically speaking, even if our interventions exhibit only small effects on CB (Cohen's $d = 0.2$), this still means that 58% of donations from people from the *congruent appeal + norm* intervention will be higher than the mean donation from people from the *no intervention* condition. In case we find medium effects ($d = 0.5$), as much as 69% of the *intervention* donations will be higher than the mean of the *no intervention* donations (see <http://rpsychologist.com/d3/cohend/> for the visualization and interpretation of Cohen's d effect sizes).

Thus, taking into account the relative ease and low cost of the implementation of these interventions, we believe that any kind of improvement in terms of donations can be considered as practically significant for charity organizations.

Limitations

Despite the potential benefits, the present project is not without limitations. Specifically, the hypothetical bias is one potential problem that is well known in the literature. Hypothetical bias is the difference between the stated and revealed value of a certain good (Murphy et al., 2005). In terms of charitable giving, people often state they are willing to donate significantly more money than they would really donate. One reason for this could be related to loss aversion which is more pronounced in case of giving real money than hypothetical money. Although we acknowledge that this is a problem, we believe that our approach at least attenuates it. Specifically, in line with the findings of Murphy et al.'s (2005) meta-analysis, stating that choice-based elicitation could be important in reducing the bias, we are giving our participants the choice between several different ways of allocating the 50\$ Amazon card. Furthermore, we are providing real, valid gift cards from which at least some of the participants will actually donate money. Although this is still far from representing a real situation in which people are giving away their own money, we believe that it is a step away from a purely hypothetical situation in which no money would be donated. In their meta-analysis, Murphy et al. (2005) found that the median hypothetical bias fell between 1 and 1.5 (i.e., in hypothetical situations people donated between 1 and 1.5 times more than in real situations). Although impossible to know, we hope that with this approach the hypothetical bias in studies based on this protocol would not be much higher than this estimation.

However, it is important to note that, although loss aversion and hypothetical bias should affect the amount people would be willing to donate in absolute terms, it should not affect the relative differences among groups within the study. Specifically, as the participants will be randomly selected into different groups, we have no basis to believe that some groups will have significantly different hypothetical bias from others. Therefore, although absolute amounts would certainly be inflated, we believe that loss aversion and hypothetical bias should not play a significant role in our study, as the relative differences between the groups will still be observable.

A second limitation of the current study is grouping. Although we can expect the majority of our participants to fit into one of the two categories – those who rely more and those who rely less on binding foundations – some participants will not fit into this binary classification. For example, Iyer et al. (2012) showed that libertarians' pattern of moral foundations significantly differs from those of liberals and conservatives. This means that our (in)congruent calls-to-action will not be adequately tailored toward these participants, and this may impact the overall effectiveness of our manipulations. However, most participants are expected to fit into our two categories, and those who do not will be randomly distributed between conditions. Future studies could benefit from constructing a wider range of calls-to-action to cover different types of moral foundation profiles.

Finally, we would like to address what we see as probably the most significant problem with following this protocol, and that is the Facebook app. Namely, not all researchers will be in a position to collect the data with a custom made app in order to gain access to participants' Facebook data. However, we already noted that the part of the protocol regarding accessing Facebook data and predicting moral foundations from it is purely exploratory, and independent of the former, confirmatory part. In other words, one can reconstruct all the steps included in the confirmatory part of the protocol (i.e., presentations of the questionnaires, calls-to-action, attitudes and norms measures and dependent variables of donating time and money) using only readily available online survey tools. We believe that such a study would be easily implemented, interesting and beneficial on its own.

Ethics

There are several important ethical concerns regarding some aspects of this protocol. First, using tailored persuasive messages in a marketing setting can yield both positive and negative outcomes. For example, it seems that people are much happier and more satisfied when spending money on things that are more congruent with their personality or needs (Matz et al., 2016). Thus, it might be expected that people will feel much better about themselves when making donations to charities whose causes and appearance are in line with their moral worldviews. However, some grimmer scenarios readily come to mind, too. Namely, if it is possible to manipulate people's decisions and behaviors just by changing few words, this knowledge could be used for less noble purposes. For example, there are suggestions that the 2016 US presidential campaign used profiles of millions of US citizens and approached them with specifically tailored messages in order to make them stay home on the election day, possibly even swinging the election outcome in this way⁷. We fully agree that things like this one happen and will probably continue to happen. Even more, we can be sure that many companies that are not constrained by strict ethical guidelines put large amounts of money into testing different kinds of persuasive messages on a daily basis. The problem is that this mainly happens outside of user's awareness and the knowledge gained remains private. Therefore, in our view, much more serious, ethically minded scientific research is needed in the area, because only in this way will knowledge, along with all the benefits, drawbacks and ethical concerns, enter into the public sphere and be useful for both the scientific community and the general public.

There are also certain concerns related to using data from social media to predict individual characteristics. For example, Lambiotte and Kosinski (2014) warn that it is possible for companies or individuals to use others' social media data to infer their personal characteristics (e.g., intelligence or sexual orientation), without them noticing or without their consent. It is especially troublesome that many users may be unaware of how revealing the information available on social media can

be (Kosinski et al., 2014). Youyou et al. (2015) point out that personal information can also be used to manipulate or influence people for illicit purposes. Similarly, Matz and Netzer (2017) state that social media data could be used to target individuals who are prone to impulsive or addictive behavior with ads for online casinos. These concerns are real and relevant in today's digital societies.

However, the outlook does not have to be so bleak. Kosinski et al. (2015) argue that, for instance, Facebook users have far more control over their data than is usually assumed. According to the authors, Facebook requires its users to give consent to applications that want to use their data, and allows them to limit or revoke access to their data after it has been granted. Furthermore, users can be informed of the specific data that is being collected and the way it is being used, allowing them to make an informed decision (Kosinski et al., 2015). Also, once collected, the data can be anonymised and made unrelated to the specific person prior to conducting analyses. We believe that these, and other steps, can be taken to make the data collection and analysis processes transparent and ethical, safeguarding the participants' privacy.

Finally, this protocol obtained an official approval from the ethical board of the Department of Psychology, University of Zagreb. Each participant will provide consent for participation and data sharing (just the Facebook likes and no other personal information such as name, residence, friends etc.), will be familiarized with the overall goals of the study and will receive feedback detailing the exact goals of the study as well as the description and explanation of their own scores on the measures.

AUTHOR CONTRIBUTIONS

All authors listed have made substantial contributions to this project, and have approved the manuscript for publication. NE provided the initial idea and design for the study. All other authors contributed equally to the research design and to the preparation of this protocol.

ACKNOWLEDGMENTS

This study was conducted as part of the Junior Researcher Programme (JRP). We would like to thank Lindsey van Bokhorst and Dr. Kai Ruggeri for their helpful comments on an earlier draft, and would like to thank the entire JRP team for their support and dedication.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpsyg.2018.01913/full#supplementary-material>

⁷ https://motherboard.vice.com/en_us/article/mg9vvn/how-our-likes-helped-trump-win

REFERENCES

- Ajzen, I. (1985). "From intentions to actions: a theory of planned behavior," in *Action Control: From Cognition to Behavior*, eds J. Kuhl and J. Beckman (Heidelberg: Springer), 11–39.
- Ajzen, I. (1991). The theory of planned behavior. *Organ. Behav. Hum. Decis. Process.* 50, 179–211. doi: 10.1016/0749-5978(91)90020-T
- Ajzen, I. (2006). *Constructing a Theory of Planned Behavior Questionnaire*. Available at: <http://people.umass.edu/ajzen/pdf/tpb.measurement.pdf>
- Aquino, K., and Reed, A. (2002). The self-importance of moral identity. *J. Pers. Soc. Psychol.* 83, 1423–1440. doi: 10.1037/0022-3514.83.6.1423
- Armitage, C. J., and Conner, M. (2001). Efficacy of the theory of planned behaviour: a meta-analytic review. *Br. J. Soc. Psychol.* 40, 471–499. doi: 10.1348/014466601164939
- Baglama, J., Reichel, L., and Lewis, B. W. (2017). *Irlba: Fast Truncated Singular Value Decomposition and Principal Components Analysis for Large Dense and Sparse Matrices. R package version 2.2.1*.
- Baron, R. M., and Kenny, D. A. (1986). The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *J. Pers. Soc. Psychol.* 51, 1173–1182. doi: 10.1037/0022-3514.51.6.1173
- Bobbio, A., Nencini, A., and Sarrica, M. (2011). Il Moral Foundation Questionnaire: analisi della struttura fattoriale della versione italiana. *G. Di Psicol.* 5, 7–18.
- Cesario, J., Higgins, E. T., and Scholer, A. A. (2008). Regulatory fit and persuasion: basic principles and remaining questions. *Soc. Personal. Psychol. Compass* 2, 444–463. doi: 10.1111/j.1751-9004.2007.00055.x
- Charities Aid Foundation [CAF] (2017). *World Giving Index*. Available at: <http://www.cafonline.org>
- Cialdini, R. B., Reno, R. R., and Kallgren, C. A. (1990). A focus theory of normative conduct: recycling the concept of norms to reduce littering in public places. *J. Pers. Soc. Psychol.* 58, 1015–1026. doi: 10.1037/0022-3514.58.6.1015
- Conover, M. D., Goncalves, B., Ratkiewicz, J., Flammini, A., and Menczer, F. (2011). "Predicting the political alignment of twitter users," in *Proceedings of the Privacy, Security, Risk and Trust (PASSAT) and 2011 IEEE 3th International Conference on Social Computing (SocialCom)*, (Piscataway, NJ: IEEE), 192–199. doi: 10.1109/PASSAT/SocialCom.2011.34
- Croson, R., Handy, F., and Shang, J. (2009). Keeping up with the joneses: the relationship of perceived descriptive social norms, social information, and charitable giving. *Nonprofit Manag. Leadersh.* 19, 467–489. doi: 10.1002/nml
- Faul, F., Erdfelder, E., Lang, A. G., and Buchner, A. (2007). G* Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behav. Res. Methods* 39, 175–191. doi: 10.3758/BF03193146
- Feinberg, M., and Willer, R. (2013). The moral roots of environmental attitudes. *Psychol. Sci.* 24, 56–62. doi: 10.1177/0956797612449177
- Field, A. (2009). *Discovering statistics using SPSS*. London: Sage Publications.
- Friedman, J., Hastie, T., and Tibshirani, R. (2010). Regularization paths for generalized linear models via coordinate descent. *J. Stat. Softw.* 33, 1–22. doi: 10.18637/jss.v033.i01
- Graham, J., Haidt, J., Koleva, S., Motyl, M., Iyer, R., Wojcik, S., et al. (2012). Moral foundations theory: the pragmatic validity of moral pluralism. *Adv. Exp. Soc. Psychol.* 47, 55–130. doi: 10.1016/B978-0-12-407236-7.00002-4
- Graham, J., Haidt, J., and Nosek, B. A. (2008). *The Moral Foundations Questionnaire*. Available at: <http://www.MoralFoundations.org>
- Graham, J., Haidt, J., and Nosek, B. A. (2009). Liberals and conservatives rely on different sets of moral foundations. *J. Pers. Soc. Psychol.* 96, 1029–1046. doi: 10.1037/a0015141
- Graham, J., Nosek, B. A., Haidt, J., Iyer, R., Koleva, S., and Ditto, P. H. (2011). Mapping the moral domain. *J. Pers. Soc. Psychol.* 101, 366–385. doi: 10.1037/a0021847
- Grün, B., and Hornik, K. (2011). Topicmodels: an R package for fitting topic models. *J. Stat. Softw.* 40, 1–30. doi: 10.18637/jss.v040.i13
- Haidt, J., and Graham, J. (2007). When morality opposes justice: conservatives have moral intuitions that liberals may not recognize. *Soc. Justice Res.* 20, 98–116. doi: 10.1007/s11211-007-0034-z
- Haidt, J., and Joseph, C. (2004). Intuitive ethics: how innately prepared intuitions generate culturally variable virtues. *Am. Acad. Arts Sci.* 133, 55–66.
- Hayes, A. F. (2013). *An Introduction to Mediation, Moderation, and Conditional Process Analysis*. New York, NY: The Guilford Press.
- Higgins, E. T. (2005). Value from regulatory fit. *Curr. Dir. Psychol. Sci.* 14, 209–213. doi: 10.1111/j.0963-7214.2005.00366.x
- Hirsh, J. B., Kang, S. K., and Bodenhausen, G. V. (2012). Personalized persuasion: tailoring persuasive appeals to recipients' personality traits. *Psychol. Sci.* 23, 578–581. doi: 10.1177/0956797611436349
- Iyer, R., Koleva, S., Graham, J., Ditto, P., and Haidt, J. (2012). Understanding libertarian morality: the psychological dispositions of self-identified libertarians. *PLoS One* 7:e42366. doi: 10.1371/journal.pone.0042366
- James, G., Witten, D., Hastie, T., and Tibshirani, R. (2013). *An Introduction to Statistical Learning*. New York, NY: Springer. doi: 10.1007/978-1-4614-7138-7
- Kandori, M. (1992). Social norms and community enforcement. *Rev. Econ. Stud.* 59, 63–80. doi: 10.2307/2297925
- Kidwell, B., Farmer, A., and Hardesty, D. M. (2013). Getting liberals and conservatives to go green: political ideology and congruent appeals. *J. Consum. Res.* 40, 350–367. doi: 10.1086/670610
- Kosinski, M., Bachrach, Y., Kohli, P., Stillwell, D., and Graepel, T. (2014). Manifestations of user personality in website choice and behaviour on online social networks. *Mach. Learn.* 95, 357–380. doi: 10.1007/s10994-013-5415-y
- Kosinski, M., Matz, S. C., Gosling, S. D., Popov, V., and Stillwell, D. (2015). Facebook as a research tool for the social sciences: opportunities, challenges, ethical considerations, and practical guidelines. *Am. Psychol.* 70, 543–556. doi: 10.1037/a0039210
- Kosinski, M., Stillwell, D., and Graepel, T. (2013). Private traits and attributes are predictable from digital records of human behavior. *Proc. Natl. Acad. Sci. U.S.A.* 110, 5802–5805. doi: 10.1073/pnas.1218772110
- Kosinski, M., Wang, Y., Lakkaraju, H., and Leskovec, J. (2016). Mining big data to extract patterns and predict real-life outcomes. *Psychol. Methods* 21, 493–506. doi: 10.1037/met0000105
- Kosugi, E., Fujisawa, X., and Shimizu, N. (2014). "Verification of the moral foundations theory in Japan," in *Poster at the European Congress of Psychology 2011*, (Istanbul),
- Lambiotte, R., and Kosinski, M. (2014). Tracking the digital footprints of personality. *Proc. IEEE* 102, 1934–1939. doi: 10.1109/JPROC.2014.2359054
- MacKinnon, D. P., Lockwood, C. M., Hoffman, J. M., West, S. G., and Sheets, V. (2002). A comparison of methods to test mediation and other intervening variable effects. *Psychol. Methods* 7, 83–104. doi: 10.1037/1082-989X.7.1.83
- MacLaughlin, S. (2012). *Charitable Giving Report: How NonProfit Fundraising Performed in 2012*. Blackbaud Institute for Philanthropic Impact. Available at: <https://www.blackbaud.com/files/resources/downloads/2012.CharitableGivingReport.pdf>
- MacLaughlin, S. (2014). *Charitable Giving Report: How NonProfit Fundraising Performed in 2014*. Blackbaud Institute for Philanthropic Impact. Available at: <http://www.associations.gov.fr/IMG/pdf/how-nonprofit-fundraising-performed-in-2014.pdf>
- MacLaughlin, S. (2017). *Charitable Giving Report: How NonProfit Fundraising Performed in 2016*. Blackbaud Institute for Philanthropic Impact. Available at: <https://institute.blackbaud.com/asset/2016-charitable-giving-report/>
- Matz, S. C., Gladstone, J. J., and Stillwell, D. (2016). Money buys happiness when spending fits our personality. *Psychol. Sci.* 27, 715–725. doi: 10.1177/0956797616635200
- Matz, S. C., and Netzer, O. (2017). Using big data as a window into consumers' psychology. *Curr. Opin. Behav. Sci.* 18, 7–12. doi: 10.1016/j.cobeha.2017.05.009
- Murphy, J. J., Allen, P. G., Stevens, T. H., and Weatherhead, D. (2005). A meta-analysis of hypothetical bias in stated preference valuation. *Environ. Res. Econ.* 30, 313–325. doi: 10.1007/s10640-004-3332-z
- Nilsson, A., Erlandsson, A., and Västfjäll, D. (2016). The congruency between moral foundations and intentions to donate, self-reported donations, and actual donations to charity. *J. Res. Personal.* 65, 22–29. doi: 10.1016/j.jrp.2016.07.001
- Ponemon Institute (2018). *Cost of a Data Breach Study*. Available at: <https://securityintelligence.com/series/ponemon-institute-cost-of-a-data-breach-2018/>
- Quercia, D., Kosinski, M., Stillwell, D., and Croccroft, J. (2011). "Our twitter profiles, our selves: predicting personality with twitter," in *Proceedings of the 2011 IEEE Third International Conference on Privacy, Security, Risk and Trust (PASSAT) and 2011 IEEE Third International Conference on Social Computing*

- (*SocialCom*), (Piscataway: IEEE), 180–185. doi: 10.1109/PASSAT/SocialCom.2011.26
- R Core Team (2017). *R: A Language and Environment for Statistical Computing*. Available at: <https://www.r-project.org/>
- Reed, A. I. I., Aquino, K., and Levy, E. (2007). Moral identity and judgments of charitable behaviors. *J. Mark.* 71, 178–193. doi: 10.1509/jmkg.71.1.178
- Shang, J., and Croson, R. (2009). A field experiment in charitable contribution: the impact of social information on the voluntary provision of public goods. *Econ. J.* 119, 1422–1439. doi: 10.1111/j.1468-0297.2009.02267.x
- Smith, J. R., and McSweeney, A. (2007). Charitable giving: the effectiveness of a revised theory of planned behavior model in predicting donating intentions and behavior. *J. Comm. Appl. Soc. Psychol.* 17, 363–386. doi: 10.1002/casp.906
- The National Council for Voluntary Organisations [NCVO] (2017). *UK Civil Society Almanac 2017*. Available at: <https://data.ncvo.org.uk/almanac17/>
- Updegraff, J. A., Sherman, D. K., Luyster, F. S., and Mann, T. L. (2007). The effects of message quality and congruency on perceptions of tailored health communications. *J. Exp. Soc. Psychol.* 43, 249–257. doi: 10.1016/j.jesp.2006.01.007
- van der Linden, S. (2011). Charitable intent: a moral or social construct? A revised theory of planned behavior model. *Curr. Psychol.* 30, 355–374. doi: 10.1007/s12144-011-9122-1
- Winterich, K. P., Zhang, Y., and Mittal, V. (2012). How political identity and charity positioning increase donations: insights from moral foundations theory. *Int. J. Res. Mark.* 29, 346–354. doi: 10.1016/j.ijresmar.2012.05.002
- Wolsko, C., Ariceaga, H., and Seiden, J. (2016). Red, white, and blue enough to be green: effects of moral framing on climate change attitudes and conservation behaviors. *J. Exp. Soc. Psychol.* 65, 7–19. doi: 10.1016/j.jesp.2016.02.005
- Youyou, W., Kosinski, M., and Stillwell, D. (2015). Computer-based personality judgments are more accurate than those made by humans. *Proc. Natl. Acad. Sci. U.S.A.* 112, 1036–1040. doi: 10.1073/pnas.1418680112
- Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.
- Copyright © 2018 Erceg, Burghart, Cottone, Lorimer, Manku, Pütz, Vlašiček and Willems. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Cross-Validation of the Spanish HP-Version of the Jefferson Scale of Empathy Confirmed with Some Cross-Cultural Differences

Adelina Alcorta-Garza^{1†}, Montserrat San-Martín^{2,3†}, Roberto Delgado-Bolton^{4,5}, Jorge Soler-González⁶, Helena Roig⁷ and Luis Vivanco^{3,5,8*}

¹ Service of Oncology, University Hospital Jose E. Gonzalez – Autonomous University of Nuevo León, Monterrey, Mexico,

² Scientific Computing Group, Logroño, Spain, ³ Ibero-American University Foundation, Barcelona, Spain, ⁴ Hospital San Pedro, Logroño, Spain, ⁵ Center for Biomedical Research of La Rioja, Logroño, Spain, ⁶ Department of Medicine (Gesec and Gerds Group), Faculty of Medicine, University of Lleida, Lleida, Spain, ⁷ Borja Institute of Bioethics, Ramon Llull University, Barcelona, Spain, ⁸ National Centre of Documentation on Bioethics, Logroño, Spain

OPEN ACCESS

Edited by:

Pietro Cipresso,
IRCCS Istituto Auxologico Italiano, Italy

Reviewed by:

Leonard Bliss,
Florida International University, USA
Dianna Theadora Kenny,
The University of Sydney, Australia
Joshua Fredrick Wiley,
Mary MacKillop Institute for Health
Research at Australian Catholic
University, Australia

*Correspondence:

Luis Vivanco
vivanco@riojasalud.es

[†]These authors have contributed
equally to this work.

Specialty section:

This article was submitted to
Quantitative Psychology and
Measurement,
a section of the journal
Frontiers in Psychology

Received: 29 March 2016

Accepted: 17 June 2016

Published: 12 July 2016

Citation:

Alcorta-Garza A, San-Martín M,
Delgado-Bolton R, Soler-González J,
Roig H and Vivanco L (2016)
Cross-Validation of the Spanish
HP-Version of the Jefferson Scale of
Empathy Confirmed with Some
Cross-Cultural Differences.
Front. Psychol. 7:1002.
doi: 10.3389/fpsyg.2016.01002

Context: Medical educators agree that empathy is essential for physicians' professionalism. The Health Professional Version of the Jefferson Scale of Empathy (JSE-HP) was developed in response to a need for a psychometrically sound instrument to measure empathy in the context of patient care. Although extensive support for its validity and reliability is available, the authors recognize the necessity to examine psychometrics of the JSE-HP in different socio-cultural contexts to assure the psychometric soundness of this instrument. The first aim of this study was to confirm its psychometric properties in the cross-cultural context of Spain and Latin American countries. The second aim was to measure the influence of social and cultural factors on the development of medical empathy in health practitioners.

Methods: The original English version of the JSE-HP was translated into International Spanish using back-translation procedures. The Spanish version of the JSE-HP was administered to 896 physicians from Spain and 13 Latin American countries. Data were subjected to exploratory factor analysis using principal component analysis (PCA) with oblique rotation (promax) to allow for correlation among the resulting factors, followed by a second analysis, using confirmatory factor analysis (CFA). Two theoretical models, one based on the English JSE-HP and another on the first Spanish student version of the JSE (JSE-S), were tested. Demographic variables were compared using group comparisons.

Results: A total of 715 (80%) surveys were returned fully completed. Cronbach's alpha coefficient of the JSE for the entire sample was 0.84. The psychometric properties of the Spanish JSE-HP matched those of the original English JSE-HP. However, the Spanish JSE-S model proved more appropriate than the original English model for the sample in this study. Group comparisons among physicians classified by gender, medical specialties, cultural and cross-cultural backgrounds yielded statistically significant differences ($p < 0.001$).

Conclusions: The findings support the underlying factor structure of the Jefferson Scale of Empathy (JSE). The results reveal the importance of culture in the development of medical empathy. The cross-cultural differences described could open gates for further lines of medical education research.

Keywords: empathy, physician, cross-culture comparison, Spanish, psychometrics

INTRODUCTION

Medical empathy is defined as a predominantly cognitive (rather than emotional) attribute that involves the ability to understand (rather than feel) patient's experiences, concerns and perspectives, and communicate this understanding (Hojat et al., 2002). Empathy has been listed consistently as one of the key elements of professionalism. The importance of empathy, as key element of professionalism, has been discussed in medical education and health care research (Veloski and Hojat, 2006), and in global bioethics (Vivanco and Delgado-Bolton, 2015).

Despite its importance in enhancing these relationships and improving patient care, research on physician empathy has been limited for two main reasons. Firstly, the theoretical investigation of physician empathy has been hampered by ambiguity in its conceptualization and definition. Secondly, empirical research in this area has been limited by a lack of tools to gauge the empathy of medical students and physicians. Nevertheless, the development of standardized instruments currently allows assessing the empathy in the interactions that take place in the context of healthcare (Hemmerdinger et al., 2007). One of the most popular instruments for this purpose is the Jefferson Scale of Empathy (JSE). Medical education researchers of the Jefferson Medical College, in the United States, developed this tool. The generic version of the scale was originally developed to measure medical students' orientations or attitudes toward empathic relationships in the context of patient care (Hojat et al., 2001). However, very soon there was a demand to use the scale for administration not only to medical students, but also to physicians and other health professionals involved in patient care, and all health professions students other than medical students. Thus, the authors decided to slightly modify the content of the generic scale so that three versions would be available (Hojat, 2006): one version for administration to medical students (the S-Version); a second version for administration to physicians and other practicing healthcare professionals (the HP-Version); and the third version for administration to students of all healthcare professions other than medical students (the HPS-Version).

Construct validity refers to the extent to which a test measures the theoretical constructs of the attribute that it aims to measure. In this sense, factor analysis of the JSE helps to determine whether the underlying factors of the scale are consistent with the theoretical constructs of the concept measured, being in this case empathy. Following this principle, the factor analysis of the generic scale of the JSE revealed four preliminary factors, which were consistent with the multifaceted concept of empathy reported in the literature (Spiro et al., 1993). The first of those factors included 10 items. This factor was called "the physician's

view of the patient's perspective." The second factor included five items, and it was called "understanding patient's experiences." The third factor, composed by two items, was called "ignoring emotions in patient care" (that refers to the opposite pole of standing in a patient's shoes). Finally, the fourth factor, composed by two items, was called "thinking like the patient." Following the recommendation of a minimum number of three items per factor (Velicer and Fava, 1998), the authors considered the last two factors had less stable factor pattern than the first two. Subsequent analysis showed that the first factor was the most salient among all other extracted factors.

The factor analysis of the HP-Version of the JSE showed three definitive underlying factors:

- (i) A first major factor that is composed by 10 positively worded items. This factor, described as the core cognitive ingredient of the empathy and as the stepping-stone in empathic engagement, was called "perspective taking." It is similar to the first factor described in the generic version.
- (ii) A second factor that is composed by eight negatively worded items. This factor was called "compassionate care." According with the authors, this construct is conceptually similar to the two factors that emerged in the generic version: "emotions in patients care" and "understanding patient's experiences."
- (iii) Finally, a third factor that is composed by two other negatively worded items. This factor was called "standing/walking in the patient's shoes" (the positive pole of the contents of the negatively worded but reverse-scored items). This last factor was described as a trivial component of the JSE, and similar to the factor "thinking like the patient" of the generic version.

According to the authors, these findings suggest that the factor structure of the JSE is consistent with the notion of the multidimensionality of empathy (Davis, 1983; Kunyk and Olson, 2001). In addition, the stability and the similarity between the factor structure and components across different samples (students and professionals) and across different versions of the scale provide, according to the authors, further support for the JSE's validity (Hojat, 2016).

Since its creation, both researchers and medical educators at international level have acknowledged the validity of the JSE. Its first cross-cultural adaptation was designed by Mexican researchers who administered the S-Version to medical students (Alcorta-Garza et al., 2005). Subsequently, the JSE was translated into 42 languages and is currently used worldwide in 60 countries located in Europe, the Middle East, Africa, Asia, North America, Latin America, and New Zealand (Hojat et al., 2011a). In order to improve the clarity of the scale for an international

audience, minor revisions were made in the wording of verbatim translation of some items that created some confusion in Italian and Spanish translations (Hojat, 2016). This is the case of the item 18: “I do not allow myself to be touched by intense emotional relationships between my patients and their family members” (a negatively worded item),” in the generic version. The symbolic meaning of “to be touched by” (to be affected or emotionally stirred) was not apparent in the translated versions. Therefore, the authors decided to replace “to be touched” by “to be influenced” (Hojat, 2016).

However, it is difficult to say whether due to these changes, to some unresolved translation issues, or due to cultural differences, the factorial position of this item remains still problematic in the factor analysis of some translations (Alcorta-Garza et al., 2005; Magalhaes et al., 2011; Tavakol et al., 2011b; Paro et al., 2012; Shariat and Habibi, 2013; Wen et al., 2013; Leombruni et al., 2014). Despite this issue, most of the studies using the JSE conducted in different countries report evidence supporting construct validity, criterion-related validity, predictive validity, internal consistency reliability, and test-retest reliability. In most of the cases, exploratory factor analysis using principal component analysis (PCA) with orthogonal rotation was used to determine the factor structure of the JSE. Exploratory factor analysis studies have often resulted in the three aforementioned factors (Alcorta-Garza et al., 2005; Paro et al., 2012; Wen et al., 2013). There are only a few adaptations where the factor structure of the JSE was studied using confirmatory factor analysis (CFA). In some cases, CFA was used to confirm a factor structure resulting from a previous PCA (Magalhaes et al., 2011; Tavakol et al., 2011b), and in others to confirm whether if the sample studied fitted the original theoretical model (Shariat and Habibi, 2013; Leombruni et al., 2014).

All these studies provide clues about the underlying components of the JSE, not only in samples from different disciplines, but also in a wide variety of cultural contexts. However, despite cumulative evidence, in a recent publication some of authors recognize the need to undertake additional research using samples from different professional and cultural contexts (Hojat and Lanoue, 2014).

Spanish is the second most widely spoken language in the world in terms of native speakers after Chinese. It is the official language of more than 20 countries, most of them in Latin America (Otero and Powell-Davies, 2011), which is where almost 90% of the population of native Spanish speakers lives. However, its different varieties along this territory involve significant cultural differences (Mato, 2008), which are even more noticeable when compared to the variety spoken in Spain (Oesterreicher, 2013). Conversely, Spain’s cross-cultural diversity is higher because of constant migratory flows. This is also reflected in the structure of the Spanish Healthcare System, with one of the highest levels of cultural diversity in the European area (Sánchez-Sagrado, 2013).

Both the cultural diversity resulting from the language, and the cross-cultural characteristics of Spain and Latin America, provide an ideal scenario to test the psychometric properties of the JSE (Delgado-Bolton et al., 2015). In addition, a better understanding of the role of culture in the development of communication skills,

professional behaviors, and lifelong learning abilities of health care practitioners is fundamental for the improvement of medical education, health management, and bioethics from a global scope (Vivanco and Delgado-Bolton, 2015).

This study has three purposes: to develop a validated translation of the JSE (HP-Version) that may be used by Spanish and Latin American researchers; to confirm the psychometric properties of the JSE in the Spanish language context; and to achieve a better understanding of the role of culture in the development of the medical empathy.

MATERIALS AND METHODS

Participants

The study is based on a sample of 896 healthcare professionals (physicians and physicians-in-training) involved in direct patient care in 13 healthcare institutions from Spain, Mexico, Colombia, Bolivia, and Argentina, who were invited to participate voluntarily and anonymously.

Instrument

The participants completed the JSE (HP-Version). This questionnaire is a psychometrically sound instrument developed specifically to measure physicians’ empathetic orientation in the context of patient care. The JSE includes 20 items, each answered on a 7-point Likert-type scale (1 = strongly disagree, 7 = strongly agree). Possible scores range from 20 to 140 and the higher the score, the greater the empathic orientation. The JSE identifies three factors: “perspective taking,” “compassionate care,” and “standing/walking in the patient’s shoes” (Hojat, 2016).

Complementary Information

Information about age, gender, professional status, medical specialty, country of birth, country of studies, and country of current residence was collected through a complementary survey.

Procedures

The original version of the JSE was translated into international Spanish, adapted, and reviewed using a cross-cultural back-translation procedure (Geisinger, 1994). Between 2014 and 2015, the translated version was administered to physicians and physicians-in-training from 13 institutions. The questionnaires consisted of paper forms provided together with an information letter in enclosed envelopes that were returned to the local researchers following a general protocol previously approved by an Independent Ethics Committee (Ref. CEICLAR PI 199). The work was carried out in accordance with the Declaration of Helsinki. There was no potential risk for participants, and anonymity was guaranteed throughout the process.

Statistical Assessment

Internal consistency reliability was calculated using Cronbach’s alpha coefficient. Following the guidelines suggested by the American Educational Research Association, values higher than 0.7 were considered satisfactory.

To consider the underlying factors, the data obtained for the 20 items of the JSE were subjected to exploratory factor analysis.

The purpose of this was to explore the association between the observed variables (items) and the latent variables (factors) using PCA with oblique rotation (promax) to allow for correlations among the extracted factors. The retained factors were limited to three so that the findings could be compared to the previously reported results of factor analysis (Hojat et al., 2002; Hojat and Lanoue, 2014). Retained factors were considered satisfactory when their eigenvalues were greater than one (Henson and Roberts, 2006).

The aim of the CFA tests was to discover whether if the observed data fitted a previously postulated model. In this study, the agreement of two models was tested using a CFA. In this regard, as opposed to a theory-generating model such as PCA, CFA is a theory-testing model that begins with a hypothesis prior to the analysis (Brown, 2014). This hypothesis can be based on theory, research, or both (Suhr, 2005). The first model tested in this study, Model A, is based on the 3-factor structure of the original HP-Version of the JSE, with American physicians (Hojat et al., 2002). The second model, Model B, is based on the 3-factor structure of the Spanish validated S-Version of the JSE, with Mexican students (Alcorta-Garza et al., 2005). The only difference between both models is the factor distribution of worded item 18: in Model A it is included in Factor 2 (compassionate care), whereas in Model B, it appears in Factor 3 (standing/walking in the patient's shoes). A third model, artificial Model AB, is based on a two-factor distribution hypothesis (Factors 2 and 3) for worded item 18, which was tested conducting preliminary CFA. This preliminary analysis was also used to test whether the underlying factors should be treated as correlated or uncorrelated.

Robust WLS is an estimation method used for structural equation modeling with ordinal observed variables with non-normality extremes ($Asymmetry > 3$ and $Kurtosis > 8$) (Muthén et al., 1997). Since the nature of the data meets these criteria, this was the estimation method used for the CFA. The goodness of fit indexes calculated to assess each model's fit were χ^2 statistics and its subsequent ratio with degrees of freedom (χ^2/df), Comparative Fit Index (CFI), Tucker-Lewis Index (TLI), Root Mean Square Error of Approximation (RMSEA), and Standardized Root Mean Square Residual (SRMR) (Muthén et al., 1997; Kline, 2005).

Group comparisons of empathy scores were performed. Gender, professional status (physicians and physicians-in-training), place of birth (Latin America and Spain), place of professional studies (Latin America and Spain), and residence (Latin America and Spain) were treated as dichotomous variables. Medical specialties were divided into the following groups: "non-hospital speciality" (this group included family medicine and occupational medicine specialties), "hospital speciality," "medical-surgical speciality," "surgical speciality," and "other specialties." For physicians without specialization, a "no speciality" group was created for non-specialist physicians. According to their migratory condition, physicians were divided into three groups: "Spaniards living in Spain," "Latin Americans living in Spain," and "Latin Americans living in Latin America."

All analyses were performed using R statistical software, version 3.1.1 for Windows. The statistical analyses of the data

also included multilevel (Bliese, 2013), nortest (Gross, 2012), and lavaan (Rosseel, 2012) packages.

RESULTS

Of the 896 participants who received the JSE, 715 were returned fully completed, giving an overall effective response rate of 80%. This response rate was higher than the minimum recommended to ensure the representativeness of the sample for mailed surveys to professionals (Gough and Hall, 1977).

The mean age was 35 years old with a 24–71 year-old age range ($SD = 10.8$). Three hundred and fifty-one (48%) of the physicians reported to be born in Spain and 347 (47%) of the physicians were born in Latin America. Thirteen countries were reported in this group (Mexico, Colombia, Bolivia, Argentina, Dominican Republic, Venezuela, Peru, Ecuador, Chile, Honduras, Cuba, El Salvador, and Uruguay). Seventeen physicians (2%) were born in non-Spanish-speaking territories. Eleven countries were reported in this group (Brazil, Italy, Ukraine, Morocco, Andorra, Belgium, Canada, France, Haiti, Moldova, and Ruanda). Finally, 18 (3%) physicians did not specify their country of birth.

The empathy score distribution, descriptive statistics, and reliability for the JSE in this study are described in **Table 1**.

Components of the JSE

The three meaningful factors yielded by PCA had eigenvalues > 1 , a result that is in accordance with the factor structure described for the original version. The first factor, which reflected the original first factor, "perspective taking," included 10 items with factor loadings higher than 0.30, accounting for 15.5% of the total variance. The second factor, which reflected the original second factor "compassionate care," included seven items (one less than the original English version) with factor loadings higher than 0.30, accounting for 11.2% of the total variance. The third factor, which reflected the original third factor "standing/walking in the patient's shoes," included two items with factor loadings higher than 0.30, accounting for 5.9% of the total variance.

TABLE 1 | Descriptive statistics and psychometric properties of the Spanish JSE-HP version.

Statistics	Value
N	715
Possible range	20–140
Actual range	59–140
Mean	116
Standard deviation	14
PERCENTILE	
25th	108
50th (Median)	119
75th	127
CRONBACH'S ALPHA COEFFICIENT	
Entire group	0.84
Physicians	0.83
Physicians-in-training	0.84

Worded item 18 (originally associated with Factor 2 in the English version) showed a low factor loading (0.24), associated with Factor 3.

A preliminary CFA revealed a good data fit for correlated Model AB. All items, with the exception of worded item 18, were significant for the three underlying factors ($p < 0.001$). Item 18 was significant for Factor 3 ($p = 0.005$), but not for Factor 2 ($p = 0.075$). Uncorrelated Model AB revealed poor data fit (ratio $\chi^2/df > 13$, $CFI = 0.57$, $TLI = 0.52$, $RMSEA = 0.13$, and $SRMR = 0.14$). Goodness of fit indexes for the correlated

model AB, the correlated model A, and the correlated model B revealed good data fit for all cases. However, the item 18 was not statistically significant ($p = 0.075$) in factor 2 of model AB. Goodness of fit indexes for the three correlated 3-factor models, including p -values for item 18, are reported in **Table 2**.

Based on these findings, a final factor structure of the JSE is reported in **Table 3**. Goodness of fit indexes of this factor structure model was tested according to gender, professional status, place of birth, place of studies, and residence. The report of this analysis is shown in **Table 4**.

TABLE 2 | Goodness of fit indexes for the three correlated 3-factor models of the Spanish JSE-HP version including p -values for item 18.

Model	χ^2	df	Ratio χ^2/df	CFI	TLI	RMSEA	SRMR	Item 18	
								F2	F3
Model AB	236	166	1.422	0.984	0.982	0.025	0.046	0.075	0.005
Model A	244	167	1.462	0.984	0.982	0.025	0.048	0.000	–
Model B	251	167	1.508	0.983	0.981	0.027	0.048	–	0.000

χ^2 , Chi-square statistic; df, degrees of freedom; CFI, comparative fit index; TLI, Tucker-Lewis index; RMSEA, root mean square error of approximation; SRMR, standardized root mean square residual.

TABLE 3 | Items' measures and factor structure of the Spanish JSE-HP version.

Item	Statement	M (SD)	PCA ^b	CFA ^c	r^d
FACTOR 1: "PERSPECTIVE TAKING"					
2	My patients feel better when I understand their feelings	6.4 (1.1)	0.56	0.67	0.55
4	I consider understanding my patients' body language as important as verbal communication in caregiver-patient relationships	6.3 (1.1)	0.34	0.55	0.51
5	I have a good sense of humor that I think contributes to a better clinical outcome	5.4 (1.3)	0.46	0.49	0.34
9	I try to imagine myself in my patients' shoes when providing care to them	5.7 (1.4)	0.58	0.79	0.59
10	My patients value my understanding of their feelings which is therapeutic in its own right	5.7 (1.3)	0.59	0.76	0.51
13	I try to understand what is going on in my patients' minds by paying attention to their non-verbal cues and body language	5.9 (1.3)	0.57	0.83	0.59
15	Empathy is a therapeutic skill without which my success in treatment is limited	6.0 (1.4)	0.55	0.72	0.66
16	An important component of the relationship with my patients is my understanding of their emotional status, as well as that of their families	6.1 (1.2)	0.68	0.86	0.66
17	I try to think like my patients in order to render better care	5.4 (1.4)	0.53	0.57	0.45
20	I believe that empathy is an important therapeutic factor in medical or surgical treatment	6.4 (1.0)	0.61	0.63	0.58
FACTOR 2: "COMPASSIONATE CARE"					
1	My understanding of how my patients and their families feel does not influence my medical or surgical treatment ^a	5.9 (1.9)	0.38	0.66	0.50
7	I try not to pay attention to my patients' emotions in history taking ^a	5.8 (1.6)	0.34	0.82	0.56
8	Attentiveness to my patients' personal experiences does not influence treatment outcomes ^a	5.8 (1.7)	0.57	1.16	0.67
11	Patients' illnesses can be cured only by medical or surgical treatment; therefore, emotional ties to my patients do not have a significant influence on medical or surgical outcomes ^a	6.1 (1.3)	0.66	0.94	0.61
12	Asking patients about what is happening in their personal lives is not helpful in understanding their physical complaints ^a	5.8 (1.8)	0.48	0.92	0.58
14	I believe that emotion has no place in the treatment of medical illness ^a	6.3 (1.4)	0.75	1.01	0.63
19	I do not enjoy reading non-medical literature or the arts ^a	6.3 (1.4)	0.45	0.58	0.37
FACTOR 3: "STANDING/WALKING IN THE PATIENT'S SHOES"					
3	It is difficult for me to view things from my patients' perspectives ^a	5.2 (1.6)	0.60	0.81	0.45
6	Because people are different, it is difficult for me to see things from my patients' perspectives ^a	5.5 (1.5)	0.72	1.19	0.56
18	I do not allow myself to be influenced by strong personal bonds between my patients and their family members ^a	3.7 (1.7)	0.24	0.48	0.31

^aResponses were reverse-scored on these items; otherwise, items were scored directly (strongly disagree = 1, strongly agree = 7).

^bFactor loadings for the principal components analysis.

^cFactor loadings for the confirmatory factor analysis.

^dItem-total correlation Spearman's coefficient.

TABLE 4 | Goodness of fit indexes of the correlated 3-factor model B of the Spanish JSE-HP version by gender, professional status, place of birth, studies, and residence.

Model	χ^2	df	Ratio χ^2/df	CFI	TLI	RMSEA	SRMR
GENDER							
Men	150	167	0.898	1.000	1.000	0.000	0.050
Women	170	167	1.018	0.998	0.998	0.007	0.056
PROFESSIONAL STATUS							
Physician	141	167	0.844	1.000	1.000	0.000	0.046
Physician-in-training	203	167	1.216	0.984	0.982	0.026	0.061
PLACE OF BIRTH							
Spain	137	167	0.822	1.000	1.000	0.000	0.051
Latin America	196	167	1.171	0.990	0.989	0.022	0.056
PLACE OF STUDIES							
Spain	131	167	0.782	1.000	1.000	0.000	0.052
Latin America	193	167	1.155	0.991	0.990	0.022	0.056
RESIDENCE							
Spain	143	167	0.856	1.000	1.000	0.000	0.045
Latin America	210	167	1.260	0.980	0.977	0.032	0.067

χ^2 , Chi-square statistic; df, degrees of freedom; CFI, comparative fit index; TLI, Tucker-Lewis index; RMSEA, root mean square error of approximation; SRMR, standardized root mean square residual.

Group Comparisons

Neither the entire sample, nor the sub-groups studied showed normal distribution empathy scores. Consequently, comparisons were made based on empathy global scores using non-parametric tests. The comparisons revealed statically significant differences ($p < 0.001$) among all the studied groups. The complete report of this analysis is shown in **Table 5**.

Gender comparisons revealed that female physicians scored higher in empathetic interaction than male physicians. When participants were compared according to their professional status, physicians obtained lower global empathy scores than physicians-in-training. In both cases, the group differences were statistically significant ($p < 0.001$).

Statistically significant differences also appeared in speciality comparisons. Physicians qualified in surgical specialities obtained the lowest global scores. When this group was excluded from the group comparison, the differences were not statistically significant ($p = 0.41$). On the other hand, physicians qualified in non-hospital specialities obtained the highest global empathy global scores. This group included 240 family physicians, and 3 occupational physicians.

Socio-demographic group comparisons revealed important cross-cultural differences ($p < 0.001$). The Spanish group obtained the highest global empathy scores, followed by the group of Latin American physicians with cross-cultural exchange experience in Spain. The group of physicians who had never left Latin America obtained the lowest global empathy scores. The comparison of the underlying factors of the JSE revealed significant differences in descending order of magnitude for “compassion care” ($p < 0.001$), “standing/walking in the patient’s shoes” ($p < 0.001$), and “perspective taking” ($p = 0.04$), as can be observed in **Figure 1**.

TABLE 5 | Comparisons of global score of the Spanish JSE-HP version according to the variables studied.

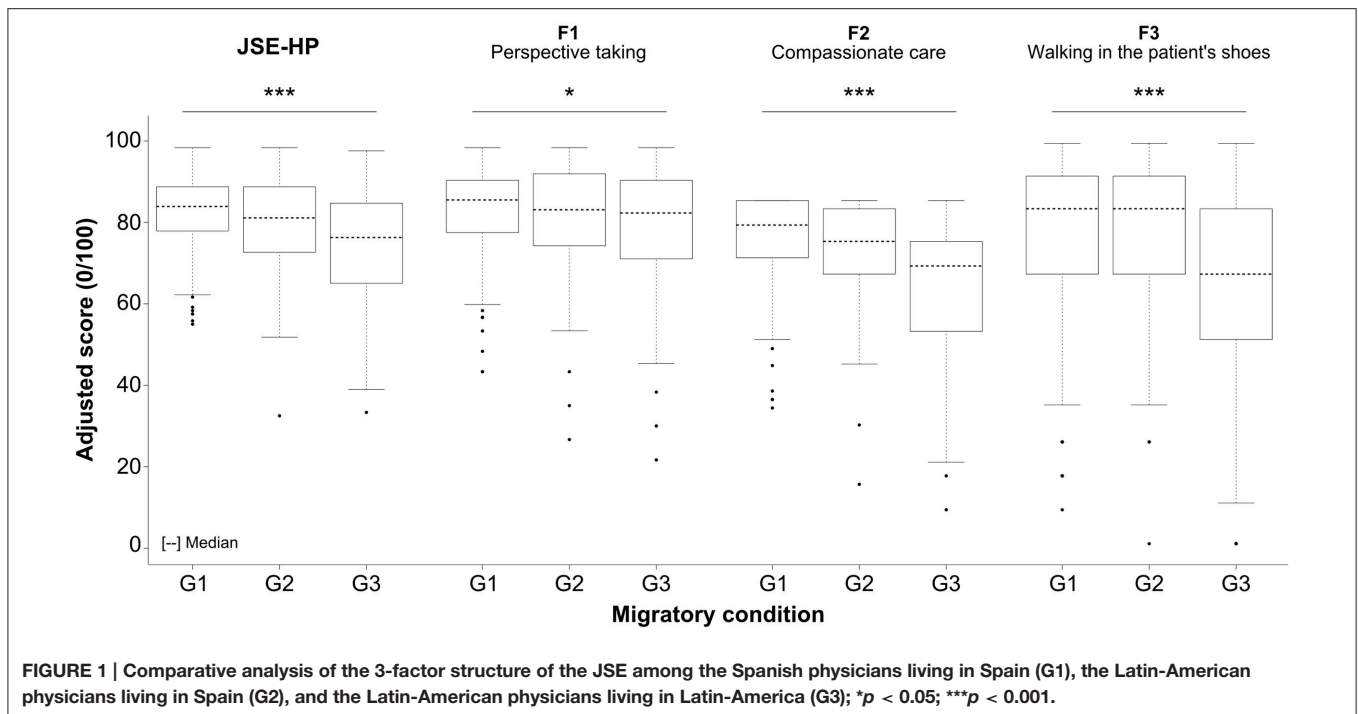
Characteristics	N (%)	Median	Mean	SD
Global score (JSE-HP)	715	119	116	14
GENDER^{a***}				
Men	332 (47)	117	114	15
Women	382 (54)	121	119	13
PROFESSIONAL STATUS^{a***}				
Physician	311 (42)	117	114	13
Physician-in-training	422 (58)	120	118	15
SPECIALITY^{b***}				
No speciality	73 (10)	117	115	14
Non-hospital speciality	243 (34)	120	118	13
Hospital speciality	252 (36)	120	117	14
Medical-Surgical speciality	48 (7)	119	117	14
Surgical speciality	65 (9)	109	108	17
Other	28 (4)	119	115	13
PLACE OF BIRTH^{a***}				
Spain	351 (50)	122	121	11
Latin America	347 (50)	115	112	16
PLACE OF STUDIES^{a***}				
Spain	363 (51)	122	121	11
Latin America	344 (49)	115	112	16
RESIDENCE^{a***}				
Spain	480 (66)	121	120	12
Latin America	253 (35)	113	110	16
MIGRATORY CONDITION^{b***}				
Spaniard living in Spain	351 (50)	122	121	11
Latin-American living in Spain	94 (14)	119	117	14
Latin-American living in Latin-America	253 (36)	113	110	16

SD, Standard deviation; ^aU Mann-Withney test; ^bKruskal Wallis test; *** $p < 0.001$.

DISCUSSION

The JSE was originally designed as a research tool to measure the development of empathy within the specific context of medical care and interaction with patients. Since it was first published in 2002, the JSE has been widely accepted at the international level. The first publications reported consisted mainly of exploratory studies with PCA (Alcorta-Garza et al., 2005; Di Lillo et al., 2009; Kataoka et al., 2009; Paro et al., 2012; Suh et al., 2012; Wen et al., 2013). However, in recent years there has been an increase in the number of publications including studies based on CFA. The authors of one of the latest articles (Hojat and Lanoue, 2014) use CFA to validate the JSE’s psychometric properties. They also recommend the use of the CFA, and suggest maintenance of all the 20 items in the instrument, not only for goodness of fit of the 3-factor model, but also to obtain significant item-total correlations and substantial item discrimination effect size indexes for all items.

The results obtained from this study prove the stability of the JSE’s most relevant characteristics: high reliability of the instrument, need for the inclusion of all the items and the existence of a 3-factor model composed by two main factors, one



major cognitive and one emotional, and a third trivial factor. However, for research based on samples of Spanish and Latin American individuals, this study shows that it is preferable to include item 18 within the “standing/walking in the patient’s shoes” factor, rather than within “compassionate care.” The differences observed for item 18 are in agreement with the findings of other authors whose works are not necessarily focused on the Spanish cultural context, but also on others. The study also reveals that the factors that make up the JSE follow an “oblique,” rather than “orthogonal” model. This difference is explained by the fact that empathy is understood as a cognitive-emotional unit where each of the factors influences the development of the remaining two, rather than as a sum of statistically independent units. Certain researchers (Magalhaes et al., 2011; Leombruni et al., 2014) and, more recently, authors themselves (Hojat and Lanoue, 2014), have pointed out the care that must be taken when approaching this topic.

Group comparisons yield differences that are consistent with those of previous studies, both for the categories of gender (Hojat et al., 2002; Alcorta-Garza et al., 2005; Kataoka et al., 2009; Magalhaes et al., 2011; Tavakol et al., 2011b; Suh et al., 2012; Shariat and Habibi, 2013; Wen et al., 2013; Leombruni et al., 2014) and speciality (Hojat et al., 2002; Suh et al., 2012). In agreement with the observations of other researchers, this study provides evidence of a positive association between the development of empathy and work in professional specialities involving greater roles in patient care (Tavakol et al., 2011a). In the case of students, variations may be explained by differences related to personality or emotional intelligence (Costa et al., 2014; Hojat et al., 2015), or by the influence of psycho-social factors (Hojat, 2006). With regard to practicing physicians, together with the differences mentioned above, the development of empathy

might be influenced by professional burnout caused by exposure to adverse work environments (Almeida, 2002; Brazeau et al., 2010; Delgado-Bolton et al., 2015).

In the healthcare area, Latin America’s complex economic, political, social, and cultural network poses a constant challenge (Cotlear et al., 2015). Added to professional limitations in disease treatment are inequities of access to the healthcare system, scarcity and misuse of resources, corruption in the sector and a high social demand. Even though there have been significant improvements in the sector (Atun et al., 2015), they are not consistent throughout the different regions and the challenges are hardly met. Proof of this is the deterioration in the physician-patient relationship, which is still far too common (Correa and Javier, 2011; Delgado-Bolton et al., 2015; San-Martín et al., 2016). The analysis of cultural variables carried out in this study proves the latency of this issue, while also providing clues about the importance of the role played by professional, cultural, and social surroundings in the development and improvement of medical empathy.

On the other hand, there are very important evidences in support of the validity and utility of the JSE in the context of patient care. The use of the JSE as indicator of this professional ability to predict clinical and patient outcomes in adult patients with diabetes was proved in two studies (Hojat et al., 2011b; Del Canale et al., 2012). In both cases an association between empathy in patient care and positive outcomes was confirmed. This is utterly important because the ultimate goal of medical education and all other health professions is to optimize clinical outcomes. In two studies, carried out with French physicians, a protective role of empathy for burnout (Lamothe et al., 2014), and a positive relationship between empathy and overall better clinical practice was demonstrated (Zenasni et al., 2012). Moreover,

in Latin American and Spanish physicians-in-training with higher empathetic scoring in the JSE presented better and more effective learning abilities, and abilities toward inter-professional collaboration (San-Martín et al., 2016). In medical and healthcare education contexts, there are important evidences in support of the importance of the JSE (Hojat, 2016). Nevertheless, the authors are cautious and recognize that further research is needed to investigate the relationship between scores on the JSE and educational outcomes. Furthermore, large-scale research is also needed with national samples to develop national norm tables and cutoff scores for the JSE to identify low and high scorers in different populations and health professions students (Hojat and Gonnella, 2015).

In general, all these findings underlying the importance that empathy has in medical and healthcare education, in clinical practice, and in practitioners' health and welfare. Understanding the role of empathy is an issue with special relevance in geographical contexts where practitioners have to address daily social needs with scarce resources, as it happens in many public health institutions of Latin American countries.

AUTHOR CONTRIBUTIONS

LV and AA undertook the translation and linguistic adaptation of the scale. MS and LV performed the statistical processing

of data. LV was in charge of the study's overall design, coordination with the participating institutions, and drafting of the manuscript. RD, HR, and JS were in charge of the coordination with Spanish Healthcare institutions and LV, MS, and AA were in charge of the coordination with Latin American Healthcare institutions. All authors contributed to the presented work. All authors participated during the interpretation process of the results, and approved the final manuscript.

FUNDING

This study was supported by the Rioja Salud Foundation (FRS), Spain.

ACKNOWLEDGMENTS

We would like to acknowledge the contribution of the following institutions: In Spain, Consejería de Salud del Gobierno de La Rioja, Instituto Catalán de la Salud, Instituto Borja de Bioética. In Mexico, Universidad Autónoma de Nuevo León, Sistema Mexicano de Seguridad Social; in Bolivia, Departamento de Salud de La Paz. In Argentina, Hospital Regional Rio Grande del Sistema Argentino de Salud. In Colombia, Grupo de Salud ESIMED de Bogotá.

REFERENCES

- Alcorta-Garza, A., González-Guerrero, J. F., Tavitas-Herrera, S. E., Rodríguez-Lara, F. J., and Hojat, M. (2005). Validación de la Escala de Empatía Médica de Jefferson en estudiantes de medicina mexicanos. *Salud Men.* 28, 57–63. Available online at: <http://www.redalyc.org/articulo.oa?id=58252808> (Accessed June 28, 2016).
- Almeida, C. (2002). Reforma de sistemas de servicios de salud y equidad en América Latina y el Caribe: algunas lecciones de los años 80 y 90 Health systems reform and equity in Latin America and the Caribbean: lessons from. *Cad. Saúde Pública* 18, 905–925. doi: 10.1590/S0102-311X2002000400002
- Atun, R., De Andrade, L. O. M., Almeida, G., Cotlear, D., Dmytraczenko, T., Frensz, P., et al. (2015). Health-system reform and universal health coverage in Latin America. *Lancet* 385, 1230–1247. doi: 10.1016/S0140-6736(14)61646-9
- Bliese, P. (2013). *Multilevel Functions. R package Version 2.5* [Online]. Available online at: <http://CRAN.R-project.org/package=multilevel> (Accessed March 08, 2016).
- Brazeau, C. M., Schroeder, R., Rovi, S., and Boyd, L. (2010). Relationships between medical student burnout, empathy, and professionalism climate. *Acad. Med.* 85, S33–36. doi: 10.1097/ACM.0b013e3181ed4c47
- Brown, T. A. (2014). *Confirmatory Factor Analysis for Applied Research*. New York, NY: Guilford Publications.
- Correa, L., and Javier, F. (2011). Pobreza, vulnerabilidad y calidad de vida en América Latina: retos para la bioética. *Acta Bioeth.* 17, 19–29. doi: 10.4067/S1726-569X2011000100003
- Costa, P., Alves, R., Neto, I., Marvao, P., Portela, M., and Costa, M. J. (2014). Associations between medical student empathy and personality: a multi-institutional study. *PLoS ONE* 9:e89254. doi: 10.1371/journal.pone.0089254
- Cotlear, D., Gomez-Dantes, O., Knaul, F., Atun, R., Barreto, I. C., Cetrangolo, O., et al. (2015). Overcoming social segregation in health care in Latin America. *Lancet* 385, 1248–1259. doi: 10.1016/S0140-6736(14)61647-0
- Davis, M. H. (1983). Measuring individual differences in empathy: evidence for a multidimensional approach. *J. Pers. Soc. Psychol.* 44, 113–126 doi: 10.1037/0022-3514.44.1.113
- Del Canale, S., Louis, D. Z., Maio, V., Wang, X., Rossi, G., Hojat, M., et al. (2012). The relationship between physician empathy and disease complications: an empirical study of primary care physicians and their diabetic patients in Parma, Italy. *Acad. Med.* 87, 1243–1249. doi: 10.1097/ACM.0b013e31826288bf
- Delgado-Bolton, R., San-Martín, M., Alcorta-Garza, A., and Vivanco, L. (2015). Medical empathy of physicians-in-training who are enrolled in professional training programs. A comparative intercultural study in Spain. *Aten. Primaria* doi: 10.1016/j.aprim.2015.10.005. [Epub ahead of print].
- Di Lillo, M., Cicchetti, A., Lo Scalzo, A., Taroni, F., and Hojat, M. (2009). The jefferson scale of physician empathy: preliminary psychometrics and group comparisons in Italian physicians. *Acad. Med.* 84, 1198–1202. doi: 10.1097/ACM.0b013e3181b17b3f
- Geisinger, K. F. (1994). Cross-cultural normative assessment: translation and adaptation issues influencing the normative interpretation of assessment instruments. *Psychol. Assess.* 6:304. doi: 10.1037/1040-3590.6.4.304
- Gough, H. G., and Hall, W. B. (1977). A comparison of physicians who did or did not respond to a postal questionnaire. *J. Appl. Psychol.* 62, 777. doi: 10.1037/0021-9010.62.6.777
- Gross, J. (2012). *Nortest: Tests for Normality. R Package Version 1.0-2* [Online]. Available online at: <http://cran.r-project.org/web/packages/nortest/index.html> (Accessed March 08, 2016).
- Hemmerding, J. M., Stoddart, S. D., and Lilford, R. J. (2007). A systematic review of tests of empathy in medicine. *BMC Med. Educ.* 7:24. doi: 10.1186/1472-6920-7-24
- Henson, R. K., and Roberts, J. K. (2006). Use of exploratory factor analysis in published research common errors and some comment on improved practice. *Educ. Psychol. Meas.* 66, 393–416. doi: 10.1177/0013164405282485
- Hojat, M. (2006). *Empathy in Patient Care: Antecedents, Development, Measurement, and Outcomes*. New York, NY: Springer.
- Hojat, M. (2016). *Empathy in Health Professions Education and Patient Care*. Cham: Springer International.
- Hojat, M., and Gonnella, J. S. (2015). Eleven years of data on the Jefferson Scale of Empathy-Medical Student Version (JSE-S): proxy norm data and tentative cutoff scores. *Med. Prin. Pract.* 24, 344–350. doi: 10.1159/000381954

- Hojat, M., Gonnella, J. S., Nasca, T. J., Mangione, S., Vergare, M., and Magee, M. (2002). Physician empathy: definition, components, measurement, and relationship to gender and specialty. *Am. J. Psychiatry* 159, 1563–1569. doi: 10.1176/appi.ajp.159.9.1563
- Hojat, M., and Lanoue, M. (2014). Exploration and confirmation of the latent variable structure of the Jefferson scale of empathy. *Int. J. Med. Educ.* 5, 73–81. doi: 10.5116/ijme.533f.0c41
- Hojat, M., Louis, D. Z., Markham, F. W., Wender, R., Rabinowitz, C., and Gonnella, J. S. (2011b). Physicians' empathy and clinical outcomes in diabetic patients. *Acad. Med.* 86, 359–364. doi: 10.1097/ACM.0b013e3182086fe1
- Hojat, M., Louis, M., Daniel, Z., Maxwell, B., Gonnella, M., and Joseph, S. (2011a). The Jefferson Scale of Empathy (JSE): an update. *Population Health Matters (Formerly Health Policy Newsletter)* 24, 5.
- Hojat, M., Mangione, S., Nasca, T. J., Cohen, M. J. M., Gonnella, J. S., Erdmann, J. B., et al. (2001). The Jefferson scale of physician empathy: development and preliminary psychometric data. *Educ. Psychol. Meas.* 61, 349–365. doi: 10.1177/00131640121971158
- Hojat, M., Vergare, M., Isenberg, G., Cohen, M., and Spandorfer, J. (2015). Underlying construct of empathy, optimism, and burnout in medical students. *Int. J. Med. Educ.* 6, 12–16. doi: 10.5116/ijme.54c3.60cd
- Kataoka, H. U., Koide, N., Ochi, K., Hojat, M., and Gonnella, J. S. (2009). Measurement of empathy among Japanese medical students: psychometrics and score differences by gender and level of medical education. *Acad. Med.* 84, 1192–1197. doi: 10.1097/ACM.0b013e3181b180d4
- Kline, R. B. (2005). *Principles and Practice of Structural Equation Modeling*. New York, NY: Guilford.
- Kunyk, D., and Olson, J. K. (2001). Clarification of conceptualizations of empathy. *J. Adv. Nurs.* 35, 317–325. doi: 10.1046/j.1365-2648.2001.01848.x
- Lamothe, M., Boujut, E., Zenasni, F., and Sultan, S. (2014). To be or not to be empathic: the combined role of empathic concern and perspective taking in understanding burnout in general practice. *BMC Fam. Pract.* 15:15. doi: 10.1186/1471-2296-15-15
- Leombruni, P., Di Lillo, M., Miniotti, M., Picardi, A., Alessandri, G., Sica, C., et al. (2014). Measurement properties and confirmatory factor analysis of the Jefferson Scale of Empathy in Italian medical students. *Perspect. Med. Educ.* 3, 419–430. doi: 10.1007/s40037-014-0137-9
- Magalhaes, E., Salgueira, A. P., Costa, P., and Costa, M. J. (2011). Empathy in senior year and first year medical students: a cross-sectional study. *BMC Med. Educ.* 11:52. doi: 10.1186/1472-6920-11-52
- Mato, D. (2008). *Diversidad Cultural e Interculturalidad en Educación Superior. Experiencias en América Latina*. Caracas: UNESCO; IESALC.
- Muthén, B., Du Toit, S. H., and Spisic, D. (1997). Robust inference using weighted least squares and quadratic estimating equations in latent variable modeling with categorical and continuous outcomes. *Psychometrika* 75, 1–45.
- Oesterreicher, W. (2013). El español, lengua pluricéntrica: perspectivas y límites de una autoafirmación lingüística nacional en Hispanoamérica. El caso mexicano. *Lexis* 26, 275–304.
- Otero, J., and Powell-Davies, P. (2011). *Word for Word: The Social, Economic and Political Impact of Spanish and English*. Madrid: Santillana Español.
- Paro, H. B., Daud-Gallotti, R. M., Tiberio, I. C., Pinto, R. M., and Martins, M. A. (2012). Brazilian version of the Jefferson Scale of Empathy: psychometric properties and factor analysis. *BMC Med. Educ.* 12:73. doi: 10.1186/1472-6920-12-73
- Rosseel, Y. (2012). lavaan: an R Package for structural Equation Modeling. *J. Stat. Softw.* 48, 1–36. doi: 10.18637/jss.v048.i02
- Sánchez-Sagrado, T. (2013). ¿ Sobran o faltan médicos en España? La migración, el eterno recurso. *Rev. Clin. Esp. (Barc)* 213, 347–353. doi: 10.1016/j.rce.2013.03.011
- San-Martín, M., Roig-Carrera, H., Villalonga-Vadell, R. M., Benito-Sevillano, C., Torres-Salinas, M., Claret-Teruel, G., et al. (2016). Empathy, inter-professional collaboration, and lifelong medical learning in Spanish and Latin-American physicians-in-training who start their postgraduate training in hospitals in Spain. Preliminary outcomes. *Aten. Primaria*. doi: 10.1016/j.aprim.2016.02.007. [Epub ahead of print].
- Shariat, S. V., and Habibi, M. (2013). Empathy in Iranian medical students: measurement model of the Jefferson scale of empathy. *Med. Teach.* 35, e913–e918. doi: 10.3109/0142159X.2012.714881
- Spiro, H. M., McCrea Curnen, M. G. M., Peschel, E., and St. James, D. (1993). *Empathy and the Practice of Medicine: Beyond Pills and the Scalpel*. New Haven, CT: Yale University Press.
- Suh, D. H., Hong, J. S., Lee, D. H., Gonnella, J. S., and Hojat, M. (2012). The Jefferson Scale of Physician Empathy: a preliminary psychometric study and group comparisons in Korean physicians. *Med. Teach.* 34, e464–468. doi: 10.3109/0142159X.2012.668632
- Suhr, D. D. (2005). “Principal component analysis vs. exploratory factor analysis,” in *Proceedings of the Thirtieth Annual SAS(R) User Group International Conference* (Cary, NC: SAS Institute Inc.), 203–230.
- Tavakol, S., Dennick, R., and Tavakol, M. (2011a). Empathy in UK medical students: differences by gender, medical year and specialty interest. *Educ. Prim. Care* 22, 297–303. doi: 10.1080/14739879.2011.11494022
- Tavakol, S., Dennick, R., and Tavakol, M. (2011b). Psychometric properties and confirmatory factor analysis of the Jefferson Scale of Physician Empathy. *BMC Med. Educ.* 11:54. doi: 10.1186/1472-6920-11-54
- Velicer, W. F., and Fava, J. L. (1998). Effects of variables and subject sampling on factor pattern recovery. *Psychol. Methods* 3, 231–251. doi: 10.1037/1082-989X.3.2.231
- Veloski, J., and Hojat, M. (2006). “Measuring specific elements of Professionalism: Empathy, Teamwork, and Lifelong learning,” in *Measuring Medical Professionalism*, ed D. T. Stern (New York, NY: Oxford University Press), 117–146.
- Vivanco, L., and Delgado-Bolton, R. (2015). “Professionalism,” in *Encyclopedia of Global Bioethics*, ed H. Ten Have (Dordrecht: Springer and Business Media). doi: 10.1007/978-3-319-05544-2_353-1
- Wen, D., Ma, X., Li, H., Liu, Z., Xian, B., and Liu, Y. (2013). Empathy in Chinese medical students: psychometric characteristics and differences by gender and year of medical education. *Med. Educ.* 13:130. doi: 10.1186/1472-6920-13-130
- Zenasni, F., Boujut, E., du Vaure, B., Catu-Pinault, A., Tavani, J. L., Rigal, L., et al. (2012). Development of a French language version of the Jefferson Scale of Physician Empathy and association with practice characteristics and burnout in a sample of general practitioners. *Int. J. Person Centered Med.* 2, 759–766. doi: 10.5750/ijpcm.v2i4.295

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2016 Alcorta-Garza, San-Martín, Delgado-Bolton, Soler-González, Roig and Vivanco. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Depressive Symptomatology among Norwegian Adolescent Boys and Girls: The Patient Health Questionnaire-9 (PHQ-9) Psychometric Properties and Correlates

Jasmina Burdzovic Andreas* and Geir S. Brunborg

Department of Substance Use, Norwegian Institute of Public Health, Oslo, Norway

OPEN ACCESS

Edited by:

Kai Ruggeri,
University of Cambridge,
United Kingdom

Reviewed by:

David Daniel Ebert,
Friedrich-Alexander University
Erlangen-Nuremberg, Germany
Tanja Gabriele Baudson,
Technische Universität Dortmund,
Germany

*Correspondence:

Jasmina Burdzovic Andreas
jabu@fhi.no

Specialty section:

This article was submitted to
Quantitative Psychology and
Measurement,
a section of the journal
Frontiers in Psychology

Received: 15 November 2016

Accepted: 15 May 2017

Published: 08 June 2017

Citation:

Burdzovic Andreas J and
Brunborg GS (2017) Depressive
Symptomatology among Norwegian
Adolescent Boys and Girls: The
Patient Health Questionnaire-9
(PHQ-9) Psychometric Properties and
Correlates. *Front. Psychol.* 8:887.
doi: 10.3389/fpsyg.2017.00887

This study explored the potential contribution of the Diagnostic and Statistical Manual for Mental Disorders (DSM-IV)-based Patient Health Questionnaire-9 item (PHQ-9) instrument to the developmental epidemiology research in Norway, by examining depressive symptoms in a school sample of adolescents ($N = 846$). The average PHQ-9 scores were 6.89 ($SD = 5.13$) for girls, and 4.57 ($SD = 3.98$) for boys; 8.5% of girls and 2.6% of boys were classified into the originally proposed categories indicative of Major Depressive Disorder (MDD; PHQ-9 scores ≥ 15). Multi-group confirmatory factor analysis (CFA) confirmed a single-factor structure for the PHQ-9 with solid psychometric properties and high internal consistency for both genders. However, even though configural equality was observed, there was no evidence for metric or scalar equality across genders, warranting further investigation of measurement equivalence for the current Norwegian version of the PHQ-9. We observed no major associations between the PHQ-9 scores and adolescent religion or immigrant background. Further, school grade, not living together with both biological parents, and diagnosed chronic illness were differently associated with elevated depressive symptoms for boys and girls. Finally, high residential instability, perceived low SES, school dissatisfaction, lack of close friendships, history of suicide attempts and self-harm, and elevated emotional problems were all significantly and consistently associated with greater depression for both genders. Overall, the PHQ-9 appears to be a promising research tool, potentially offering clinically-relevant classification of adolescent self-reported depressive symptomatology in addition to the symptom severity captured by continuous scores. Nevertheless, further investigation concerning the observed measurement non-equivalence, as well as the comprehensive validation and comparison against the gold standard is required before the PHQ-9 is to be used for diagnostic screening in Norway.

Keywords: depression, adolescents, PHQ-9, Norway, cross-cultural comparison

INTRODUCTION

Depression in children and adolescents is far from infrequent or inconsequential (Birmaher et al., 1996; Lewinsohn et al., 1998; Merikangas et al., 2010; Rohde et al., 2012; Thapar et al., 2012). Even though many children recover, early-onset depression remains a potent risk factor for subsequent mental health problems and other negative outcomes (Harrington et al., 1990; Weissman et al., 1999; Kovacs et al., 2016). In addition, early gender differentiation has also been observed, where adolescent girls tend to have both elevated symptoms and a different developmental course of depression (Hankin et al., 1998; Twenge and Nolen-Hoeksema, 2002; Dunn and Goodyear, 2006; Dekker et al., 2007; Essau et al., 2010). Accurately detecting early depression symptoms, and detecting them accurately for boys and girls is thus a public health imperative. In that regard, brief self-reports may be especially useful, as such screeners can rapidly identify “at-risk” youth in need of further evaluation and possibly treatment.

This may especially be true for Norway, where the 2014 Public Health report states that “mental disorders are a major health problem for children and adolescents in Norway today” (Norwegian Institute of Public Health, 2014). Despite these clearly identified issues and the general focus on adolescent health and development, the use of diagnostically-informative measures in research practice has been somewhat limited in Norway. For example, multiple Norwegian studies investigating early depression (Sund et al., 2003, 2011; Lundervold et al., 2013; Larsson et al., 2016) have utilized various versions of the Mood and Feelings Questionnaire (MFQ; Angold et al., 1995). Its wide use in studies of developmental epidemiology notwithstanding, the MFQ is fairly extensive (34 items for the full and 13 items for the short version), rendering it not necessarily the best *brief* instrument. And even though the MFQ can theoretically aid in clinical screening, its cut-offs are not fully established or necessarily even recommended¹. Other reports have used shorter and thus more practical instruments; for example the 12-item and 5-item Symptom Checklists (SCL-12, and SCL-5) (Heyerdahl et al., 2004; Derdikman-Eiron et al., 2012; Mykkestad et al., 2012), and a 6-item Depressed Mood Inventory (Wichstrom, 1999; Abebe et al., 2016), all of which appear to measure anxiety and depression and to have been derived from the 25-item Hopkins Symptom Checklist (HSCL) for adults (Derogatis et al., 1974). However, it is not entirely clear how appropriate these derivations may be for the assessment of adolescents, or how well they differentiate between anxiety and depression given that they tend to conflate the two into “anxiety-depression” (Heyerdahl et al., 2004) and are meant to measure general “psychological” or “global mental distress” (Tambs and Moum, 1993; Strand et al., 2003; Mykkestad et al., 2012). In addition, the items, responses, cut-off scores, and clinical interpretation of these HSCL derivations appear to vary from study to study in Norway (i.e., from version to version; Heyerdahl et al., 2004; Derdikman-Eiron et al.,

2012; Abebe et al., 2016). Finally, psychiatric symptoms among Norwegian adolescents have also been evaluated using the Strengths and Difficulties Questionnaire (SDQ) and its various subscales (Rønning et al., 2004; Indredavik et al., 2005; Goodman et al., 2011). Despite its many advantages and widespread international use, the 25-item SDQ is also somewhat long, and it ideally requires the children’s, parental, and teachers’ reports for a complete evaluation. Thus, any potential large-scale screening based on multiple SDQ reports/informants may be both impractical and costly. The use of SDQ for such purposes may additionally be questionable, given the related cultural considerations (Heiervang et al., 2007, 2008), as well as its somewhat limited ability to detect mental health disorders in general (Brøndbo et al., 2011), and depressive symptomatology in particular. For example, it is not clear how well the SDQ Emotional Problems subscale differentiates between anxiety and depression, as both classes of problems appear to be assessed under the general heading of “emotional” disorders (Goodman et al., 2000a,b).

Thus, the public health initiatives and the related research and clinical practice in Norway could benefit from a self-report instrument designed to swiftly and effectively screen specifically for early-onset depression based on the common and internationally-validated criteria. The 9-item Patient Health Questionnaire (PHQ-9) adolescent version may be especially well-suited for such purposes (Kroenke et al., 2001; Johnson et al., 2002). First, the PHQ-9 has only 9 items, and it requires only the youth self-reports. The instrument was originally developed to assess symptoms of depression in accordance to the *Diagnostic and Statistical Manual of Mental Disorders, 4th* edition (DSM-IV) criteria, but it also corresponds to the more recent DSM-5 criteria (American Psychiatric Association, 2016). Further, the PHQ-9 has been shown to be a valid tool in detecting depression among adolescents across various cultures and settings (Adewuya et al., 2006; Richardson et al., 2010; Fatiregun and Kumapayi, 2014; Tsai et al., 2014). Most importantly, the PHQ-9 also captures depression severity, as it provides both continuous scores and clinically-meaningful classification of depressive symptomatology. Availability of continuous scores ensures no loss of variability, and thus may be both recommended and preferred in research, whereas categorical classification may be more meaningful in clinical practice and assessment. Ideally, a self-report instrument should serve both purposes, and the PHQ-9 for adolescents theoretically does so at the international level. For example, the PHQ-9 was successfully used to estimate the prevalence of Major Depressive Disorder (MDD) in Chinese and Nigerian school samples (Fatiregun and Kumapayi, 2014; Tsai et al., 2014), in addition to the US community and primary care samples (Richardson et al., 2010; Rhew et al., 2016). Such apparent cultural robustness strengthens the potential advantages of this instrument in multiple large scale national and cross-cultural comparisons.

In conclusion, there is a need for a brief and internationally validated instrument that captures both the severity and corresponding clinical categories of self-reported depressive symptomatology in Norwegian adolescent populations. Such an instrument would strengthen both the research and clinical

¹Angold, A. MFQ Message from the Author. Available online at <http://devepi.duhs.duke.edu/%5Cinstruments%5CMFQ%20user.pdf>

practice in Norway, while simultaneously opening the door to cross-cultural and cross-national monitoring, evaluation, and comparison. Perhaps the PHQ-9—with its brief format, simple self-report administration, and attractive scoring features—can fill this need. This report offers the first step in that direction, by examining the PHQ-9 basic properties and correlates in a sample of Norwegian adolescents. Specifically, we examined: (1) the basic psychometric properties of the PHQ-9, including measurement equivalence (invariance) by gender, and (2) severity and correlates of depressive symptoms among Norwegian middle- and high-school students as measured by the PHQ-9.

METHODS

Sample and Procedures

The sample comprised middle- and high-school students enrolled in a mixed-methods short-term longitudinal study primarily focusing on substance use among Norwegian youth. Seven schools in the vicinity of the Norwegian capital were approached for study participation, with the goal of complete enrollment in grades 8 through 12. A total of 1,326 students from the five assenting schools were approached for survey participation. Middle-school students' participation was predicated upon their own assent and parental consent, whereas high-school students (i.e., those older than 16) consented themselves. A modest contribution was made to each participating classroom (approximately 100–120 Euros), while the teachers who helped out with data collection were reimbursed with a modest honorarium. Of the consented 943 (71.1%) students, 884 (93.7%) participated in the baseline assessment conducted in the Fall of 2014, where they completed a computer-administered questionnaire during their regular class time under teacher supervision.

Students were relatively evenly distributed across middle school (17.9% in grade 8, 15.0% in grade 9, and 18.3% in grade 10) and high-school grades (23.9% in grade 11 and 24.9% in grade 12). Approximately half of the participants were boys (46.3%), and the majority had no immigrant background (80.7% reported both parents born in Norway, and 91.8% were Norwegian-born themselves). This report utilized data from the baseline assessment. Outliers ($n = 9$) and cases with incomplete responses on the depression scale items ($n = 29$) were excluded, resulting in an analytical sample of 846 students.

The study was approved by the Data Protection Official for Research/Norwegian Centre for Research Data (NSD, case #39513). Additional descriptions of the sample and study procedures are provided elsewhere (Brunborg et al., 2017).

Measures

The student questionnaire assessed a range of developmentally-relevant characteristics from all levels of human ecology (Bronfenbrenner, 1979). All instruments were based on the internationally validated and commonly used measures, which were translated and modified for the Norwegian context as needed.

Depressive Symptomatology

Students reported their symptoms of depression during the last 7 days on the 9-item Patient Health Questionnaire (PHQ-9) adapted for use with adolescents (Kroenke et al., 2001; Johnson et al., 2002) and as recommended for research and clinical evaluation by the American Psychiatric Association (2016). The PHQ-9 uses the DSM-IV diagnostic criteria to assess depressive symptomatology (i.e., sleep, concentration, and energy problems, low self-esteem, anhedonia, etc.) on a 4-point scale ranging from 0 (“not at all”) to 3 (“nearly every day”), (Kroenke et al., 2001; Kroenke and Spitzer, 2002).

In addition to its utility as a short screener, the PHQ-9 also captures depression severity. Overall scale scores are computed as a sum of the 9 items (possible range 0–27), and the prorated scores can be obtained as long as there are at least 7 items with valid responses (American Psychiatric Association, 2016). The corresponding severity categories were originally defined as

TABLE 1 | Sample characteristics; by gender.

Variables	Boys ($n = 387$)	Girls ($n = 459$)
	<i>M</i> (SD) or %	<i>M</i> (SD) or %
DEMOGRAPHICS		
School grade (8 through 12)	10.2 (1.4)	10.2 (1.4)
Religion		
None declared	38.0	22.6
Christianity	53.0	70.2
Islam	4.1	4.6
Other	4.9	2.6
Residential instability		
1 move/school change	14.7	13.1
2 or more moves/school changes	13.7	12.2
Intact biological family (Yes)	66.7	66.4
Born in Norway (Yes)	90.4	90.0
Immigrant background (Yes)	19.1	18.3
Perceived low SES	3.7 (1.8)	4.1 (1.8)
PSYCHO-SOCIAL CHARACTERISTICS		
School dissatisfaction	1.6 (0.61)	1.7 (0.71)
Close friend (Yes)	93.3	90.4
Ever attempted suicide		
Yes	2.7	4.2
Don't know/Will not respond	4.9	5.6
Ever self-harmed		
Yes	8.4	19.4
Don't know/Will not respond	3.3	5.6
Chronic illness diagnosis (Yes)	21.6	21.5
SDQ Emotional problems (continuous scores)	2.0 (1.8)	3.8 (2.4)
SDQ Emotional problems (clinical positive)	5.94	25.3

Shown are means (standard deviations) for continuous variables, and proportions (%) for categorical variables.

none (PHQ-9 scores 0–4), mild (PHQ-9 scores 5–9), moderate (PHQ-9 scores 10–14), moderately severe (PHQ-9 scores 15–19), and severe (PHQ-9 scores 20–27; Richardson et al., 2010). Adolescents with the PHQ-9 scores of 15 or above (i.e., those classified as exhibiting moderately severe, or severe depressive symptomatology) may be of particular clinical concern, as they are likely to meet the diagnostic criteria for Major Depressive Disorder (MDD) with 95% specificity (Kroenke et al., 2001; Tsai et al., 2014).

Demographics

Students reported their gender, school grade (8 through 12), religion, and whether they and their parents were born in Norway. Participants were classified as native-born if they were born in Norway, and without immigrant background if both of their parents were also Norwegian-born. In addition, students reported their residence circumstances, including residential instability (i.e., the number of school changes due to the family move) and whether they currently live with their intact biological

TABLE 2 | The PHQ-9 item and scale descriptives.

Patient Health Questionnaire 9 items (PHQ-9) Adolescent version		Item M (SD)		
		Total (N = 846)	Boys (n = 387)	Girls (n = 459)
1.	Feeling down, depressed, irritable, or hopeless	0.69 (0.76)	0.42 (0.61)	0.92 (0.80)
2.	Little interest or pleasure in doing things	0.90 (0.79)	0.79 (0.78)	1.00 (0.79)
3.	Trouble falling asleep, staying asleep, or sleeping too much	1.00 (0.98)	0.82 (0.93)	1.15 (0.99)
4.	Poor appetite, weight loss, or overeating	0.35 (0.70)	0.23 (0.59)	0.43 (0.77)
5.	Feeling tired, or having little energy	1.05 (0.86)	0.87 (0.80)	1.20 (0.88)
6.	Feeling bad about yourself, or that you're a failure or that you've let yourself or your family down	0.66 (0.85)	0.43 (0.69)	0.86 (0.92)
7.	Trouble concentrating on things like school work, reading, or watching TV	0.71 (0.81)	0.63 (0.79)	0.78 (0.82)
8.	Moving or speaking so slowly that other people could have noticed	0.21 (0.52)	0.20 (0.50)	0.21 (0.54)
9.	Thought that you would be better off dead, or of hurting yourself in some way	0.23 (0.61)	0.17 (0.52)	0.28 (0.67)

PHQ-9 overall scores		Scale Descriptives		
		M (SD)	Boys (n = 387)	Girls (n = 459)
	M (SD)	5.83 (4.78)	4.57 (3.98)	6.89 (5.13)
	Range	0–27	0–24	0–27
	Skewness (s.e.)	1.48 (0.08)	1.55 (0.12)	1.35 (0.11)
	Kurtosis (s.e.)	2.86 (0.17)	3.69 (0.25)	2.16 (0.23)

Response options: 0, "not at all"; 1, "some days"; 2, "more than half the days"; 3, "nearly every day."

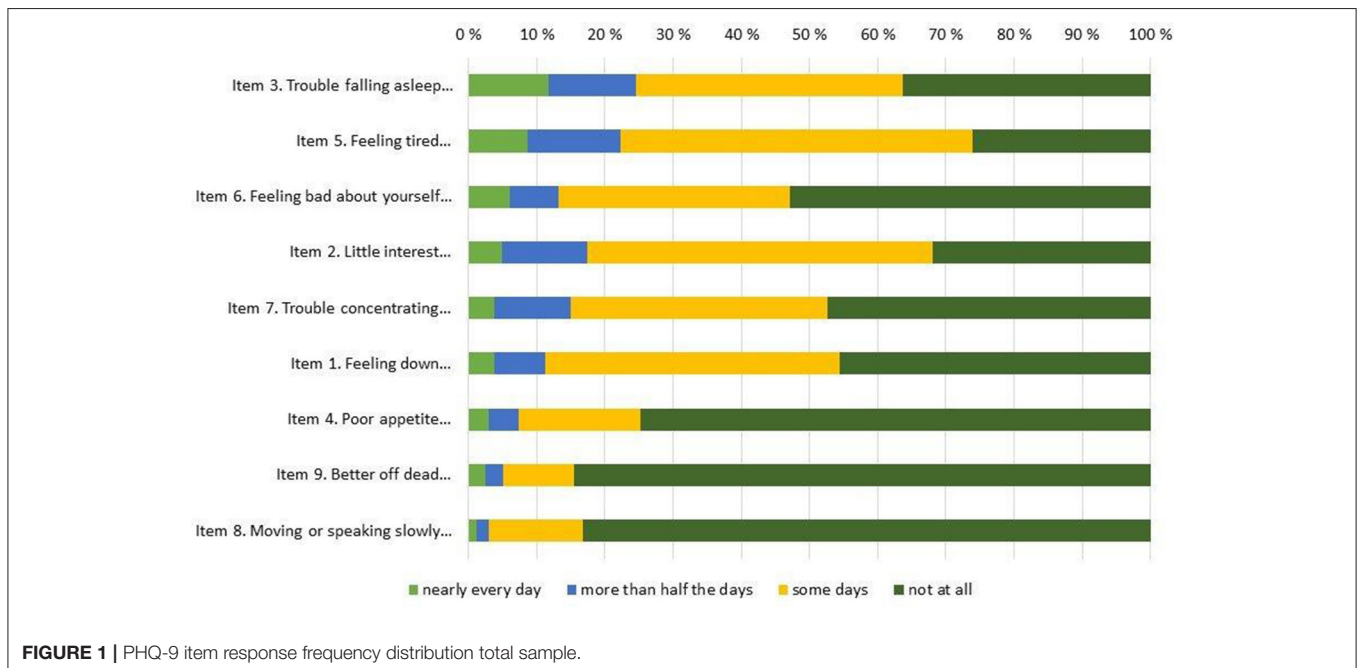


FIGURE 1 | PHQ-9 item response frequency distribution total sample.

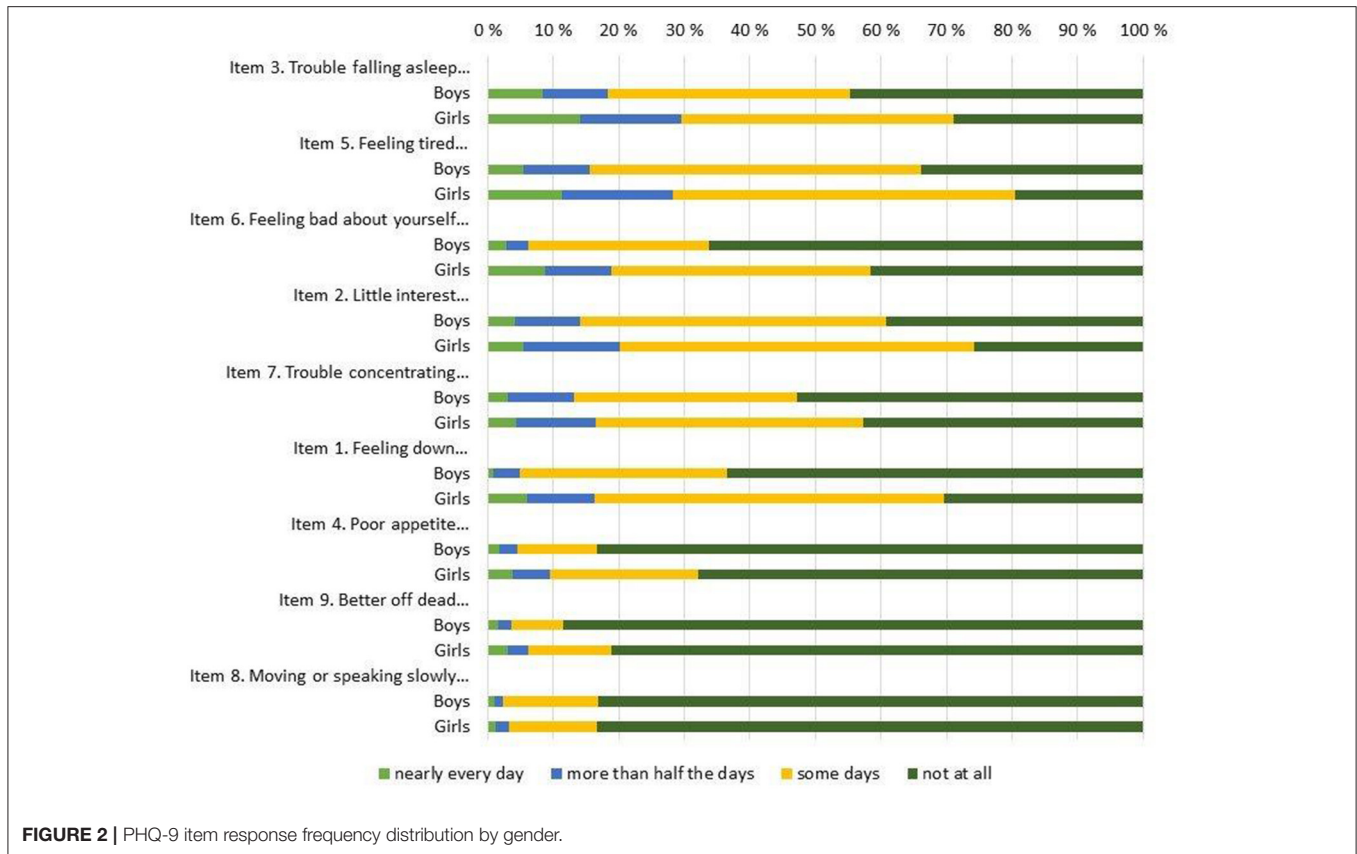


FIGURE 2 | PHQ-9 item response frequency distribution by gender.

TABLE 3 | Originally proposed PHQ-9 severity categories; total sample and by gender.

PHQ-9 Depression severity categories	Total (N = 846) %	Boys (n = 387) %	Girls (n = 459) %
None (PHQ-9: 0–4)	46.8	57.9	37.8
Mild (PHQ-9: 5–9)	36.1	32.8	38.8
Moderate (PHQ-9: 10–14)	11.3	6.7	15.3
Moderately severe (PHQ-9: 15–19)	3.4	1.6	5.0
Severe (PHQ-9: 20–27)	2.4	1.0	3.5
PHQ-9 clinical-level (PHQ-9 ≥ 15)	5.8	2.6	8.5
Moderately severe/severe depression			

family. Finally, the perceived low social status was measured by the MacArthur Scale of Subjective Status—Youth version (Goodman et al., 2001), where the participants placed their family along the Norwegian socio-economic ladder ranging from those families who “have it best” (coded “1”) to those who are “the worst off” (coded “10”).

Psycho-Social Characteristics

Students completed the 5-item School Connectedness Scale (McNeely et al., 2002). The original items (e.g., “I feel like I am part of this school”) utilized Likert-type response options ranging from 1 (“completely agree”) to 5 (“completely disagree”). The items were averaged to compute the scale score (Cronbach’s α

= 0.83, possible range 1–5) such that greater scores reflected the risk factor of greater school disconnectedness. Students also reported if they feel they have at least one close friend. Additional health problems were assessed with two items asking about lifetime suicide attempts and self-harm, and with a single item asking about the presence of a diagnosed chronic illnesses. Finally, participants also completed the 5-item Emotional Problems subscale from the Strengths and Difficulties Questionnaire (SDQ; Goodman and Goodman, 2009) previously used in Norwegian samples (Heiervang et al., 2008; Goodman et al., 2011; Bøe et al., 2016). The original SDQ 3-point responses were summed up to compute the Emotional Problems scores (Cronbach’s α = 0.74, possible range 0–10). In addition, participants were classified into those with clinical-level emotional problems (i.e., scoring at or above the cut-off score of 6) vs. rest, using the SDQ norms for Norwegian adolescents (Rønning et al., 2004; Van Roy et al., 2006).

Statistical Analyses

The initial set of analyses focused on the basic psychometric properties and structure of the PHQ-9, which we examined for the entire sample, and separately for boys and girls. Next, before an instrument is used to compare levels of a latent variable (e.g., depression) between groups, it is important that the instrument is established as measurement equivalent for such groups (also referred to as measurement invariant). To that extent, we used multi-group confirmatory factor analysis (CFA) to examine the

PHQ-9 measurement equivalence for boys and girls as described by Byrne (2012).

Specifically, we examined: (a) configural equality to test whether the factor structure is equal for boys and girls by fitting a model where factor loadings and intercepts were allowed to vary between the two groups; (b) metric equality, to test whether items are interpreted in the same way for both boys and girls by restricting factor loadings to be equal, but letting intercepts vary between the two groups; and (c) scalar equality, to test whether the response scale is used in the same way by boys and girls by restricting factor loadings to be equal and restricting all but one intercept to be equal for the two groups, and full scalar equality by restricting all factor loadings and all intercepts to be equal for the two groups. Direct comparisons (e.g., tests of differences in means) between two groups are valid only if scalar equality holds. The robust maximum likelihood estimator was used because we did not assume multivariate normality for the items. The Satorra-Bentler Chi-square test (S-B χ^2) was used to test statistically whether the CFA models were different. In addition, the root-mean square error of approximation (RMSEA), the comparative fit index (CFI), the standardized root mean square residual (SRMR), and Akaike information criterion (AIC) were used to assess model fit. Suggested cut-off points indicating adequate fit for the RMSEA, CFI and SRMR are ≤ 0.08 , > 0.90 , and ≤ 0.05 , respectively (Byrne, 2012). The AIC has no cut-off points, but lower AIC suggests better fit.

Finally, we examined divergent/convergent validity and the associations between the adolescent PHQ-9 measures and other psycho-social characteristics. These analyses were based on simple, unadjusted regression models. CFA was performed using Mplus, and all other analyses were performed in STATA statistical software.

RESULTS

Sample Characteristics

The results shown in **Table 1** indicate that even though ours was a convenience sample, it appeared highly representative of the Norwegian adolescent population. This is not surprising, given the relative homogeneity of Norwegian society. For example, the basic socio-demographic characteristics, such as residing with both biological parents (approximately 2/3 in our sample vs. “62 per cent for 17 year-olds” for Norway as a whole) and having immigrant background (i.e., 18.6% with at least one parent born outside of Norway from our sample vs. 16.3% “born in Norway of two foreign-born parents and four foreign-born grandparents” for Norway as a whole) appear reflective of the official Norwegian population estimates (Statistics Norway, 2016, 2017). However, even though approximately 2.4% of the Norwegian population self-identifies as Muslim, this proportion was 4.4% in our sample. Whether this was a realistic departure from the Muslim representation among Norwegian adolescents specifically is not known. Most importantly, in terms of psychological adjustment, our sample appears congruent with other youth community-samples from Norway. For example, the lifetime prevalence of suicidal attempts (defined as those who reported such attempts plus those who “refused to answer”) was 8.8% in our sample, vs. 8.2% observed in a representative sample of high-school students (Wichstrøm, 2000). Finally, the average SDQ Emotional Problems scores from our sample, as well as the proportion of clinical-level cases were comparable to the estimates obtained from several representative samples in Norway (Rønning et al., 2004; Van Roy et al., 2006). For example, Van Roy et al. (2006) classified a total of 12.2% of younger adolescents and 13.4% of older adolescents into the SDQ Emotional Problems clinical range in 2006, as compared to 16.4% of our sample in early-to-middle adolescence in 2014.

was 8.8% in our sample, vs. 8.2% observed in a representative sample of high-school students (Wichstrøm, 2000). Finally, the average SDQ Emotional Problems scores from our sample, as well as the proportion of clinical-level cases were comparable to the estimates obtained from several representative samples in Norway (Rønning et al., 2004; Van Roy et al., 2006). For example, Van Roy et al. (2006) classified a total of 12.2% of younger adolescents and 13.4% of older adolescents into the SDQ Emotional Problems clinical range in 2006, as compared to 16.4% of our sample in early-to-middle adolescence in 2014.

Item Statistics and Distributional Properties

The overall response rate was very high for all items, with 6 or fewer omissions on all items, save for item #8 (i.e., “Moving or speaking so slowly that other people could have noticed...”) where 14 participants failed to respond. **Table 2** shows basic descriptive statistics for all 9 individual items (top of **Table 2**) and for the entire scale (bottom of **Table 2**), both for the entire sample and for boys and girls separately. In addition, the response distributions for individual items are shown in **Figure 1** (for the entire sample) and **Figure 2** (by gender), ordered by the prevalence of the most severe response category (i.e., “nearly every day”). The **Figure 1** pattern indicates that the sleep problems (item #3), energy loss (item #5), low self-esteem (#6), and anhedonia (item #2) were the items endorsed with greatest frequency, whereas movement problems (item #8) and suicidal ideation (item #9) were the items endorsed with lowest frequency. **Figure 2** shows discrepancies in the item response patterns between boys and girls. For example, nearly 60% of girls endorsed the low self-esteem item (#6) in some form, as opposed to only 34% of boys. In general, girls appeared more likely to endorse all items, save for item #8, as evident in the average scores (shown in **Table 2**) and response distribution (shown in **Figure 2**).

Distributions for the originally proposed PHQ-9 diagnostic categories are shown in **Table 3**. As would be expected, the overall PHQ-9 scores did not follow the normal distribution, as the majority of our participants reported no or only mild depressive symptomatology (also see **Figure 1, 2** for individual item distribution). The average PHQ-9 score was 6.89 (5.13) for girls and 4.57 (3.98) for boys, while 8.5% of girls and 2.6% of boys were classified into the original PHQ-9 categories indicative of MDD (i.e., PHQ-9 scores ≥ 15).

Confirmatory Factor Analysis and Measurement Equivalence

The results from the confirmatory factor analysis (CFA)—where all PHQ items were set to load on one latent factor (i.e., “depression”) according to its theoretical conceptualization—are shown in **Table 4**, including standardized factor loadings and fit indices. The CFA confirmed a single-factor solution, with standardized factor loadings ranging from 0.51 (item #8, “Moving or speaking so slowly that other people could have noticed”) to 0.77 (item #6, “Feeling bad about yourself, or that you’re a failure or that you’ve let yourself or your family down”)

TABLE 4 | Confirmatory factor analysis for a single-factor solution for the PHQ-9 items; total sample and by gender.

PHQ-9 items	Standardized factor loadings		
	Total (N = 846)	Boys (n = 387)	Girls (n = 459)
1. Feeling down, depressed, irritable, or hopeless	0.73	0.63	0.75
2. Little interest or pleasure in doing things	0.56	0.46	0.61
3. Trouble falling asleep, staying asleep, or sleeping too much	0.62	0.56	0.62
4. Poor appetite, weight loss, or overeating	0.63	0.63	0.61
5. Feeling tired, or having little energy	0.67	0.55	0.70
6. Feeling bad about yourself, or that you're a failure, or that you've let yourself or your family down	0.77	0.71	0.77
7. Trouble concentrating on things like school work, reading, or watching TV	0.61	0.57	0.64
8. Moving or speaking so slowly that other people could have noticed	0.51	0.40	0.60
9. Thought that you would be better off dead, or of hurting yourself in some way	0.62	0.64	0.63
FIT INDICES			
S-B χ^2 (df)	148.8 (27)	74.6 (27)	98.6 (27)
RMSEA	0.073	0.067	0.076
CFI	0.93	0.91	0.94
SRMR	0.04	0.049	0.04

TABLE 5 | Tests of PHQ-9 measurement equivalence across gender.

Model	Comparison model	S-B χ^2 (df)	S-B $\Delta \chi^2$ (df)	P	RMSEA	CFI	SRMR	AIC
A. Basic model (configural equality)	–	171.6 (54)			0.072	0.926	0.045	14714
B. Full metric equality	A.	187.3 (63)	8.75 (9)	0.44	0.068	0.922	0.093	14733
C. Partial scalar equality	B.	238.5 (70)	65.33 (7)	<0.001	0.075	0.894	0.100	14783
D. Full scalar equality	B.	247.8 (71)	77.95 (8)	<0.001	0.077	0.889	0.100	14793

Shown are results from the CFA analyses examining PHQ-9 equivalence between boys and girls. As we did not assume multivariate normality, the robust maximum likelihood estimator (MLR) was used. Model fit was assessed by the Satorra-Bentler (S-B) adjusted χ^2 difference test, root mean square error of approximation (RMSEA), the comparative fit index (CFI), standardized root mean square residual (SRMR), and Akaike's information criterion (AIC).

for the entire sample. This single-factor solution displayed acceptable fit to the data (bottom of **Table 4**). Cronbach's alpha for the items was 0.86. Conceptually identical CFA results were obtained for boys and girls when analyzed separately (**Table 4**), including the single factor solutions, and the poorest and best performance exhibited by the #8 "movement" item (factor loading_{Boys} = 0.40, factor loading_{Girls} = 0.60) and #6 "self-esteem" item (factor loading_{Boys} = 0.71, factor loading_{Girls} = 0.77). The model fit the data adequately for both girls and boys (bottom of **Table 4**); Cronbach's alpha_{Boys} = 0.81; Cronbach's alpha_{Girls} = 0.88.

The results from multi-group equality testing are presented in **Table 5**. Metric equality was assessed by comparing the configural model (Model A) with a model where factor loadings were restricted to be equal (Model B). The S-B χ^2 test was not statistically significant. The RMSEA and CFI values did not change substantially and remained within the acceptable range. The AIC was higher for Model B compared to Model A, and the SRMR was higher and outside of the acceptable range for Model B, indicating worse absolute fit compared to Model A. Overall, the results do not support the assumption of metric equality across genders. This also means that scalar equality was not supported. In practice, such a set of results implies that further formal tests of gender differences in our sample should not be

conducted without caution. This multi-group equality testing was repeated with the weaker item #8 and/or the highly skewed item #9 omitted from the CFA. The results showed again that metric and scalar equivalence was not supported in either case.

Adolescent Psycho-Social Characteristics and Depressive Symptomatology

Table 6 documents the associations between the adolescents' psycho-social characteristics and the PHQ-9 continuous scores, which were investigated separately for boys and girls because of the aforementioned results concerning measurement non-equivalence. These associations indicate that the basic demographic characteristics—including religious affiliation and parental or adolescent immigrant background—appeared to have minimal associations with depressive symptomatology among adolescents from our sample. Demographic characteristics associated with greater depressive symptomatology were the school grade (for girls only), high residential instability and the low perceived SES (for both boys and girls). A set of more specific risk-factors and health characteristics—such as school dissatisfaction, lack of close friendships, history of suicide attempts and self-harms, and elevated emotional problems as measured by the SDQ subscale—were consistently and significantly associated with depressive symptomatology across

TABLE 6 | Adolescent depressive symptomatology as a function of psycho-social characteristics; crude associations by gender.

Variables	Boys (n = 387)	Girls (n = 459)
	b (s.e.)	b (s.e.)
DEMOGRAPHICS		
School grade (8 through 12)	0.18 (0.14)	0.50 (0.16)***
Religion		
None declared ^a	–	–
Christianity	–0.75 (0.43)	–0.74 (0.58)
Islam	–0.13 (1.04)	–0.21 (1.23)
Other	–0.58 (0.97)	–2.33 (1.56)
Residential Instability		
0 moves/school changes ^a	–	–
1 move/school change	0.97 (0.56)	1.07 (0.71)
2 or more moves/school changes	2.27 (0.58)***	2.57 (0.73)***
Intact biological family		
No ^a	–	–
Yes	–1.18 (0.42)**	–0.24 (0.51)
Born in Norway		
No ^a	–	–
Yes	–0.41 (0.69)	–0.12 (0.79)
Immigrant background		
No ^a	–	–
Yes	0.19 (0.51)	–0.21 (0.62)
Perceived low SES	0.27 (0.11)*	0.32 (0.13)*
PSYCHO-SOCIAL CHARACTERISTICS		
School dissatisfaction	2.47 (0.30)***	3.48 (0.29)***
Close friend		
Other ^a	–	–
At least one close friend	–3.44 (0.79)***	–3.81 (0.79)***
Ever attempted suicide		
No ^a	–	–
Yes	9.16 (1.18)***	10.24 (1.07)***
Don't know/Will not respond	2.03 (0.89)*	5.00 (0.94)***
Ever self-harmed		
No ^a	–	–
Yes	5.20 (0.69)***	5.61 (0.54)***
Don't know/Will not respond	3.22 (1.08)**	5.77 (0.93)***
Chronic illness diagnosis		
No ^a	–	–
Yes	1.49 (0.49)**	0.98 (0.57)
Emotional problems (SDQ scores)	1.17 (0.09)***	1.33 (0.07)***
Emotional problems (SDQ clinical)		
No ^a	–	–
Yes	6.09 (0.79)***	6.22 (0.47)***

Shown are the unstandardized regression coefficients (i.e., b) from crude regression models examining the associations between each individual characteristic and adolescent depressive symptomatology as measured by the PHQ-9 continuous scores. For categorical predictors, the reference group is noted by superscript ^a.

* $p < 0.05$.

** $p < 0.01$.

*** $p < 0.001$.

both genders (Table 6). Specifically, for both boys and girls, the explicit report of lifetime suicide attempt was associated with a roughly 10-point increase in the PHQ-9 scores (Table 6). Similarly, each 1-point increase in the SDQ Emotional Problems scores was associated with the significant increases of 1.17-point and 1.33-point in the PHQ-9 scores, or standardized regression coefficients of $r = 0.54$ and $r = 0.63$, $p < 0.001$ for boys and girls, respectively. An identical pattern of results was obtained when the SDQ clinical categories were examined: membership in the SDQ Emotional Problems category was associated with an approximately 6-point increase in the PHQ-9 scores for both genders. Among these psycho-social characteristics, the only one exhibiting gender differential was the self-reported diagnosis of chronic illness (Table 6), such that poorer physical health was significantly associated with greater symptoms of depression for boys but not for girls from our sample.

DISCUSSION

Our results appear consistent with previous international reports examining depression in adolescent school samples using the PHQ-9 instrument, including the school samples of Chinese and Nigerian adolescents (prevalence of *moderately severe/severe depression* = 5.2 and 5.1%, respectively; Fatiregun and Kumapayi, 2014; Tsai et al., 2014) and the community samples of American adolescents (*moderately severe/severe depression* prevalence = 6.5%, even though this study used a somewhat different clinical classification algorithm; Rhew et al., 2016). Most importantly, our estimate of approximately 6% prevalence of clinically-elevated symptoms is relatively congruent with other recent reports of current depressive symptomatology among Norwegian adolescents, ranging from 2.6% for MDD and 6.3% for depressive disorder not otherwise specified during the 2-month window (Sund et al., 2011), to the 11% prevalence of the less specific “high depressive symptoms” during the past week (Abebe et al., 2016). Nevertheless, these estimates should be interpreted as preliminary given that we used the cut-off scores established internationally but not in Norway, and the sample which was not necessarily representative of all Norwegian adolescents.

Our CFA results confirmed that the PHQ-9 measures adolescent depression as a unidimensional theoretical construct. Relatively high factor loadings (i.e., all loadings ≥ 0.40) and solid fit indices were observed for the entire sample, and for boys and girls separately. One item (e.g., “Moving or speaking so slowly that other people could have noticed”) consistently showed poorer, yet still acceptable performance. It is possible that the wording and/or meaning were difficult for younger participants to comprehend, as evident in the relatively high number of missing responses on this particular item. Specifically, this question asked about what other people could have noticed, which may be confusing for youngest participants. Refinement or alternative wording of this item may improve the scale performance.

In agreement with previous evidence that girls tend to exhibit greater depressive symptomatology than boys starting around the age of 13 (Hankin et al., 1998; Wichstrøm, 1999; Twenge and Nolen-Hoeksema, 2002), and with other Norwegian reports

demonstrating substantive gender differences in measurements of adolescent depression (Lundervold et al., 2013), girls from our sample appeared to have greater depression problems. However, we did not proceed to perform formal tests of gender differences because we could not fully establish metric or scalar equality for the PHQ-9 instrument across genders. This means that girls and boys cannot be directly compared in terms of the PHQ-9 scores without reservations, contrasting previous international reports of the PHQ-9 measurement invariance by gender and age (Yu et al., 2012; Petersen et al., 2015). However, our results may not be out of place in a Norwegian context, where only partial measurement-equivalence for gender was reported for the SDQ instrument in adolescent populations (Bøe et al., 2016). Clearly, the causes and implications of these non-equivalence results should be investigated further when it comes to the use of PHQ-9 (and possibly other instruments for which measurement equivalence was not fully tested before use) in Norway. Nevertheless, it should also be noted that these findings do not preclude utilization of the PHQ-9 in research practice, as long as the analyses are stratified by gender (Bøe et al., 2016).

Finally, we examined divergent and convergent validity, as well as the associations between the PHQ-9 scores for boys and girls separately. Given that the PHQ-9 was not previously used in Norway, we only had general expectations that the direction and magnitude of these associations would be comparable to those from other Norwegian studies on adolescent depression. This was generally the case. We found no strong evidence for the association between depression in adolescents and their basic demographic characteristics such as religion or immigrant status. Such a pattern may be reflecting the somewhat inconclusive state of knowledge regarding the mental health among adolescent immigrants in Norway (Abebe et al., 2014). Other putative risk factors—including age, not living together with both biological parents, and diagnosed chronic illness—were associated with depressive symptomatology differently for boys and girls. Older age appeared to be a risk factor for depression among girls only, while not living with both biological parents or chronic illness appeared to be risk factors only among boys. In contrast, high number of moves and school changes, perceptions of one's family as poor by Norwegian standards, social isolation and lack of close friendships, history of suicide attempts and self-harms, and elevated SDQ emotional problems were uniformly and significantly associated with elevated depression for both genders. Similar patterns—for example, the associations between elevated depression and residential instability, low SES, not living with both biological parents, school dissatisfaction and lack of close friendships (Sund et al., 2003; Myklestad et al., 2012)—were observed in other Norwegian adolescents samples. More importantly, our results demonstrated the high convergence between the PHQ-measures with other theoretically related constructs such as suicidality or impairments in emotional adjustment (Goodman and Goodman, 2009; Hawton et al., 2012; Silverstone et al., 2015), where the standardized regression coefficients with the SDQ Emotional problems scores exceeded 0.5 for both genders. They also suggest the potential effectiveness of the PHQ-9 in preliminary identification of youth at high risk for depression using the large-scale epidemiological surveys.

The current study is limited by several factors, including its convenience sample and reliance on self-reports for all indicators. We also used a somewhat shorter time reference and assessed the PHQ-9 depressive symptomatology during the last 7 days as recommended by the American Psychiatric Association (2016). Most importantly, our adolescent PHQ-9 self-reports were not validated against the external diagnostic criteria, such as the official psychiatric diagnoses for example. However, it should be noted that the PHQ-9 has been internationally validated against various diagnostic interviews in multiple adolescent studies (Richardson et al., 2010; Allgaier et al., 2012; Ganguly et al., 2013; Tsai et al., 2014), with some reports even using the PHQ-9 itself as a gold standard against which to validate other measures of depression in community youth samples (Rhew et al., 2016). Nevertheless, external validation of the PHQ-9 against diagnostic interviews would have strengthened our study, especially because we used somewhat conservative criteria and because there may be cultural, ethnic, and national variations in depressive symptomatology and corresponding clinical cut-offs (Kroenke and Spitzer, 2002; Richardson et al., 2010; Allgaier et al., 2012; Ganguly et al., 2013; Jaber et al., 2015). Full validation, including the validation of the originally suggested cut-offs and associated diagnostic categories, is therefore necessary before any clinical evaluation is to be undertaken using the PHQ-9 among Norwegian youth.

Despite these limitations, the PHQ-9 appears to be a promising research tool, potentially offering clinically-relevant classification of adolescent depressive symptomatology in addition to the symptom severity captured by continuous scores. Most importantly, given the internationally validated and streamlined clinical criteria, the use of PHQ-9 has the potential to advance not only national, but also cross-national assessments and comparisons of depressive symptomatology among youth populations. As such, we encourage further exploration of the PHQ-9 adolescent instrument in developmental epidemiology research and in studies of general adolescent health and development. Nevertheless, further study of the hereby observed measurement non-equivalence, as well as a comprehensive validation against the proper diagnostic criteria are required before the PHQ-9 is to be used for youth psychiatric screening in Norway.

ETHICS STATEMENT

This study was carried out in accordance with the recommendations of the Norwegian Centre for Research Data (NSD; <http://www.nsd.uib.no/nsd/english/pvo.html>; Protocol#39513) with informed consent from all participants or guardians.

AUTHOR CONTRIBUTIONS

The present report was drawn from a larger adolescent development project, directed by JBA and GSB. Both JBA and GSB contributed to study design, development of research questions, data analyses, and writing. Both authors approved the final manuscript.

REFERENCES

- Abebe, D. S., Frøyland, L. R., Bakken, A., and von Soest, T. (2016). Municipal-level differences in depressive symptoms among adolescents in Norway: results from the cross-national Ungdata study. *Scand. J. Public Health* 44, 47–54. doi: 10.1177/1403494815604764
- Abebe, D. S., Lien, L., and Hjelde, K. H. (2014). What we know and don't know about mental health problems among immigrants in Norway. *J. Immigr. Minor. Health* 16, 60–67. doi: 10.1007/s10903-012-9745-9
- Adewuya, A. O., Ola, B. A., and Afolabi, O. O. (2006). Validity of the Patient Health Questionnaire (PHQ-9) as a screening tool for depression amongst Nigerian university students. *J. Affect. Disord.* 96, 89–93. doi: 10.1016/j.jad.2006.05.021
- Allgair, A.-K., Pietsch, K., Frühe, B., Sigl-Glückner, J., and Schulte-Körne, G. (2012). Screening for depression in adolescents: validity of the Patient Health Questionnaire in pediatric care. *Depress. Anxiety* 29, 906–913. doi: 10.1002/da.21971
- American Psychiatric Association (2016). *DSM V Severity measure for Depression, Child Age 11 to 17 (PHQ-9 Modified for Adolescents (PHQ-A), Adapted)*. Available online at: https://www.psychiatry.org/File%20Library/Psychiatrists/Practice/DSM/APA_DSM5_Severity-Measure-For-Depression-Child-Age-11-to-17.pdf
- Angold, A., Costello, E. J., Messer, S. C., Pickles, A., Winder, F., and Silver, D. (1995). The development of a short questionnaire for use in epidemiological studies of depression in children and adolescents. *Int. J. Methods Psychiatr. Res.* 5, 237–249.
- Birmaher, B., Ryan, N. D., Williamson, D. E., Brent, D. A., Kaufman, J., Dahl, R. E., et al. (1996). Childhood and adolescent depression: a review of the past 10 years. Part I. *J. Am. Acad. Child Adolesc. Psychiatry* 35, 1427–1439. doi: 10.1097/00004583-199611000-00011
- Bøe, T., Hysing, M., Skogen, J. C., and Breivik, K. (2016). The Strengths and Difficulties Questionnaire (SDQ): factor structure and gender equivalence in Norwegian adolescents. *PLoS ONE* 11:e0152202. doi: 10.1371/journal.pone.0152202
- Brondbo, P. H., Mathiassen, B., Martinussen, M., Heiervang, E., Eriksen, M., Moe, T. F., et al. (2011). The Strengths and Difficulties Questionnaire as a screening instrument for Norwegian child and adolescent mental health services, application of UK scoring algorithms. *Child Adolesc. Psychiatry Ment. Health* 5:32. doi: 10.1186/1753-2000-5-32
- Bronfenbrenner, U. (1979). *The Ecology of Human Development: Experiments by Nature and Design*. Cambridge, MA: Harvard University Press.
- Brunborg, G. S., Burdzovic Andreas, J., and Kvaavik, E. (2017). Social media use and episodic heavy drinking among adolescents. *Psychol. Rep.* 120, 475–490. doi: 10.1177/0033294117697090
- Byrne, B. M. (2012). *Structural Equation Modeling with Mplus: Basic concepts, Applications, and Programming*. New York, NY: Routledge.
- Dekker, M. C., Ferdinand, R. F., van Lang, N. D., Bongers, I. L., van der Ende, J., and Verhulst, F. C. (2007). Developmental trajectories of depressive symptoms from early childhood to late adolescence: gender differences and adult outcome. *J. Child Psychol. Psychiatry* 48, 657–666. doi: 10.1111/j.1469-7610.2007.01742.x
- Derdikman-Eiron, R., Indredavik, M. S., Bakken, I. J., Bratberg, G. H., Hjemdal, O., and Colton, M. (2012). Gender differences in psychosocial functioning of adolescents with symptoms of anxiety and depression: longitudinal findings from the Nord-Trøndelag Health Study. *Soc. Psychiatry Psychiatr. Epidemiol.* 47, 1855–1863. doi: 10.1007/s00127-012-0492-y
- Derogatis, L. R., Lipman, R. S., Rickels, K., Uhlenhuth, E. H., and Covi, L. (1974). The Hopkins Symptom Checklist (HSCL): a self-report symptom inventory. *Behav. Sci.* 19, 1–15. doi: 10.1002/bs.3830190102
- Dunn, V., and Goodyear, I. M. (2006). Longitudinal investigation into childhood and adolescence-onset depression: psychiatric outcome in early adulthood. *Br. J. Psychiatry* 188, 216–222. doi: 10.1192/bjp.188.3.216
- Essau, C. A., Lewinsohn, P. M., Seeley, J. R., and Sasagawa, S. (2010). Gender differences in the developmental course of depression. *J. Affect. Disord.* 127, 185–190. doi: 10.1016/j.jad.2010.05.016
- Fatiregun, A. A., and Kumapayi, T. E. (2014). Prevalence and correlates of depressive symptoms among in-school adolescents in a rural district in southwest Nigeria. *J. Adolesc.* 37, 197–203. doi: 10.1016/j.adolescence.2013.12.003
- Ganguly, S., Samanta, M., Roy, P., Chatterjee, S., Kaplan, D. W., and Basu, B. (2013). Patient Health Questionnaire-9 as an effective tool for screening of depression among Indian adolescents. *J. Adolesc. Health* 52, 546–551. doi: 10.1016/j.jadohealth.2012.09.012
- Goodman, A., and Goodman, R. (2009). Strengths and Difficulties Questionnaire as a dimensional measure of child mental health. *J. Am. Acad. Child Adolesc. Psychiatry* 48, 400–403. doi: 10.1097/CHI.0b013e3181985068
- Goodman, A., Heiervang, E., Collishaw, S., and Goodman, R. (2011). The 'DAWBA bands' as an ordered-categorical measure of child mental health: description and validation in British and Norwegian samples. *Soc. Psychiatry Psychiatr. Epidemiol.* 46, 521–532. doi: 10.1007/s00127-010-0219-x
- Goodman, E., Adler, N. E., Kawachi, I., Frazier, A. L., Huang, B., and Colditz, G. A. (2001). Adolescents' perceptions of social status: development and evaluation of a new indicator. *Pediatrics* 108:E31. doi: 10.1542/peds.108.2.e31
- Goodman, R., Ford, T., Simmons, H., Gatward, R., and Meltzer, H. (2000a). Using the Strengths and Difficulties Questionnaire (SDQ) to screen for child psychiatric disorders in a community sample. *Br. J. Psychiatry* 177, 534–539. doi: 10.1192/bjp.177.6.534
- Goodman, R., Renfrew, D., and Mullick, M. (2000b). Predicting type of psychiatric disorder from Strengths and Difficulties Questionnaire (SDQ) scores in child mental health clinics in London and Dhaka. *Eur. Child Adolesc. Psychiatry* 9, 129–134. doi: 10.1007/s007870050008
- Hankin, B. L., Abramson, L. Y., Moffitt, T. E., Silva, P. A., McGee, R., and Angell, K. E. (1998). Development of depression from preadolescence to young adulthood: emerging gender differences in a 10-year longitudinal study. *J. Abnorm. Psychol.* 107, 128–140. doi: 10.1037/0021-843X.107.1.128
- Harrington, R., Fudge, H., Rutter, M., Pickles, A., and Hill, J. (1990). Adult outcomes of childhood and adolescent depression: I. psychiatric status. *Arch. Gen. Psychiatry* 47, 465–473. doi: 10.1001/archpsyc.1990.01810170065010
- Hawton, K., Saunders, K. E., and O'Connor, R. C. (2012). Self-harm and suicide in adolescents. *Lancet* 379, 2373–2382. doi: 10.1016/S0140-6736(12)60322-5
- Heiervang, E., Goodman, A., and Goodman, R. (2008). The Nordic advantage in child mental health: separating health differences from reporting style in a cross-cultural comparison of psychopathology. *J. Child Psychol. Psychiatry* 49, 678–685. doi: 10.1111/j.1469-7610.2008.01882.x
- Heiervang, E., Stormark, K. M., Lundervold, A. J., Heimann, M., Goodman, R., Posserud, M.-B., et al. (2007). Psychiatric Disorders in Norwegian 8- to 10-Year-Olds. *J. Am. Acad. Child Adolesc. Psychiatry* 46, 438–447. doi: 10.1097/chi.0b013e31803062bf
- Heyerdahl, S., Kvernmo, S., and Wichstrøm, L. (2004). Self-reported behavioural/emotional problems in Norwegian adolescents from multiethnic areas. *Eur. Child Adolesc. Psychiatry* 13, 64–72. doi: 10.1007/s00787-004-0359-1
- Indredavik, M. S., Vik, T., Heyerdahl, S., Kulseng, S., and Brubakk, A. M. (2005). Psychiatric symptoms in low birth weight adolescents, assessed by screening questionnaires. *Eur. Child Adolesc. Psychiatry* 14, 226–236. doi: 10.1007/s00787-005-0459-6
- Jaber, R. M., Farroukh, M., Ismail, M., Najda, J., Sobh, H., Hammad, A., et al. (2015). Measuring depression and stigma towards depression and mental health treatment among adolescents in an Arab-American community. *Int. J. Cult. Ment. Health* 8, 247–254. doi: 10.1080/17542863.2014.953188
- Johnson, J. G., Harris, E. S., Spitzer, R. L., and Williams, J. B. (2002). The Patient Health Questionnaire for adolescents: validation of an instrument for the assessment of mental disorders among adolescent primary care patients. *J. Adolesc. Health* 30, 196–204. doi: 10.1016/S1054-139X(01)00333-0
- Kovacs, M., Obrosky, S., and George, C. (2016). The course of major depressive disorder from childhood to young adulthood: recovery and recurrence in a longitudinal observational study. *J. Affect. Disord.* 203, 374–381. doi: 10.1016/j.jad.2016.05.042
- Kroenke, K., and Spitzer, R. L. (2002). The PHQ-9: a new depression and diagnostic severity measure. *Psychiatr. Ann.* 32, 509–515. doi: 10.3928/0048-5713-20020901-06
- Kroenke, K., Spitzer, R. L., and Williams, J. B. (2001). The PHQ-9: validity of a brief depression severity measure. *J. Gen. Intern. Med.* 16, 606–613. doi: 10.1046/j.1525-1497.2001.016009606.x
- Larsson, B., Ingul, J., Jozefiak, T., Leikanger, E., and Sund, A. M. (2016). Prevalence, stability, 1-year incidence and predictors of depressive symptoms among Norwegian adolescents in the general population as measured by the

- Short Mood and Feelings Questionnaire. *Nord. J. Psychiatry* 70, 290–296. doi: 10.3109/08039488.2015.1109137
- Lewinsohn, P. M., Rohde, P., and Seeley, J. R. (1998). Major depressive disorder in older adolescents: prevalence, risk factors, and clinical implications. *Clin. Psychol. Rev.* 18, 765–794. doi: 10.1016/S0272-7358(98)00010-5
- Lundervold, A. J., Breivik, K., Posserud, M. B., Stormark, K. M., and Hysing, M. (2013). Symptoms of depression as reported by Norwegian adolescents on the Short Mood and Feelings Questionnaire. *Front. Psychol.* 4:613. doi: 10.3389/fpsyg.2013.00613
- McNeely, C. A., Nonnemaker, J. M., and Blum, R. W. (2002). Promoting school connectedness: evidence from the national longitudinal study of adolescent health. *J. School Health* 72, 138–146. doi: 10.1111/j.1746-1561.2002.tb06533.x
- Merikangas, K. R., He, J. P., Brody, D., Fisher, P. W., Bourdon, K., and Koretz, D. S. (2010). Prevalence and treatment of mental disorders among US children in the 2001–2004 NHANES. *Pediatrics* 125, 75–81. doi: 10.1542/peds.2008-2598
- Myklestad, L., Røysamb, E., and Tambs, K. (2012). Risk and protective factors for psychological distress among adolescents: a family study in the Nord-Trøndelag Health Study. *Soc. Psychiatry Psychiatr. Epidemiol.* 47, 771–782. doi: 10.1007/s00127-011-0380-x
- Norwegian Institute of Public Health (2014). *Public Health Report 2014: Mental Health Disorders among Children and Adolescents in Norway*. Norwegian Institute of Public Health, Oslo, Norway. Available online at: In: <https://www.fhi.no/en/op/public-health-report-2014/health--disease/mental-health-children-adolescents/>
- Petersen, J. J., Paulitsch, M. A., Hartig, J., Mergenthal, K., Gerlach, F. M., and Gensichen, J. (2015). Factor structure and measurement invariance of the Patient Health Questionnaire-9 for female and male primary care patients with major depression in Germany. *J. Affect. Disord.* 170, 138–142. doi: 10.1016/j.jad.2014.08.053
- Rhew, I. C., Monahan, K. C., Oesterle, S., and Hawkins, J. D. (2016). The Communities that Care Brief Depression Scale: psychometric properties and criterion validity. *J. Commun. Psychol.* 44, 391–398. doi: 10.1002/jcop.21766
- Richardson, L. P., McCauley, E., Grossman, D. C., McCarty, C. A., Richards, J., Russo, J. E., et al. (2010). Evaluation of the Patient Health Questionnaire (PHQ-9) for detecting major depression among adolescents. *Pediatrics* 126, 1117–1123. doi: 10.1542/peds.2010-0852
- Rohde, P., Lewinsohn, P. M., Klein, D. N., Seeley, J. R., and Gau, J. M. (2012). Key characteristics of Major Depressive Disorder occurring in childhood, adolescence, emerging adulthood, and adulthood. *Clin. Psychol. Sci.* 1, 41–53. doi: 10.1177/2167702612457599
- Rønning, J. A., Handegaard, B. H., Sourander, A., and Mørch, W. T. (2004). The Strengths and Difficulties Self-Report Questionnaire as a screening instrument in Norwegian community samples. *Eur. Child Adolesc. Psychiatry* 13, 73–82. doi: 10.1007/s00787-004-0356-4
- Silverstone, P. H., Bercov, M., Suen, V. Y., Allen, A., Cribben, I., Goodrick, J., et al. (2015). Initial findings from a novel school-based program, empathy, which may help reduce depression and suicidality in youth. *PLoS ONE* 10:e0125527. doi: 10.1371/journal.pone.0125527
- Statistics Norway (2016). *Families and households*, 1 January 2016. Statistics Norway (Statistisk sentralbyrå). Available online at: <https://www.ssb.no/en/familie/>
- Statistics Norway (2017). *Immigrants and Norwegian-Born to Immigrant Parents*, 1 January 2016. Statistics Norway (Statistisk sentralbyrå). Available online at: <https://www.ssb.no/en/befolkning/statistikker/innvbf/aar/2016-03-03>
- Strand, B. H., Dalgard, O. S., Tambs, K., and Rognerud, M. (2003). Measuring the mental health status of the Norwegian population: a comparison of the instruments SCL-25, SCL-10, SCL-5 and MHI-5 (SF-36). *Nord. J. Psychiatry* 57, 113–118. doi: 10.1080/08039480310000932
- Sund, A. M., Larsson, B., and Wichstrøm, L. (2003). Psychosocial correlates of depressive symptoms among 12-14-year-old Norwegian adolescents. *J. Child Psychol. Psychiatry* 44, 588–597. doi: 10.1111/1469-7610.00147
- Sund, A. M., Larsson, B., and Wichstrøm, L. (2011). Prevalence and characteristics of depressive disorders in early adolescents in central Norway. *Child Adolesc. Psychiatry Ment. Health* 5, 28–28. doi: 10.1186/1753-2000-5-28
- Tambs, K., and Moum, T. (1993). How well can a few questionnaire items indicate anxiety and depression? *Acta Psychiatr. Scand.* 87, 364–367. doi: 10.1111/j.1600-0447.1993.tb03388.x
- Thapar, A., Collishaw, S., Pine, D. S., and Thapar, A. (2012). Depression in adolescence. *Lancet* 379, 1056–1067. doi: 10.1016/S0140-6736(11)60871-4
- Tsai, F. J., Huang, Y. H., Liu, H. C., Huang, K. Y., Huang, Y. H., and Liu, S. I. (2014). Patient Health Questionnaire for school-based depression screening among Chinese adolescents. *Pediatrics* 133, e402–e409. doi: 10.1542/peds.2013-0204
- Twenge, J. M., and Nolen-Hoeksema, S. (2002). Age, gender, race, socioeconomic status, and birth cohort difference on the children's depression inventory: a meta-analysis. *J. Abnorm. Psychol.* 111, 578–588. doi: 10.1037/0021-843X.111.4.578
- Van Roy, B., Grøholt, B., Heyerdahl, S., and Clench-Aas, J. (2006). Self-reported strengths and difficulties in a large Norwegian population 10–19 years: age and gender specific results of the extended SDQ-questionnaire. *Eur. Child Adolesc. Psychiatry* 15, 189–198. doi: 10.1007/s00787-005-0521-4
- Weissman, M. M., Wolk, S., Goldstein, R. B., Moreau, D., Adams, P., Greenwald, S., et al. (1999). Depressed adolescents grown up. *J. Am. Med. Assoc.* 281, 1707–1713. doi: 10.1001/jama.281.18.1707
- Wichstrøm, L. (1999). The emergence of gender difference in depressed mood during adolescence: the role of intensified gender socialization. *Dev. Psychol.* 35, 232–245. doi: 10.1037/0012-1649.35.1.232
- Wichstrøm, L. (2000). Predictors of adolescent suicide attempts: a nationally representative longitudinal study of Norwegian adolescents. *J. Am. Acad. Child Adolesc. Psychiatry* 39, 603–610. doi: 10.1097/00004583-200005000-00014
- Yu, X., Tam, W. W., Wong, P. T., Lam, T. H., and Stewart, S. M. (2012). The Patient Health Questionnaire-9 for measuring depressive symptoms among the general population in Hong Kong. *Compr. Psychiatry* 53, 95–102. doi: 10.1016/j.comppsy.2010.11.002

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2017 Burdzovic Andreas and Brunborg. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Advantages of publishing in Frontiers



OPEN ACCESS

Articles are free to read for greatest visibility and readership



FAST PUBLICATION

Around 90 days from submission to decision



HIGH QUALITY PEER-REVIEW

Rigorous, collaborative, and constructive peer-review



TRANSPARENT PEER-REVIEW

Editors and reviewers acknowledged by name on published articles

Frontiers

Avenue du Tribunal-Fédéral 34
1005 Lausanne | Switzerland

Visit us: www.frontiersin.org

Contact us: info@frontiersin.org | +41 21 510 17 00



REPRODUCIBILITY OF RESEARCH

Support open data and methods to enhance research reproducibility



DIGITAL PUBLISHING

Articles designed for optimal readership across devices



FOLLOW US

[@frontiersin](https://twitter.com/frontiersin)



IMPACT METRICS

Advanced article metrics track visibility across digital media



EXTENSIVE PROMOTION

Marketing and promotion of impactful research



LOOP RESEARCH NETWORK

Our network increases your article's readership