COGNITION AND INTERACTION: FROM COMPUTERS TO SMART OBJECTS AND AUTONOMOUS AGENTS

EDITED BY: Amon Rapp, Maurizio Tirassa and Tom Ziemke

frontiers Research Topics

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

COGNITION AND INTERACTION: FROM COMPUTERS TO SMART OBJECTS AND AUTONOMOUS AGENTS

Topic Editors:
Amon Rapp, University of Turin, Italy
Maurizio Tirassa, University of Turin, Italy
Tom Ziemke, Linköping University, University of Skövde, Sweden
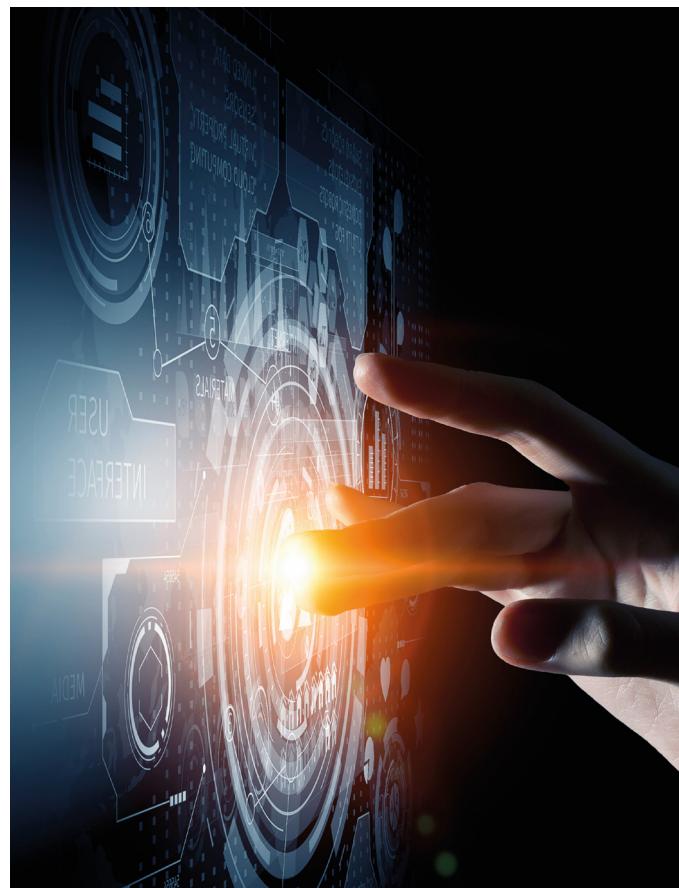
Image: Sergey Nivens/Shutterstock.com

Cognitive sciences have been involved under numerous accounts to explain how humans interact with technology, as well as to design technological instruments tailored to human needs. As technological advancements in fields like wearable and ubiquitous computing, virtual reality, robotics and artificial intelligence are presenting novel modalities for interacting with technology, there are opportunities for deepening, exploring, and even rethinking the theoretical foundations of human technology use.

This volume entitled "Cognition and Interaction: From Computers to Smart Objects and Autonomous Agents" is a collection of articles on the impacts that novel

interactive technologies are producing on individuals. It puts together 17 works, spanning from research on social cognition in human-robot interaction to studies on neural changes triggered by Internet use, that tackle relevant technological and theoretical issues in human-computer interaction, encouraging us to rethink how we conceptualize technology, its use and development.

The volume addresses fundamental issues at different levels. The first part revolves around the biological impacts that technologies are producing on our bodies and brains. The second part focuses on the psychological level, exploring how our psychological characteristics may affect the way we use, understand and perceive technology, as well as how technology is changing our cognition. The third part addresses relevant theoretical problems, presenting reflections that aim to reframe how we conceptualize ourselves, technology and interaction itself. Finally, the last part of the volume pays attention to the factors involved in the design of technological artifacts, providing suggestions on how we can develop novel technologies closer to human needs.

Overall, it appears that human-computer interaction will have to face a variety of challenges to account for the rapid changes we are witnessing in the current technology landscape.

Citation: Rapp, A., Tirassa, M., Ziemke, T., eds. (2019). Cognition and Interaction: From Computers to Smart Objects and Autonomous Agents. Lausanne: Frontiers Media. doi: 10.3389/978-2-88963-002-8

Table of Contents

CHAPTER 4
TECHNOLOGY AND DESIGN

frontiers
in Psychology

Editorial: Cognitive Aspects of Interactive Technology Use: From Computers to Smart Objects and Autonomous Agents

Amon Rapp [1,2]*, Maurizio Tirassa [3] and Tom Ziemke [4,5]

[1] Computer Science Department, University of Turin, Turin, Italy, [2] ICxT - ICT and Innovation for Society and Territory, University of Turin, Turin, Italy, [3] Psychology Department, University of Turin, Turin, Italy, [4] Cognition and Interaction Lab, Department of Computer and Information Science, Linköping University, Linköping, Sweden, [5] Interaction Lab, School of Informatics, University of Skövde, Skövde, Sweden

Editorial on the Research Topic

Cognitive Aspects of Interactive Technology Use: From Computers to Smart Objects and Autonomous Agents

The current advancements in many interactive technology fields, and the consequent spread of digital and intelligent devices in the consumer market, give the opportunity of deepening, exploring, and even rethinking the foundations of human technology use.

The increasing adoption of wearable and self-tracking technologies (Rapp et al., 2015), for instance, is changing how people reflect on themselves (Rapp and Tirassa, 2017), think of their past (Matassa et al., 2013; Elsden et al., 2016), perform physical activity and sport (Rapp and Tirabeni, 2018), and manage their health (Schroeder et al., 2018), encouraging us to explore the impacts of such technologies on mind and behavior. Likewise, people spend more and more time in digital environments, like video games (Rapp, 2017), social media (Lu et al., 2018), and virtual organizations (Reinecke et al., 2013), whereby these virtual and augmented realities are blurring the boundaries between the digital and the material world, potentially affecting how we experience and perceive what we call the reality. The miniaturization of sensors and the rise of the Internet of Things (Atzori et al., 2010) are further making traditional human-computer interfaces disappear (Console et al., 2013), at the same time modifying the affordances that are commonly associated to everyday objects (Rapp and Cena, 2015). Lastly, the increasing ubiquity of different types of interactive robots and autonomous agents, suggests that we investigate in-depth how we humanize artificial entities (Warshaw et al., 2015), how we socially interact with them (Rapp, 2018), and how we understand their behavior (Thellman et al., 2017).

These represent only a few examples of technological changes that are reconfiguring how we interact with "tools," which can inspire a renewed discussion on human-technology interaction. This volume precisely explores how humans create, interact, account for, and are impacted by emerging interactive technologies. It puts together 17 high-quality works, spanning from research on social cognition in human-robot interaction to studies on neural changes triggered by Internet use, also tackling relevant technological and theoretical issues that foster us to rethink how we conceptualize technology, its use, and its development. In other words, this volume addresses relevant issues at different levels, including biological, psychological, theoretical, and technological/design levels.

As for the *biological level*, Liu et al. investigate the neural mechanisms underlying Internet search, discovering that it impacts human brain functions: their study results suggest that Internet search enhances the spatial information processing, but also may alter the memory system, making individuals less engaged in remembering information. Smyk et al. conduct a study with 20 participants investigating how sensorimotor oscillatory electroencephalogram (EEG) activity can be affected by the perceived nature of a task partner, human or robot, during a novel "reciprocal touch" paradigm. The results provide evidence for differences in attentional and tactile processing when interacting with human and robotic partners.

With reference to the *psychological level*, Hou et al. try to understand individual differences in the use of social network sites by surveying 714 Chinese students in order to assess how personality traits relate to excessive use of WeChat and Weibo. They find that neuroticism, loneliness, and external locus of control have positive correlations with excessive use of Weibo and WeChat, while agreeableness, social support, and social interaction negatively correlate to their excessive use. Further, they discover that the two social network sites fulfill different needs and thus attract people with different personality traits. Reichenberger et al. explore the potentiality of virtual environments in conducting social fear conditioning related experiments. They show that social fear can successfully be induced and extinguished using virtual reality, providing insights into learning and unlearning of social fear. Whereas, Kätsyri explores the reasons underlying the sense of eeriness and lack of familiarity that we may experience when we observe virtual characters. Results of an experiment with 64 participants, asking them to learn and recognize a set of virtual, and real faces, seem to suggest that lesser perceptual expertise may contribute to the lack of subjective familiarity with virtual faces.

Böffel and Müsseler investigate an important phenomenon in human-avatar interaction: when spatial dissociation between the user's and the avatar's orientations arises as a consequence of the task handled, the user has to adopt the avatar's perspective and identify herself with it. A study is then set to identify the conditions that benefit this change of perspective: the finding is that perceived ownership, elicited by interaction instructions leading to effector congruency between the participant's hands and the hands of the avatar, benefits perspective taking. Morganti involves 61 participants in order to understand if different embodied affordances could provide different knowledge organization during wayfinding by using distinct spatial simulations. The results show that different embodiments afforded by different environments and the increasing complexity in turn types result in different spatial outcomes.

As for the *theoretical level*, Osiurak et al. attempt to provide a structured way of organizing the literature about the cognitive processes involved in the different interactions we have with tools by proposing a theoretical framework organized into three levels. The first level describes how we interact when using physical technologies which increase our sensorimotor abilities; the second level pertains to sophisticated technologies, for which we do not systematically understand the underlying working principles; the third level tackles symbiotic technologies, which

link our brain directly to machines. In doing so, they highlight the key role of technical and practical reasoning, which could be undermined by the increasing use of sophisticated and symbiotic tools. Likewise, Duus et al. use the extended mind theory to explore how human-tech hybrids, represented by humans interacting with mobile phones, smart watches, and wearable activity trackers, gain and enact collective skills, how agency is expressed and affects the interaction, and what the darker sides are of being a human-tech hybrid. The proposed concept of agency pendulum, which seen agency swinging between the human and the device depending on the situation, reflects the dynamism of agency in these hybrid entities. In his article, Baber retraces the historical roots of the concept of affordance and how it has been applied to interaction design. In reaffirming its fundamental role for understanding "interactivity," Baber further develops the concept by extending it to the interaction with "smart objects," which sense how they are being used, communicate with each other, and provide prompts to solicit certain actions. Here, the human-object-environment system pursues shared intentions and goals, and affordances become both the means by which actions are encouraged, and the manner in which intentions are identified and agreed.

Honig and Oron-Gilad present a literature review of 52 studies that explore when people perceive and resolve robot failures, how robots communicate failure, how failures influence people's perceptions and feelings toward robots, and how such effects can be mitigated. On the basis of this review, they develop a model of information processing for robotic failures describing how individuals perceive, process, and act on failures in human-robot interaction. Musetti and Corsano present an interesting perspective for conceptualizing Internet not as a tool, but as a social environment. As people are part of an information society and can access whatever information they lack whenever they want, no boundaries between their online and offline lives can still be traced clearly. As a consequence of this shift in theorizing, main models of Internet-related pathologies, like Internet addiction, need to be rethought so to avoid pathologizing normal behaviors. In the same vein, Veissière and Stendel, Veissière and Stendel aim to recast current understandings of the mechanisms involved in the addictive use of smartphones in a broader evolutionary focus, suggesting that it is the social expectations and rewards of connecting with other people and seeking to learn from them that yield and sustain addictive behaviors. They thus propose a hypernatural monitoring model of smartphone addiction grounded in a general social rehearsal theory of human cognition.

Finally, as for the *technological/design level*, Triberti et al. focus on the role of emotions in designing interactive technologies, highlighting that designers can not only rely on aesthetic and engagement aspects of interaction, but also on emotions as cognitive processes and active agents of interaction, in order to create innovative and effective devices. van der Kuil et al. evaluate the usability of a serious game addressed to aid patients in the development of compensatory navigation strategies after a brain injury. Results show that mouse controlled interaction in 3D environments is more effective than keyboard interaction, that patients prefer video-based instructions over

text-based instructions, and that feedback timing has no effect on performance and motivation. This may provide useful insight for the design of serious games aiming to transfer skills from virtual environments to real-life situations. Lovato and Piper acknowledge the growing availability of voice interfaces, making it possible for children to ask questions via Internet search even before they have learned to read and write. Drawing on human-computer interaction research, they thus review studies of how children look for information, and of how they perceive and understand the informational and social roles of technology. This review leads to important considerations for the design of future voice-based search interfaces. Lastly, Macedonia et al. emphasize how guided embodiment is an essential feature in intelligent tutoring systems for second language learning and aphasia rehabilitation, as it increases efficiency in the learning process. To enable the system of guiding the user through embodiment, the authors suggest that the system tracks users' gestures and provide corrective feedback, so that sensor technologies are paramount. The authors thus provide an overview of the sensor technologies that can be used to this aim, ranging from camera-based systems to sensing textiles.

These 17 articles give a snapshot of the current perspectives on the foundations of human interaction with tools and technology, proposing opportunities for debating emerging issues about the design and the understanding of novel interactive devices. We hope that readers will find the articles thought-provoking and insightful, encouraging them to move the debate forward.

AUTHOR CONTRIBUTIONS

AR drafted this editorial. All authors provided critical comments and approved it for publication.

REFERENCES

Atzori, L., Iera, A., and Morabito, G. (2010). The internet of things: a survey. *Comp. Netw.* 54, 2787–2805. doi: 10.1016/j.comnet.2010.05.010

Console, L., Antonelli, F., Biamino, G., Carmagnola, F., Cena, F., Chiabrando, E., et al. (2013). Interacting with social networks of intelligent things and people in the world of gastronomy. *ACM Transac. Interact. Intellig. Syst.* 3:38. doi: 10.1145/2448116.2448120

Elsden, C., Kirk, D. S., and Durrant, A. C. (2016). A quantified past: toward design for remembering with personal informatics. *Hum. Comput. Interact.* 31, 518–557. doi: 10.1080/07370024.2015.1093422

Lu, Z., Xia, H., Heo, S., and Wigdor, D. (2018). "You watch, you give, and you engage: a study of live streaming practices in China," in *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18)* (New York, NY: ACM; Paper 466), 13.

Matassa, A., Rapp, A., and Simeoni, R. (2013). "Wearable accessories for cycling: tracking memories in urban spaces," in *Proceedings of the 2013 ACM Conference on Pervasive and Ubiquitous Computing Adjunct Publication (UbiComp '13 Adjunct)* (New York, NY: ACM), 415–424. doi: 10.1145/2494091.2495973

Rapp, A. (2018). Social game elements in World of Warcraft: Interpersonal relations, groups and organizations for gamification design. *Inter. J. Hum. Comput. Interact.* 34, 759–773. doi: 10.1080/10447318.2018.

Rapp, A., and Cena, F. (2015). "Affordances for self-tracking wearable devices," in *Proceedings of International Symposium on Wearable Computers, ISWC 2015* (New York, NY: ACM), 141–142. doi: 10.1145/2802083.2802090

Rapp, A., Cena, F., Kay, J., Kummerfeld, B., Hopfgartner., F., Plumbaum, T., et al. (2015). "New frontiers of Quantified Self: finding new ways for engaging users in collecting and using personal data," in *Adjunct Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2015 ACM International Symposium on Wearable Computers (UbiComp/ISWC'15 Adjunct)* (New York, NY: ACM), 969–972. doi: 10.1145/2800835.2807947

Rapp, A., and Tirabeni, L. (2018). Personal informatics for sport: meaning, body, and social relations in amateur and elite athletes. *ACM Transac. Comput. Hum. Interact.* 25, 1–30. doi: 10.1145/3196829

Rapp, A., and Tirassa, M. (2017). Know Thyself: A theory of the self for personal informatics. *Hum. Comput. Interact.* 32, 335–380. doi: 10.1080/07370024.2017.1285704

Rapp., A. (2017). Drawing inspiration from world of warcraft: gamification design elements for behavior change technologies. *Interact. Comput.* 29, 648–678. doi: 10.1093/iwc/iwx001

Reinecke, K., Nguyen, M. K., Bernstein, A., Näf, M., and Gajos, K. Z. (2013). "Doodle around the world: online scheduling behavior reflects cultural differences in time perception and group decision-making," in *Proceedings of the 2013 Conference on Computer Supported Cooperative Work (CSCW'13)*, February 23–27 (San Antonio, TX), 45–54.

Schroeder, J., Chung, C. F., Epstein, D. A., Karkar, R., Parsons, A., Murinova, N., et al. (2018). "Examining self-tracking by people with migraine: goals, needs, and opportunities in a chronic health condition," in *Proceedings of the 2018 Designing Interactive Systems Conference (DIS '18)* (New York, NY: ACM), 135–148.

Thellman, S., Silvervarg, A., and Ziemke, T. (2017). Folk-psychological interpretation of human vs. humanoid robot behavior: exploring the intentional stance toward robots. *Front. Psychol.* 14:1962. doi: 10.3389/fpsyg.2017.01962

Warshaw, J., Matthews, T., Whittaker, S., Kau, C., Bengualid, M., and Smith, B. A. (2015). "Can an algorithm know the "Real You"?: understanding people's reactions to hyper-personal analytics systems," in *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (CHI '15)* (New York, NY:ACM), 797–806. doi: 10.1145/2702123.2702274

Check for
updates

Internet Search Alters Intra- and Inter-regional Synchronization in the Temporal Gyrus

Xiaoyue Liu[1], Xiao Lin[2], Ming Zheng[1], Yanbo Hu[3], Yifan Wang[1], Lingxiao Wang[1], Xiaoxia Du[4] and Guangheng Dong[1,5]*

[1] Department of Psychology, Zhejiang Normal University, Jinhua, China, [2] Center for Life Sciences, Peking University, Beijing, China, [3] Department of Psychology, London Metropolitan University, London, United Kingdom, [4] Department of Physics, Shanghai Key Laboratory of Magnetic Resonance, East China Normal University, Shanghai, China, [5] Institute of Psychological and Brain Sciences, Zhejiang Normal University, Jinhua, China

Internet search changed the way we store and recall information and possibly altered our brain functions. Previous studies suggested that Internet search facilitates the information-acquisition process. However, this process may cause individuals to lose the ability to store and recollect specific contents. Despite the numerous behavioral studies conducted in this field, little is known about the neural mechanisms underlying Internet searches. The present study explores potential brain activity changes induced by Internet search. The whole paradigm includes three phases, namely, pre-resting state fMRI (rs-fMRI) scan, 6-day Internet search training, and post rs-fMRI scan. We detected the functional integrations induced by Internet search training by comparing post- with pre-scan. Regional homogeneity (ReHo) and functional connectivity (FC) were used to detect intra- and interregional synchronized activity in 42 university students. Compared with pre-scan, post-scan showed decreased ReHo in the temporal gyrus, the middle frontal gyrus, and the postcentral gyrus. Further seed-based FC analysis showed that the temporal gyrus exhibited decreased FC in the parahippocampal cortex and the temporal gyrus after training. Based on the features of current task and functions exhibited by these brain regions, results indicate that short-term Internet search training changed the brain regional activities involved in memory retrieval. In general, this study provides evidence that supports the idea that Internet search can affect our brain functions.

Keywords: internet-search, short-term training, regional homogeneity, functional connectivity, long-term memory

INTRODUCTION

Finding information through Internet search engines has become a common daily activity for people (Small et al., 2009). The widespread use of the Internet changed the way we find and store information. "Google effect" indicates that when people use the Internet as an external storage, they need to remember "where" it is instead of the information itself (Sparrow et al., 2011; Ward, 2013). These studies suggested that Internet search reduced the need for effort to process and remember information (Carr, 2010).

Recent studies explored the influence of Internet search on the brain. Nicholas suggested that the younger "Google generation" spends less time on individual questions and searches quicker, but the

members of this generation show poorer working memory and are less confident about the answers they provided compared with the older generation (Nicholas, 2013). Small et al. (2009) showed that prior experience of using Internet search increased brain responsiveness in neural circuits involved in decision-making and complex reasoning in aged adults after short-term training. The brain activities of experienced Google users were broader than that of novices during searches (Small et al., 2009). A previous study found that people who obtained information through Internet-based search memory task would show lower accuracy when recalling information and are highly impulsive in novel trials (Dong and Potenza, 2015). However, 6 days of practicing Internet search improved their efficiency, but it reduced their dependency on their long-term memory (Dong and Potenza, 2016; Dong et al., 2017).

Brain and behavior are a dynamic system that influences each other and form the basis of brain plasticity (Mechelli et al., 2004). A large number of studies proved that learning induced brain plasticity (Kolb and Whishaw, 2003; Vartanian et al., 2013). In addition, various tasks generated differentiated functional response patterns through modulation of training (Koeneke et al., 2004; Kelly and Garavan, 2005; Erickson et al., 2007; Thomas et al., 2009). Previous studies revealed that 3 months of juggling training could lead to a transient bilateral expansion in gray matter in the mid-temporal area and in the left posterior intraparietal sulcus (Draganski et al., 2004). Five hours of meditation training can change the activity of default mode network and connectivity of the white matter (Brewer et al., 2011; Ding et al., 2015). The neural circuitry of the frontal pole, anterior temporal cortex, and anterior and posterior cingulate was activated in Internet-savvy subjects after 5 days of practice (Small et al., 2009).

Researchers recently realized that the effect of training/expertise-specific on brain functions may extend beyond task state to resting state. Resting state refers to the state when subjects relax, stationary, eyes closed, and avoid any systematic thinking (Mazoyer et al., 2001). Resting-state fMRI was also used to study intrinsic functional connectivity (FC) and has been widely used as a tool to assess large-scale networks in the human brain in both clinical and healthy populations (Wink et al., 2006; Tian et al., 2007; Buckner et al., 2013).

Regional homogeneity (ReHo) and FC are often used to evaluate brain activity synchronization in resting-state of healthy subjects and patients. ReHo is a rank-based non-parametric data-driven approach that reflects the temporal homogeneity of the regional BOLD signal (Sepulcre et al., 2010). ReHo measures the functional coherence of a given voxel with its nearest neighbors by calculating Kendall's coefficient concordance (KCC) (Zang et al., 2004; Zuo et al., 2012). The test–retest reliability of ReHo has been found very high despite the physiological noise and preprocessing effect (Zuo et al., 2012; Zuo and Xing, 2014). FC measures the similarity of the time series of two relatively remote brain regions (Biswal et al., 1995). These two measures are often used together to

detect local and remote brain activity synchronizations. Hence, combining the ReHo and FC analysis could provide additional information about brain activity synchronization induced by Internet search.

Numerous studies revealed that short-term training could alter the resting-state features of our brain. An existing study found resting-state coherence in the right medial motor cortex was increased by brief sensory motor intervention (Verrel et al., 2015). A study on acupuncturists found that training/expertise could modulate resting-state activity by increasing regional clustering strength (Dong et al., 2014). In a short-term simulated microgravity study, 72 h of −6° head down tilt (HDT) resulted in decreased ReHo in the right inferior frontal gyrus (Liao et al., 2013). Modulation of resting-state functional connectivity (rs-FC) in the parietal circuit was found after 4 weeks of daily training of an explicit sequence learning task (Ma et al., 2011). A study found that 4 weeks of working memory training increased rs-FC between the medial prefrontal cortex (mPFC) and the precuneus, but decreased rs-FC between the mPFC and the right posterior parietal cortex (Takeuchi et al., 2013). These studies suggest that the resting-state features of the brain could be altered by short-term training.

Given these findings, we speculated that the brain activity in resting-state would be affected by search engines usage. In the present study, we first explore abnormal brain activity using ReHo analysis and investigate the FC between regions with altered ReHo and other brain regions. Previous studies found that people using Internet search as a tool for remembering new information showed lower brain activations in the middle temporal gyrus (Dong and Potenza, 2015) and regions along the ventral stream (Knutson et al., 2012). In addition, people using Internet search tools just need to remember where the information is stored instead of the information itself (Sparrow et al., 2011). Hence, we hypothesized that short-term training would make people to be better at remembering where information is stored (higher brain activities in regions along the dorsal stream) than the specific content (lower brain activities in regions along the ventral stream). We compared data on resting brain states between pre- and post-scan from 42 college student volunteers to examine the changes of brain activity induced by Internet search by measuring ReHo and FC.

METHOD AND PROCEDURE

Participants

The experiment complied with the Code of Ethics of the World Medical Association (Declaration of Helsinki). The Human Investigations Committee of Zhejiang Normal University approved this study. This study was conducted in accordance with the approved guidelines. Forty-two university students were recruited through advertisements (22 males; 20 females; age: 21.4 ± 1.2 years). All participants provided written informed consent and underwent structured psychiatric interviews using the Mini-International Neuropsychiatric Interview

(MINI) (Lecrubier et al., 1997) performed by an experienced psychiatrist. According to the MINI assessment, they were free of psychiatric disorders, including major depression, anxiety disorders, schizophrenia, and substance dependence disorders. All participants were medication-free and were instructed not to use any substances, including coffee, on the day of scanning. To obtain information regarding their Internet search behaviors, all subjects were assessed using an Internet-search-use questionnaire (Supplementary Material) (Dong and Potenza, 2015; Wang et al., 2016). Results showed that all participants were familiar with Internet search and used it regularly. We included Internet search experience as a covariate to exclude the impact on experimental results (Wang et al., 2016).

Experiment Procedure

The whole experiment consists of three steps, namely, pre-scan, 6 days of training, and post-scan.

Subjects were "trained" for at least 1 h per day for six consecutive days. In the experiment, subjects were asked to finish one of six search tasks randomly without repetition. Each search task consisted of 80 fill-in-the-blank items that required subjects to seek answers using an Internet search engine. Participants were informed that they will receive 20 Chinese yuan for their everyday participation. To elicit motivation in searching, the subjects were advised that the daily reward would be paid based on their real performance [20 * accuracy rates (%)]. This reward premise was approved by the ethics committee. Participants who took the work seriously and finished with an accuracy rate of over 80% passed the training.

Image Acquisition

Functional MRI was performed on a 3.0 Tesla Siemens Trio scanner. The functional scan was acquired using gradient echo planar imaging sequence with the following parameters: [repeat time (TR) = 2 s, echo time (TE) = 30 ms, flip angle = 90°; interleaved sequence; 33 slice per volume; 3 mm thickness; field of view (FOV) = 220 mm × 220 mm^2, matrix 64 cm × 64 cm, acquisition matrix = 64 × 64]. Each functional run included 210 imaging volumes for each participant and the scan lasted 7 min. All subjects were instructed to rest quietly in the scanner without falling asleep. Post-scan conforms to the same standard and parameter.

Data Pre-processing

Data pre-processing was conducted using Data Processing Assistant for Resting-State fMRI (DPARSFA[1]), a MATLAB toolbox for "pipeline" data analysis of resting-state fMRI (Yan and Zang, 2010; Song et al., 2011). DPARSFA is based on some functions in Statistical Parametric Mapping[2] and Resting-State (REST[3]). The main preprocessing steps and parameters are listed as follows. The first 10 volumes of each functional time series were abandoned to avoid the instability of the initial fMRI signal, thereby leaving 200 volumes.

[1]www.restfmri.net
[2]http://www.fil.ion.ucl.ac.uk/spm
[3]http://www.restfmri.net/forum/

Preprocessing included slice timing, head motion correction, and spatial normalization to a standard template. Participants with maximum translation that exceeds 2.5 mm or maximum rotation that exceeds 2.5° were excluded from further analysis. To reduce the effects of confounding factors, a regression of nuisance signals including cerebral spinal fluid, white matter, six motion vectors was performed. Following regression, detrending was performed and temporal filtering (0.01 to 0.08 Hz) was applied to the time series of each voxel to reduce low-frequency drift and high-frequency noise. Data for one subject were excluded according to the head-motion parameter (2.5 mm; 2.5°).

ReHo Analysis

Resting-state fMRI data without spatial smoothing were used for ReHo analysis with DPARSFA. Individual ReHo maps were generated by calculating the KCC of the time series of a given voxel with its nearest neighbors (26 voxels) in all directions on a voxel-wise basis. The calculated formula of ReHo is defined as follows:

$$W = \frac{\sum (R_i)^2 - n(\bar{R})^2}{\frac{1}{12} K^2 (n^3 - n)},$$

where W is the KCC for a given voxel that ranged from 0 to 1. R_i is sum rank of the ith time point and n is the number of ranks; $\bar{R} = [(n++1) K]/2$ is the mean of the R_i; K is the voxel number among time series (27 voxels, one given voxel plus the number of its neighbors). To reduce the influence of individual variations in the KCC value, each standardized ReHo map was generated by dividing the raw ReHo map by the global mean ReHo. Spatial smoothing was conducted on the ReHo maps with a Gaussian kernel of 4 mm × 4 mm × 4mm full-width at half-maximum (Liao et al., 2013; Chen et al., 2015).

FC Analysis

A seed-based correlation approach was used for FC analysis. The seed was defined in the regions' existing difference and analyzed by ReHo. We calculated the temporal correlation between these seed regions and every other voxel within the brain. These procedures were executed using DPARSFA software.

Post–Pre Analysis

To explore the differences between the pre- and post-training, a pair-sample t-test was performed on the normalized ReHo and FC maps with REST software. The result and statistical map were set at a combined threshold of $p < 0.05$ (AlphaSim corrected) and a minimum cluster size of 110 voxels.

RESULTS

Regional Homogeneity (ReHo)

Compared with the pre-scan data, the neuroimaging data from post-scan were associated with the decreased ReHo values in the temporal gyrus, which include the

TABLE 1 | Brain areas showing Regional homogeneity (ReHo) difference between post- and pre-training.

Areas	Hemisphere	Voxels	Peak t-value	Coordinates
Decreased ReHo after training				
Superior temporal gyrus	R	147	−5.3319	54–24 6
Middle temporal gyrus	L	157	−4.7822	−48–72 18
Middle frontal gyrus	L	345	−4.7366	−33–6 63
Postcentral gyrus	L	211	−4.4208	−60–12 0

Voxel size = 3 × mm3 mm × 3 mm, p < 0.05 AlphaSim corrected and at least 110 voxels. L, left; R, right; t: t-values from a paired two tailed t-test of the statistical different clusters.

superior temporal gyrus and the middle temporal gyrus, the middle frontal gyrus and the postcentral gyrus. **Table 1** shows the detailed information for the brain regions with ReHo difference after training. **Figure 1** shows the brain areas.

Seed-Based FC of Altered ReHo Regions
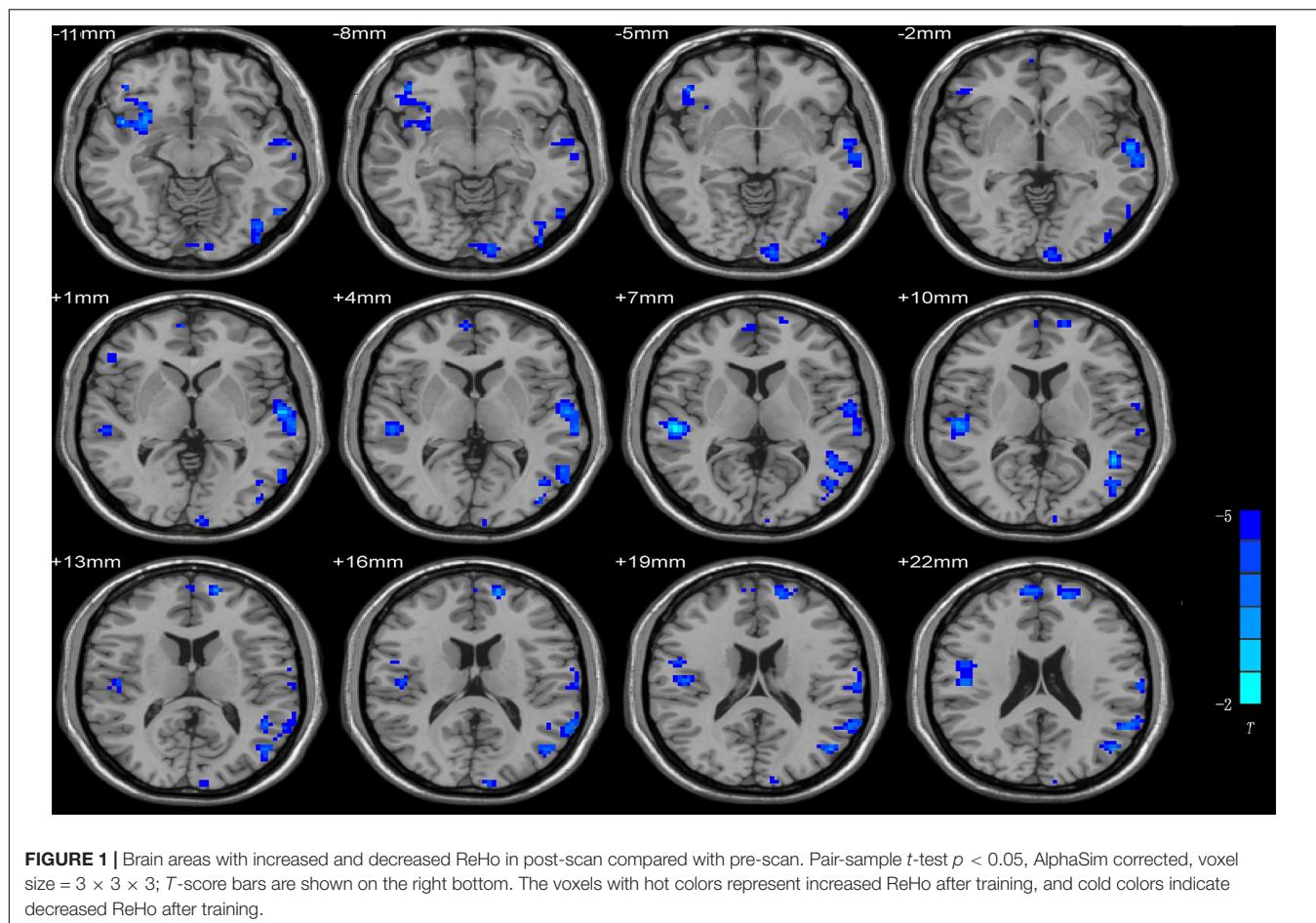
Based on the ReHo results, the temporal gyrus and the middle frontal gyrus were selected as seed regions of interest for FC analysis. The temporal gyrus exhibited decreased FC with the parahippocampal cortex and the temporal gyrus after training whereas the temporal gyrus showed increased FC with the parietal gyrus. The middle frontal gyrus exhibited decreased FC with the middle temporal gyrus and the middle frontal gyrus after training (**Table 2** and **Figure 2**).

DISCUSSION

This study finds decreased ReHo in the temporal gyrus, the middle frontal gyrus and the postcentral after training. The temporal gyrus exhibited decreased FC with the parahippocampal cortex and the temporal gyrus after training, whereas the temporal gyrus showed increased FC with the parietal gyrus. The middle frontal gyrus exhibited decreased FC with the middle temporal gyrus and the middle frontal gyrus after training. This evidence of local functional homogeneity and interregional FC contributes to our understanding of the changed brain activity produced by Internet search.

The middle temporal gyrus located at the end of the ventral stream, which was described as the "what" stream, is involved in object identification and recognition (Mishkin and

FIGURE 1 | Brain areas with increased and decreased ReHo in post-scan compared with pre-scan. Pair-sample t-test p < 0.05, AlphaSim corrected, voxel size = 3 × 3 × 3; T-score bars are shown on the right bottom. The voxels with hot colors represent increased ReHo after training, and cold colors indicate decreased ReHo after training.

Ungerleider, 1982; Goodale and Milner, 1992). The Decreased ReHo and FC in the temporal gyrus supports our hypothesis that short-term training decreases brain activities in regions along the ventral stream. The parahippocampal cortex is a part of the limbic system, which plays an important role in memory encoding and retrieval (Ekstrom and Bookheimer, 2007). Evidence from patient studies suggested that the middle temporal gyrus is associated with long-term memory (Squire and Zolamorgan, 1991; Meulemans and Van der Linden, 2003). Functional neuroimaging study also found that the middle temporal gyrus is activated when individuals participate in memory encoding and retrieval processes of memory (Onoda et al., 2009). Previous studies found that Internet search decreases brain activation in the temporal gyrus (Dong and Potenza, 2016). Decreased ReHo and FC in the temporal gyrus might suggest that Internet search causes people to rely less on their long-term memory.

Results showed that short-term training increased FC between the temporal gyrus and the parietal lobe. The posterior parietal cortex is referred to as the dorsal stream or "where" stream, which is involved in visuospatial processing (Mishkin and Ungerleider, 1982). Therefore, increased FC might suggest that Internet search enhances the spatial information processing (Dong et al., 2017). The result is consistent with our hypothesis. The middle frontal gyrus plays an important role in selective attention, executive control, and working memory (Fuster, 2002).

TABLE 2 | Pre–post differences in seed-based FC in altered ReHo regions.

Areas	Hemisphere	Voxels	Peak t-value	Coordinates
Post-training – pre-training				
FC of R Temporal gyrus				
Parietal gyrus	R	110	3.2562	36–57 45
Temporal gyrus	R	406	−4.2347	39 6–21
Parahippocampal cortex	R	446	−3.8151	24–45 0
Superior frontal gyrus	R	548	−4.9860	9–18 78
FC of L Middle frontal gyrus				
Superior temporal gyrus	L	374	−4.0656	−33 9–18
Middle temporal gyrus	R	417	−3.8514	42–57 60
Middle frontal gyrus	R	147	−3.4775	45 6 51

Voxel size = 3 mm × 3 mm × 3 mm, p < 0.05 AlphaSim corrected and at least 110 voxels. L, left; R, right; FC, functional connectivity; t: t-values from a paired two tailed t-test of the statistical different clusters.

Decreased ReHo and FC in the middle frontal gyrus may suggest that individuals engage less in remembering something when faced with information that can be found in the Internet. According to Sparrow, people who used Internet search engines to access information show worse recall rates of information (Sparrow et al., 2011). Our results indicated that short-term Internet search training could alter the memory system, which further provided evidence for the hypothesis that Internet search
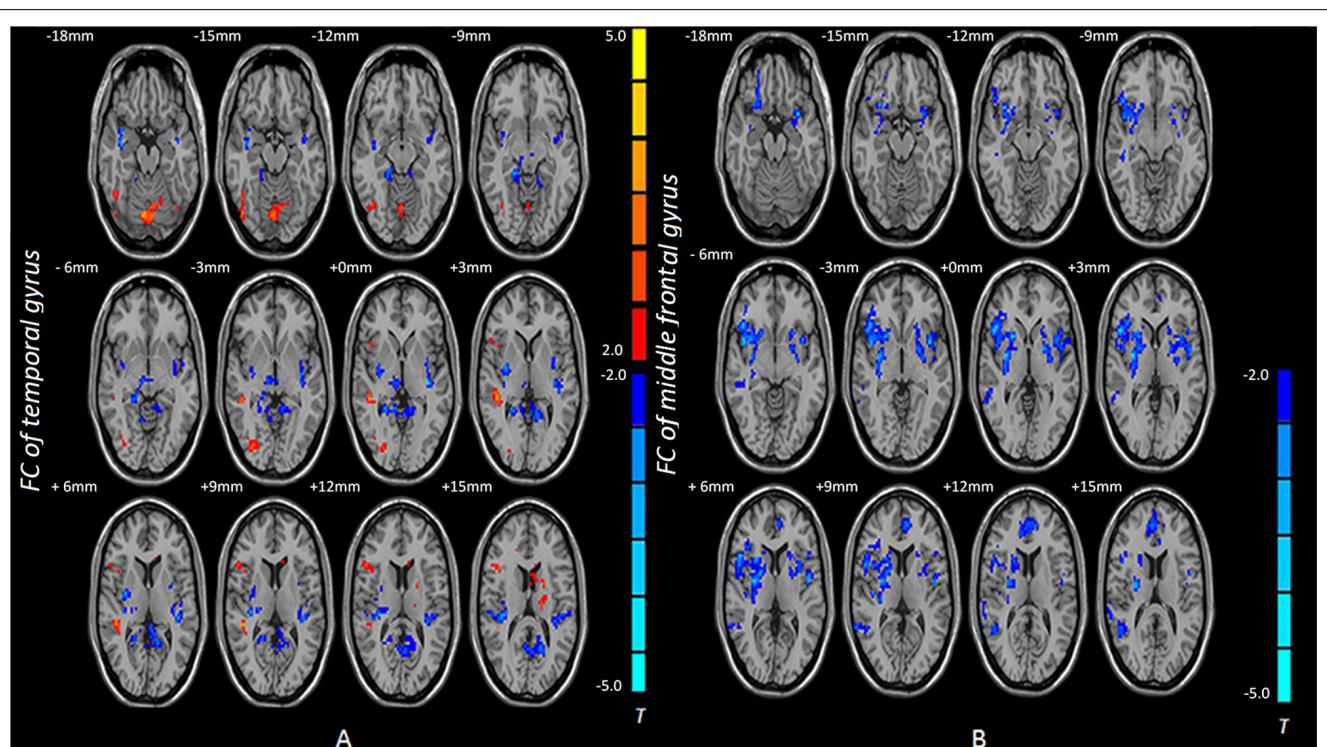
FIGURE 2 | Brain areas with increased and decreased FC in post-scan compared with pre-scan. **(A)** The voxels with hot colors represent increased FC between the temporal gyrus and the parietal lobe after training, and cold colors indicate that the temporal gyrus shows decreased FC with the parahippocampal cortex and the temporal gyrus after training. **(B)** The middle frontal gyrus exhibited decreased FC with the middle temporal gyrus and the middle frontal gyrus after training. Pair-sample t-test p < 0.05, AlphaSim corrected; FC, functional connectivity.

enables people to be better at remembering where information is stored than the specific content.

As the Internet evolved into a useful tool in our daily life, people become susceptible to the unprecedented Internet search environment (Zickuhr and Madden, 2012; Loh and Kanai, 2016). Some researchers found that many people are increasingly relying on Internet search (Zickuhr and Madden, 2012; Yang et al., 2014; Wang Y. et al., 2017). People show irritability and depression when they cannot immediately find what they want (Block, 2008). These symptoms are similar with withdrawal symptoms of pathological Internet use (Davis, 2001). The executive function of Internet addiction subjects is impaired, whereas the inhibition control of Internet-using behaviors of those people is weakened (Wang L. et al., 2017). Our study found that the brain activities of Internet search users behaved differently. These people might engage in less effort in remembering something after short-term training.

CONCLUSION

This study found that the aberrant ReHo and FC were mainly distributed in the temporal gyrus and the middle frontal gyrus, which are responsible for long-term memory. Results provide evidence that support the idea that Internet search can affect our brain functions.

Limitations

Several limitations should be addressed. First, given financial and time constraints, we canceled the control group and pre–post resting-state data were not collected from the control group. We paid more attention to pre–post difference in one group of subjects. Second, we found it impossible to recruit university students that have no experience in Internet search. Thus, all subjects were familiar with Internet search. This sample group might affect training effect. Given the limitation of rs-fMRI studies, the explanations of the results rely on the brain functions of relevant brain regions, which lack direct support from behavioral or task data. Thus, the explanations are based on reasoning and the interpretation of the results is uncertain. However, the findings provide evidence that Internet search can affect our brain functions.

AUTHOR CONTRIBUTIONS

XLiu and XLin analyzed the data and wrote the first draft of the manuscript. LW and YW contributed to experimental programming and data collection. XD contributed to fMRI data collection. GD and YH designed this research. GD and MZ revised and improved the manuscript. All authors contributed to and had approved the final manuscript.

FUNDING

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2018.00260/full#supplementary-material

REFERENCES

Biswal, B., Yetkin, F. Z., Haughton, V. M., and Hyde, J. S. (1995). Functional connectivity in the motor cortex of resting human brain using echo-planar MRI. *Magn. Reson. Med.* 34, 537–541. doi: 10.1002/mrm.1910340409

Block, J. J. (2008). Issues for DSM-V: internet addiction. *Am. J. Psychiatry* 165, 306–307. doi: 10.1176/appi.ajp.2007.07101556

Brewer, J. A., Worhunsky, P. D., Gray, J. R., Tang, Y. Y., Weber, J., and Kober, H. (2011). Meditation experience is associated with differences in default mode network activity and connectivity. *Proc. Natl. Acad. Sci. U.S.A.* 108, 20254–20259. doi: 10.1073/pnas.1112029108

Buckner, R. L., Krienen, F. M., and Yeo, B. T. T. (2013). Opportunities and limitations of intrinsic functional connectivity MRI. *Nat. Neurosci.* 16, 832–837. doi: 10.1038/nn.3423

Carr, N. (2010). *The Shallows: What the Internet Is Doing to Our Brains*. New York, NY: W. W. Norton & Company.

Chen, C., Xiu, D., Chen, C., Moyzis, R., Xia, M., He, Y., et al. (2015). Regional homogeneity of resting-state brain activity suppresses the effect of dopamine-related genes on sensory processing sensitivity. *PLoS One* 10:e0133143. doi: 10.1371/journal.pone.0133143

Davis, R. A. (2001). A cognitive-behavioral model of pathological Internet use. *Comput. Hum. Behav.* 17, 187–195. doi: 10.1016/S0747-5632(00)00041-8

Ding, X., Tang, Y. Y., Cao, C., Deng, Y., Wang, Y., Xin, X., et al. (2015). Short-term meditation modulates brain activity of insight evoked with solution cue. *Soc. Cogn. Affect. Neurosci.* 10, 43–49. doi: 10.1093/scan/nsu032

Dong, G., Li, H., and Potenza, M. N. (2017). Short-term internet-search training is associated with increased fractional anisotropy in the superior longitudinal fasciculus in the parietal lobe. *Front. Neurosci.* 11:372. doi: 10.3389/fnins.2017.00372

Dong, G., and Potenza, M. N. (2015). Behavioural and brain responses related to Internet search and memory. *Eur. J. Neurosci.* 42, 2546–2554. doi: 10.1111/ejn.13039

Dong, G., and Potenza, M. N. (2016). Short-term internet-search practicing modulates brain activity during recollection. *Neuroscience* 335, 82–90. doi: 10.1016/j.neuroscience.2016.08.028

Dong, M., Wei, Q., Zhao, L., Yang, X., Yuan, K., Zeng, F., et al. (2014). Expertise modulates local regional homogeneity of spontaneous brain activity in the resting brain: an fMRI study using the model of skilled acupuncturists. *Hum. Brain Mapp.* 35, 1074–1084. doi: 10.1002/hbm.22235

Draganski, B., Gaser, C., Busch, V., Schuierer, G., Bogdahn, U., and May, A. (2004). Neuroplasticity: changes in grey matter induced by training. *Nature* 427, 311–312. doi: 10.1038/427311a

Ekstrom, A. D., and Bookheimer, S. Y. (2007). Spatial and temporal episodic memory retrieval recruit dissociable functional networks in the human brain. *Learn. Mem.* 14, 645–654. doi: 10.1101/lm.575107

Erickson, K. I., Colcombe, S. J., Wadhwa, R., Bherer, L., Peterson, M. S., Scalf, P. E., et al. (2007). Training-induced functional activation changes in dual-task processing: an fMRI study. *Cereb. Cortex* 17, 192–204. doi: 10.1093/cercor/bhj137

Fuster, J. M. (2002). "Physiology of executive functions: the perception-action cycle," in *Principles of Frontal Lobe Function*, eds D. T. Stuss and R. T. Knight (New York, NY: Oxford University Press), 96–109. doi: 10.1093/acprof:oso/9780195134971.003.0006

Goodale, M. A., and Milner, A. D. (1992). Separate visual pathways for perception and action. *Trends Neurosci.* 15, 20–25. doi: 10.1016/0166-2236(92)90344-8

Kelly, A. M. C., and Garavan, H. (2005). Human functional neuroimaging of brain changes associated with practice. *Cereb. Cortex* 15, 1089–1102. doi: 10.1093/cercor/bhi005

Knutson, A. R., Hopkins, R. O., and Squire, L. R. (2012). Visual discrimination performance, memory, and medial temporal lobe function. *Proc. Natl. Acad. Sci. U.S.A.* 109, 13106–13111. doi: 10.1073/pnas.1208876109

Koeneke, S., Lutz, K., Wüstenberg, T., and Jäncke, L. (2004). Long-term training affects cerebellar processing in skilled keyboard players. *Neuroreport* 15, 1279–1282. doi: 10.1097/01.wnr.0000127463.10147.e7

Kolb, B., and Whishaw, I. Q. (2003). Brain plasticity and behavior. *Curr. Dir. Psychol. Sci.* 49, 1–5. doi: 10.1111/1467-8721.01210

Lecrubier, Y., Sheehan, D. V., Weiller, E., Amorim, P., Bonora, I., Sheehan, K. H., et al. (1997). The mini international neuropsychiatric interview (MINI). A short diagnostic structured interview: reliability and validity according to the CIDI. *Eur. Psychiatry* 12, 224–231. doi: 10.1016/S0924-9338(97)83296-8

Liao, Y., Miao, D., Huan, Y., Yin, H., Xi, Y., and Liu, X. (2013). Altered regional homogeneity with short-term simulated microgravity and its relationship with changed performance in mental transformation. *PLoS One* 8:e64931. doi: 10.1371/journal.pone.0064931

Loh, K. K., and Kanai, R. (2016). How has the internet reshaped human cognition? *Neuroscientist* 22, 506–520. doi: 10.1177/1073858415595005

Ma, L., Narayana, S., Robin, D. A., Fox, P. T., and Xiong, J. (2011). Changes occur in resting state network of motor system during 4 weeks of motor skill learning. *Neuroimage* 58, 226–233. doi: 10.1016/j.neuroimage.2011.06.014

Mazoyer, B., Zago, L., Mellet, E., Bricogne, S., Etard, O., Houdé, O., et al. (2001). Cortical networks for working memory and executive functions sustain the conscious resting state in man. *Brain Res. Bull.* 54, 287–298. doi: 10.1016/S0361-9230(00)00437-8

Mechelli, A., Crinion, J. T., Noppeney, U., O'Doherty, J., Ashburner, J., Frackowiak, R. S., et al. (2004). Neurolinguistics: structural plasticity in the bilingual brain. *Nature* 431:757. doi: 10.1038/431757a

Meulemans, T., and Van der Linden, M. (2003). Implicit learning of complex information in amnesia. *Brain Cogn.* 52, 250–257. doi: 10.1016/S0278-2626(03)00081-2

Mishkin, M., and Ungerleider, L. G. (1982). Contribution of striate inputs to the visuospatial functions of parieto-preoccipital cortex in monkeys. *Behav. Brain Res.* 6, 57–77. doi: 10.1016/0166-4328(82)90081-X

Nicholas, D. (2013). Google Generation II: web behaviour experiments with the BBC. *Aslib Proc.* 63, 28–45. doi: 10.1108/00012531111103768

Onoda, K., Okamoto, Y., and Yamawaki, S. (2009). Neural correlates of associative memory: the effects of negative emotion. *Neurosci. Res.* 64, 50–55. doi: 10.1016/j.neures.2009.01.008

Sepulcre, J., Liu, H., Talukdar, T., Martincorena, I., Yeo, B. T., and Buckner, R. L. (2010). The organization of local and distant functional connectivity in the human brain. *PLoS Comput. Biol.* 6:e1000808. doi: 10.1371/journal.pcbi.1000808

Small, G. W., Moody, T. D., Siddarth, P., and Bookheimer, S. Y. (2009). Your brain on Google: patterns of cerebral activation during internet searching. *Am. J. Geriatr. Psychiatry* 17, 116–126. doi: 10.1097/JGP.0b013e3181953a02

Song, X. W., Dong, Z. Y., Long, X. Y., Li, S. F., Zuo, X. N., Zhu, C. Z., et al. (2011). REST: a toolkit for resting-state functional magnetic resonance imaging data processing. *PLoS One* 6:e25031. doi: 10.1371/journal.pone.0025031

Sparrow, B., Liu, J., and Wegner, D. M. (2011). Google effects on memory: cognitive consequences of having information at our fingertips. *Science* 333, 776–778. doi: 10.1126/science.1207745

Squire, L. R., and Zolamorgan, S. (1991). The medial temporal lobe memory system. *Science* 253, 1380–1386. doi: 10.1126/science.1896849

Takeuchi, H., Taki, Y., Rui, N., Hashizume, H., Sekiguchi, A., Kotozaki, Y., et al. (2013). Effects of working memory training on functional connectivity and cerebral blood flow during rest. *Cortex* 49, 2106–2125. doi: 10.1016/j.cortex.2012.09.007

Thomas, A. G., Marrett, S., Saad, Z. S., Ruff, D. A., Martin, A., and Bandettini, P. A. (2009). Functional but not structural changes associated with learning: an exploration of longitudinal Voxel-Based Morphometry (VBM). *Neuroimage* 48, 117–125. doi: 10.1016/j.neuroimage.2009.05.097

Tian, L., Jiang, T., Liu, Y., Yu, C., Wang, K., Zhou, Y., et al. (2007). The relationship within and between the extrinsic and intrinsic systems indicated by resting state correlational patterns of sensory cortices. *Neuroimage* 36, 684–690. doi: 10.1016/j.neuroimage.2007.03.044

Vartanian, O., Jobidon, M. E., Bouak, F., Nakashima, A., Smith, I., Lam, Q., et al. (2013). Working memory training is associated with lower prefrontal cortex activation in a divergent thinking task. *Neuroscience* 236, 186–194. doi: 10.1016/j.neuroscience.2012.12.060

Verrel, J., Almagor, E., Schumann, F., Lindenberger, U., and Kühn, S. (2015). Changes in neural resting state activity in primary and higher-order motor areas induced by a short sensorimotor intervention based on the Feldenkrais method. *Front. Hum. Neurosci.* 9:232. doi: 10.3389/fnhum.2015.00232

Wang, L., Shen, H., Lei, Y., Zeng, L. L., Cao, F., Su, L., et al. (2017). Altered default mode, fronto-parietal and salience networks in adolescents with Internet addiction. *Addict. Behav.* 70, 1–6. doi: 10.1016/j.addbeh.2017.01.021

Wang, Y., Wu, L., Luo, L., Zhang, Y., and Dong, G. (2017). Short-term Internet search using makes people rely on search engines when facing unknown issues. *PLoS One* 12:e0176325. doi: 10.1371/journal.pone.0176325

Wang, Y., Wu, L., Zhou, H., Xu, J., and Dong, G. (2016). Development and validation of a self-reported questionnaire for measuring internet search dependence. *Front. Public Health* 4:274. doi: 10.3389/fpubh.2016.00274

Ward, A. F. (2013). Supernormal: how the internet is changing our memories and our minds. *Psychol. Inq.* 24, 341–348. doi: 10.1080/1047840X.2013.850148

Wink, A. M., Bernard, F., Salvador, R., Bullmore, E., and Suckling, J. (2006). Age and cholinergic effects on hemodynamics and functional coherence of human hippocampus. *Neurobiol. Aging* 27, 1395–1404. doi: 10.1016/j.neurobiolaging.2005.08.011

Yan, C., and Zang, Y. (2010). DPARSF: a MATLAB toolbox for "pipeline" data analysis of resting-state fMRI. *Front. Syst. Neurosci.* 4:13. doi: 10.3389/fnsys.2010.00013

Yang, H., Chen, Y., Zheng, L., Xu, X., and Cao, X. (2014). Analysis of internet use behaviors among clinical medical students in China. *BMC Med Educ.* 14:67. doi: 10.1186/1472-6920-14-67

Zang, Y., Jiang, T., Lu, Y., He, Y., and Tian, L. (2004). Regional homogeneity approach to fMRI data analysis. *Neuroimage* 22, 394–400. doi: 10.1016/j.neuroimage.2003.12.030

Zickuhr, K., and Madden, M. (2012). *Older Adults and Internet Use*. Pew Internet & American Life Project. Available at: http://www.pewinternet.org/?/media/files/Reports/2012/PIP-Digital-differences-041312

Zuo, X. N., and Xing, X. X. (2014). Test-retest reliabilities of resting-state FMRI measurements in human brain functional connectomics: a systems neuroscience perspective. *Neurosci. Biobehav. Rev.* 45, 100–118. doi: 10.1016/j.neubiorev.2014.05.009

Zuo, X. N., Xu, T., Jiang, L., Yang, Z., Cao, X. Y., He, Y., et al. (2012). Toward reliable characterization of functional homogeneity in the human brain: preprocessing, scan duration, imaging resolution and computational space. *Neuroimage* 65, 374–386. doi: 10.1016/j.neuroimage.2012.10.017

Sensorimotor Oscillations During a Reciprocal Touch Paradigm With a Human or Robot Partner

Nathan J. Smyk*, Staci Meredith Weiss and Peter J. Marshall

Department of Psychology, Temple University, Philadelphia, PA, United States

Robots provide an opportunity to extend research on the cognitive, perceptual, and neural processes involved in social interaction. This study examined how sensorimotor oscillatory electroencephalogram (EEG) activity can be influenced by the perceived nature of a task partner – human or robot – during a novel "reciprocal touch" paradigm. Twenty adult participants viewed a demonstration of a robot that could "feel" tactile stimulation through a haptic sensor on its hand and "see" changes in light through a photoreceptor at the level of the eyes; the robot responded to touch or changes in light by moving a contralateral digit. During EEG collection, participants engaged in a joint task that involved sending tactile stimulation to a partner (robot or human) and receiving tactile stimulation back. Tactile stimulation sent by the participant was initiated by a button press and was delivered 1500 ms later via an inflatable membrane on the hand of the human or on the haptic sensor of the robot partner. Stimulation to the participant's finger (from the partner) was sent on a fixed schedule, regardless of partner type. We analyzed activity of the sensorimotor mu rhythm during anticipation of tactile stimulation to the right hand, comparing mu activity at central electrode sites when participants believed that tactile stimulation was initiated by a robot or a human, and to trials in which "nobody" received stimulation. There was a significant difference in contralateral mu rhythm activity between anticipating stimulation from a human partner and the "nobody" condition. This effect was less pronounced for anticipation of stimulation from the robot partner. Analyses also examined beta rhythm responses to the execution of the button press, comparing oscillatory activity when participants sent tactile stimulation to the robot or the human partner. The extent of beta rebound at frontocentral electrode sites following the button press differed between conditions, with a significantly larger increase in beta power when participants sent tactile stimulation to a robot partner compared to the human partner. This increase in beta power may reflect greater predictably in event outcomes. This new paradigm and the novel findings advance the neuroscientific study of human–robot interaction.

Keywords: human–robot interaction, mu desynchronization, beta synchronization, social robotics, tactile perception

INTRODUCTION

As automation and technology become more ubiquitous in society, it is increasingly commonplace for interactions that have typically occurred between humans to also occur between humans and robots. Social-cognitive neuroscience offers a novel window into these interactions. Robots are traditionally constructed as highly complex tools, a design approach that persists in the present discourse on robotics and society (Šabanović, 2010). Increasingly, robots are designed as social agents, capable of interacting with humans in varied natural settings (Fong et al., 2003). Social and interactive robots have been developed for healthcare applications (Wada and Shibata, 2007; Broadbent et al., 2010; Robinson et al., 2014; Mann et al., 2015), educational settings (Tanaka et al., 2007; Toh et al., 2016), and mental health treatments (Begum et al., 2016). Social robots designed for these domains are often embodied, with varying degrees of human likeness; there is evidence that embodied social agents are more judged more favorably than disembodied social agents (Lee et al., 2006), especially within the context of social touch (Cramer et al., 2009a,b). Disembodied robots may either be simulated via a computer program or presented remotely through a screen. In either case, people empathize more with robots when they are physically embodied and present, compared with agents that are disembodied (Kwak et al., 2013). Given the embodied nature of social robots, it is likely that our interactions with such machines will increasingly involve tactile experiences (Huisman et al., 2013).

Humans use touch to communicate a wide range of social and emotional information (Knapp et al., 2013), and there is a good deal of current interest in this channel of communication in the context of human–robot interaction (HRI) (Gallace and Spence, 2010; Van Erp and Toet, 2015). We suggest that methods from social cognitive neuroscience can be applied to questions within the field of HRI, informing the design of robot entities to make human engagement with technology more fluid. Additionally, robots provide an opportunity to study human social behavior in various ways (Broadbent, 2017). There are many factors affecting how we perceive robots, from physical organization and appearance, to more subtle influences based on function and perceived intent. Functional affordances of a robot are the actions it is able to do, be it physical actions, gestures, or utterances (Awaad et al., 2015). In a study examining responses to multiple kinds of social robots, participants were more likely to report stronger engagement with a robot and intention to use it if it had sufficient affordances to complete a physical task, while physical appearance was rated as less important for engagement (Paauwe et al., 2015). Humanoid social robots with the ability to communicate through arm and hand gestures are rated as more anthropomorphic and likeable than physically identical robots without these capabilities (Salem et al., 2013).

Behavioral research within HRI has uncovered multiple factors and contexts that influence the ways in which people interact with and perceive robotic agents, specifically in the context of touch. Participants in a collaborative virtual reality environment rated a virtual agent capable of social touch through vibrotactile feedback more positively on affective adjectives than they did a non-touching agent (Huisman et al., 2014). Touch to (and from) a robot was shown to encourage participants' effort on a simple motor task (Shiomi et al., 2017). More specifically, active touch from a robot has been shown to be a stronger motivator than passive touch (Nakagawa et al., 2011). When interacting with robotic hands, participants report increased feelings of trust and friendship when the hands are warm, compared to cold robot hands or holding no hand at all (Nie et al., 2012); a similar study with a robotic social dinosaur found people liked warmer versions compared to tepid or cold conditions (Park and Lee, 2014).

One goal of research in HRI is to quantify whether robotic partners are sufficient analogs for human contact across different domains; perhaps unsurprisingly, there are contexts in which people prefer the contact of humans. In a study conducted in a nursing home, patients were comfortable with a robot touching their arm when they believed the robots intention was to clean them, but responded less positively when they believed the robot intended to provide them comfort through touch (Chen et al., 2014). A robot completing menial physical tasks fits well within our conception of what robots ought to do, but people tend to have reservations when imagining robots in social roles (Fong et al., 2003). In individuals with preexisting negative feelings toward robots, physical contact was shown to increase those negative attitudes (Wang and Quadflieg, 2015; Wullenkord et al., 2016); the inverse was true for those with positive attitudes. While a massage is both a functional and social task, participants receiving a head massage from a robot rated it worse than an equivalent massage delivered by a human (Walker and Bartneck, 2013). There is evidence that people may feel arousal or embarrassment when asked to touch the intimate parts of a robot (Li et al., 2016), and that humans also feel they are able to convey comforting and affectionate emotional states to a haptic machine through the action of touch (Yohanan and MacLean, 2012). These examples serve to highlight the complex nature of touch between humans and robots, and the many ways it differs across context and application.

Beyond the preceding review, little research exists on how humans respond to tactile interactions with robots, and even less work has considered the neural processes associated with these interactions. There is a robust literature within social cognitive neuroscience on the neural underpinnings of social touch, and it is the goal of the present study to combine those approaches with the goals and methods of HRI. Our interest lies in how examining neural activity related to touch can inform the psychological differences in interactions with robots rather than humans. To probe this question, we investigated whether oscillatory neural activity in anticipation of receiving tactile stimulation is influenced by participants' beliefs about whether the tactile stimulation was initiated by either a human or a robot agent. We were also interested in whether sensorimotor brain potentials associated with triggering tactile stimulation to another entity differ when the entity is a human or a robot. To examine these questions we employed a novel "reciprocal touch" paradigm and utilized measures derived from the electroencephalogram (EEG), specifically the sensorimotor mu and beta rhythms. Within the domain of social cognitive neuroscience, these rhythms have

often been studied in the context of the connections between the actions of the self and the actions of others, including tactile aspects of these linkages (Marshall and Meltzoff, 2015; Shen et al., 2017). Combining this line of work with behavioral insights from HRI research can help to forge new directions in the study of human responses to interacting with machines.

Changes in EEG brain rhythms have proven useful as reliable indicators of attentional orienting to touch, predicting perception of a subsequent weak tactile stimulus when that stimulus can reliably be expected (Zhang and Ding, 2010). Recent work in this area has focused interest on alpha-range rhythms, particularly in relation to anticipatory attention. Anticipatory desynchronization of alpha oscillations appears to be an index of local sensory cortex excitability, with heightened desynchronization associated with the perceptual salience of upcoming stimuli (Zhang and Ding, 2010; Foxe and Snyder, 2011). While this phenomenon has been studied in various sensory modalities, the focus here is on the sensorimotor mu rhythm in the EEG in relation to touch. The mu rhythm is an alpha-range oscillation that occurs at 8–13 Hz in adults and is typically observed over central electrode sites. During anticipation of impending tactile stimulation of the hand, there is a reduction of mu rhythm amplitude over contralateral somatosensory cortex (Anderson and Ding, 2011; Haegens et al., 2012; Shen et al., 2017). This desynchronization of the mu rhythm appears to reflect an increase in local field potentials of neurons in somatosensory cortex (Gomez-Ramirez et al., 2016). Shen et al. (2017) observed mu desynchronization in anticipation of tactile stimulation to one's own hand, which was not present when a partner or "nobody" received tactile stimulation. However, it is unknown whether the perceived origin of tactile stimulation delivered to the self (e.g., tactile stimulation initiated by a human vs. a machine) influences the amplitude of mu rhythm modulation during anticipation of touch.

While much research on the mu rhythm has concerned anticipatory attention, the EEG beta rhythm (14–30 Hz) has mainly been examined in the context of action production (Puzzo et al., 2011). Beta rhythm responses are modulated by motor movement and imagery (McFarland et al., 2000), and appear to reflect various spatial and temporal attentional mechanisms (van Ede et al., 2011). The beta response to the initiation of movement of the hands has been localized to the contralateral sensorimotor cortex, and takes the form of an event-related desynchronization (ERD) (Miller et al., 2007) followed by an event-related synchronization (ERS) that appears to reflect activity around the precentral gyrus (MI) (Gaetz and Cheyne, 2006). The increase in beta amplitude (i.e., beta rebound) following motor movement initiation is believed to reflect decreased cortical excitability and reduced processing of afferent sensory information involved in motor feedback (Pfurtscheller, 2001), and is also related to greater predictability of events and maintenance of the sensorimotor set (Engel and Fries, 2010). Alpha and beta oscillatory responses have also been widely used in the development and implementation of brain-computer interfaces (Yuan and He, 2014), with particular interest in the beta rhythm due to the range of human behaviors that engender this rhythm; the beta rhythm is responsive to both overt and imagined motor movement, and can be used to control machines (Neuper et al., 2009).

The present study introduces a novel paradigm in which participants carried out a joint tactile task with a robot or a human partner. The study aimed to answer the following questions: (1) Are different anticipatory neural responses seen to impending tactile stimulation to the self if it is believed that this stimulation is initiated by a human versus a robot partner?; (2) Is the sensorimotor EEG response to initiating tactile stimulation different when the target of the stimulation is a human or robot partner? In order to address these questions, we collected EEG from adult participants while they engaged in a turn-taking task with a robot or human partner. The overarching aim of the study was to contribute to the development of new HRI protocols in which human brain activity is monitored during interaction with robotic agents in a controlled setting.

MATERIALS AND METHODS

Participants

Twenty undergraduates (18 females; mean age = 19.70 years; $SD = 2.34$) received course credit in return for participation. This study was carried out with approval from the Institutional Review Board at Temple University, with informed consent being obtained from each participant. All participants were right handed according to the Oldfield handedness questionnaire (Oldfield, 1971), had normal or corrected vision, and reported no history of neurological illness or abnormality.

Stimuli and Materials
Tactile Stimulation
Tactile stimuli were delivered using an inflatable membrane (10 mm diameter) mounted in a plastic casing and attached to the finger by a flexible plastic clip. The membrane was attached to the right index finger of the participants and their human partner; the membrane was attached to the haptic sensor of the robotic task partner (see below). The membrane was inflated by a short burst of compressed air delivered via flexible polyurethane tubing (3 m length, 3.2 mm outer diameter). The compressed air delivery was controlled by STIM stimulus presentation software in combination with a pneumatic stimulator unit (both from James Long Company) and an adjustable regulator that restricted the airflow to 60 psi. The pneumatic stimulator and regulator were located in an adjacent room to the participant. To generate each tactile stimulus, the STIM software delivered a TTL trigger (10 ms duration) that served to open and close a solenoid in the pneumatic stimulator. Expansion of the membrane started 15 ms after trigger onset and peaked 20 ms later (i.e., 35 ms after trigger onset). The total duration of membrane movement was around 100 ms. This stimulation method has been used previously in a number of EEG and MEG studies (Pihko and Lauronen, 2004; Saby et al., 2015; Shen et al., 2018).

Task Partner
Prior to participating in the experimental procedure, participants were shown a demonstration of the robot that they would be

FIGURE 1 | Experimental setup during the robot condition. Participants responded to visual stimuli presented on the monitor. The robot used for this study is on the right side of the barrier. The embedded image in the top right shows how the stimulation device is attached to the index finger of the participant. Written informed consent was obtained from the individual on the left for publication of this image.

interacting with. The robot was implemented via an Arduino UNO board. The robot was comprised of left and right "hands," a torso, and a head (see **Figure 1**). The left hand contained a single haptic sensor; the right hand had a single point of articulation at the position of the index finger, which was movable via a servo embedded within the hand. The index finger was programmed to move downward to touch the surface in front of it in response to either a flash of light, or a touch to the left hand during the demonstration. Two red LEDs served as the "eyes," while a small photoreceptor was placed between these LEDs. Participants were asked to shine a light from a cellphone over the photoreceptor of the robot, which triggered movement of the right index finger. Participants were shown a small LED light and were told that during the study, the robot would know when it was its turn to press the button based on this LED flashing toward its photoreceptor. All participants were given the same introduction to the robot, and told they would be carrying out a joint task involving reciprocal tactile stimulation.

Following the demonstration of the robot, participants were all given the same introduction to their human task partners, and told they would be carrying out the same joint task as with the robot. Participants were shown that the human partner would see the same visual cues as them, wear the same inflatable membrane, and press an identical button to initiate tactile stimulation.

Design, Task, and Procedure
Procedure
Participants were seated 60 cm from a flat panel monitor (40 cm viewable), on which visual cues relating to the onset of tactile stimulation were presented. Participants held a small box in their left hand on which a single response button was mounted. Seated across from the participant was either a human partner or the robot, depending on condition and order. During each block,

the participant was aware whether their partner was a human or the robot, but could not see them (**Figure 1**): Participants were separated from their task partner by a divider, in order to control for visual influences during data collection. To mask any subtle sounds associated with delivery of the tactile stimuli, participants wore earplugs during EEG collection, and ambient white noise was broadcast in the testing room.

Task Conditions
Participants engaged in three blocks of trials with a human partner and three blocks of trials with a robot partner. All blocks within a condition (human/robot partner) occurred together, and the order of presentation (human first/robot first) was counterbalanced between participants. Prior to beginning the protocol, a practice trial was conducted by an experimenter, who demonstrated each of the three trial types shown in **Figure 2**: (1) *Nobody* trials: during these trials, an initial fixation point was replaced with a black diamond, which then turned green indicating that a tactile event was being sent to "nobody." Neither the participant nor the partner were required to press a button, and an air pulse was sent to an inflatable membrane (not attached to anyone or anything) in the testing room; (2) *Self* trials: during these trials, the fixation point was replaced by a black arrow facing downward, indicating that the participant could expect tactile stimulation to delivered to his or her right hand following a button press by the partner. Participants were told that the arrow turns green when their partner presses the button. Importantly, the partner (human or robot) was not actually triggering stimulation; the arrow turned green at a fixed interval of 400 ms following the black arrow across both conditions. 1500 ms after the arrow turned green, tactile stimulation was delivered to the participant's finger. The trial timing was held constant across conditions, in order to keep the human/robot conditions as similar as possible aside from the type of partner; (3) *Other* trials: During these trials, a black arrow facing upward replaced the fixation, indicating that the participant could now press the button with his or her left hand, which then triggered the arrow to turn green and initiated the tactile pulse to be delivered to the partner's hand 1500 ms later (See **Figure 2**).

Each of these three trial types was presented 80 times within each condition (human/robot), resulting in six different conditions and a total of 480 trials; since the *nobody* conditions did not differ in any way between partner types, these were collapsed into a single condition, resulting in five conditions for analyses: *nobody*: tactile stimulation is triggered and felt by no one; *self-human*: tactile stimulation is sent from a human partner to the participant; *self-robot*: tactile stimulation is sent from a robot partner to the participant; *other-human*: tactile stimulation is sent from the participant to a human partner; *other-robot*: tactile stimulation is sent from the participant to a robot partner. Trials within the *nobody* and *self* conditions were 4900 ms in length, while trials in the *other* conditions varied in duration due to variation in reaction time for the button press. The mean reaction time for the button press was 280 ms, resulting in an average trial length of 4780 ms for this condition. Data collection lasted approximately 45 min, including breaks between each block. *Nobody* trials were randomly presented during the blocks,
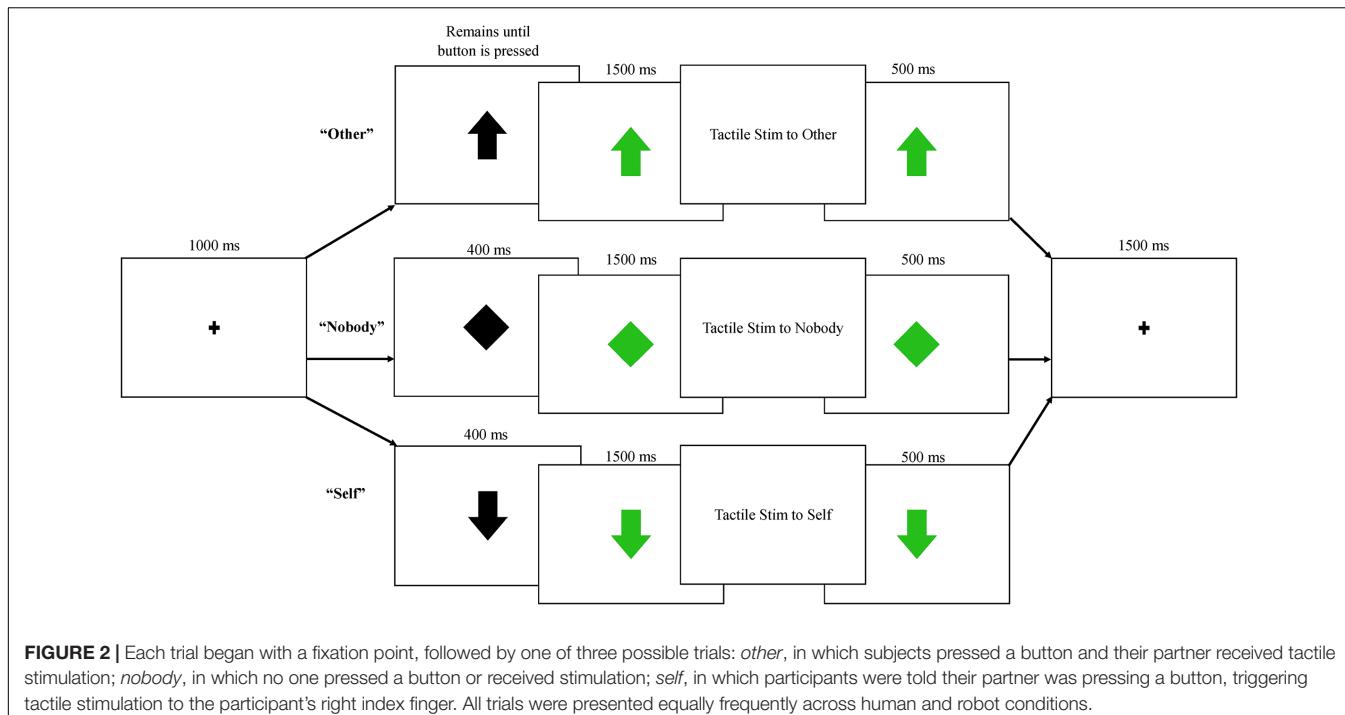
FIGURE 2 | Each trial began with a fixation point, followed by one of three possible trials: *other*, in which subjects pressed a button and their partner received tactile stimulation; *nobody*, in which no one pressed a button or received stimulation; *self*, in which participants were told their partner was pressing a button, triggering tactile stimulation to the participant's right index finger. All trials were presented equally frequently across human and robot conditions.

while self and other trials were always presented in a self-other-self or other-self-other fashion, as a way to keep the reciprocal nature of the task salient. The presentation of these three trial units was randomized across all blocks.

Following EEG collection, a brief questionnaire was given as a manipulation check, consisting of 13 questions about their performance with the robot partner, rated on a Likert scale from 1 (strongly disagree) to 5 (strongly agree). Overall, participants believed that their human partner chose when to push the button ($M = 4.85$, $SD = 0.95$), while participants were less likely to agree that the robot partner chose when to push the button ($M = 3.00$, $SD = 1.3$).

Joint Task

Participants were given instructions for completing the tactile attention task in cooperation with their partners during each of the six blocks. Within each block, there were trials during which the participant received two tactile pulses rather than one. Participants were instructed to count the number of these double pulses within each block. Before beginning the experiment, each participant received practice in distinguishing the double pulses from the regular single pulses. For each block, a predetermined number of double pulses was sent to the partner, and between 3 to 12 pulses were sent to the participant. After each block, an experimenter entered the room and asked the participant how many double pulses he or she had felt, and then either asked the human partner or checked a small LCD screen on the robot. The respective totals were summed and compared to a total that was unknown to the participant, but was known to the experimenter. The researcher would report the correct total, and as appropriate, would state the number of missed double pulses. Trials with

double pulses were excluded from EEG analyses. Participants' mean performance on the attentional task of detecting double pulses was 92%; performance did not differ between condition or order.

Data Acquisition

The EEG signal was acquired from 32 electrodes secured in a Lycra stretch cap (ANT Neuro, Germany) according to the International 10–20 format. Each electrode casing was filled with a small amount of conductive gel. Preparation of the EEG cap took place after participants had been given the demonstration of the robot, before the practice trials. The EEG signals were collected referenced to Cz with an AFz ground, and were re-referenced offline to the average of the left and right mastoids. Eye blinks were monitored via EOG electrodes placed above and below the left eye. Scalp impedance at each electrode site was kept under 25 kΩ. All EEG and EOG signals were amplified by optically isolated, high input impedance (>1 GΩ) bio amplifiers from SA Instrumentation (San Diego, CA, United States) and were digitized using a 16-bit A/D converter (+/− 2.5 V input range) at a sampling rate of 512 Hz using Snap-Master data acquisition software (HEM Data Corp., Southfield, MI, United States). Hardware filter settings were 0.1 Hz (high-pass) and 100 Hz (low-pass) with a 12 dB/octave rolloff. Bioamplifier gain was 4000 for the EEG channels and 1000 for the EOG channels.

Data Analysis
Preprocessing of EEG Data

Electroencephalogram analysis was performed using the EEGLAB 13.6.5b toolbox (Delorme and Makeig, 2004)

implemented in MATLAB. Epochs were extracted from the continuous EEG data. For the analysis of the *nobody/self* conditions, each extracted epoch was 3500 ms in duration, beginning 600 ms before visual cue onset and ending 1000 ms after the onset of the tactile stimulus. For analysis of the *other* conditions, each epoch was time-locked to the moment when the participant pressed the button, which triggered stimulation to their participant that occurred 1500 ms later. For these trials, analysis began at -2900 ms relative to the 0 ms point of tactile stimulation delivery, with a baseline period of the 500 ms prior to the visual cue, during the display of the fixation point.

Independent component analysis was conducted to remove eye movement artifacts (Hoffmann and Falkenstein, 2008). Visual inspection of the EEG signal was used to reject epochs containing movement artifact. Across all participants, 92.43% of trials were retained. There was no significant difference in the number of rejected epochs between trial type, $p = 0.387$.

Time Frequency Analysis
Time-frequency decompositions of single trial data were conducted using event-related spectral perturbation (ERSP) analysis (Makeig, 1993), for a 2500 ms window that ran from -2000 ms prior to the onset of the tactile stimulus to 500 ms after tactile stimulation onset. ERSP was computed using a Morlet wavelet decomposition over a frequency range of 5–30 Hz, with 100 overlapping windows starting with a 3-cycle wavelet at the lowest frequency. Event-related desynchronization (ERD) was taken as an ERSP decrease relative to the baseline.

Statistical Analyses
Mean ERSP in the mu band (8–13 Hz) over the 1000 ms epoch leading up to the onset of the tactile stimulus was computed for centroparietal electrodes overlying the contralateral (CP1, CP5, P3 and C3) and ipsilateral (CP2, CP6, P4 and C4) sensorimotor cortex. In order to assess anticipatory effects induced by the different conditions over this region of interest, mean ERSP was submitted to a 3 × 2 repeated-measures ANOVA involving condition (*human, robot, nobody*) and hemisphere (contralateral, ipsilateral). To confirm whether any condition effects were regionally specific, we also executed a mass univariate analysis comparing alpha ERSP amplitudes between conditions for each of the 32 electrodes.

In order to assess differences in beta band responses following the button press by the participants, mean beta (14–22 Hz) ERSP over the 1000 ms epoch following the button press was compared between the two relevant conditions (*other-human, other-robot*) across three target electrode regions within a 2 × 3 repeated-measures ANOVA. The beta band of 14–22 Hz was chosen based on previous research on post-movement beta responses, in which modulation of power is most frequently seen in the 15–20 Hz range (Pfurtscheller, 2001). The three regions encompassed frontocentral (Fz, FC1, FC2, Cz) and centroparietal electrodes overlying the contralateral (CP1, CP5, P3 and C3) and ipsilateral (CP2, CP6, P4 and C4) sensorimotor cortex. To confirm whether any condition differences were regionally

specific, we also executed a mass univariate analysis comparing beta ERSP amplitudes between conditions at each of the 32 electrodes.

RESULTS

Tactile Anticipation
Mu Rhythm (8–13 Hz)
The ANOVA for mu ERSP indicated significant main effects of condition, $F(2,18) = 18.63$, $p < 0.001$ and hemisphere, $F(1,19) = 22.88$, $p < 0.001$. Further, the ANOVA indicated a significant interaction of condition and hemisphere, $F(1,19) = 14.32$, $p < 0.001$. Follow-up analyses indicated that mu ERSP over the contralateral (i.e., left) centroparietal region was significantly reduced (indicating greater desynchronization) when participants expected tactile stimulation to self (whether initiated by a *robot*, $M = -0.89$, $SD = 0.701$, or a *human*, $M = -1.02$, $SD = 0.635$), compared to trials when no tactile stimulation was expected (*nobody*, $M = -0.94$, $SD = 0.411$). There was no significant difference between conditions over the ipsilateral centroparietal region.

The mass univariate analyses confirmed regional specificity of effects by showing that anticipatory mu ERD was significantly different at the $p < 0.01$ threshold at various centroparietal electrodes, but not over other scalp regions, when comparing stimulation from the *human* to the *nobody* condition. These differences were apparent at C3, $t(18) = 5.31$, $p < 0.001$, CP1, $t(18) = 3.56$, $p = 0.002$, P3, $t(18) = 3.98$, $p < 0.001$, and CP5, $t(18) = 4.65$, $p < 0.001$, such that mu desynchronization at these electrodes was greater for the *human* condition compared to the *nobody* condition. Compared with the *nobody* condition, there was also significantly greater mu desynchronization in the *robot* condition at C3, $t(18) = 2.79$, $p = 0.012$, and CP5, $t(18) = 2.67$, $p = 0.015$ at the $p < 0.05$ threshold (see **Figure 3**). For the direct comparison of mu ERSP in relation to the source of stimulation to the self (*human vs. robot*), there was only a marginal difference in amplitude observed at one electrode (CP1), $F(1,19) = 2.081$, $p = 0.061$; ERSP at all other electrodes did not differ significantly between the human and robot conditions.

Execution of Action
Behavioral Measures
Mean reaction time for participants' button presses following the visual cue did not differ between conditions (mean RT with human partner = 278.68 ms, $SD = 16.94$; mean RT with robot partner = 282.62, $SD = 14.22$), $t(19) = -0.74$, $p = 0.463$.

Beta Band (14–22 Hz)
The ANOVA for beta ERSP indicated significant main effects of condition, $F(1,18) = 7.45$, $p < 0.001$ and region, $F(1,19) = 21.96$, $p < 0.001$. The ANOVA further indicated a significant interaction of condition and region, $F(1,19) = 14.32$, $p < 0.001$. Follow up analyses indicated that there was a significantly greater beta ERSP after the button press over frontocentral regions when participants delivered tactile stimulation to a robot ($M = 1.02$, $SD = 0.67$), compared to trials where participants delivered
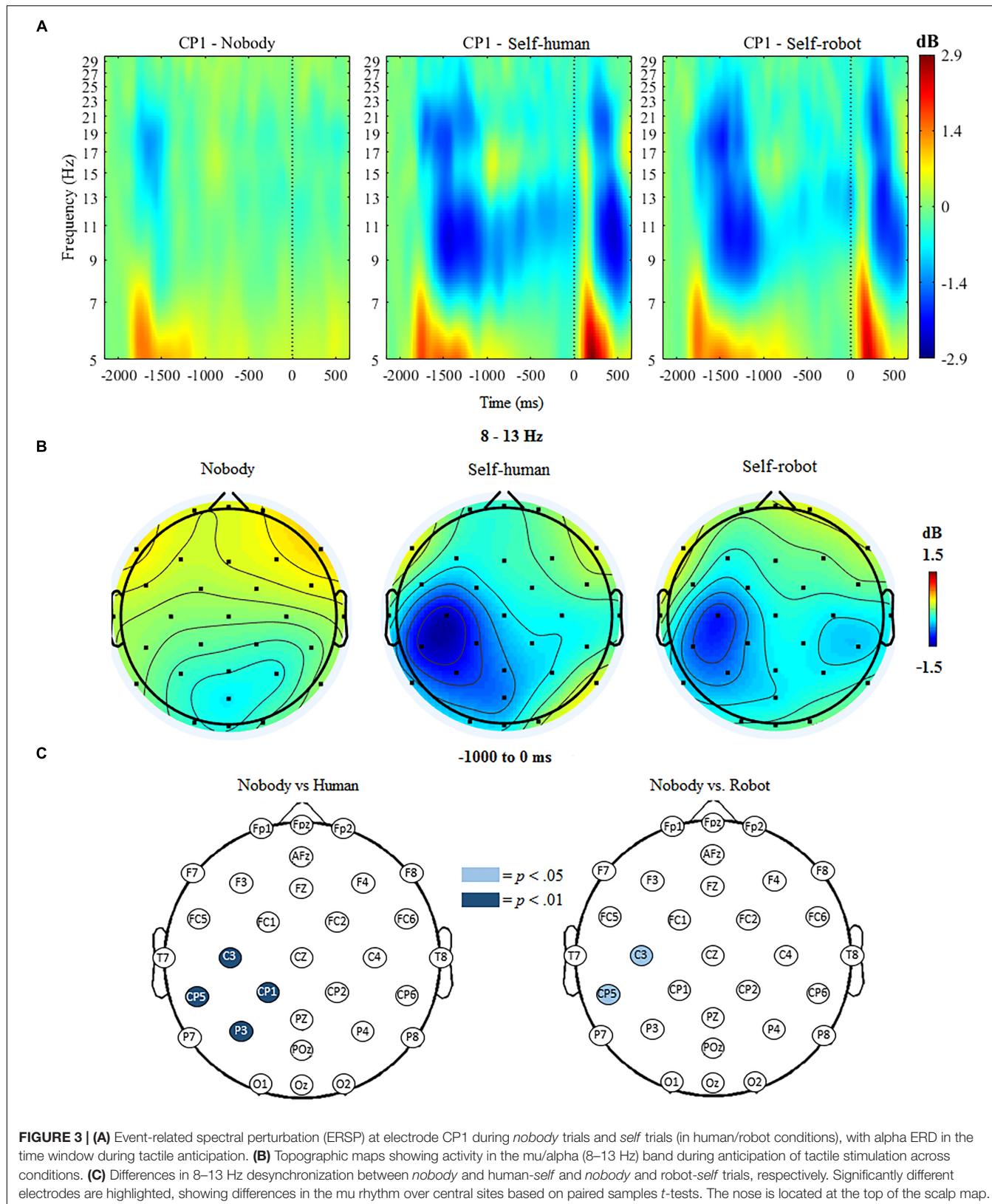
FIGURE 3 | (A) Event-related spectral perturbation (ERSP) at electrode CP1 during *nobody* trials and *self* trials (in human/robot conditions), with alpha ERD in the time window during tactile anticipation. **(B)** Topographic maps showing activity in the mu/alpha (8–13 Hz) band during anticipation of tactile stimulation across conditions. **(C)** Differences in 8–13 Hz desynchronization between *nobody* and human-*self* and *nobody* and robot-*self* trials, respectively. Significantly different electrodes are highlighted, showing differences in the mu rhythm over central sites based on paired samples *t*-tests. The nose is located at the top of the scalp map.
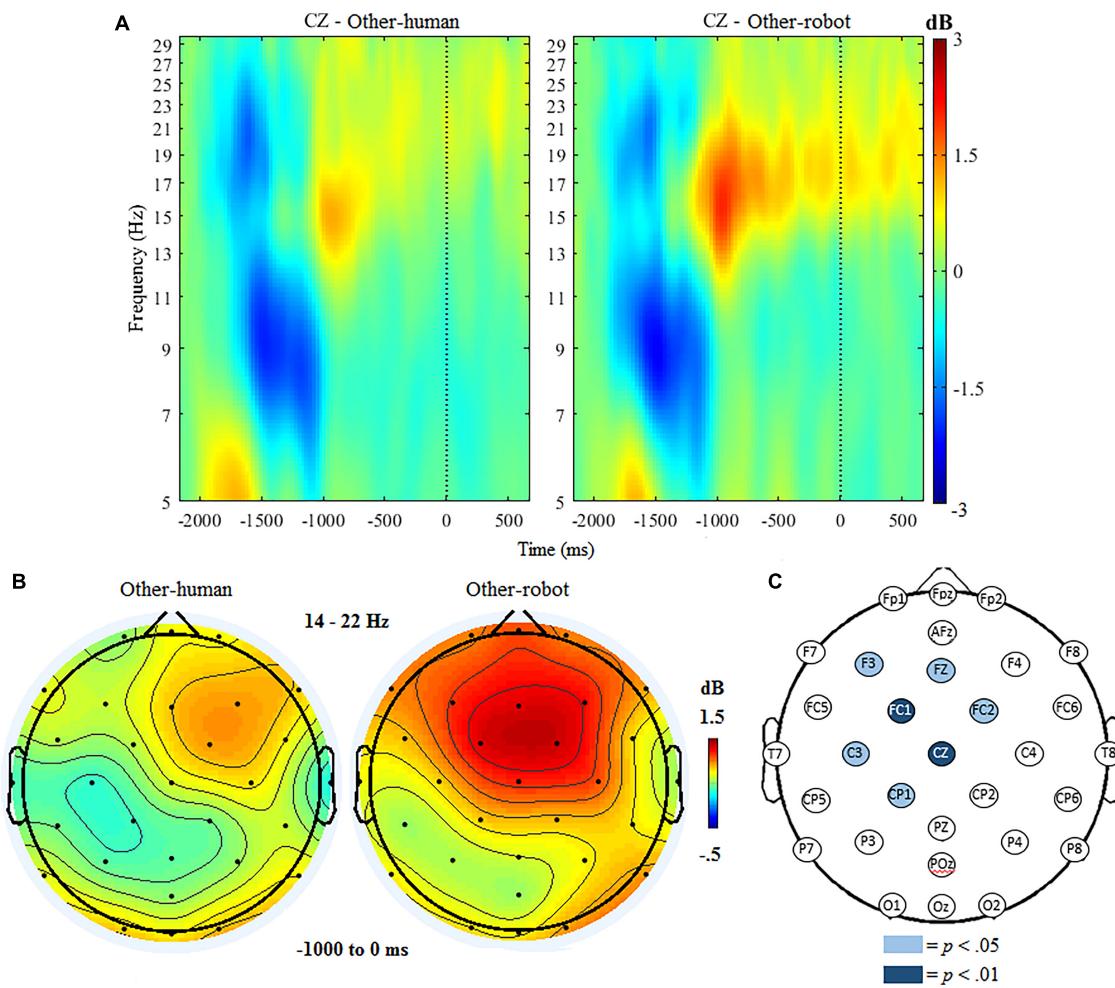
FIGURE 4 | (A) Event-related spectral perturbation at electrode Cz during *other* trials, with beta ERS present the time window following the button press (−1000 to 0 ms). *Human-other* and *robot-other* conditions are shown separately. **(B)** Topographic maps showing condition differences for post-button press activity in the beta band (14–22 Hz) from −1000 to 0 ms, with 0 ms being time locked to the participant's button press. **(C)** Differences in beta power increase between human and robot conditions in the *other* trials. Significantly different electrodes are highlighted, showing differences in beta ERSP over central sites based on paired samples *t*-tests. The nose is located at the top of the scalp map.

stimulation to a human ($M = 0.02$, $SD = 0.59$). There were no significant differences between conditions at centroparietal sites.

Mass univariate analyses further indicated that the above effect was specific to frontocentral sites, with beta ERSP being significantly greater in *other-robot* trials than for *other-human* trials at electrodes FC1, $t(18) = 6.31$, $p < 0.001$, and Cz, $t(18) = 5.95$, $p < 0.001$, at the $p < 0.01$ threshold. Beta ERSP was significantly greater in *other-robot* trials at F3, $t(18) = 2.80$, $p = 0.011$, Fz, $t(18) = 2.68$ $p = 0.015$, FC2, $t(18) = 2.60$ $p = 0.018$, CP1, $t(18) = 2.53$, $p = 0.021$, and C3, $t(18) = 2.49$, $p = 0.022$) at the $p < 0.05$ threshold (see **Figure 4**).

DISCUSSION

We investigated sensorimotor oscillations during a reciprocal touch paradigm, using EEG measures to compare aspects of

brain oscillatory responses to receiving and initiating tactile stimulation during a joint task involving either a human or robot partner. Our specific questions were twofold: First, whether desynchronization of the sensorimotor mu rhythm during anticipation of tactile stimulation differed according to the perceived origin of the stimulation (as being initiated by a human vs. a robot). Second, whether EEG beta band responses to the act of initiating delivery of a tactile stimulus to another entity differed according to whether that entity is a human or a robot.

Anticipation of Tactile Stimulation

In line with previous research (Shen et al., 2017), there was a clear desynchronization of the EEG mu rhythm over the contralateral central region during the anticipation of tactile stimulation to self. A similar desynchronization was not present during the "nobody" condition in which a cue was present and a stimulus was triggered, but the stimulus was not directed toward anyone.

In terms of the central question, we found little evidence for a differential modulation of mu rhythm activity when participants anticipated tactile stimulation that they believed was initiated by a button press from a robot or a human partner. The extent of anticipatory mu desynchronization did not meaningfully differ in amplitude when the source of the tactile stimulation was the action of a human as opposed to a robot. Given that anticipatory mu desynchronization is considered an index of selective attention in the tactile modality (van Ede et al., 2012; Weiss et al., 2018), these results suggest that participants were equally attentive in monitoring for upcoming tactile stimulation from human and robot partners. While the directions of the means suggested that mu ERD was somewhat greater when participants expected stimulation initiated by a human rather than a robot partner, only trend-level differences in amplitude were apparent, and only at one electrode site.

One strength of our task protocol was that visual cues were constant across conditions, in order to isolate the influence of participant's beliefs about the nature of their partner on the brain responses during the task. However, it is also possible that the salience of the manipulation could be increased by allowing participants to observe the human or robot partner press the button. The subtle differences we observed can be further investigated by providing participants with contingent visual information about the nature of the partner, or providing a more "social" rather than physical interaction with the robot partner.

Although it is understood that the extent of anticipatory mu rhythm desynchronization is related to subsequent perceptual processing of the target stimulus (Zhang and Ding, 2010), little is known about the determinants of the anticipatory mu response, including individual differences. There remains sustained interest in the neural processes underlying the mapping of somatosensory experience from one own body to that of another (Keysers et al., 2010; Marshall and Meltzoff, 2015). One study found no clear evidence for mu desynchronization during the anticipation of tactile stimulation delivered to another person (Shen et al., 2017), and we did not examine this question here. Instead, the current study took a novel approach by examining how the perceived origin of tactile stimulation modulated the anticipatory mu response to self. With this in mind, we considered anticipatory processes in the context of a sustained interactive task, which also allowed us to examine processes related to the sending of tactile stimulation to the partner.

Execution of Action

One well-studied electrophysiological correlate of action production (particularly finger movements, such as a button press) is the beta rebound response, which takes the form of an increase in beta band power after a brief reduction of power immediately following the action (Cheyne, 2013). Here we found modulation of the beta rebound response in the period after participants initiated tactile stimulation to a partner, prior to the actual delivery of the tactile stimulus. When participants initiated tactile stimulation that was directed toward a robot, there was greater beta ERS across central and frontal electrode sites. The exact function and mechanism of beta ERS following action production is not entirely understood, but it is believed to

partly reflect motor inhibition (Heinrichs-Graham et al., 2017). In certain contexts, post-movement beta ERS relates to increases in cortical deactivation (Pfurtscheller, 2001), particularly in the lower beta (14–22 Hz) range used in the current study. The greatest level of beta ERS was seen over frontoparietal sites, which is expected based on the localization of post-movement ERS to motor cortex (Jurkiewicz et al., 2006). The meaning of these differences between conditions in beta ERS still needs to be elucidated. One line of reasoning relates to the idea that reduced beta band activity increases the capability for cognitive and motor flexibility in terms of upcoming or future responses (Engel and Fries, 2010). As such, enhancement of beta band activity (i.e., a larger rebound effect) in the context of HRI may reflect the perceived greater predictability of robot compared to human partners. Further work can examine this speculation as well as investigate other possible influences on the beta response (e.g., differing button press force between conditions).

In behavioral studies within the domain of HRI, reactions to robots vary greatly across studies, due in part to the wide range of robotic forms implemented across this area of research (Hoffmann et al., 2010). In some contexts, artificial agents may be more engaging than human counterparts (Gratch et al., 2007), but within most natural contexts, people tend to prefer the company of humans to machines. Increased beta ERS while interacting with the robot could reflect a different attentional state, or a decrease in uncertainty regarding the outcome of the present action (Engel and Fries, 2010). This speculation warrants further investigation, as do alternative explanations for the differential beta rebound responses. The nature of the present paradigm was limited in how much it immersed the participant in interactions with the partner; beyond the initial introductions, participants only interacted with the human or robot partner through the delivery of tactile stimulation, without any visual, auditory, or direct physical contact. The non-visual nature of the present study was intentional, given the strong visual effects found in previous research derived from aesthetic differences between human and robotic stimuli (Press, 2011). Furthermore, the embodiment of the robot in the current study was somewhat limited. The addition of different sensory modalities and the use of more sophisticated robot platforms could help to develop a richer picture of brain responses during interactions with robots.

Implications

The results of this study provide some of the first evidence for differences in attentional and tactile processing when interacting with human and robotic partners. Past research on HRI has focused largely on differences in physical appearance and abilities (Hoffmann et al., 2010; Paauwe et al., 2015; Strait et al., 2017), while the present study removed visual information from the experimental procedure; differences between conditions are therefore likely to be the result of a participant's beliefs, rather than visual input during the task. Previous work in HRI has begun to identify factors which influence how people respond to touch from robots (Nakagawa et al., 2011; Wullenkord et al., 2016), and future work on the cognitive neuroscience of HRI will need to incorporate these factors into the design of robots used

in this line of research. Research in this area could also draw on emerging work in the area of "sociomotor action control" which has clear implications for progress in HRI (Kunde et al., 2017).

Future work with this and similar paradigms will continue to shed light on sensory processing in the context of stimuli that are delivered by a non-human agent. A potential follow-up to the present study could include an additional condition in which neither a human nor robot is present during tactile stimulation to the participant; such a condition would allow further exploration of the effects of other agents on sensory processing. Given the controlled nature of the present study, the tasks has little resemblance to a typical social interaction. Further studies with a similar paradigm could be situated in naturalistic contexts, such as physical therapy, wherein touch is an integral and natural part of the interaction. Additional work could also examine differences between protocols that involve passive touch (as in the task used in the current study) and the more active kinds of touch that characterizes typical human–environment interactions.

Within research on brain-computer interfaces (BCI), sensorimotor oscillations are most frequently targeted as a source of input to control various machines or devices (Yuan and He, 2014). In this line of work, the beta rhythm has been targeted most frequently, specifically post-movement beta rebound, due to the well-timed relation between motor movements and the corollary oscillations measurable through EEG (Pfurtscheller and Solis-Escalante, 2009). These oscillations are of particular interest for BCI researchers due to the range of human behaviors which can engender them, with and without overt motor movement. Through examination of sensorimotor activity in response to the use of BCI, feedback loops can be created which form a sort of continuous connection to brain-controlled machines (Neuper et al., 2009), allowing for the control of machines through motor imagery alone (Pfurtscheller and Neuper, 2010). The present study can inform the creation of BCI platforms by showing how beta rebound, and sensorimotor oscillations in general, may be influenced by the nature of the machine being acted upon. Follow-up work on this issue could be conducted across a range of different types of machines from humanoid robots and androids to simple mechanical machines. Algorithms that are robust to psychological perturbations on sensorimotor rhythms would be ideal in the application to BCI.

In addition to helping us understand how to better integrate robots in social contexts, a social-cognitive neuroscience approach to robotics can provide insights beyond the field of HRI (Broadbent, 2017). Robots provide a unique control in social paradigms, as various levels of intentionality, autonomy, and humanoid appearance can be manipulated. Human reactions to robotic bodies varies greatly depending on the nature of the machine and context in which it is experienced (Chaminade and Cheng, 2009), but there appears to be something unique about the way in which we process information about a thing when that thing is a fellow human (Saygin et al., 2011; Urgen et al., 2013).

ETHICS STATEMENT

AUTHOR CONTRIBUTIONS

ACKNOWLEDGMENTS

REFERENCES

Anderson, K., and Ding, M. (2011). Attentional modulation of the somatosensory mu rhythm. *Neuroscience* 180, 165–180. doi: 10.1016/j.neuroscience.2011. 02.004

Awaad, I., Kraetzschmar, G. K., and Hertzberg, J. (2015). The role of functional affordances in socializing robots. *Int. J. Soc. Robot.* 7, 421–438. doi: 10.1007/s12369-015-0281-3

Begum, M., Serna, R. W., and Yanco, H. A. (2016). Are robots ready to deliver autism interventions? a comprehensive review. *Int. J. Soc. Robot.* 8, 157–181. doi: 10.1007/s12369-016-0346-y

Broadbent, E. (2017). Interactions with robots: the truths we reveal about ourselves. *Annu. Rev. Psychol.* 68, 627–652. doi: 10.1146/annurev-psych-010416-043958

Broadbent, E., Kuo, I. H., Lee, Y. I., Rabindran, J., Kerse, N., Stafford, R., et al. (2010). Attitudes and reactions to a healthcare robot. *Telemed. E-Health* 16, 608–613. doi: 10.1089/tmj.2009.0171

Chaminade, T., and Cheng, G. (2009). Social cognitive neuroscience and humanoid robotics. *J. Physiol. Paris* 103, 286–295. doi: 10.1016/j.jphysparis.2009.08.011

Chen, T. L., King, C.-H. A., Thomaz, A. L., and Kemp, C. C. (2014). An investigation of responses to robot-initiated touch in a nursing context. *Int. J. Soc. Robot.* 6, 141–161. doi: 10.1007/s12369-013-0215-x

Cheyne, D. O. (2013). MEG studies of sensorimotor rhythms: a review. *Exp. Neurol.* 245, 27–39. doi: 10.1016/j.expneurol.2012.08.030

Cramer, H., Kemper, N., Amin, A., and Evers, V. (2009a). "The effects of robot touch and proactive behaviour on perceptions of human-robot interactions," in *Proceedings of the Human-Robot Interaction (HRI), 2009 4th ACM/IEEE International Conference on Human robot interaction* (San Diego, CA: IEEE), 275–276.

Cramer, H., Kemper, N., Amin, A., Wielinga, B., and Evers, V. (2009b). 'Give me a hug': the effects of touch and autonomy on people's responses to embodied social agents. *Comput. Animat. Virt. Worlds* 20, 437–445. doi: 10.1002/cav.317

Delorme, A., and Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *J. Neurosci. Methods* 134, 9–21. doi: 10.1016/j.jneumeth.2003.10.009

Engel, A. K., and Fries, P. (2010). Beta-band oscillations—signalling the status quo? *Curr. Opin. Neurobiol.* 20, 156–165. doi: 10.1016/j.conb.2010.02.015

Fong, T., Nourbakhsh, I., and Dautenhahn, K. (2003). A survey of socially interactive robots. *Robot. Autonom. Syst.* 42, 143–166. doi: 10.1016/S0921-8890(02)00372-X

Foxe, J. J., and Snyder, A. C. (2011). The role of alpha-band brain oscillations as a sensory suppression mechanism during selective attention. *Front. Psychol.* 2:154. doi: 10.3389/fpsyg.2011.00154

Gaetz, W., and Cheyne, D. (2006). Localization of sensorimotor cortical rhythms induced by tactile stimulation using spatially filtered MEG. *Neuroimage* 30, 899–908. doi: 10.1016/j.neuroimage.2005.10.009

Gallace, A., and Spence, C. (2010). The science of interpersonal touch: an overview. *Neurosci. Biobehav. Rev.* 34, 246–259. doi: 10.1016/j.neubiorev.2008.10.004

Gomez-Ramirez, M., Hysaj, K., and Niebur, E. (2016). Neural mechanisms of selective attention in the somatosensory system. *J. Neurophysiol.* 116, 1218–1231. doi: 10.1152/jn.00637.2015

Gratch, J., Wang, N., Okhmatovskaia, A., Lamothe, F., Morales, M., van der Werf, R. J., et al. (2007). "Can virtual humans be more engaging than real ones?," in *Proceedings of the International Conference on Human-Computer Interaction* (Berlin: Springer), 286–297. doi: 10.1007/978-3-540-73110-8_30

Haegens, S., Luther, L., and Jensen, O. (2012). Somatosensory anticipatory alpha activity increases to suppress distracting input. *J. Cogn. Neurosci.* 24, 677–685. doi: 10.1162/jocn_a_00164

Heinrichs-Graham, E., Kurz, M. J., Gehringer, J. E., and Wilson, T. W. (2017). The functional role of post-movement beta oscillations in motor termination. *Brain Struct. Funct.* 222, 3075–3086. doi: 10.1007/s00429-017-1387-1

Hoffmann, S., and Falkenstein, M. (2008). The correction of eye blink artefacts in the EEG: a comparison of two prominent methods. *PLoS One* 7:e3004. doi: 10.1371/journal.pone.0003004

Hoffmann, M., Marques, H., Arieta, A., Sumioka, H., Lungarella, M., and Pfeifer, R. (2010). Body schema in robotics: a review. *IEEE Trans. Autonom. Ment. Dev.* 2, 304–324. doi: 10.1109/TAMD.2010.2086454

Huisman, G., Bruijnes, M., Kolkmeier, J., Jung, M., Frederiks, A. D., and Rybarczyk, Y. (2013). "Touching virtual agents: embodiment and mind," in *Proceedings of the International Summer Workshop on Multimodal Interfaces* (Berlin: Springer), 114–138.

Huisman, G., Kolkmeier, J., and Heylen, D. (2014). "Simulated social touch in a collaborative game," in *Proceedings of the International Conference on Human Haptic Sensing and Touch Enabled Computer Applications* (Berlin: Springer), 248–256. doi: 10.1007/978-3-662-44193-0_32

Jurkiewicz, M. T., Gaetz, W. C., Bostan, A. C., and Cheyne, D. (2006). Post-movement beta rebound is generated in motor cortex: evidence from neuromagnetic recordings. *Neuroimage* 32, 1281–1289. doi: 10.1016/j.neuroimage.2006.06.005

Keysers, C., Kaas, J. H., and Gazzola, V. (2010). Somatosensation in social perception. *Nat. Rev. Neurosci.* 11, 417–428. doi: 10.1038/nrn2833

Knapp, M. L., Hall, J. A., and Horgan, T. G. (2013). *Nonverbal Communication in Human Interaction*. Boston, MA: Cengage Learning.

Kunde, W., Weller, L., and Pfister, R. (2017). Sociomotor action control. *Psychon. Bull. Rev.* 25, 917–931. doi: 10.3758/s13423-017-1316-6

Kwak, S. S., Kim, Y., Kim, E., Shin, C., and Cho, K. (2013). What makes people empathize with an emotional robot? The impact of agency and physical embodiment on human empathy for a robot. *Paper Presented in 22nd IEEE International Symposium on Robot and Human Interactive Communication:*

Living Together, Enjoying Together, and Working Together with Robots, IEEE RO-MAN, Gyeongju, 180–185. doi: 10.1109/ROMAN.2013.6628441

Lee, K. M., Jung, Y., Kim, J., and Kim, S. R. (2006). Are physically embodied social agents better than disembodied social agents?: the effects of physical embodiment, tactile interaction, and people's loneliness in human–robot interaction. *Int. J. Hum. Comput. Stud.* 64, 962–973. doi: 10.1016/j.ijhcs.2006.05.002

Li, J., Ju, W., and Reeves, B. (2016). Touching a mechanical body: tactile contact with intimate parts of a humanoid robot is physiologically arousing. *Paper presented at the 66th Annual Conference of the International Communication Association*, Fukuoka.

Makeig, S. (1993). Auditory event-related dynamics of the EEG spectrum and effects of exposure to tones. *Electroencephalogr. Clin. Neurophysiol.* 86, 283–293. doi: 10.1016/0013-4694(93)90110-H

Mann, J. A., MacDonald, B. A., Kuo, I.-H., Li, X., and Broadbent, E. (2015). People respond better to robots than computer tablets delivering healthcare instructions. *Comput. Hum. Behav.* 43, 112–117. doi: 10.1016/j.chb.2014.10.029

Marshall, P. J., and Meltzoff, A. N. (2015). Body maps in the infant brain. *Trends Cognit. Sci.* 19, 499–505. doi: 10.1016/j.tics.2015.06.012

McFarland, D. J., Miner, L. A., Vaughan, T. M., and Wolpaw, J. R. (2000). Mu and beta rhythm topographies during motor imagery and actual movements. *Brain Topogr.* 12, 177–186. doi: 10.1023/A:1023437823106

Miller, K. J., Leuthardt, E. C., Schalk, G., Rao, R. P., Anderson, N. R., Moran, D. W., et al. (2007). Spectral changes in cortical surface potentials during motor movement. *J. Neurosci.* 27, 2424–2432. doi: 10.1523/JNEUROSCI.3886-06.2007

Nakagawa, K., Shiomi, M., Shinozawa, K., Matsumura, R., Ishiguro, H., and Hagita, N. (2011). "Effect of robot's active touch on people's motivation," in *Proceedings of the 6th International Conference on Human-Robot Interaction* (Lausanne: ACM), 465–472. doi: 10.1145/1957656.1957819

Neuper, C., Scherer, R., Wriessnegger, S., and Pfurtscheller, G. (2009). Motor imagery and action observation: modulation of sensorimotor brain rhythms during mental control of a brain-computer interface. *Clin. Neurophysiol.* 120, 239–247. doi: 10.1016/j.clinph.2008.11.015

Nie, J., Park, M., Marin, A. L., and Sundar, S. S. (2012). "Can you hold my hand? Physical warmth in human-robot interaction," in *Proceedings of the Human-Robot Interaction (HRI), 2012 7th ACM/IEEE International Conference on Human robot interaction* (Boston: IEEE), 201–202.

Oldfield, R. C. (1971). The assessment and analysis of handedness: the Edinburgh inventory. *Neuropsychologia* 9, 97–113. doi: 10.1016/0028-3932(71)90067-4

Paauwe, R. A., Hoorn, J. F., Konijn, E. A., and Keyson, D. V. (2015). Designing robot embodiments for social interaction: affordances topple realism and aesthetics. *Int. J. Soc. Robot.* 7, 697–708. doi: 10.1007/s12369-015-0301-3

Park, E., and Lee, J. (2014). I am a warm robot: the effects of temperature in physical human–robot interaction. *Robotica* 32, 133–142. doi: 10.1017/S026357471300074X

Pfurtscheller, G. (2001). Functional brain imaging based on ERD/ERS. *Vis. Res.* 41, 1257–1260. doi: 10.1016/S0042-6989(00)00235-2

Pfurtscheller, G., and Neuper, C. (2010). "Neurofeedback training for BCI control," in *Brain–Computer Interfaces – Revolutionizing Human-Computer Interaction* eds B. Graimann, B. Z. Allison, and G. Pfurtscheller (New York, NY: Springer), 65–78.

Pfurtscheller, G., and Solis-Escalante, T. (2009). Could the beta rebound in the EEG be suitable to realize a "brain switch"? *Clin. Neurophysiol.* 120, 24–29. doi: 10.1016/j.clinph.2008.09.027

Pihko, E., and Lauronen, L. (2004). Somatosensory processing in healthy newborns. *Exp. Neurol.* 190, 2–7. doi: 10.1016/j.expneurol.2004.01.024

Press, C. (2011). Action observation and robotic agents: learning and anthropomorphism. *Neurosci. Biobehav. Rev.* 3, 1410–1418. doi: 10.1016/j.neubiorev.2011.03.004

Puzzo, I., Cooper, N. R., Cantarella, S., and Russo, R. (2011). Measuring the effects of manipulating stimulus presentation time on sensorimotor alpha and low beta reactivity during hand movement observation. *Neuroimage* 57, 1358–1363. doi: 10.1016/j.neuroimage.2011.05.071

Robinson, H., MacDonald, B., and Broadbent, E. (2014). The role of healthcare robots for older people at home: a review. *Int. J. Soc. Robot.* 6, 575–591. doi: 10.1007/s12369-014-0242-2

Šabanović, S. (2010). Robots in society, society in robots. *Int. J. Soc. Robot.* 2, 439–450. doi: 10.1007/s12369-010-0066-7

Saby, J. N., Meltzoff, A. N., and Marshall, P. J. (2015). Neural body maps in human infants: somatotopic responses to tactile stimulation in 7-month-olds. *Neuroimage* 118, 74–78. doi: 10.1016/j.neuroimage.2015.05.097

Salem, M., Eyssel, F., Rohlfing, K., Kopp, S., and Joublin, F. (2013). To err is human (-like): Effects of robot gesture on perceived anthropomorphism and likability. *Int. J. Soc. Robot.* 5, 313–323. doi: 10.1007/s12369-013-0196-9

Saygin, A. P., Chaminade, T., Ishiguro, H., Driver, J., and Frith, C. (2011). The thing that should not be: predictive coding and the uncanny valley in perceiving human and humanoid robot actions. *Soc. Cogn. Affect. Neurosci.* 7, 413–422. doi: 10.1093/scan/nsr025

Shen, G., Meltzoff, A. N., and Marshall, P. J. (2018). Touching lips and hearing fingers: effector-specific congruency between tactile and auditory stimulation modulates N1 amplitude and alpha desynchronization. *Exp. Brain Res.* 236, 13–29. doi: 10.1007/s00221-017-5104-3

Shen, G., Saby, J. N., Drew, A. R., and Marshall, P. J. (2017). Exploring potential social influences on brain potentials during anticipation of tactile stimulation. *Brain Res.* 1659, 8–18. doi: 10.1016/j.brainres.2017.01.022

Shiomi, M., Nakagawa, K., Shinozawa, K., Matsumura, R., Ishiguro, H., and Hagita, N. (2017). Does a robot's touch encourage human effort? *Int. J. Soc. Robot.* 9, 5–15. doi: 10.1007/s12369-016-0339-x

Strait, M. K., Floerke, V. A., Ju, W., Maddox, K., Remedios, J. D., Jung, M. F., et al. (2017). Understanding the uncanny: both atypical features and category ambiguity provoke aversion toward humanlike robots. *Front. Psychol.* 8:1366. doi: 10.3389/fpsyg.2017.01366

Tanaka, F., Cicourel, A., and Movellan, J. R. (2007). Socialization between toddlers and robots at an early childhood education center. *Proc. Natl. Acad. Sci. U.S.A.* 104, 17954–17958. doi: 10.1073/pnas.0707769104

Toh, L. P. E., Causo, A., Tzuo, P. W., Chen, I., and Yeo, S. H. (2016). A review on the use of robots in education and young children. *J. Educ. Technol. Soc.* 19:148.

Urgen, B. A., Plank, M., Ishiguro, H., Poizner, H., and Saygin, A. P. (2013). EEG theta and Mu oscillations during perception of human and robot actions. *Front. Neurorobot.* 7:19. doi: 10.3389/fnbot.2013.00019

van Ede, F., de Lange, F., Jensen, O., and Maris, E. (2011). Orienting attention to an upcoming tactile event involves a spatially and temporally specific modulation of sensorimotor alpha-and beta-band oscillations. *J. Neurosci.* 31, 2016–2024. doi: 10.1523/JNEUROSCI.5630-10.2011

van Ede, F., de Lange, F. P., and Maris, E. (2012). Attentional cues affect accuracy and reaction time via different cognitive and neural processes. *J. Neurosci.* 32, 10408–10412. doi: 10.1523/JNEUROSCI.1337-12.2012

Van Erp, J. B., and Toet, A. (2015). Social touch in human–computer interaction. *Front. Digital Humanit.* 2:2. doi: 10.3389/fdigh.2015.00002

Wada, K., and Shibata, T. (2007). Living with seal robots—its sociopsychological and physiological influences on the elderly at a care house. *IEEE Trans. Robot.* 23, 972–980. doi: 10.1109/TRO.2007.906261

Walker, R., and Bartneck, C. (2013). "The pleasure of receiving a head massage from a robot," in *Proceedings of the RO-MAN, 2013 IEEE* (New York, NY: IEEE), 807–813. doi: 10.1109/ROMAN.2013.6628412

Wang, Y., and Quadflieg, S. (2015). In our own image? Emotional and neural processing differences when observing human–human vs human-robot interactions. *Soc. Cognit. Affect. Neurosci.* 10, 1515–1524. doi: 10.1093/scan/nsv043

Weiss, S. M., Meltzoff, A. N., and Marshall, P. J. (2018). Neural measures of anticipatory bodily attention in children: Relations with executive function. *Dev. Cogn. Neurosci.* 34, 148–145. doi: 10.1016/j.dcn.2018.08.002

Wullenkord, R., Fraune, M. R., Eyssel, F., and Šabanović, S. (2016). "Getting in touch: How imagined, actual, and physical contact affect evaluations of robots," in *Proceedings of the Robot and Human Interactive Communication (RO-MAN), 2016 25th IEEE International Symposium on Robot and Human Interactive Communication* (New York, NY: IEEE), 980–985. doi: 10.1109/ROMAN.2016.7745228

Yohanan, S., and MacLean, K. E. (2012). The role of affective touch in human-robot interaction: human intent and expectations in touching the haptic creature. *Int. J. Soc. Robot.* 4, 163–180. doi: 10.1007/s12369-011-0126-7

Yuan, H., and He, B. (2014). Brain–computer interfaces using sensorimotor rhythms: current state and future perspectives. *IEEE Trans. Biomed. Eng.* 61, 1425–1435. doi: 10.1109/TBME.2014.2312397

Zhang, Y., and Ding, M. (2010). Detection of a weak somatosensory stimulus: role of the prestimulus mu rhythm and its top–down modulation. *J. Cogn. Neurosci.* 22, 307–322. doi: 10.1162/jocn.2009.21247

Weibo or WeChat? Assessing Preference for Social Networking Sites and Role of Personality Traits and Psychological Factors

Juan Hou[1†], Yamikani Ndasauka[2,3†], Xuefei Pan[1], Shuangyi Chen[1], Fei Xu[2] and Xiaochu Zhang[2,4,5]*

[1] Department of Philosophy, Anhui University, Hefei, China, [2] School of Humanities and Social Science, University of Science and Technology of China, Hefei, China, [3] Department of Philosophy, Chancellor College, University of Malawi, Zomba, Malawi, [4] Anhui Mental Health Center, Hefei, China, [5] CAS Key Laboratory of Brain Function and Disease, School of Life Science, University of Science and Technology of China, Hefei, China

Research trying to understand individual difference in the use of different social networking sites (SNSs) is minimal. In the present study, we collected data from 714 college students in China (273 males) to assess how personality traits and psychological factors relate to excessive use of WeChat and Weibo. We found that excessive use of Weibo and WeChat correlated positively with neuroticism, loneliness, and external locus of control and negatively with agreeableness, social support, and social interaction. Furthermore, people that scored high on loneliness, lack of social support, and poor social interaction skills excessively used Weibo more than WeChat. These results entail that by fulfilling different needs, WeChat and Weibo attract different kinds of people; significant lesson for future development of SNSs.

Keywords: social networking sites, Weibo, WeChat, preference, personality traits

INTRODUCTION

The internet is an essential component of our life, which influences many aspects of human behavior. Through the internet, people play games, shop, communicate, socialize, search-for, and spread information (Amichai-Hamburger and Ben-Artzi, 2000). Over the years, social networking sites (SNSs) have become the most popular and fastest web sites for information dissemination and private social interaction (Hughes et al., 2012). Popular SNSs such as Facebook and Twitter in US and WeChat and Weibo in China, allow individuals to express themselves, exchange information, and socialize. These sites attract and host millions of users.

An interesting question arises: considering that most SNSs attract millions of users, do these different sites attract different users? Joinson (2004) pointed out that internet behavior is a product of both the user and the specific tool, such that individual differences and personality can influence computer-media choices (Amiel and Sargent, 2004; Ryan and Xenos, 2011; Hou et al., 2014). Just as "Uses and Gratifications Theory" points out that "the audience is conceived as active" (Katz et al., 1974, p. 20). This means that, a user will use specific text to gain the valuable knowledge that they are seeking from a program or text (Katz et al., 1974). Because in mass communication much initiative to link media choice lies with users (Katz et al., 1974). This implies that people use the media to fulfill individual's specific different needs and gratifications.

Therefore, under the "Uses and Gratification Theory," which seeks to understand the use of media from the perspective of the individual and not the media, individual experiences are important factors in determining the choice of SNSs. So, this study seeks to understand if particular individual differences are related to the preference for the excessive use of SNSs. We investigated the role of individual difference in the excessive use of SNSs in this study. This is an exploratory study that aimed to find out the relationship between personality traits/psychological factors and excessive use of Weibo and WeChat. Thus, we assessed how Big-Five personality traits, loneliness, social support, external locus of control, and social interaction relate to the excessive use of two largest SNSs in China- WeChat and Weibo.

WeChat and Weibo

WeChat is one of the most popular SNSs in China and currently boosts of over 600 million active users worldwide (CNNIC, 2015). WeChat is a cross-platform communication application combining popular features of Facebook and WhatsApp (Wu, 2014). It allows users to create a profile, search for friends or find new friends within one's geographical location. On this profile, users can send instant text messages, voice notes and make free voice calls. Further, WeChat allows users to post information, pictures and videos of interest, and comment on friends' posts. All of these features make WeChat popular for online socializing.

Another popular SNS in China is Weibo, created in 2009, and has more than 204 million users (CNNIC, 2015). Weibo allows users to post 140-character information. Similar to Twitter, but unlike WeChat, Weibo focuses on sharing of opinions and information exchange rather than on social interaction (Kwak et al., 2010); and offers some anonymity in online communication (Huberman et al., 2008). Weibo does not need users to post their private information to find "friends" and it focuses less on "who you are" and more on "what you say" (Huberman et al., 2008).

The reduction of social pressure brought about by anonymity (Hughes et al., 2012) may mean different motivation for using Weibo from WeChat. Further, a previous study of interest found that only 2 out of 10 interviewed people had used WeChat to search for new friends (Hou et al., 2017). This entails that most people use WeChat to keep in touch with friends they made in real life. It is hence expected that these differences will be evident in the relationships between individual traits and excessive use of WeChat and Weibo.

Personality

A number of personality traits appear to be associated with the use of SNSs. The Big-Five personality test, developed by Goldberg (1999), is the most commonly used model in investigating the relationship between internet use and individual personality (Landers and Lounsbury, 2006; Ehrenberg et al., 2008). The Big-Five personality test consists of five factors: neuroticism, extraversion, openness, agreeableness, and conscientiousness (McCrae and John, 1992). Several of the factors are associated with problematic use of online social media, such as blogs (Guadagno et al., 2008) and SNSs (Ross et al., 2009; Amichai-Hamburger and Vinitzky, 2010; Correa et al., 2010).

Neuroticism is characterized by anxiety, hostility, depression, self-consciousness, impulsivity, and fragility. Individuals who are low in this trait tend to be more stable and emotionally resilient. Butt and Phillips (2008) found that people with high scores in neuroticism use the Internet frequently, spent more time on Facebook (Ryan and Xenos, 2011) and instant messenger (Correa et al., 2010). Neurotics prefer to use the Internet to relieve loneliness and find a sense of group belonging (Amichai-Hamburger and Ben-Artzi, 2003; Butt and Phillips, 2008). In the current study, we hypothesized that those with high scores in neuroticism will excessively use Weibo more than WeChat (H1).

Extraversion is characterized by excitability, sociability, talkativeness, assertiveness, and high amounts of emotional expressiveness. People who are high in extraversion (also called extroverts) are like to be in touch with people, are full of energy and often feel positive emotions. People who are low in extraversion (also called introverts) are quiet, cautious, and don't like excessive contact with the outside world. Extraverts have been found to be excessive users of instant messaging and SNSs (Correa et al., 2010). They have more friends online (Amichai-Hamburger and Vinitzky, 2010) and tend to make even more friends offline (Ross et al., 2009). Thus, extraverts like socializing, but they don't take online socializing as a substitute for real life social interaction. Thus, extraverts use SNSs mainly for social enhancement. People with only a few offline contacts compensate for their introversion, low self-esteem, and low life-satisfaction by using Facebook for online popularity (Ellison et al., 2007; Barker, 2009; Pollet et al., 2011). In this study, we hypothesized that extraversion will positively correlate more with excessive use of WeChat than Weibo (H2).

Openness has the characteristics of imagination, aesthetics, rich emotions, differences, creativity, intelligence, etc. People who score high in this trait prefer abstract thinking and have a wide range of interests. People who score low in this trait are practical, preferring conventions, and more traditional and conservative. Individuals who score high on openness have been found to excessively use instant messaging and SNSs (Correa et al., 2010). They have wide interests and curiosity (McCrae and Costa, 1987), so they prefer to use Internet for information seeking (McElroy et al., 2007). So, since Weibo can provide more new information than WeChat, in this study, we expected that the excessive use of Weibo would be more in open people than WeChat (H3).

Agreeableness has qualities such as trust, altruism, outspokenness, compliance, modesty, empathy, and so on. People who score high in agreeableness are considerate, friendly, generous, and helpful and willing to give up their own interests for others. Agreeableness has been found to be unrelated to Internet and SNSs' use in many studies (Amichai-Hamburger and Vinitzky, 2010; Correa et al., 2010). However, Ross et al. (2009) pointed out that less agreeable people interact more online and take Internet as a tool to improve social skills and build friendships. La et al. (2009) found that females scoring high on this trait posted significantly more pictures than females scoring low, with the opposite being true for males. In the

current study, we hypothesized that agreeableness will negatively correlate more with excessive use of Weibo than WeChat (H4).

Conscientiousness shows the characteristics of competence, impartiality, coherence, due diligence, achievement, self-discipline, discretion, and restraint. Those who score high on conscientiousness are responsibility, dedication to work, and seriousness. Numerous studies have found significant negative correlation between conscientiousness and SNSs' use time (Amichai-Hamburger and Vinitzky, 2010; Ryan and Xenos, 2011). Butt and Phillips (2008) suggested that conscientious people do not allow SNSs to disrupt their important work. They may prefer Twitter to Facebook because "tweets" are limited to 140 characters, which mean just a temporal distraction for them (Hughes et al., 2012). In addition to this, people with high conscientiousness were found to have significantly more friends and to upload significantly fewer pictures than those scoring low on this personality trait (La et al., 2009). Thus, conscientious people tend to cultivate their online and offline contacts more without the necessity to share too much personal information publicly. We hence hypothesized that conscientiousness will negatively correlate with the excessive use of WeChat (H5), but will not correlate with the excessive use of Weibo.

Ross et al. (2009) argued that the Big Five might be too broad when assessing individual differences in SNS usage. Taking into account the characteristics of SNSs, this study included different psychological factors, namely loneliness, social support, external locus of control, and social interaction.

Loneliness and Social Support

Loneliness is considered as one of the most important predictors of internet addiction (Baumeister et al., 2005; Wang, 2006; Bozoglan et al., 2013). Lonely people usually report less support from their social network in real life (Routasalo et al., 2006). McKenna et al. (2002) found that people who feel lonely are more likely to prefer online social interactions than face-to-face settings (Clerkin et al., 2013; Ye and Lin, 2015). Further, people with low real life social support and high virtual social support tend to draw support from online communication (Yeh et al., 2008).

Admittedly, causal direction of this relation is not clear. There is a two-way relationship between loneliness/social support and SNS use. On the one hand, lonely individuals are attracted to SNS to relieve loneliness or get support; on the other hand, excessive use of SNS has been found to increase sense of loneliness (van den Eijnden et al., 2008). Further, through internet communication, particularly communication with known people, lonely people can increase sense of social support (Shaw and Gant, 2002). Based on socializing features of WeChat, which mainly features people one interacts with in real life and the anonymity of Weibo, we hypothesized that lonely people and those with lack of real life social support will excessively use Weibo more than WeChat (H6, H7).

External Locus of Control and Social Interaction

Locus of control refers to the extent to which individuals believe that they can control things that affect them (Rotter, 1966). People with high external locus of control believe that their

lives are controlled by luck, fate and chance (Ndasauka et al., 2016) and not by their effort or ability. Such people are more likely to engage in problematic internet use (Chak and Leung, 2004), have more online social interaction (Koo, 2009; Ye and Lin, 2015) and have low social interaction skills (Cloitre et al., 1992; Stipek, 1993). With diminished social interaction skills, they prefer communicating through the internet, where they can contact with others without face-to-face interaction (Ndasauka et al., 2016). In the present study, we expected that external locus of control will positively correlate with excessive use of both WeChat and Weibo (H8). We also hypothesized that social interaction in real life will negatively correlate with excessive use of Weibo more than WeChat (H9).

METHODS AND MATERIALS

Participants

Total number of participants was 714; 273 males (38.2%), 441 females (61.8%), and were recruited from 3 college campuses in Anhui province, East China. The mean age was 19.8 years ($SD = 1.3$) ranging from 17 to 21, and 11 participants were under 18 years old.

Ethics Statement

The study was approved by the Human Research Ethics Committee of the University of Science and Technology of China (USTC). All participants gave consent to participate in the study and principles expressed in the Declaration of Helsinki were closely followed. Participants were undergraduate students.

We did not obtain informed consent from the next of kin, caretakers, or guardians on behalf of the minors/children (under 17 years) enrolled in our study. These young college students were considered to have comparable intelligence and ability to adult students, and able to take charge of their behaviors. According to the General principles of the Civil Law of the People's Republic of China; "A minor aged 10 or over shall be a person with limited capacity for civil conduct and may engage in civil activities appropriate to his age and intellect; in other civil activities, he shall be represented by his agent ad litem or participate with the consent of his agent ad litem" (Article 12, Chapter II). Therefore, we obtained the same consent from these participants between 17 and 18 as those above 18 years old, which was also approved by the Human Research Ethics Committee of the University of Science and Technology of China (USTC). Small gifts (keychain and nail-cutter of not more than $1.5) were given as incentive to participate in the study. Written informed consent was obtained from all participants.

Measures

Demographic Data

Participants answered two questions regarding their gender and age. We also asked participants to rate themselves with regard to their preference of WeChat and Weibo. They were asked to choose one between Weibo and WeChat by stating which one they preferred.

WeChat Excessive Use Scale (WEUS)

The scale was developed to assess excessive use of WeChat (Hou et al., 2017). It includes items such as "I check my WeChat before something else that I need to do," "I have used WeChat to relieve of loneliness and stress," and "There are times when I would rather play on WeChat than go out with my friends." The 10-item scale showed good internal consistency in the initial study, with Cronbach's alpha of 0.907. In the current study, the Cronbach's alpha was 0.899. The scale is scored on a Likert-type scale ranging from 1 (*never*) to 5 (*always*).

Microblog Excessive Use Scale (MEUS)

The scale was developed to measure excessive use of Weibo (Hou et al., 2014). It includes items such as "How often do you find yourself saying "just a few more minutes" when using microblogs?," "How often would you try to increase your followers unconsciously by all means?," and "How often do you feel depressed, moody, or nervous when you are off microblogs?" MEUS has 10 items rated on a 6 point Likert scale from "1 = *never*" to "6 = *always*." In the current study the scale showed good internal consistency with Cronbach's alpha of 0.908.

Big Five Personality Questionnaire

We used a 60-item personality questionnaire developed by Leung (2011) to assess five different personalities, namely, Neuroticism, Extraversion, Openness, Agreeableness, and Conscientiousness. Each factor featured 12 items rated on 5-point scale ranging from *strongly disagree* to *strongly agree* (McCrae and Costa, 1991). Four personalities factors of Neuroticism ($\alpha = 0.708$), Extraversion ($\alpha = 0.648$), Agreeableness ($\alpha = 0.573$), and Conscientiousness ($\alpha = 0.697$) showed adequate to good reliability and internal consistency in our sample. However, Openness factor showed poor reliability and internal consistency ($\alpha = 0.371$). As such, this personality factor was not included in our analyses.

UCLA Loneliness Scale

This scale, originally developed in 1978 by Russell, Peplau, and Ferguson, has 20 items rated on a 4-point scale from "1 = *never*" to "4 = *often*." In the current study, the Cronbach's alpha was 0.760.

Social Support Scale

We used the Interpersonal Support/Social Support Scale; a 12-item measure of perceptions of social support. This measure is a short version of the original ISEL (40 items; Cohen and Hoberman, 1983). Items are rated on a 4-point scale ranging from "*definitely true*" to "*definitely false*." In the current study, the Cronbach's alpha was 0.770.

Locus of Control Scale (LOC)

We used the multidimensional locus of control scale developed by Levenson (1981). The scale has three dimensions: Internal scale, which measures internal locus of control; Powerful Others scale and Chance scale, which measure external locus of control. In this study, we utilized the latter two scales to measure external locus of control. The scale had 16 items and were scored on a 6-point Likert scale from "1 = *strongly disagree*" to "6 = *strongly*

agree." The scale showed adequate internal consistency in our study ($\alpha = 0.738$).

Social Interaction Scale (SIS)

We employed the Social Interaction Scale developed by Yan (2011). The Social Interaction Scale contains 24 questions divided into two parts namely- Real Life Scale (14 items) and Online Scale (10 items). In our study, we only used the Real Life Scale to measure interpersonal communication with classmates, friends, parents, and other people in real life. The items are rated on a 4-point Likert scale from "1 = *never*" to "4 = *always*." In our study, the Real Life scale showed good internal consistency and reliability ($\alpha = 0.808$).

Data Analysis

We analyzed the data using the statistics software package SPSS 23.0. We calculated correlations between the scales using Pearson's r (Pearson product-moment correlation coefficient). To test significant differences of different levels in total scores, we used the Kruskal–Wallis test (Kruskal–Wallis one-way analysis of variance) and t-tests. We also used Scheffe's *post-hoc* tests to analyze significant differences between different levels. Further, we used a method by Lee and Preacher (2013) to calculate the difference in correlations between WEUS ↔ personality traits and MEUS ↔ personality traits. The significance level in this study was $p \leq 0.05$.

RESULTS

Weibo vs. WeChat

With regard to preference, 386 participants (females = 261) reported that they preferred Weibo to WeChat, while 328 participants (females = 180) reported that they preferred WeChat to Weibo. We used the chi-square test to calculate the gender proportion, and there was significant difference between the usage proportion of Weibo and WeChat ($\chi^2 = 12.184$, $p < 0.001$), with more females preferring Weibo to WeChat.

We analyzed the correlation between MEUS and WEUS. Results showed significant positive correlation between the two variables ($r = 0.462$, $p < 0.001$).

Weibo, WeChat, Conscientiousness, Extraversion, Neuroticism, and Agreeableness

We analyzed the correlation between MEUS and the four personality traits (see **Table 1**). MEUS positively correlated with Neuroticism ($r = 0.173$, $p < 0.001$), negatively with Agreeableness ($r = -0.234$, $p < 0.001$), and Conscientiousness ($r = -0.083$, $p = 0.026$), but did not significantly correlate with Extraversion ($r = 0.007$, $p = 0.852$).

We then analyzed the correlation between WEUS and the four personality traits (see **Table 1**). WEUS positively correlated with Neuroticism ($r = 0.118$, $p < 0.001$), negatively with Agreeableness ($r = -0.153$, $p < 0.001$) but did not significantly correlate with Conscientiousness ($r = -0.025$, $p = 0.503$) and Extraversion ($r = 0.057$, $p = 0.128$). For the significant correlations, we also draw the scatter plot on the relationship

	Neuroticism	Extraversion	Agreeableness	Conscientiousness	Loneliness	Social support	Social interaction skills in real life	External locus of control
MEUS	0.173***	0.007	−0.234***	−0.083*	0.139***	−0.111**	−0.084*	0.179***
WEUS	0.118**	0.057	−0.153***	−0.025	0.025	−0.048	−0.008	0.208***

*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$.

between MEUS/WEUS and the four personality traits (see **Figure 1**).

Further, we found significant difference between correlations of MEUS ↔ Agreeableness and WEUS ↔ Agreeableness ($z = −2.14$, $p = 0.033$). These results were consistent with H4 and H1, H2, H3, H5 were not supported.

Weibo, WeChat, Loneliness, and Social Support

Our results showed that loneliness positively correlated with MEUS ($r = 0.139$, $p < 0.001$) but did not significantly correlate with WEUS ($r = 0.025$, $p = 0.508$; see **Table 1**). We analyzed the difference between these two correlations and found that correlation between loneliness and MEUS was significantly higher than correlation between loneliness and WEUS ($z = 2.931$, $p = 0.003$).

Further, we found that social support negatively correlated with MEUS ($r = −0.111$, $p = 0.003$), but did not significantly correlate with WEUS ($r = −0.048$, $p = 0.200$; see **Table 1**). For the significant correlations, we also draw the scatter plot on the relationship between MEUS/WEUS and loneliness/social support (see **Figure 1**).

We then analyzed the difference between the two correlations and found that the two correlations were not significantly different ($z = 1.627$, $p = 0.104$). These results were consistent with H6 and H7 is not supported.

Weibo, WeChat, External Locus of Control, and Social Interaction in Real Life

We analyzed the correlation between MEUS and external locus of control/real life social interaction (see **Table 1**). MEUS correlated positively with external locus of control ($r = 0.178$, $p < 0.001$), and negatively with real life social interaction ($r = −0.084$, $p = 0.024$).

We then analyzed the correlation between WEUS and the two variables (external locus of control and real life social interaction; see **Table 1**). WEUS correlated positively with external locus of control ($r = 0.208$, $p < 0.001$), but did not significantly correlate with real life social interaction ($r = −0.008$, $p = 0.822$). For the significant correlations, we also draw the scatter plot on the relationship between MEUS/WEUS and external locus of control/social interaction in real life (see **Figure 1**).

While the correlation of MEUS and external locus of control did not significantly differ with correlation of WEUS and external locus of control ($z = 0.79$, $p = 0.429$), correlation of MEUS and real life social interaction was significantly higher than correlation of WEUS and real life social interaction ($z = 2.243$,

$p = 0.025$). These results were consistent with H9 and H8 is not supported.

Correcting for Multiple Comparisons

Applying the Bonferroni correction, we divided $p = 0.05$ by the number of tests (10) to get the Bonferroni critical value. This resulted in change of p-value to $p < 0.005$ to be significant. Under that criterion, all tests were found to be significant except for correlation between MEUS and Conscientiousness whose $p = 0.026$ and MEUS and real life social interaction whose $p = 0.024$.

DISCUSSION

The aim of the current study was to investigate some of the individual differences associated with the use of the two largest SNSs in China- Weibo and WeChat. We found that different personality traits were influential in explaining the excessive use of the two SNSs, and some correlations between these traits and Weibo and WeChat were also significantly different.

Weibo vs. WeChat

From the item- "which one do you prefer, Weibo or WeChat?" results showed that more females preferred Weibo to WeChat than males. This is consistent with previous studies that females scored significantly higher on the Weibo scale than males (Hou et al., 2014), and women are more likely to use Twitter than men (Smith and Rainie, 2010). By broadcasting to everyone, Weibo allows young female students to express themselves to and seek attention from a larger audience than they would on WeChat, which is limited in its broadcasting and sharing of moments.

Also, we found that Weibo and WeChat had significant moderate correlation with each other. These results may mean that users of one also tend to use another. However, the results may also entail that there is some difference between Weibo and WeChat.

The Use of Weibo and WeChat
Personality Differences in the Use of Weibo and WeChat

Neuroticism positively correlated with excessive use of both Weibo and WeChat. This result is consistence with previous studies that neurotic people are more likely to use SNSs for socializing (Amichai-Hamburger and Ben-Artzi, 2000; Butt and Phillips, 2008). However, contrary to our hypothesis, there was no significant difference between correlations of neuroticism ↔ use of Weibo and neuroticism ↔ use of WeChat. These results imply that neurotic people use both two SNSs in similar manner. One factor that may help explain this is Weibo's anonymity
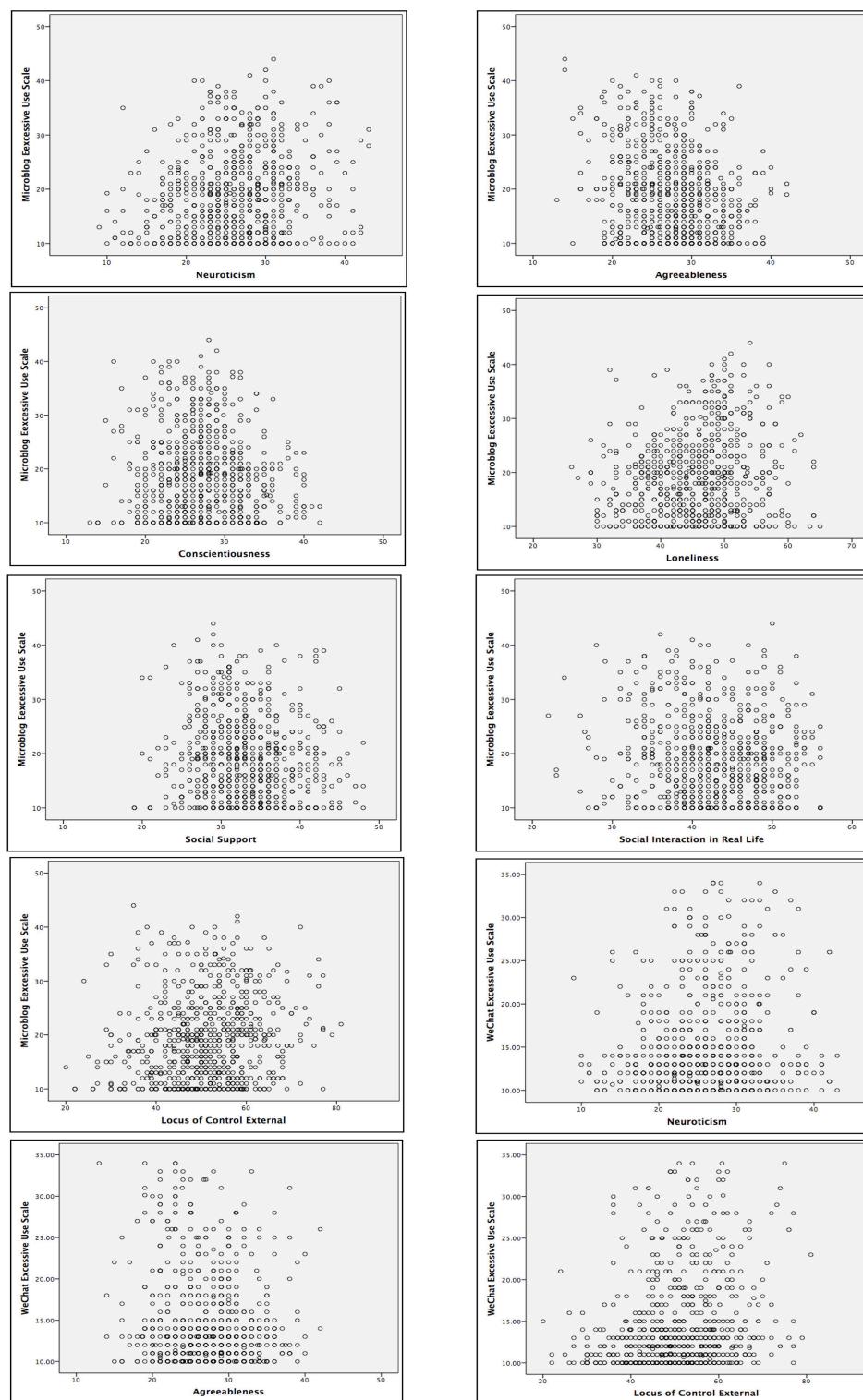
FIGURE 1 | Scatter plot of the correlation between MEUS/WEUS and personality traits, loneliness/social support and social interaction skills in real life/ external locus of control.

and other anonymity features of WeChat, like "Shake" and "drift bottle." This allows neurotics to interact with people online, because these forms of interaction do not require face-to-face contact. As Amichai-Hamburger et al. (2002, p. 127–128) reported: "It would appear that the social services provided on the Internet, with their anonymity, lack of need to reveal physical appearance, rigid control of information revealed in the interaction… provide an excellent answer to people who experience great difficulty in forming social contacts due to their introverted personality."

Agreeableness negatively correlated with both Weibo and WeChat. This result also is supported by Ross et al. (2009) that less agreeable people are inclined to use SNSs more often and sometimes in an excessive way. Further, consistent with our hypothesis, we found significant difference between correlations of agreeableness ↔ use of Weibo and agreeableness ↔ use of WeChat. Thus, less agreeable people are more likely to excessively use Weibo than WeChat. Since less agreeable people are considered less friendly, they may find Weibo a better alternative to fulfill their social needs than WeChat. Although they use WeChat, Weibo provides them a wider scope of social fulfillment because they can interact with many people with whom they are not friends in real life.

Conscientiousness negatively correlated with the use of Weibo in our study, but did not significantly correlate with WeChat. These results are a direct contrast to our hypotheses. This may be due to the other attributes of both Weibo and WeChat. Weibo is a half-open platform, and for highly conscientious people, it may increase sense of insecurity, which may not be the case with WeChat because they may feel in control of their friends' circle and hence feel secure to post pictures and socialize without fear. A previous study of interest found that despite having more friends than those scoring low in the trait, conscientious people tend to upload significantly fewer pictures on Weibo (Amichai-Hamburger and Vinitzky, 2010).

Loneliness and Social Support Differences in the Use of Weibo and WeChat

Loneliness positively correlated with the use of Weibo while social support negatively correlated with the use of Weibo. Surprisingly however, the two factors did not significantly correlate with WeChat. Furthermore, we found significant difference between correlations of the two psychological factors ↔ use of Weibo and the psychological factors ↔ use of WeChat. The results suggest that individuals who are lonely and people who lack social support in real life tend to use Weibo more than WeChat. Thus, because Weibo is half-open platform, people can engage in identity experimentation, which brings more gratification to lonely people than those or not lonely (Leung, 2011). Further, the open and anonymity attributes of Weibo (Hughes et al., 2012) offer people an opportunity to jump out of the real-life circle of friends. Through Weibo, lonely people can make new friends, seek novelty and information of interest. This is unlike WeChat; whose posts are mainly from relatively the same people one interacts with in real life. As such, WeChat may bring the same isolated feeling for people that lack social support and feel lonely.

External Locus of Control and Social Interaction in Real Life Differences in the Use of Weibo and WeChat

Our results showed that excessive use of both Weibo and WeChat is associated with external locus of control. Thus, people who have faith in that environment causes their life events, excessively use Weibo and WeChat. These findings are similar to results of previous studies (Karatas and Tagay, 2012; Ndasauka et al., 2016). Thus, externals are lonelier (Hojat, 1982) and feel less confident in control of their lives and behaviors (Ye and Lin, 2015) hence they excessively use WeChat and Weibo.

Further, we also found negative correlation between Weibo and social interaction in real life. However, we found no significant difference between correlations of problematic use of Weibo ↔ social interaction in real life and the use of WeChat ↔ social interaction in real life. This entails that people who spend less time in socializing in real life or lack social skills in real life tend to choose online social interaction on Weibo. Compared to WeChat, Weibo provides them a bigger "circle of friends," where they can practice social skills with strangers. In addition, the reduction of social pressure on Weibo (Hughes et al., 2012) may be the one of the factors that attracts people with poor real life social skills.

CONCLUSION AND LIMITATIONS

Overall, the study investigated how individual difference are associated with the use of Weibo and WeChat. Results demonstrate that personality traits are linked to excessive use of Weibo and WeChat. We found that personality such as neuroticism, loneliness, and external locus of control had positive correlations with excessive use of Weibo and WeChat, while agreeableness, social support, and social interaction negatively correlated to excessive use of Weibo and WeChat. Furthermore, we compared the difference between correlations of the personal traits and excessive use of Weibo and WeChat, and we found that lonely people, people that lack social support and those with poor social interaction skills tend to excessively use Weibo more than WeChat. These results are pertinent because they entail that people who experience loneliness or social frustration in real life choose sites that impose less social pressure to relieve loneliness and maybe gain confidence for real life social interaction.

The study had some limitations that merit consideration. Firstly, the selected respondents were only from Eastern China, and the sample representation may not be completely correct. Secondly, the study exclusively used self-report questionnaires in data collection. Due to social desirability or understanding problem, self-reporting sometimes affects reliability and validity of the answers. Therefore, in future research, we need to combine a variety of research methods to draw a complete picture of the use of Weibo and WeChat in China.

RELEVANCE AND CONTRIBUTION OF STUDY

Although WeChat is considered as a more socializing platform than Weibo, which is viewed as platform for sharing information,

our study has shown that people scoring high on loneliness and neuroticism are more likely to use Weibo than WeChat. The results of this study add to the uses and gratification theory, which states that people engage in some activity, in this case social networking, to meet certain psychological needs. As such, by meeting different needs of people, WeChat and Weibo tend to attract different kinds of people.

People using WeChat tend to mainly transfer their offline social interaction with friends to online environment. For people lacking social support and social interaction skills, transferring to WeChat does not often meet their social needs in real life. So, when dealing with people experiencing social problems, it is more meaningful to provide them with a broader social space and open platform than to let them practice their social skills with known friends. As such in helping people who are struggling with excessive use of Weibo, WeChat, or other social network applications, it is important to focus on improving their social skills, and reducing their social pressure and loneliness.

Further, medical practitioners dealing with people struggling with excessive or addictive use of SNSs should pay attention to the particular sites in which their patients are overusing. This is because, as shown in this study, different people are attracted and motivated to use different SNSs in that those with some psychological traits are more likely to use one more than another

or others. Finally, these results should encourage SNSs developers to rethink the function of their sites. When developing social networking platforms, developers should consider how best to meet psychological needs of different groups.

AUTHOR CONTRIBUTIONS

JH, YN, XZ, and FX: Conceived and designed the experiments; JH, YN, XP, and SC: Performed the experiments; JH, YN, XP, and SC: Analyzed the data; JH, YN, XP, and SC: Contributed reagents, materials, analysis tools; JH, YN, and XZ: Wrote the paper; JH, YN, XZ, and FX: Discussed the result; JH, YN, XZ, and FX: Final approval of the version to be published.

ACKNOWLEDGMENTS

REFERENCES

Amichai-Hamburger, Y., and Ben-Artzi, E. (2000). The relationship between extraversion and neuroticism and the different uses of the internet. *Comput. Hum. Behav.* 16, 441–449. doi: 10.1016/S0747-5632(00)00017-0

Amichai-Hamburger, Y., and Ben-Artzi, E. (2003). Loneliness and Internet use. *Comput. Hum. Behav.* 19, 71–80. doi: 10.1016/S0747-5632(02)00014-6

Amichai-Hamburger, Y., and Vinitzky, G. (2010). Social network use and personality. *Comput. Hum. Behav.* 26, 1289–1295. doi: 10.1016/j.chb.2010.03.018

Amichai-Hamburger, Y., Wainapel, G., and Fox, S., (2002). On the Internet no one knows i'm an introvert: extroversion, neuroticism, and internet interaction. *Cyber Psychol. Behav.* 5, 125–128. doi: 10.1089/109493102753770507

Amiel, T., and Sargent, S. L. (2004). Individual differences in Internet usage motives. *Comput. Hum. Behav.* 20, 711–726. doi: 10.1016/j.chb.2004.09.002

Bozoglan, B., Demirer, V., and Sahin, I. (2013). Loneliness, self-esteem, and life satisfaction as predictors of internet addiction: a cross-sectional study among turkish university students. *Scand. J. Psychol.* 54, 313–319. doi: 10.1111/sjop.12049

Barker, V. (2009). Older adolescents5 motivations for social network site use: the influence of gender, group identity, and collective self-esteem. *Cyberpsychol. Behav.* 12, 209–213. doi: 10.1089/cpb.2008.0228

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

Butt, S., and Phillips, J. (2008). Personality and self-reported mobile phone use. *Comput. Hum. Behav.* 24, 346–360. doi: 10.1016/j.chb.2007.01.019

Chak, K., and Leung, L. (2004). Shyness and locus of control as predictors of internet addiction and internet use. *Cyberpsychol. Behav.* 7, 559–570. doi: 10.1089/cpb.2004.7.559

Clerkin, E. M., Smith, A. R., and Hames, J. L. (2013). The interpersonal effects of facebook reassurance seeking. *J. Affect. Disord.* 151, 525–530. doi: 10.1016/j.jad.2013.06.038

Cloitre, M., Heimberg, R. G., Liebowitz, M. R., and Gitow, A. (1992). Perceptions of control in panic disorder and social phobia. *Cogn. Ther. Res.* 16, 569–577. doi: 10.1007/BF01175142

CNNIC (2015). *China Internet Network Information Center The Thirty-Sixth Statistics Report of China Internet Development.* Beijing: China Internet.

Cohen, S., and Hoberman, H. M. (1983). Positive events and social supports as buffers of life change stress. *J. Appl. Soc. Psychol.* 13, 99–125. doi: 10.1111/j.1559-1816.1983.tb02325.x

Correa, T., Hinsley, A., and de Zúñiga, H. (2010). Who interacts on the web?: the intersection of users' personality and social media use. *Comput. Hum. Behav.* 26, 247–253. doi: 10.1016/j.chb.2009.09.003

Ehrenberg, A., Juckes, S., White, K., and Walsh, S. (2008). Personality and self-esteem as predictors of young people's technology use. *Cyberpsychol. Behav.* 11, 739–741. doi: 10.1089/cpb.2008.0030

Ellison, N. B., Steinfield, C., and Lampe, C. (2007). The benefits of Facebook "friends": social capital and college students! use of online social network sites. *J. Comput. Mediat. Commun.* 12, 1143–1168. doi: 10.1111/j.1083-6101.2007.00367.x

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.

Guadagno, R., Okdie, B., and Eno, C. (2008). Who blogs? Personality predictors of blogging. *Comput. Hum. Behav.* 24, 1993–2004. doi: 10.1016/j.chb.2007.09.001

Hojat, M. (1982). Psychometric characteristics of the ucla loneliness scale: a study with iranian college students. *Educ. Psychol. Meas.* 42, 917–925. doi: 10.1177/001316448204200328

Hou, J., Huang, Z., Li, H., Liu, M., Zhang, W., Ma, N., et al. (2014). Is the excessive use of microblogs an internet addiction? Developing a scale for assessing the excessive use of microblogs in Chinese college students. *PLoS ONE* 9:e110960. doi: 10.1371/journal.pone.0110960

Hou, J., Ndasauka, Y., Jiang, Y., Ye, Z., Wang, Y., and Yang, L., et al. (2017). Excessive use of wechat, social interaction and locus of control among college students in china. *PLoS ONE* 12:e0183633. doi: 10.1371/journal.pone.0183633

Huberman, B. A., Romero, D. M., and Wu, F. (2008). Social networks that matter: twitter under the microscope. *First Mondy* 14:2009. doi: 10.2139/ssrn.1313405

Hughes, D. J., Rowe, M., Batey, M., and Lee, A. (2012). A tale of two sites: twitter vs. facebook and the personality predictors of social media usage. *Comput. Hum. Behav.* 28, 561–569. doi: 10.1016/j.chb.2011.11.001

Joinson, A. N. (2004). Self-esteem, interpersonal risk, and preference for e-mail to face-to-face communication. *Cyberpsychol. Behav.* 7, 472–478. doi: 10.1089/cpb.2004.7.472

Karatas, Z., and Tagay, O. (2012). Self-esteem, locus of control and multidimensional perfectionism as the predictors of subjective well-being. *Int. Educ. Stud.* 5, 141–145. doi: 10.5539/ies.v5n6p131

Katz, E., Blumler, J. G., and Gurevitch, M. (1974). "Utilization of mass communication by the individual," in *The Uses of Mass Communications: Current Perspectives on Gratifications Research*, eds J. G. Blumler and E. Katz (Beverly Hills, CA: Sage Publications, Inc.), 19–32.

Koo, D. M. (2009). The moderating role of locus of control on the links between experiential motives and intention to play online games. *Comput. Hum. Behav.* 25, 466–474. doi: 10.1016/j.chb.2008.10.010

Kwak, H., Lee, C., Park, H., and Moon, S. (2010). "What is twitter, a social network or a news media?" in *Proceedings of the 19th International Conference on World Wide Web* (Raleigh), 591–600.

La, B. D., La, P. F., and Valsavoia, R. (2009). Social network and addiction. *Stud. Health Technol. Inf.* 144, 33–36.

Landers, R. N., and Lounsbury, J. W. (2006). An investigation of big five and narrow personality traits in relation to internet usage. *Comput. Hum. Behav.* 22, 283–293. doi: 10.1016/j.chb.2004.06.001

Lee, I. A., and Preacher, K. J. (2013). *Calculation for the Test of the Difference Between Two Dependent Correlations With One Variable in Common [Computer Software]*. Available online at http://quantpsy.org

Leung, L. (2011). Loneliness, social support, and preference for online social interaction: the mediating effects of identity experimentation online among children and adolescents. *Chin. J. Commun.* 4, 381–399. doi: 10.1080/17544750.2011.616285

Levenson, H. (1981). "Differentiating among internality, powerful others, and chance," in *Research with the Locus of Control Construct*, ed H. Lefcourt (New York, NY: Academic Press), 15–63.

McCrae, R. R., and Costa, P. T. Jr. (1987). Validation of the five-factor model of personality across instruments and observers. *J. Pers. Soc. Psychol.* 52, 81–90. doi: 10.1037/0022-3514.52.1.81

McCrae, R. R., and Costa, P. T., Jr. (1991). Adding liebe und arbeit: The full five-factor model and well-being. *Personal. Soc. Psychol. Bull.* 17, 227–232.

McCrae, R. R., and John, O. P. (1992). An introduction to the five-factor model and its applications. *J. Personal.* 113, 630–633. doi: 10.1111/j.1467-6494.1992.tb00970.x

McElroy, J. C., Hendrickson, A. R., Townsend, A. M., and Demarie, S. M. (2007). Dispositional factors in internet use: personality versus cognitive style. *Manag. Inf. Syst. Quart.* 31, 809–820. doi: 10.2307/25148821

McKenna, K. Y., Green, A. S., and Gleason, M. E. (2002). Relationship formation on the Internet: what's the big attraction? *J. Soc. Issues* 58, 9–31. doi: 10.1111/1540-4560.00246

Ndasauka, Y., Hou, J., Wang, Y., Yang, L., Yang, Z., Ye, Z., et al. (2016). Excessive use of Twitter among college students in the UK: validation of the microblog excessive use scale and relationship to social interaction and loneliness. *Comput. Hum. Behav.* 55, 963–971. doi: 10.1016/j.chb.2015.10.020

Pollet, T. V., Roberts, S. G., and Dunbar, R. I. (2011). Use of social network sites and instant messaging does not lead to increased offline social network size, or to emotionally closer relationships with offline network members. *Cyberpsychol. Behav. Soc. Netw.* 14, 253–258. doi: 10.1089/cyber.2010.0161

Ross, C., Orr, E. S., Sisic, M., Arseneault, J. M., Simmering, M. G., and Orr, R. R. (2009). Personality and motivations associated with Facebook use. *Comput. Hum. Behav.* 25, 578–586. doi: 10.1016/j.chb.2008.12.024

Rotter, J. B. (1966). Generalizad expectancies for internal versus external control of reinforcement. *Psychol. Monogr.* 80, 1–28. doi: 10.1037/h0092976

Routasalo, P. E., Savikko, N., Tilvis, R. S., Strandberg, T. E., and Pitkälä, K. H. (2006). Social contacts and their relationship to loneliness among aged people–a population-based study. *Gerontology* 52, 181–187. doi: 10.1159/000091828

Ryan, T., and Xenos, S. (2011). Who uses Facebook? An investigation into the relationship between the big five, shyness, narcissism, loneliness, and Facebook usage. *Comput. Hum. Behav.* 27, 1658–1664. doi: 10.1016/j.chb.2011.02.004

Shaw, L. H., and Gant, L. M. (2002). In defense of the internet: the relationship between internet communication and depression, loneliness, self-esteem, and perceived social support. *CyberPsychol. Behav.* 5, 157–171. doi: 10.1089/109493102753770552

Smith, A., and Rainie, L. (2010). *8% of online Americans use Twitter*. Subreport of research from Pew Internet & American Life Project (misc).

Stipek, D. J. (1993). *Motivation to Learn: From Theory to Practice*. Englewood Cliffs, NJ: Prentice-Hall.

van den Eijnden, R. J., Meerkerk, G. J., Vermulst, A. A., Spijkerman, R., and Engels, R. C. (2008). Online communication, compulsive Internet use, and psychosocial well-being among adolescents: a longitudinal study. *Dev. Psychol.* 44, 655–665. doi: 10.1037/0012-1649.44.3.655

Wang, B. (2006). A study on the relationship between the loneliness and internet addiction tendency of college students. *Psychol. Sci.* 79, 427–436.

Wu, H. Y. (2014). *Analysis on the Actuality and Impact of WeChat Use among College Students*. Unpublished bachelor's thesis, Ningxia Medical University, Yinchuan, China.

Yan, M. Y. (2011). *The Research of Interpersonal and Internet Addiction in the Undergraduate Students*. Unpublished Master's thesis, Anhui Medical University, Hefei.

Yeh, Y. C., Ko, H. C., Wu, J. Y. W., and Cheng, C. P. (2008). Gender differences in relationships of actual and virtual social support to internet addiction mediated through depressive symptoms among college students in Taiwan. *Cyberpsychol. Behav.* 11, 485–487. doi: 10.1089/cpb.2007.0134

Ye, Y. H., and Lin, L. (2015). Examining relations between locus of control, loneliness, subjective well-being, and preference for online social interaction. *Psychol. Rep.* 116, 164–175. doi: 10.2466/07.09.PR0.116k14w3

Social Fear Conditioning Paradigm in Virtual Reality: Social vs. Electrical Aversive Conditioning

Jonas Reichenberger, Sonja Porsch, Jasmin Wittmann, Verena Zimmermann and Youssef Shiban*

Department of Clinical Psychology and Psychotherapy, Institute of Psychology, University of Regensburg, Regensburg, Germany

In a previous study we could show that social fear can be induced and extinguished using virtual reality (VR). In the present study, we aimed to investigate the belongingness effect in an operant social fear conditioning (SFC) paradigm which consisted of an acquisition and an extinction phase. Forty-three participants used a joystick to approach different virtual male agents that served as conditioned stimuli. Participants were randomly allocated to one of two experimental conditions. In the electroshock condition, the unconditioned stimulus (US) used during acquisition was an electric stimulation. In the social threat condition, the US consisted of an offense: a spit in the face, mimicked by a sound and a weak air blast to the participant's neck combined with an insult. In both groups the US was presented when participants were close to the agent (75% contingency for CS+). Outcome variables included subjective, psychophysiological and behavioral data. As expected, fear and contingency ratings increased significantly during acquisition and the differentiation between CS+ and CS− vanished during extinction. Furthermore, a clear difference in skin conductance between CS+ and CS− at the beginning of the acquisition indicated that SFC had been successful. However, a fast habituation to the US was found toward the end of the acquisition phase for the physiological response. Furthermore, participants showed avoidance behavior toward CS+ in both conditions. The results show that social fear can successfully be induced and extinguished in VR in a human sample. Thus, our paradigm can help to gain insight into learning and unlearning of social fear. Regarding the belongingness effect, the social threat condition benefits from a better differentiation between the aversive and the non-aversive stimuli. As next step we suggest comparing social-phobic patients to healthy controls in order to investigate possible differences in discrimination learning and to foster the development of more efficient treatments for social phobia.

Keywords: social fear conditioning, virtual reality, fear-potentiated startle, skin conductance level, avoidance behavior

INTRODUCTION

Social anxiety disorder (SAD) is one of the most relevant anxiety disorders. It is characterized by intense anxiety when faced with social interactions along with physical symptoms like blushing or trembling, and extreme avoidance behavior concerning social interaction (Fehm et al., 2005; Kessler et al., 2005; American Psychiatric Association, 2013). While learning models are relatively

well established in specific phobia, PTSD and panic disorders, learning paradigms for SAD are far less developed, both in animal models and in humans. Besides the diathesis stress model, there is evidence showing that fear conditioning may play an essential role in the development and maintenance of SAD (Mineka and Zinbarg, 2006; Mineka and Oehlberg, 2008).

Cognitive-behavioral therapy is the method of choice for the treatment of SAD; it is widely supported by current research and therefore assumed to be a reliable approach for overcoming anxiety (Arch et al., 2012). Cognitive-behavioral therapy is also often combined with exposure to feared situations in order to maximize the therapeutic success (Wolitzky-Taylor et al., 2008). Nevertheless, the effectiveness of this treatment approach is not always satisfactory and a high number of non-responders remain (Norton and Price, 2007).

Empirical findings show that conditioning mechanisms play an important role in the etiology of the elementary processes of SAD, making them essential to examine in order to maximize the impact of psychotherapeutic interventions (Mineka and Zinbarg, 2006; Mineka and Oehlberg, 2008). Classical fear conditioning (according to Pavlov) is a form of associative learning in which an organism learns to associate two stimuli with each other (Pavlov, 1927). E.g., hearing someone laugh (unconditioned stimulus: US) while giving a speech may result in the speaker showing a fear response (unconditioned response: UR). As a result, the previously neutral stimulus (giving a speech), now called conditioned stimulus (CS), triggers the newly learned fear reaction (conditioned response: CR). Classical fear conditioning is considered a central pathogenic pathway in anxiety disorders (Lissek et al., 2005; Mineka and Zinbarg, 2006; LeDoux, 2014). Operant fear conditioning (learning by consequences) may be also be relevant for the development of anxiety disorders, because it relates to stimuli that reinforced or punished the person during approach behavior. E.g., if voluntarily presenting a paper is followed by the lecturer harshly criticizing the presentation, a student might no longer report voluntarily in the future. Thus, other persons and social interactions might be prototypical stimuli involved in operant learning processes. However, until now little research has been conducted on operant fear conditioning in SAD.

Fear conditioning in mice in social as well as non-social contexts is addressed in the social fear conditioning (SFC) approach investigated by Toth et al. (2012). In this paradigm, naturally occurring preference behavior of male rodents toward an unknown conspecific was paired with an aversive US, namely an electric stimulus applied to the foot for 1 s. During acquisition phase, the rodents learned to associate the appearance of the negative stimulus with the conspecific, which induced social fear including avoidance behavior. In a following extinction phase on the next day, different male conspecifics were presented to the experimental animals in their cage without any negative US. It could be observed that avoidance and fear-driven behavior were extinguished and replaced by the naturally occurring preference behavior again. Therefore, in the course of the experiment, acquisition and extinction of fear were demonstrated. These results suggest that, using the applied paradigm, it is possible to draw conclusions about the etiology of SAD and potential

leverage points for future treatment approaches (Toth et al., 2012; Toth and Neumann, 2013; Zoicas et al., 2014).

Many uncontrollable contextual and environmental factors can play a role and therefore turn out to be confounding variables in human experimental as well as therapeutic settings. A way to circumvent this problem is conducting experiments in virtual reality (VR), which also allows for the creation of paradigms of SAD development and the exploration of potential treatment improvements. The use of an artificially designed virtual environment minimizes potentially confounding variables by presenting standardized situations to participants. Subjects are able to interact with their environment and diverse stimuli can be applied in a multimodal manner (Bohil et al., 2011). Furthermore, it is possible to directly record the participant's reactions to the stimuli in the form of verbal ratings, fear-potentiated startle or electrocardiographic data (Mühlberger et al., 2007). Thus, VR allows conducting SFC related experiments in a realistic, standardized environment in an economic and easily administrable manner. An additional advantage of VR that is particularly important in the treatment of SAD is the prevention of avoidance behavior, which often leads to the reinforcement of anxiety symptoms (American Psychiatric Association, 2013). In general, the results of conditioning processes in VR are hugely satisfying (Huff et al., 2010).

Shiban et al. (2015) implemented a procedure similar to the SFC paradigm designed for mice by Toth et al. (2012) in order to investigate SFC in humans in VR. In this experimental setting, participants had to actively approach different agents in VR using a joystick. During the acquisition phase, one of the agents, referred to as CS+, was paired with an US, a loud female scream combined with an air blast. During the extinction and the following generalization test phase, no US was administered. In line with the initial hypotheses, participants rated the CS+ as significantly less pleasant than the CS− after the acquisition phase. These results were also supported by the heart rate pattern, as the heart rate was higher for the CS+ than for the CS− after acquisition. After the extinction phase, the ratings returned to an equal level and the fear-potentiated startle response decreased. Interestingly, during the generalization test, the more socially fearful participants rated every agent as less pleasant, compared to the less socially fearful participants who only rated the CS+ as less pleasant. This indicates that more socially fearful participants tend to generalize the unpleasantness of social stimuli to a broader context. In sum, SFC could be induced and extinguished successfully, thus emphasizing the role of operant conditioning in social fear learning. Nonetheless, the study has some limitations, which could be addressed in order to potentially improve the paradigm.

For instance, it is possible to manipulate the intensity of the social contact between the agent and the participant to investigate the specificity of the paradigm for social situations. We believe that our paradigm provides the opportunity for basic social interaction between the agent and the participant (via eye contact, self-regulated movement of the avatar and movement toward the agent). In the current study we improved upon this aspect by designing a social threat condition and comparing it to a conventional electroshock condition. Furthermore, it

could be criticized that the amount of social interaction in the preliminary study was quite low, as the agent did not directly communicate with the participant. This has been taken into account in the current study, as in the social threat condition the agent verbally insults and spits at the participant is much more ecologically valid than the mere administration of an air blast or an electrical stimulation. We assume that it enables us to better use the paradigm for social fear research. In addition, the facial expressions of the agents were adjusted to the verbal utterance in order to create a more realistic and therefore more threatening experience. This also provided the opportunity to investigate the belongingness effect, since the accordance between the US and the CS plays an important role in conditioning. This concept was investigated in a study conducted by Hamm et al. (1989), in which pairs of unconditioned and neutral stimuli were rated according to their belongingness. After a classical conditioning process using rating-defined high- and low-belongingness pairs, finger pulse responses revealed significantly stronger acquisition and resistance to extinction for high-belongingness pairs.

Our current study is a further investigation of the SFC paradigm in VR in a human sample using an operant conditioning setting, which consisted of acquisition and extinction phases similar to those in the preliminary study. In the current study, we tried to maximize the immersion in VR using a head-mounted-display with a larger field of view as suggested in our first SFC study (Shiban et al., 2015). During the SFC process, fear and contingency ratings as well as physiological (fear-potentiated startle and skin conductance level) and behavioral data were collected. In order to take the above-mentioned effects of belongingness into account, a second experimental condition was added to the previous design. Besides the electroshock condition, in which an electrical stimulation to the lower arm serves as an US, an air blast combined with virtual spitting and insulting was employed as the US in the social threat condition. Because the subjective experience of (un)pleasantness was only partly in accordance with the physiological measurements in our first SFC study, we decided to use the skin conductance level (SCL) as an additional measure of distress during social interaction (e.g., Mesa et al., 2014). Moreover, we investigated the avoidance behavior quantified as the time in non-motion before the approach as well as the time in motion of the approach.

In our current study, we expected that (1) in the operant conditioning process, fear and contingency ratings for CS+ would increase after the acquisition phase compared to the baseline phase. Furthermore, (2) the amplitude of the fear-potentiated startle and the SCL as well as the time in non-motion before approaching the CS+ and time in motion of the approach toward CS+ were expected to increase. (3) After the extinction phase, fear and contingency ratings of the CS+ were supposed to return to baseline levels along with the electrophysiological reactions and the behavioral variables. (4) For the CS− and neutral stimulus (NS), no such changes were expected, i.e., the ratings and physiological measurements should remain stable. (5) The acquisition and the resistance to extinction were expected to be higher for the social threat condition than for the electroshock condition due to the belongingness effect of spitting and insulting to socially frightening situations and thus

the more realistic simulation of social interaction. Finally, (6) a stronger manifestation of the conditioning process was expected in more socially fearful participants in comparison to less socially fearful participants.

MATERIALS AND METHODS

Participants

Forty-four healthy volunteers were recruited through advertisements at the University of Regensburg. Exclusion criteria were age below 18 or above 55, a current diagnosis of psychiatric disorder, psychological treatment, history of psychotropic drug use, color blindness and uncorrected vision or hearing deficits. These criteria were assessed via a questionnaire after written informed consent had been obtained. Participants were randomly allocated to one of the two conditions. As one participant was excluded due to a technical error during data acquisition, the study comprised a total of forty-three participants (22 participants in the electroshock condition: 68.2% female, aged between 18 and 25, $M = 21.10$, $SD = 1.80$; and 21 participants in the social threat condition: 81% female, aged between 19 and 30, $M = 21.95$, $SD = 2.84$). All of the volunteers were students at the University of Regensburg and were offered credit points as compensation for their participation (see **Table 1**). The Ethics Committee of the University of Regensburg approved the study.

Apparatus

The VR environment consisted of one room (see **Figure 1A**), in which all three phases (baseline, acquisition and extinction) took place. In every phase the participant was positioned at one end of the room and could see the agent at the opposite end of the room. The agents gazed dynamically at the participant and moved their head and upper body slightly (see **Figures 1B,C**). In 75% of the conditioning trials an aversive consequence followed when the participant reached the agent. Aversive consequences consisted of an electric stimulus to the participant's lower arm in the electroshock condition or of an air blast to the right side of the participant's neck (2 bar, 10 ms) accompanied by a sound of spitting followed by an insult in the social threat condition. In addition, when the participant approached the agent a startle sound was administered with a contingency of 75% in all phases. A compressed air tank was regulated via a magnetic valve system channeled the air blast through a tube that was fixed to the participant's torso. A cuff was fixed to the participant's right lower arm to administer the electric stimulus. Each participant's individual pain threshold ($M = 2.42$ mA, $SD = 1.82$ mA) was determined before the VR session started. To this end, different strengths of electrical current were administered to the participant's lower arm and then rated on a pain scale from 0 to 10. The amperage with a mean rating of 5 was used as the US during the VR session. The VR was presented to participants via an Oculus Rift DK2 head-mounted display (HMD; Oculus VR Inc., Irvine, CA, United States; see **Figure 1D**) and was generated via Steam Source engine (Valve Corporation, Bellevue, WA, United States). The presented VR

TABLE 1 | Demographic variables and questionnaire data.

Demographics	Electric shock condition (n = 22)		Social threat condition (n = 21)				
	M	SD	M	SD	df	t	p
Age	21.10	1.80	21.95	2.84	41	−1.195	0.239
SPIN	12.95	9.23	16.67	11.71	41	−1.157	0.254
	n	%	n	%	df	χ^2	p^a
Gender [female]	15	68.2	17	81.0	1	0.920	0.337

Means (M) and Standard Deviations (SD) and also t- and p-values are given for all participants for the variables age and SPIN (German version of the Social Phobia Inventory by Stangier and Steffens, 2002). Numbers of participants (n) and percent (%) for gender is given; [a]Chi square test, two-tailed.
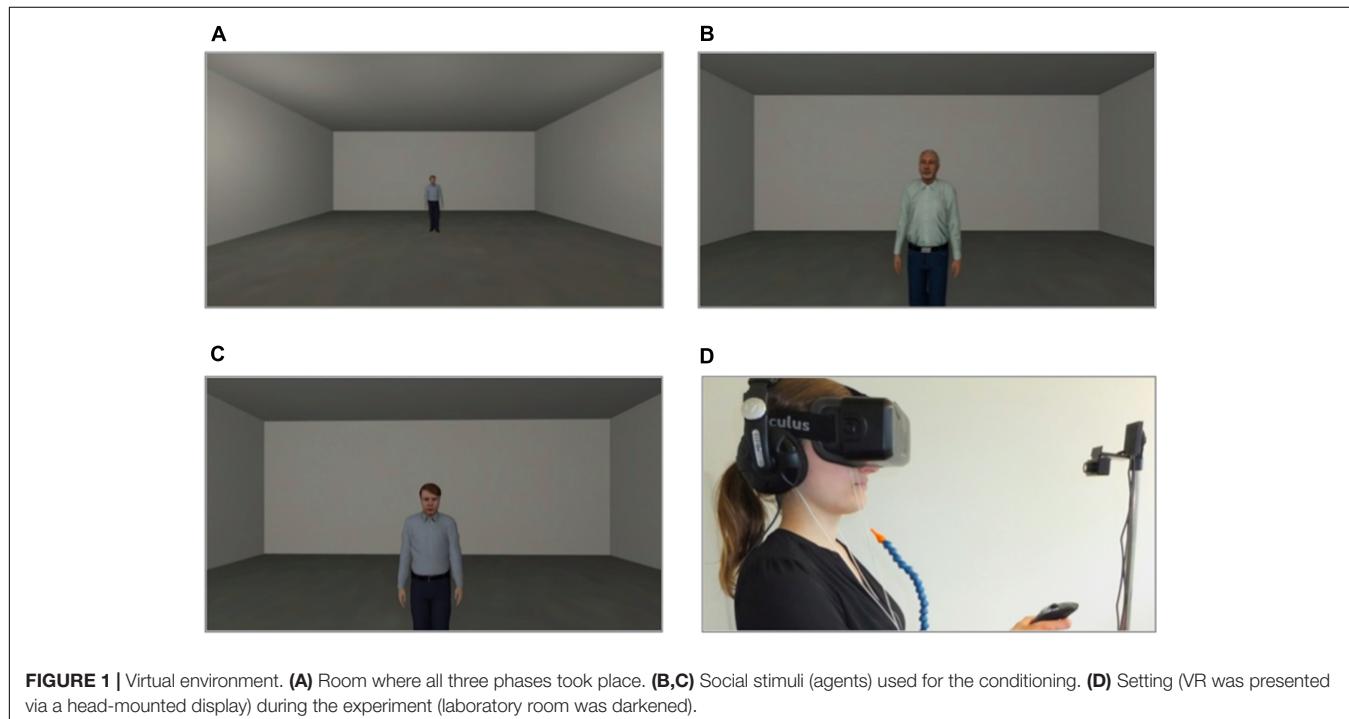
FIGURE 1 | Virtual environment. **(A)** Room where all three phases took place. **(B,C)** Social stimuli (agents) used for the conditioning. **(D)** Setting (VR was presented via a head-mounted display) during the experiment (laboratory room was darkened).
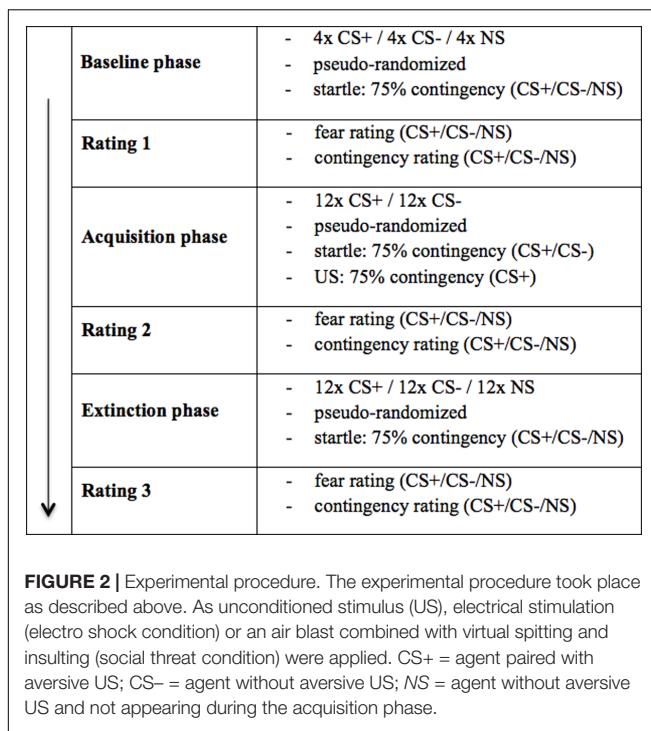
environment was controlled by "cybersession" software (VTplus GmbH, Würzburg, Germany) (see **Figure 1C**). The participant's head position was monitored via the Oculus' electromagnetic tracking device (Oculus VR Inc., Irvine, CA, United States), which adjusts the field of view to any head movements. Sounds were presented over headphones (Sennheiser HD-215, Sennheiser electronic GmbH, Germany). Participants used a joystick (Logitech Extreme 3D Pro Joystick, Logitech GmbH, Germany) to move in the VR environment. Physiological data were monitored, digitally amplified (V-Amp, Brain Products GmbH, Germany) and recorded (Brain Vision Recorder software, Version 1.20, Brain Products GmbH, Germany).

Measures
Participants filled out a demographic questionnaire (age, sex, education, and current occupation) and the Social Phobia Inventory (SPIN; Connor et al., 2000; German Version: Stangier and Steffens, 2002) to assess social fear.

The SPIN consists of 17 items that assess fear, avoidance, and physiological symptoms of social phobia in the previous week. Answers are given on a five-point Likert scale (from 0 = "not at all" to 4 = "extremely"). The German version of the SPIN was evaluated by Sosic et al. (2008). Internal consistency was excellent for a representative sample of 2043 Germans (Cronbach's Alpha = 0.95). Convergent and divergent validity are satisfactory. Furthermore, the German version of the SPIN is a sensitive and specific measure for social phobia as it distinguishes successfully between social phobia and other psychiatric disorders (Sosic et al., 2008).

In order to measure the experienced fear and contingency of the agents, ratings were assessed verbally during the presentations of the agents in the rating phase following each of the three phases ("Estimate your fear now"; "How likely would an aversive stimulus have been?"). These ratings had a range from 0 (very low fear/very unlikely) to 100 (very high fear/very likely).

FIGURE 2 | Experimental procedure. The experimental procedure took place as described above. As unconditioned stimulus (US), electrical stimulation (electro shock condition) or an air blast combined with virtual spitting and insulting (social threat condition) were applied. CS+ = agent paired with aversive US; CS– = agent without aversive US; *NS* = agent without aversive US and not appearing during the acquisition phase.

Besides the subjective measures, physiological data were collected. To record the electromyography of the musculus orbicularis oculi as a measure of fear-potentiated startle, four surface electrodes (Ag/AgCl, Ø = 8 mm) were affixed under the right eye of the participant and on the mastoid bones as reference and ground electrodes. Two additional surface electrodes (Ag/AgCl, Ø = 8 mm) were placed on the base of the thumb on the radial side of the palm of the non-dominant hand in order to record the SCL. The avoidance was measured as the time in non-motion (in s) before approaching the agents and the time in motion (in s) of the approach.

Procedure

The experiment consisted of the questionnaire phase, the baseline phase, the acquisition phase and the extinction phase [total duration was 60 min (30 min in VR); see **Figure 2**].

The baseline phase consisted of four blocks. One block consisted of three presentations of each agent (CS+, CS−, NS), resulting in a total of 12 presentations of each agent per participant. The order within each block was randomized and no US was administered. Which agent was presented as CS+/CS−/NS, was balanced across participants. A startle noise (white noise: 103 dB, 10 ms) was presented with a contingency of 75%.

Conditioning was conducted in 12 blocks. One block consisted of two presentations of both conditioned stimuli with aversive reinforcement in terms of electric stimulus or air blast combined with virtual spitting and the negative utterance "Get lost!" (CS+) and without aversive reinforcement (CS−), resulting in a total of 24 presentations per participant. The NS agent did not appear in this phase. The order within each block was randomized. The

CS-US contingency was set at 75%. As in the baseline phase, the startle noise was presented with a contingency of 75%.

The extinction phase consisted of 12 blocks designed in exactly the same way as those in the acquisition phase, except for the absence of the US and the reappearance of the NS agent. Because three agents were presented instead of two, the total number of trials was 36 in this phase. Also in the extinction phase the startle noise was presented with a contingency of 75%. After the baseline, acquisition and extinction phase, a rating phase took place in which each agent was presented (presentation 8 s, inter-stimulus interval 20 s) again without US or startle noise.

In the first session participants were briefed and the informed consent form was signed. After filling out the demographic questionnaire and the SPIN, participants were prepared for the VR part of the experiment. The electrodes, the air blast device, the cuff for the electric stimuli, the HMD and the headphones were adjusted. During the experiment the laboratory room was darkened and participants received recorded instructions via the headphones.

Before the baseline phase started, participants were able to walk around a desk standing in the middle of the room with gray walls and floor in VR. After exploring this virtual environment, the room faded into a gray background and participants relaxed for 2 min in VR. After the baseline phase, participants received the recorded instruction: "You will now meet virtual human beings. Please use the joystick to approach the person. Please try to move directly toward the person. Press the joystick forward to move straight forward and approach the person." Participants had to approach the agents actively using the joystick and as soon as they reached a specific distance to the agents (the equivalent of about 30 cm in the real world), lights faded out and the next agent was presented at the opposite wall. Each trial lasted about 10 s (depending on how fast participants approached the agents). Theoretically, participants could move laterally, diagonally or away from the agent, however, we observed no such behavior. Because the field of view was adapted to head movements, participants could theoretically look away while moving toward the agent. After the baseline phase, the first rating took place; participants approached each of the three agents and as soon as they reached the previously specified distance to the agents, lights faded out and the participants were asked to verbally rate their subjective fear and the contingency of aversive events.

During the acquisition phase, participants again received the recorded instruction to approach the agents actively via joystick and, as soon as they reached the pre-determined distance to the agents, the lights faded out. At this moment, the US was presented for CS+ agents in 75% of the trials. After the acquisition phase, participants rated the agents again as described above.

The following extinction phase differed from the acquisition only in the reappearance of the NS and the absence of aversive US. After the third rating, the experiment was complete.

Statistical Analyses

Physiological data were preprocessed with Brain Vision Analyzer 2.0 software (Brain Products GmbH, Munich, Germany) and
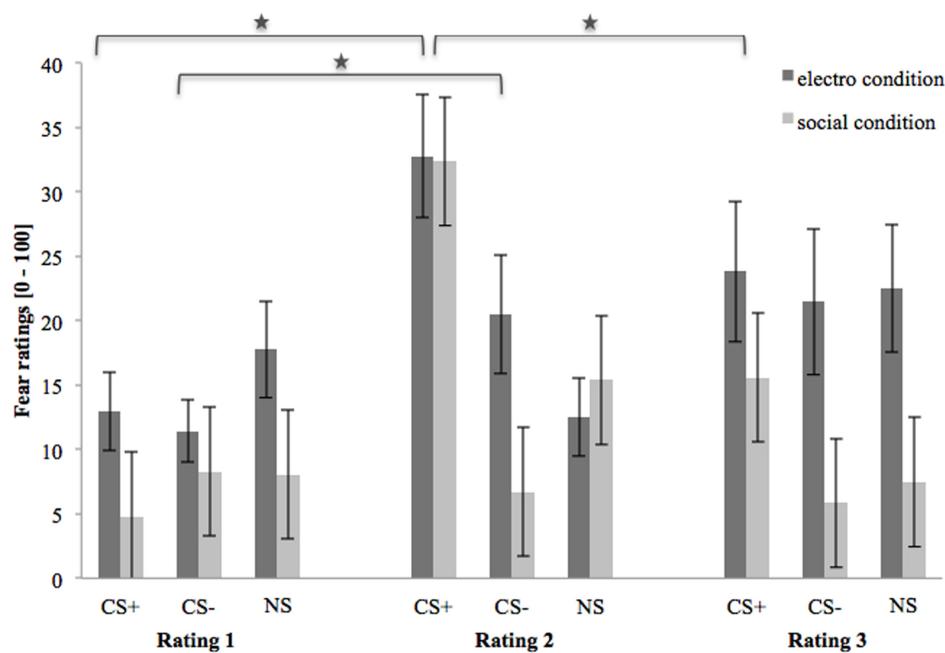
FIGURE 3 | Fear ratings (*n* = 43) for CS+, CS– and NS in the three rating phases for the electro shock and social threat condition. CS+ = agent paired with aversive unconditioned stimulus (US); CS– = agent without aversive US; *NS* = agent without aversive US and not appearing during the acquisition phase; electro condition = electrical stimulation; social condition = air blast combined with virtual spitting and insulting; Rating 1 = after baseline phase; Rating 2 = after acquisition phase; Rating 3 = after extinction phase. Mean fear ratings (0 = very low fear to 100 = very high fear) were given. Significant differences are indicated with an asterisk. Standard errors are presented by error bars.

further analyses were performed in SPSS 22.0 (IBM Corp., Armonk, NY, United States).

For each physiological outcome variable (fear-potentiated startle, SCL) and avoidance behavior, means were calculated for the baseline phase, while the first four reactions and the last four reactions in the acquisition and the extinction phase were computed as the means of the beginning and the end of the acquisition and extinction phase, respectively.

For the fear-potentiated startle, first, differences between the two electromyography electrodes were computed (see Blumenthal et al., 2005). Then, a 250 Hz high cut-off filter, a 30 Hz low cut-off filter, and a 50 Hz notch filter were applied, the data were rectified, and a moving average (50 ms) was calculated. For each fear-potentiated startle a baseline correction was conducted using the mean value of the 50 ms before each startle tone as baseline. Next, peaks were marked automatically, controlled manually and corrected if necessary. Finally, *T*-values for the startle magnitude were calculated. Due to technical errors during data acquisition, six participants had to be excluded from data analysis of the fear-potentiated startle.

For the analysis of the SCL, the difference between the two electrodes was computed, a 1 Hz high cut-off filter and a baseline correction of 1-s duration applied and the SCL exported in order to calculate *T*-values for the SCL. Due to technical errors during data acquisition, five participants had to be excluded from data analysis of the SCL.

The avoidance behavior was assessed via time in non-motion (latency) and time in motion. Time in non-motion (in s) was defined as the time before approaching the agent. Time in motion (in s) was computed subtracting the time in non-motion from the total time needed for reaching the specific distance to the agent.

The means for each agent (CS+, CS−, NS) of the subjective variable (fear and contingency ratings) measured at the three rating phases (rating 1–3) were calculated.

Participants were divided into two groups (low vs. high social anxiety) via a median split of the SPIN score (median = 13.5 in this study) in order to differentiate between highly and less socially fearful participants.

Two repeated-measures ANOVAs with the within-subject factors phase (rating 1 vs. rating 2 for acquisition and rating 2 vs. rating 3 for extinction) and stimulus (CS+ vs. CS− vs. NS) and the between-subject factors social anxiety (low vs. high) and condition (electroshock condition vs. social threat condition) were conducted for both subjective variables.

For each physiological and behavioral outcome variable, repeated-measures ANOVAs with the within-subject factors time (baseline vs. beginning vs. end of acquisition) and stimulus (CS+ vs. CS−) and the between-subject factors social anxiety (low vs. high) and condition (electroshock condition vs. social threat condition) were conducted for the acquisition phase. For the extinction phase repeated-measures ANOVAs with the within-subject factors time (beginning vs. end of extinction) and stimulus (CS+ vs. CS−) and the between-subject factors social anxiety (low vs. high) and condition

TABLE 2 | Significant results of the ANOVAs for the fear ratings of the acquisition and extinction phase.

Effect	df	F	η²	p
Acquisition				
Total				
Phase	1, 39	32.1	0.45	<0.001
Stimulus	2, 78	13.9	0.26	<0.001
Phase × Stimulus	2, 78	20.5	0.34	<0.001
Phase × Stimulus × Condition	2, 78	4.96	0.11	0.009
Electroshock condition				
Phase	1, 20	14.4	0.42	<0.001
Stimulus	2, 40	4.72	0.19	0.014
Phase × Stimulus	2, 40	11.8	0.37	<0.001
Social threat condition				
Phase	1, 19	17.6	0.48	<0.001
Stimulus	2, 38	10.2	0.35	<0.001
Time × Stimulus	2, 38	13.1	0.41	<0.001
Extinction				
Total				
Stimulus	2, 74	22.7	0.38	<0.001
Phase × Stimulus	2, 74	10.6	0.22	<0.001

df = degrees of freedom; η² = effect size; Phase = Rating 1 vs. Rating 2 for the acquisition and Rating 2 vs. Rating 3 for the extinction; Rating 1 = after baseline, Rating 2 = after acquisition, Rating 3 = after extinction; Stimulus = CS+ vs. CS− vs. NS; CS+ = agent paired with the aversive unconditioned stimulus (US), CS− = agent without aversive US, NS = agent without aversive US and not appearing during the acquisition phase; Condition = electroshock vs. social threat condition; Social Anxiety (low vs. high) was measured with the German version of the Social Phobia Inventory (SPIN; median split = 13.5, Stangier and Steffens, 2002).

(electroshock condition vs. social threat condition) were conducted.

Measuring generalization effects, ANOVAs with the within-subject factor phase (baseline vs. end of extinction) and the between-subject factors social anxiety (low vs. high) and condition (electroshock condition vs. social threat condition) were conducted for the NS as well.

In additional analyses of significant effects of time, stimulus, or social anxiety Student's *t*-tests were performed. Partial η² (η²ₚ) scores and Cohen's *d* were used as indices of effect size. The significance level was set at two-tailed alpha = 0.05.

RESULTS

Fear Ratings

Figure 3 shows the fear ratings 1–3 (after the baseline, acquisition and extinction phase, respectively). As we can see, in the beginning, (baseline) fear ratings are almost equal for all three stimuli, but slightly higher in the electroshock than in the social threat condition. After the acquisition phase, fear ratings for CS+ are clearly higher than for CS− and NS in both US conditions. Fear ratings for CS− are higher in the electroshock than in the social threat condition, while fear ratings for NS barely differ after acquisition. After the extinction phase, fear ratings for CS+ decrease in both conditions. However, fear ratings for

CS+ decreased more in the social threat condition than in the electroshock condition. CS− did not change in either condition over time, whereas the NS increased in the electroshock condition and decreased in the social threat condition. After extinction, all three stimuli are generally rated with higher fear and contingency levels in the electroshock condition than in the social threat condition.

An ANOVA comparing fear ratings before and after acquisition confirmed significant interaction effects of Phase × Stimulus and Phase × Stimulus × Condition (please see **Table 2** for all significant results of the ANOVA). A follow-up ANOVA was conducted for each condition. For the electroshock condition, a significant interaction effect of Phase × Stimulus could be detected. A follow-up *t*-test showed that the fear ratings increased significantly for CS+, $t(21) = -5.04$, $p < 0.001$, $d = 1.12$, and for CS−, $t(21) = -2.46$, $p = 0.023$, $d = 0.54$, and decreased significantly for NS, $t(21) = 2.59$, $p = 0.017$, $d = 0.31$, from pre to post acquisition. For the social condition, an interaction effect of Phase × Stimulus was also significant. Follow-up *t*-test revealed that fear ratings increased significantly only for CS+, $t(20) = -5.67$, $p < 0.001$, $d = 1.52$, from pre to post acquisition, but not for CS− or NS. Therefore, the fear rating results indicate that successful SFC took place under both conditions.

An ANOVA comparing fear ratings before and after extinction confirmed a significant interaction effect of Phase × Stimulus. Follow-up *t*-test showed that fear ratings decreased significantly for CS+, $t(40) = 3.92$, $p < 0.001$, $d = 0.60$, from pre to post extinction, but not for CS− or NS. The fear rating results indicate that social fear extinction was also successful under both conditions.

Contingency Ratings

Figure 4 shows contingency ratings 1–3 (after baseline, acquisition, and extinction phase, respectively). In the beginning, contingency ratings are almost equal for both conditions and all three stimuli. After the acquisition phase, contingency ratings for CS+ are higher than for CS− or NS in both US conditions. Regarding the CS−, contingency ratings are higher in the electroshock than in the social threat condition. In both conditions the contingency ratings for NS decrease slightly after acquisition. After the extinction phase, the contingency ratings for CS+ decrease strongly in both conditions. Contingency ratings for CS− decrease in the electroshock condition and increase slightly in the social threat condition. Conversely, contingency ratings for NS increased slightly in the electroshock condition and decreased slightly in the social threat condition.

An ANOVA comparing contingency ratings before and after acquisition confirmed significant interaction effects of Phase × Stimulus, Stimulus × Social Anxiety, and Phase × Stimulus × Condition (please see **Table 3** for all significant results of the ANOVA). Follow-up ANOVA was conducted for each condition. In the electroshock condition, significant interaction effects of Phase × Stimulus, and Stimulus × Social Anxiety could be detected. Follow-up *t*-test conducted for Phase × Stimulus interaction showed that contingency ratings increased significantly for CS+, $t(21) = -7.49$, $p < 0.001$, $d = 1.88$, and for CS−, $t(21) = -2.38$,
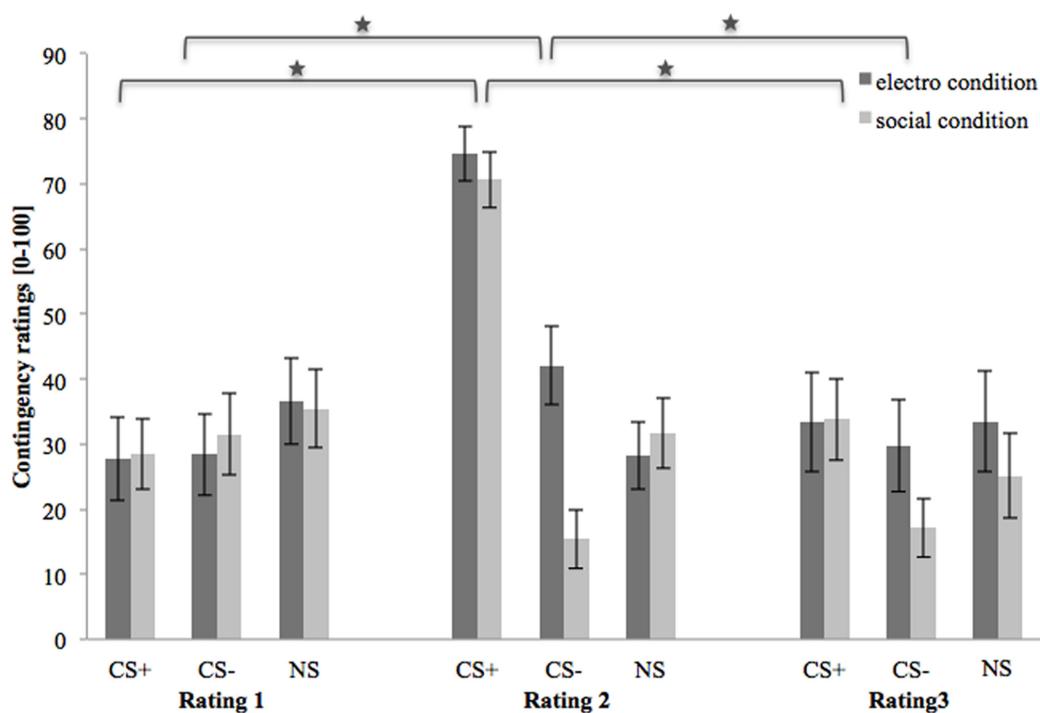
FIGURE 4 | Contingency ratings (n = 43) for CS+, CS− and NS in the three rating phases for the electro shock and social threat condition. CS+ = agent paired with aversive unconditioned stimulus (US); CS− = agent without aversive US; NS = agent without aversive US and not appearing during the acquisition phase; electro condition = electrical stimulation; social condition = air blast combined with virtual spitting and insulting; Rating 1 = after baseline phase; Rating 2 = after acquisition phase; Rating 3 = after extinction phase. Mean contingency ratings (0 = very unlikely to 100 = very likely) were given. Significant differences are indicated with an asterisk. Standard errors are presented by error bars.

$p = 0.027$, $d = 0.48$, from pre to post acquisition, but not for NS. Follow-up tests of the significant Stimulus × Social Anxiety interaction revealed a significant difference for the less socially fearful participants between CS+, CS−, and NS ($p < 0.020$), and for the higher socially fearful participants between CS+ and CS− ($p < 0.003$), but not NS. Means and standard deviations are presented in **Table 4**. In the social threat condition an interaction effect of Phase × Stimulus reached significance level. Follow-up t-test showed that contingency ratings increased significantly for CS+, $t(19) = -7.50$, $p < 0.001$, $d = 1.88$, and decreased for CS−, $t(19) = 2.47$, $p = 0.023$, $d = 0.72$, from pre to post acquisition. This pattern could not be found for NS. Thus, contingency rating results also indicate that SFC was successful.

An ANOVA on contingency ratings before and after extinction showed significant interaction effects for Stimulus × Condition, Stimulus × Social Anxiety, Phase × Stimulus, and a marginally significant interaction effect of Phase × Stimulus × Condition. Follow-up ANOVAs were conducted separately for the two conditions. In the electroshock condition, interaction effects of Phase × Stimulus and Stimulus × Social Anxiety reached significance level. Follow-up t-test conducted for the Phase × Stimulus interaction effect showed that contingency ratings decreased significantly for CS+, $t(20) = 5.88$, $p < 0.001$, $d = 1.66$, and for CS−, $t(20) = 2.66$, $p = 0.015$, $d = 0.46$, from pre to post extinction, but not for NS. Follow-up tests of the Stimulus × Social Anxiety interaction revealed a significant

difference both for the less socially fearful participants between CS+ and NS ($p < 0.020$), and for the highly socially fearful participants between CS+, CS− and NS ($p < 0.022$). In the social threat condition, interaction effects of Phase × Stimulus and Stimulus × Social Anxiety reached significance level. Follow-up t-tests of the Phase × Stimulus interaction revealed that contingency ratings decreased significantly for CS+, $t(19) = 5.91$, $p < 0.001$, $d = 1.58$, but not for CS− or NS. Follow-up tests of the significant Stimulus × Social Anxiety interaction revealed a significant difference both for the less socially fearful participants between CS+, CS− and NS ($p < 0.001$), and for the highly socially fearful participants between CS+, CS− and NS ($p < 0.030$). These results indicate that social fear extinction was successful according to the contingency ratings as well.

Fear-Potentiated Startle
Figure 5 depicts fear-potentiated startle response for the baseline, acquisition and extinction phase. In the electroshock condition fear-potentiated startle response is higher for CS− than for CS+ at the baseline and both stimuli increase at the beginning, until both decrease to the end of the acquisition. In the extinction phase CS+ response is higher than CS−, but the responses to both stimuli decreased from the beginning to the end. In the social threat condition fear-potentiated startle response is higher for CS− than for CS+ at the baseline. CS+ response increases whereby CS− do not change at the beginning, until both decrease

at the end of the acquisition. In the extinction phase both stimuli decrease from the beginning to the end.

For the acquisition phase, an ANOVA confirmed a significant main effect of time, $F(1,33) = 7.51$, $p < 0.001$, $\eta_p^2 = 0.19$, and stimulus, $F(1,33) = 5.20$, $p = 0.029$, $\eta_p^2 = 0.14$, but no significant interaction effects. **Figure 5** shows an increase of fear-potentiated startle at the beginning and a fast habituation process at the end of the acquisition phase in both conditions.

For the extinction phase, there was a significant main effect of time, $F(1,31) = 8.46$, $p = 0.007$, $\eta_p^2 = 0.21$, but no other significant

TABLE 3 | Significant results of the ANOVAs for the contingency ratings of the acquisition and the extinction phase.

Effect	df	F	η^2	p
Acquisition				
Total				
Phase	1, 38	10.8	0.22	0.002
Stimulus	2, 76	33.9	0.47	<0.001
Phase × Stimulus	2, 76	51.3	0.58	<0.001
Stimulus × Social Anxiety	2, 76	3.29	0.08	0.042
Phase × Stimulus × Condition	2, 76	5.76	0.13	0.005
Electroshock condition				
Phase	1, 20	15.5	0.44	<0.001
Stimulus	2, 40	10.8	0.35	<0.001
Phase × Stimulus	2, 40	25.4	0.56	<0.001
Stimulus × Social Anxiety	2, 40	3.98	0.17	0.027
Social threat condition				
Stimulus	2, 36	28.0	0.61	<0.001
Phase × Stimulus	2, 36	30.9	0.63	<0.001
Extinction				
Total				
Phase	1, 37	13.6	0.27	<0.001
Stimulus	2, 74	71.5	0.66	<0.001
Stimulus × Condition	2, 74	8.04	0.18	<0.001
Stimulus × Social Anxiety	2, 74	6.72	0.15	0.002
Phase x Stimulus	2, 74	31.9	0.46	<0.001
Phase × Stimulus × Condition	2, 74	3.06	0.08	0.053
Electroshock condition				
Phase	1, 19	6.49	0.26	0.020
Stimulus	2, 38	26.0	0.58	<0.001
Phase × Stimulus	2, 38	20.2	0.52	<0.001
Stimulus × Social Anxiety	2, 38	4.30	0.19	0.021
Social threat condition				
Phase	1, 18	7.71	0.30	0.012
Stimulus	2, 36	49.3	0.73	<0.001
Phase × Stimulus	2, 36	14.9	0.45	<0.001
Stimulus × Social Anxiety	2, 36	4.51	0.20	0.018

df = degrees of freedom; η^2 = effect size; Phase = Rating 1 vs. Rating 2 for the acquisition and Rating 2 vs. Rating 3 for the extinction; Rating 1 = after baseline, Rating 2 = after acquisition, Rating 3 = after extinction; Stimulus = CS+ vs. CS− vs. NS; CS+ = agent paired with the aversive unconditioned stimulus (US), CS− = agent without aversive US, NS = agent without aversive US and not appearing during the acquisition phase; Condition = electroshock vs. social threat condition; Social Anxiety (low vs. high) was measured with the German version of the Social Phobia Inventory (SPIN; median split = 13.5, Stangier and Steffens, 2002).

main or interaction effects. For NS, a significant main effect of time, $F(1,32) = 7.98$, $p = 0.008$, $\eta_p^2 = 0.20$, could be detected.

Skin Conductance Level

Figure 6 depicts SCL for the baseline, acquisition and extinction phase. In the baseline, SCL for CS+ response is slightly higher than for CS− in both conditions. In the electroshock condition, for CS+ the SCL increase from the baseline to the beginning and decrease to the end of the acquisition, whereas it decrease for CS− from the baseline to the end of the acquisition. In the beginning of the extinction, SCL for CS+ is higher than for CS−, at the end of the extinction both stimuli do not differ. In the social condition, SCL for CS+ also increase from the baseline to the beginning and decrease from the beginning to the end of the acquisition. SCL for CS− decrease from the baseline to the beginning and subsequently increase to the end of the acquisition. In the beginning of the extinction, both stimuli do not differ and both increase slightly at the end of the extinction.

For the acquisition phase, an ANOVA confirmed significant main effects of stimulus, $F(1,34) = 15.4$, $p = 0.010$, $\eta_p^2 = 0.18$, as well as significant interaction effect of Time × Stimulus, $F(2,68) = 18.5$, $p < 0.001$, $\eta_p^2 = 0.35$. Follow-up t-tests revealed that SCL for CS+ and CS− only differed at the beginning of the acquisition, $t(37) = 6.26$, $p < 0.001$, $d = 1.35$. Thus, there was a significant increase in SCL for CS+ and a significant decrease for CS− from the baseline to the beginning of the acquisition. The SCL results indicate that successful SFC took place under both condition, but also a fast habituation during acquisition.

For the extinction phase, an ANOVA showed a significant main effect of condition, $F(1,32) = 4.95$, $p = 0.033$, $\eta_p^2 = 0.13$, and a significant interaction effect of Time × Stimulus × Condition × Social Anxiety, $F(1,32) = 101.8$, $p = 0.044$, $\eta_p^2 = 0.12$. A follow-up ANOVA was conducted

TABLE 4 | Means (M) and standard deviations (SD) for contingency ratings during acquisition and extinction for high- and low-social anxious and both conditions.

	CS+		CS−		NS	
	M	**SD**	**M**	**SD**	**M**	**SD**
Acquisition						
Electroshock condition						
Low socially fear	49.9	22.5	35.4	30.5	24.6	22.4
High socially fear	53.4	18.8	35.0	15.3	45.9	24.2
Extinction						
Electroshock condition						
Low socially fear	51.7	22.2	33.9	28.6	23.4	19.3
High socially fear	57.2	22.3	38.4	28.1	43.3	21.8
Social threat condition						
Low socially fear	63.6	13.9	15.6	19.4	27.5	22.4
High socially fear	44.6	19.5	16.7	19.7	29.1	17.3

CS+ = agent paired with US, CS− = agent without aversive US, NS = agent without aversive US and not appearing during the acquisition phase; Social Anxiety (low vs. high) was measured with the German version of the Social Phobia Inventory (SPIN; median split = 13.5, Stangier and Steffens, 2002).
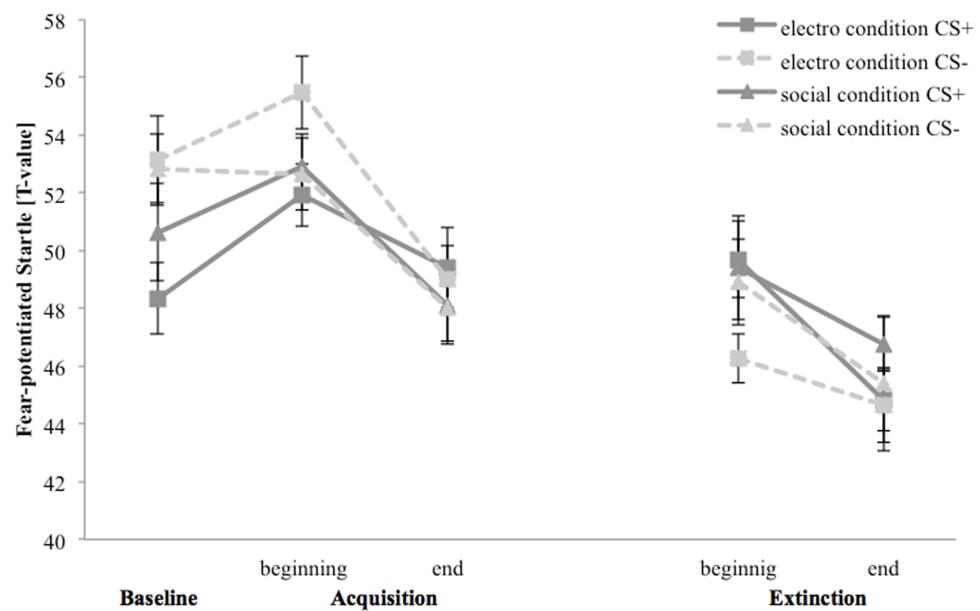
FIGURE 5 | Fear-potentiated startle response ($n = 37$) for CS+ and CS− in the three phases (baseline, acquisition, and extinction) for the electro shock and social threat condition. CS+ = agent paired with aversive unconditioned stimulus (US); CS− = agent without aversive US; electro condition = electrical stimulation; social condition = air blast combined with virtual spitting and insulting. Mean fear-potentiated startles (presented in T-values) was given. Standard errors are presented by error bars.
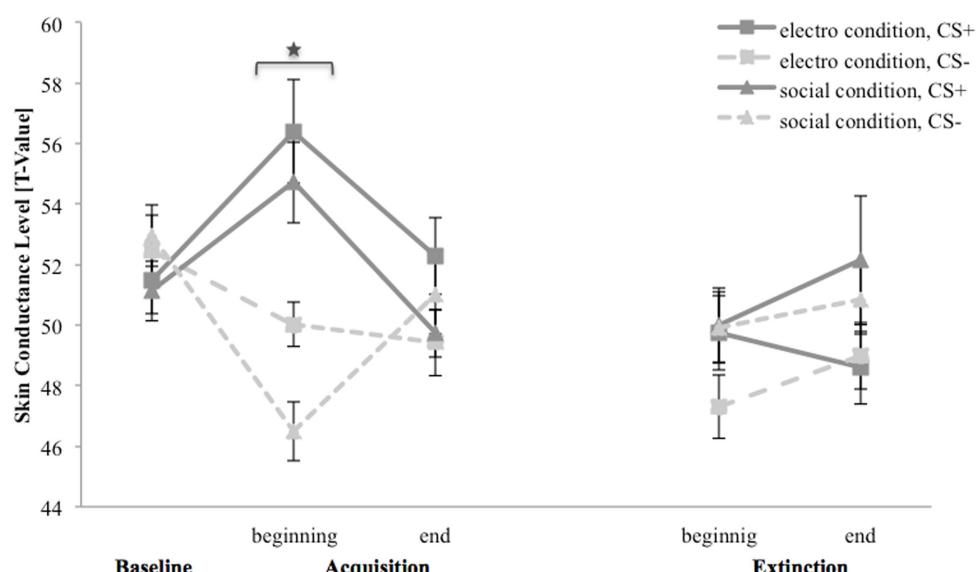
FIGURE 6 | Skin conductance level ($n = 38$) for CS+ and CS− in the three phases (baseline, acquisition, and extinction) for the electro shock and social threat condition. CS+ = agent paired with aversive unconditioned stimulus (US); CS− = agent without aversive US; electro condition = electrical stimulation; social condition = air blast combined with virtual spitting and insulting. Mean skin conductance level (presented in T-values) was given. Significant differences are indicated with an asterisk. Standard errors are presented by error bars.

separately for both conditions. In the electroshock condition, no significant main or interaction effects were found. In the social threat condition, a significant interaction effect of Time × Stimulus × Social Anxiety, $F(1,17) = 4.48$, $p = 0.049$, $\eta_p^2 = 0.21$, was detected. Follow-up t-tests conducted separately

for higher and less socially fearful participants neither showed significant differences between SCL for CS+ and CS− at the beginning nor at the end of the extinction. For NS, a significant main effect of time, $F(1,33) = 7.39$, $p = 0.010$, $\eta_p^2 = 0.18$, could be detected.
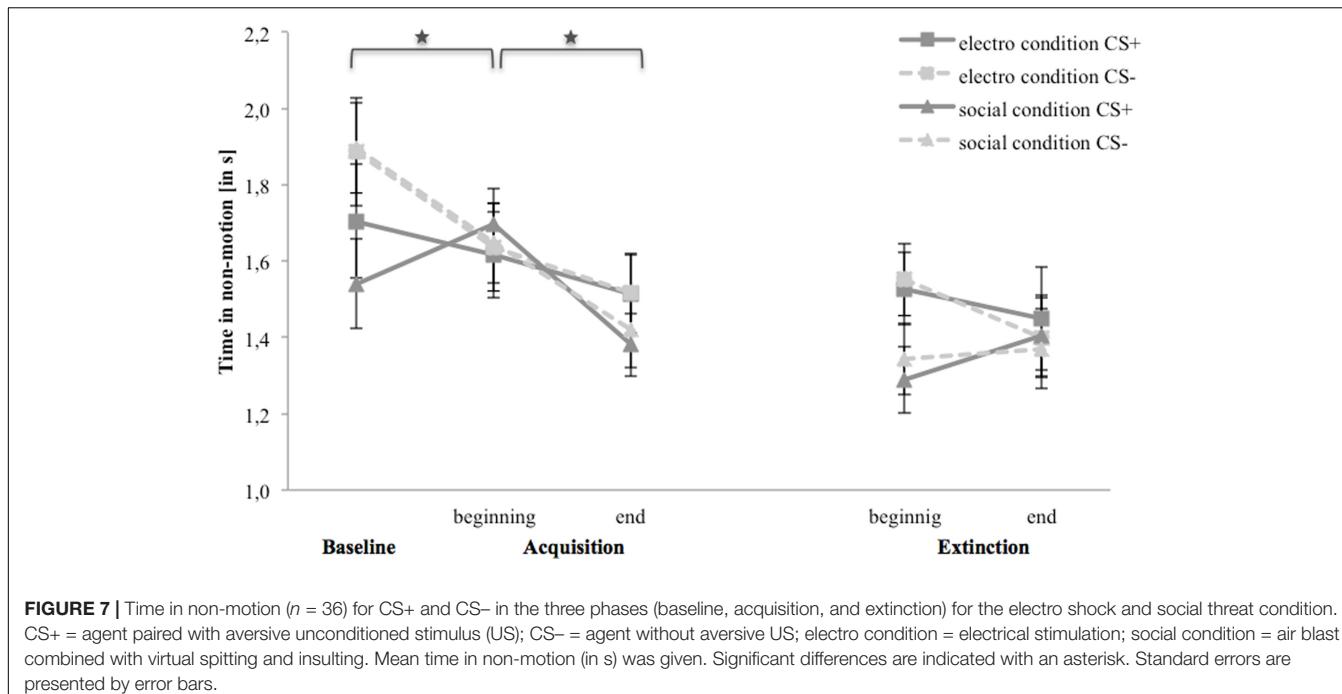
FIGURE 7 | Time in non-motion (*n* = 36) for CS+ and CS– in the three phases (baseline, acquisition, and extinction) for the electro shock and social threat condition. CS+ = agent paired with aversive unconditioned stimulus (US); CS– = agent without aversive US; electro condition = electrical stimulation; social condition = air blast combined with virtual spitting and insulting. Mean time in non-motion (in s) was given. Significant differences are indicated with an asterisk. Standard errors are presented by error bars.

Avoidance (Time in Non-motion)

Figure 7 shows time in non-motion for the baseline, acquisition and extinction phase. In the electroshock condition, avoidance for both stimuli decreases from the baseline to the end of the acquisition phase as well as from the beginning to the end of the extinction phase. In the social threat condition, avoidance for CS– is higher than for CS+ at the baseline, and to the end of the acquisition phase it decreases for CS–, whereas avoidance increases for CS+ from the baseline to the beginning until it decreases at the end of the acquisition. In the extinction phase, both stimuli do not differ at any point.

For the acquisition phase, an ANOVA confirmed significant interaction effects of Time × Stimulus, and Condition × Social Anxiety (please see **Table 5** for all significant results of the ANOVA). Follow-up ANOVA was conducted separately for both conditions. In the electroshock condition, no significant interaction effects were found. In the social threat condition, a significant interaction effect of Time × Stimulus could be detected. Follow-up *t*-tests showed that avoidance for CS+ increased from the baseline to the beginning of the acquisition phase, $t(18) = -2.13$, $p = 0.047$, $d = 0.33$, and decreased from the beginning to the end of the acquisition phase, $t(18) = 3.32$, $p = 0.004$, $d = 0.84$. Avoidance for CS– decreased from the baseline to the beginning of the acquisition, $t(18) = 2.35$, $p = 0.031$, $d = 0.53$, as well as from the beginning to the end of the acquisition, $t(18) = 2.77$, $p = 0.013$, $d = 0.51$. Therefore, the time in non-motion results indicate that successful avoidance behavior for CS+ took place in the social threat condition, but also a fast adaptation to the US occurred toward the end of the acquisition.

For the extinction phase, an ANOVA confirmed a significant interaction effect of Time × Condition. Follow-up ANOVA was

conducted separately for both conditions. In the electroshock condition, only a significant main effect of social anxiety was found. In the social threat condition, no significant effects were found. For NS, a significant main effect of time, $F(1,32) = 4.81$, $p = 0.036$, $\eta_p^2 = 0.13$, could be detected.

Avoidance (Time in Motion)

Figure 8 shows time in motion for the baseline, acquisition and extinction phase. In the electroshock condition, the avoidance of CS– is higher than of CS+ during the baseline. Avoidance toward CS– decreases from the baseline to the end of the acquisition, whereas it increases for CS+ from the baseline to the beginning and decreases to the end of the acquisition. In the extinction phase participants move faster toward CS– and slower toward CS+ from the beginning to the end of the extinction. In the social threat condition, time to approach both stimuli are equally long during baseline and increase at the beginning of the acquisition, until avoidance to both stimuli stay approximately at the same level at the end of the acquisition. In the extinction phase, the avoidance of CS+ decreases during the extinction, whereas for CS– it stays on an equal level.

For the acquisition phase, an ANOVA confirmed significant interaction effects of Time × Stimulus and Condition × Social Anxiety (please see **Table 6** for all significant results of the ANOVA). Follow-up ANOVAs were conducted separately for both conditions. In the electroshock condition, a significant interaction effect of Time × Stimulus could be detected. Follow-up *t*-tests revealed that only the CS+ significantly increased from the baseline to the beginning of the acquisition, $t(37) = -2.45$, $p = 0.026$, $d = 0.77$. In the social threat condition, no significant interaction effects were found. Therefore, time in motion results indicate a successful SFC at the beginning of the acquisition in

TABLE 5 | Significant results of the ANOVAs for avoidance (time in non-motion) of the acquisition and extinction phase.

Effect	df	F	η^2	p
Acquisition				
Total				
Time	2, 64	10.5	0.25	<0.001
Stimulus	1, 32	9.83	0.24	0.004
Time × Stimulus	2, 64	9.34	0.23	<0.001
Condition × Social Anxiety	1, 32	5.42	0.15	0.026
Electroshock condition				
Time	2, 30	4.06	0.21	0.027
Stimulus	1, 15	6.71	0.31	0.020
Social threat condition				
Time	2, 34	7.82	0.32	0.002
Stimulus	1, 17	5.27	0.24	0.035
Time × Stimulus	2, 34	7.02	0.29	0.003
Extinction				
Total				
Social Anxiety	1, 32	4.71	0.13	0.038
Time × Condition	1, 32	4.25	0.12	0.047
Electroshock condition				
Social Anxiety	1, 15	9.56	0.39	0.007

df = degrees of freedom; η^2 = effect size; Phase = Rating 1 vs. Rating 2 for the acquisition and Rating 2 vs. Rating 3 for the extinction; Rating 1 = after baseline, Rating 2 = after acquisition, Rating 3 = after extinction; Stimulus = CS+ vs. CS− vs. NS; CS+ = agent paired with the aversive unconditioned stimulus (US), CS− = agent without aversive US, NS = agent without aversive US and not appearing during the acquisition phase; Condition = electroshock vs. social threat condition; Social Anxiety (low vs. high) was measured with the German version of the Social Phobia Inventory (SPIN; median split = 13.5, Stangier and Steffens, 2002).

the electroshock condition, but also a fast adaptation to the US occurred toward the end of the acquisition.

For the extinction phase, an ANOVA confirmed significant interaction effects of Time × Stimulus × Condition, and Time × Stimulus × Condition × Social Anxiety. Follow-up ANOVA for the electroshock condition revealed a significant interaction effect of Time × Stimulus, and Time × Stimulus × Social Anxiety. Further follow-up ANOVAs were conducted separately for the low and high social fear groups, but no significant main or interaction effects were found. No significant effects were found in the social threat condition or for the NS.

DISCUSSION

The aim of this study was to replicate and extend the findings of our previous study we conducted on social fear learning (Shiban et al., 2015). In order to improve the paradigm, we investigated the "belongingness effect" (Hamm et al., 1989). To this end, we designed a social threat condition and compared it to an electroshock condition during the different phases (baseline, acquisition and extinction) of the social fear conditioning paradigm (SFC). Participants actively approached virtual agents using a joystick in a setting similar to the one used by Shiban et al. (2015). Social fear learning was examined via subjective ratings (fear and contingency ratings), physiological

TABLE 6 | Significant results of the ANOVAs for avoidance (time in motion) of the acquisition and extinction phase.

Effect	df	F	η^2	p
Acquisition				
Total				
Time	2, 64	139.8	0.81	<0.001
Social Anxiety	1, 32	7.07	0.18	0.012
Time × Stimulus	2, 64	4.68	0.13	0.013
Condition × Social Anxiety	1, 32	6.37	0.17	0.017
Electroshock condition				
Time	2, 30	75.1	0.83	<0.001
Stimulus	1, 15	5.60	0.27	0.032
Social Anxiety	1, 15	9.18	0.38	0.008
Time × Stimulus	2, 30	4.67	0.24	0.017
Social threat condition				
Time	2, 30	5.10	0.25	0.039
Extinction				
Total				
Time × Stimulus × Condition	1, 32	5.87	0.16	0.021
Time × Stimulus × Condition × Social Anxiety	1, 32	6.45	0.17	0.016
Electroshock condition				
Time × Stimulus	1, 15	5.02	0.25	0.041
Time × Stimulus × Social Anxiety	1, 15	6.91	0.32	0.019

df = degrees of freedom; η^2 = effect size; Phase = Rating 1 vs. Rating 2 for the acquisition and Rating 2 vs. Rating 3 for the extinction; Rating 1 = after baseline, Rating 2 = after acquisition, Rating 3 = after extinction; Stimulus = CS+ vs. CS− vs. NS; CS+ = agent paired with the aversive unconditioned stimulus (US), CS− = agent without aversive US, NS = agent without aversive US and not appearing during the acquisition phase; Condition = electroshock vs. social threat condition; Social Anxiety (low vs. high) was measured with the German version of the Social Phobia Inventory (SPIN; median split = 13.5, Stangier and Steffens, 2002).

(fear-potentiated startle, skin conductance level) and behavioral measures (avoidance).

Social fear acquisition was successful according to the fear and the contingency ratings. In both conditions, these measures clearly increased for CS+ compared to CS− from the baseline to the end of the acquisition phase. Interestingly, there was a higher differentiation between CS+ and CS− in the social threat compared to the electroshock condition, which might reflect a tendency toward higher belongingness in the social threat condition. Regarding the physiological outcome variables, the fear-potentiated startle results did not confirm our hypotheses, as no discrimination between CS+ and CS− could be detected. However, with respect to the SCL, successful fear conditioning took place at the beginning of the acquisition, whereas a fast habituation was found toward the end of acquisition, diminishing any discriminant effects between the CS+ and CS−. Furthermore, the avoidance behavior clearly increased for CS+ compared to CS− at the beginning of the acquisition phase for the time in non-motion in the social threat condition and the time in motion in the electroshock condition.

Fear extinction was evident in the ratings, as the differentiation in terms of fear and contingency ratings between the CS+ and the CS− that followed acquisition vanished during
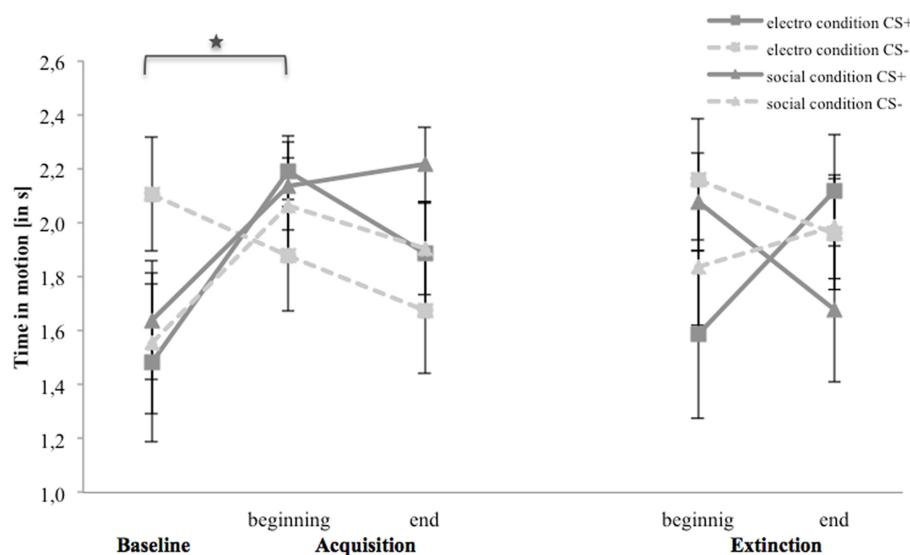
FIGURE 8 | Time in motion (*n* = 36) for CS+ and CS– in the three phases (baseline, acquisition, and extinction) for the electro shock and social threat condition. CS+ = agent paired with aversive unconditioned stimulus (US); CS– = agent without aversive US; electro condition = electrical stimulation; social condition = air blast combined with virtual spitting and insulting. Mean time in non-motion (in s) was given. Significant differences are indicated with an asterisk. Standard errors are presented by error bars.

the extinction phase for both experimental groups. However, no statistically significant extinction was found in the physiological and behavioral variables. It is possible that the physiological level had already been subject to a fast extinction process that can be expected in non-socially phobic individuals before the designated extinction phase of the experiment.

According to our data, social fear can be induced and extinguished confirming the operant conditioning paradigm. Participants did not simply explore the virtual room and the agents in our (operant) fear conditioning paradigm, but actively (using a joystick) approached the agents. They were free to decide how fast they wanted to approach the agents and to which degree they wanted to avoid them. With participants being punished while approaching the stimuli (virtual male agents), our SFC paradigm reflects operant conditioning rather than classical conditioning processes. Interestingly, less socially fearful participants differentially evaluated the contingency of CS+, CS−, and NS after extinction in the electroshock condition and only rated the contingency of the CS+ as high, whereas higher socially fearful participants rated the contingency of the CS+ and the NS on a similar level. Thus, we found a generalization effect in the contingency ratings between CS+ and NS for higher socially fearful participants. No generalization effect was reflected by the physiological measures.

Summarizing the results for the subjective ratings as well as the physiological and behavioral data, our initial hypotheses could be partially confirmed. The habituation at the end of the acquisition phase might reflect a fast adaptation to the aversive US. Possibly the US was not aversive enough to evoke long-lasting fear or the social anxiety of the sample was too low. Due to the belongingness effect, a higher differentiation in the subjective ratings between CS+ and CS− in the social threat condition was found.

Our SFC paradigm might have induced an approach-avoidance conflict. This conflict occurs when a person is faced with the decision to either pursue or avoid something that is advantageous in some respects but disadvantageous in others. In the social threat condition, the avoidance behavior (time in non-motion) clearly differed between aversive (CS+) and non-aversive (CS−) stimuli at the beginning of the acquisition. By comparison, in the electroshock condition the avoidance behavior (time in motion) clearly increased toward aversive compared to non-aversive stimuli at the beginning of the acquisition. Avoiding social situations is a core feature of SAD. Our paradigm showed increased fear and a partial increase in avoidance after the presentation of the first four aversive agents during conditioning. Besides behavioral avoidance, eye-gaze, a non-verbal social cue, is an important aspect of human social behavior. Future studies may therefore consider measuring behavioral approach-avoidance conflict via an eye-tracking method and analyze the recorded movement trajectories as an index of avoidance behavior for social anxiety. Identifying approach- and avoidance-related responses to social stimuli like emotional face stimuli (e.g., via reaction times for pressing a button or joystick responses, or through eye-gaze), has already been investigated in different studies (Mühlberger et al., 2008; Wieser et al., 2009, 2010; Radke et al., 2013). Wieser et al. (2010), e.g., reported that high anxiety was related to less gaze contact and greater backward head movement in response to male virtual agents, which showed a direct gaze. Furthermore, Dechant et al. (2017) revealed that highly fearful participants showed more avoidance in a social fear virtual paradigm than low fearful participants. It should be noted that avoidance behavior is a crucial element not only in fear learning but also in the maintenance of fear. In this study, we only focused on the

fear learning process. In order to investigate the mechanisms of avoidance behavior in SAD in its entirety, we recommend future research to also study the role of safety behaviors in the maintenance of SAD.

In past studies using stimuli of low ecological validity with regard to the nature of SAD, it remained unclear whether socially fearful persons react more sensitively to socially relevant stimuli. Our social threat condition utilizes social stimuli, which are likely to be disorder-relevant for SAD. Thus, our social threat condition might be more suitable for investigating social anxiety due to a higher belongingness between the CS and the US and consequently an enhanced ecological validity of the design. Furthermore, not using electric shocks may make the recruitment of clinical samples easier for future studies. Empirical findings indicate that successful conditioning in highly fearful individuals cannot only be induced by effective non-social US (i.e., electric shocks), but also by social stimuli, such as emotional facial expressions paired with compatible verbal feedback (Lissek et al., 2008) or isolated verbal comments (Ahrens et al., 2014). In the present study, conditioning was successful and avoidance behavior could be observed in both conditions. Still, there was a better differentiation between aversive and non-aversive stimuli in the social threat condition. One explanation for not having observed an enhanced belongingness effect in our study could be that the high social anxiety group showed a low SPIN score (median score = 13.5) as well. According to Connor et al. (2000) a SPIN score of 19 distinguishes between social phobia subjects and controls.

It is noteworthy that participants undergoing electrical stimulation typically have a more robust fear response both before and after acquisition and extinction (Schmitz and Grillon, 2012) and rate the shock as more aversive than alternative stimuli such as a female scream (Glenn et al., 2012), suggesting that they tend to overestimate the probability of aversive stimuli when being physically harmed. However, this effect could not be found in the contingency ratings, and although the subjective fear ratings before acquisition were generally higher for subjects in the electroshock condition, the fear ratings for the CS+ after acquisition barely differed. Furthermore, we found a better differentiation between the CS+ and the CS− both after the acquisition and the extinction phases in the social threat condition than in the electroshock condition, indicating that the social threat is more realistic than the electroshock condition. These findings partially confirm our hypothesis that acquisition and resistance to extinction are intensified by a sense of belongingness between the CS and the applied US. This is an important fact which should be taken into consideration in future research.

An issue regarding the experimental setting is the linguistic label of the fear ratings. Many subjects reported that it was not actually fear they had experienced, but a feeling comparable with unpleasantness or, especially in the case of the virtual spitting, even disgust. Being spat at might not only induce social fear (as expected for a socially fearful person) but also cause disgust. Still, being spat at along with hearing the agent say "go away" is a social situation that is expected to elicit emotions similar to the ones induced in a social

fearful or phobic patient. In order to investigate if conditioning had caused social fear or simply disgust, we could have asked participants which emotions had been elicited by the conditioning paradigm. Updating the understanding of SAD, future studies should measure disgust and similar emotions. Furthermore, it has to be taken into account that the three virtual agents differed in clothing, hair color and facial design, which might have led to an association of the US with the external stimuli instead of the situation. As a further limitation of the current study, our non-clinical sample was limited to young students with a high proportion of female students, which should be taken into account when generalizing the results to a broader population. However, as social phobia is twice as prevalent in women than in men, females are an interesting target group for our paradigm (Bandelow and Wedekind, 2014).

Despite these facts, all in all our paradigm has been shown to be suitable for investigating the acquisition and extinction of social fear in a VR setting similar to the paradigm used by Shiban et al. (2015). As in this previous work, results support the translation of the SFC paradigm by Toth et al. (2012) from the mice model to human studies. Further research is needed to expand these findings by increasing the sample size and by testing patients suffering from social phobia. Treatment for this widespread health issue could potentially be enhanced by optimizing the extinction process that is strived for in exposure therapy. Furthermore, it is an interesting research question if patients suffering from social phobia could benefit from extinction processes in different contexts as Dunsmoor et al. (2014) could verify for healthy humans.

ETHICS STATEMENT

This study was carried out in accordance with the recommendations of the Ethics Committee of the University of Regensburg with written informed consent from all subjects. All subjects gave written informed consent in accordance with the Declaration of Helsinki. The protocol was approved by the Ethics Committee of the University of Regensburg.

AUTHOR CONTRIBUTIONS

JR study conception, data analysis, wrote the manuscript. SP study conception, data acquisition and analysis, and contribution to the manuscript. JW and VZ data analysis, contribution to the manuscript. YS study conception, data analysis, contribution to the manuscript. All authors have approved of the final version of the manuscript and its submission.

ACKNOWLEDGMENT

REFERENCES

Ahrens, L. M., Mühlberger, A., Pauli, P., and Wieser, M. J. (2014). Impaired visuocortical discrimination learning of socially conditioned stimuli in social anxiety. *Soc. Cogn. Affect. Neurosci.* 10, 929–937. doi: 10.1093/scan/nsu140

American Psychiatric Association (2013). *Diagnostic and Statistical Manual of Mentral Disorders*, 5th Edn. Arlington, VA: American Psychiatric Publishing, doi: 10.1176/appi.books.978089042559

Arch, J. J., Eifert, G. H., Davies, C., Vilardaga, J. C. P., Rose, R. D., and Craske, M. G. (2012). Randomized clinical trial of cognitive behavioral therapy (CBT) versus acceptance and commitment therapy (ACT) for mixed anxiety disorders. *J. Consult. Clin. Psychol.* 80, 750–765. doi: 10.1037/a0028310

Bandelow, B., and Wedekind, D. (2014). Soziale Phobie. *Nervenarzt* 85, 635–644. doi: 10.1007/s00115-013-3955-9

Blumenthal, T. D., Cuthbert, B. N., Filion, D. L., Hackley, S., Lipp, O. V., and Van Boxtel, A. (2005). Committee report: guidelines for human startle eyeblink electromyographic studies. *Psychophysiology* 42, 1–15. doi: 10.1111/j.1469-8986.2005.00271.x

Bohil, C. J., Alicea, B., and Biocca, F. A. (2011). Virtual reality in neuroscience research and therapy. *Nat. Rev. Neurosci.* 12, 752–762. doi: 10.1038/nrn3122

Connor, K. M., Davidson, J. R. T., Churchill, L. E., Sherwood, A., Weisler, R. H., and Foa, E. (2000). Psychometric properties of the Social Phobia Inventory (SPIN). New self-rating scale. *Br. J. Psychiatry* 176, 379–386. doi: 10.1192/bjp.176.4.379

Dechant, M., Trimpl, S., Wolff, C., Mühlberger, A., and Shiban, Y. (2017). Potential of virtual reality as a diagnostic tool for social anxiety: a pilot study. *Comput. Hum. Behav.* 76, 128–134. doi: 10.1016/j.chb.2017.07.005

Dunsmoor, J. E., Ahs, F., Zielinski, D. J., and LaBar, K. S. (2014). Extinction in multiple virtual reality contexts diminishes fear reinstatement in humans. *Neurobiol. Learn. Mem.* 113, 157–164. doi: 10.1016/j.nlm.2014.02.010

Fehm, L., Pelissolo, A., Furmark, T., and Wittchen, H. U. (2005). Size and burden of social phobia in Europe. *Eur. Neuropsychopharmacol.* 15, 453–462.

Glenn, C. R., Lieberman, L., and Hajcak, G. (2012). Comparing electric shock and a fearful screaming face as unconditioned stimuli for fear learning. *Int. J. Psychophysiol.* 86, 214–219. doi: 10.1016/j.ijpsycho.2012.09.006

Hamm, A. O., Vaitl, D., and Lang, P. J. (1989). Fear conditioning, meaning, and belongingness: a selective association analysis. *J. Abnorm. Psychol.* 98, 395–406. doi: 10.1037/0021-843X.98.4.395

Huff, N. C., Zeilinski, D. J., Fecteau, M. E., Brady, R., and LaBar, K. S. (2010). Human fear conditioning conducted in full immersion 3-dimensional virtual reality. *J. Vis. Exp.* 42, 3–29. doi: 10.3791/1993

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

LeDoux, J. E. (2014). Coming to terms with fear. *Proc. Natl. Acad. Sci. U.S.A.* 111, 2871–2878. doi: 10.1073/pnas.1400335111

Lissek, S., Levenson, J., Biggs, A. L., Johnson, L. L., Ameli, R., Pine, D. S., et al. (2008). Elevated fear conditioning to socially relevant unconditioned stimuli in social anxiety disorder. *Am. J. Psychiatry* 165, 124–132. doi: 10.1176/appi.ajp.2007.06091513

Lissek, S., Powers, A. S., McClure, E. B., Phelps, E. A., Woldehawariat, G., Grillon, C., et al. (2005). Classical fear conditioning in the anxiety disorders: a meta-analysis. *Behav. Res. Ther.* 43, 1391–1424. doi: 10.1016/j.brat.2004.10.007

Mesa, F., Beidel, D. C., and Bunnell, B. E. (2014). An examination of psychopathology and daily impairment in adolescents with social anxiety disorder. *PLOS ONE* 9:e93668. doi: 10.1371/journal.pone.0093668

Mineka, S., and Oehlberg, K. (2008). The relevance of recent developments in classical conditioning to understanding the etiology and maintenance of anxiety disorders. *Acta Psychol.* 127, 567–580. doi: 10.1016/j.actpsy.2007.11.007

Mineka, S., and Zinbarg, R. (2006). A contemporary learning theory perspective on the etiology of anxiety disorders: it's not what you thought it was. *Am. Psychol.* 61, 10–26. doi: 10.1037/0003-066X.61.1.10

Mühlberger, A., Bülthoff, H. H., Wiedemann, G., and Pauli, P. (2007). Virtual reality for the psychophysiological assessment of phobic fear: responses during virtual tunnel driving. *Psychol. Assess.* 19, 340–346. doi: 10.1037/1040-3590.19.3.340

Mühlberger, A., Wieser, M. J., and Pauli, P. (2008). Visual attention during virtual social situations depends on social anxiety. *Cyberpsychol. Behav.* 11, 425–430. doi: 10.1089/cpb.2007.0084

Norton, P. J., and Price, E. C. (2007). A meta-analytic review of adult cognitive-behavioral treatment outcome across the anxiety disorders. *J. Nerv. Ment. Dis.* 195, 521–531. doi: 10.1097/01.nmd.0000253843.70149.9a

Pavlov, I. P. (1927). *Conditioned Reflexes. An Investigation of the Physiological Activity of the Cerebral Cortex.* London: Oxford University Press.

Radke, S., Roelofs, K., and de Bruijn, E. R. A. (2013). Acting on anger: social anxiety modulates approach avoidance tendencies after oxytocin administration. *Psychol. Sci.* 24, 1573–1578. doi: 10.1177/0956797612472682

Schmitz, A., and Grillon, C. (2012). Assessing fear and anxiety in humans using the threat of predictable and unpredictable aversive events (the NPU-threat test). *Nat. Protoc.* 7, 527–532. doi: 10.1038/nprot.2012.001

Shiban, Y., Reichenberger, J., Neumann, I. D., and Mühlberger, A. (2015). Social conditioning and extinction paradigm: a translational study in virtual reality. *Front. Psychol.* 6:400. doi: 10.3389/fpsyg.2015.00400

Sosic, Z., Gieler, U., and Stangier, U. (2008). Screening for social phobia in medical in- and outpatients with the German version of the Social Phobia Inventory (SPIN). *J. Anxiety Disord.* 22, 849–859. doi: 10.1016/j.janxdis.2007.08.011

Stangier, U., and Steffens, M. (2002). *Social Phobia Inventory (SPIN)–Deutsche Fassung.* Frankfurt am Main: Psychologisches Institut der Universität Frankfurt am Main.

Toth, I., and Neumann, I. (2013). Animal models of social avoidance and social fear. *Cell Tissue Res.* 354, 107–118. doi: 10.1007/s00441-013-1636-4

Toth, I., Neumann, I. D., and Slattery, D. A. (2012). Social fear conditioning: a novel and specific animal model to study social anxiety disorder. *Neuropsychopharmacology* 37, 1433–1443. doi: 10.1038/npp.2011.329

Wieser, M. J., Pauli, P., Alpers, G. W., and Mühlberger, A. (2009). Is eye to eye contact really threatening and avoided in social anxiety? An eye-tracking and psychophysiology study. *J. Anxiety Disord.* 23, 93–103. doi: 10.1016/j.janxdis.2008.04.004

Wieser, M. J., Pauli, P., Grosseibl, M., Molzow, I., and Mühlberger, A. (2010). Virtual social interactions in social anxiety–the impact of sex, gaze, and interpersonal distance. *Cyberpsychol. Behav. Soc. Netw.* 13, 547–554. doi: 10.1089/cyber.2009.0432

Wolitzky-Taylor, K. B., Horowitz, J. D., Powers, M. B., and Telch, M. J. (2008). Psychological approaches in the treatment of specific phobias: a meta-analysis. *Clin. Psychol. Rev.* 28, 1021–1037. doi: 10.1016/j.cpr.2008.02.007

Zoicas, I., Slattery, D. A., and Neumann, I. D. (2014). Brain oxytocin in social fear conditioning and its extinction: involvement of the lateral septum. *Neuropsychopharmacology* 39, 3027–3035. doi: 10.1038/npp.2014.156

Those Virtual People all Look the Same to me: Computer-Rendered Faces Elicit a Higher False Alarm Rate Than Real Human Faces in a Recognition Memory Task

Jari Kätsyri*

Brain and Emotion Laboratory, Department of Cognitive Neuroscience, Maastricht University, Maastricht, Netherlands

Virtual as compared with real human characters can elicit a sense of uneasiness in human observers, characterized by lack of familiarity and even feelings of eeriness (the "uncanny valley" hypothesis). Here we test the possibility that this alleged lack of familiarity is literal in the sense that people have lesser perceptual expertise in processing virtual as compared with real human faces. Sixty-four participants took part in a recognition memory study in which they first learned a set of faces and were then asked to recognize them in a testing session. We used real and virtual (computer-rendered) versions of the same faces, presented in either upright or inverted orientation. Real and virtual faces were matched for low-level visual features such as global luminosity and spatial frequency contents. Our results demonstrated a higher response bias toward responding "seen before" for virtual as compared with real faces, which was further explained by a higher false alarm rate for the former. This finding resembles a similar effect for recognizing human faces from other than one's own ethnic groups (the "other race effect"). Virtual faces received clearly higher subjective eeriness ratings than real faces. Our results did not provide evidence of poorer overall recognition memory or lesser inversion effect for virtual faces, however. The higher false alarm rate finding supports the notion that lesser perceptual expertise may contribute to the lack of subjective familiarity with virtual faces. We discuss alternative interpretations and provide suggestions for future research.

Keywords: artificial faces, face recognition, face memory, face inversion, uncanny valley hypothesis

INTRODUCTION

Virtual environments and augmented realities are not only changing the way we perceive "reality" but also the way we perceive and interact with its real and virtual inhabitants. Even though many individuals frequently encounter realistic virtual characters in video games and other media (e.g., animation films), most of our perceptual expertise is arguably still shaped by our interactions with our biological companions. For example, parents' faces are among the very first things newborns encounter after being born, and an innate interest in human faces remains characteristic to typically developing children. According to the "uncanny valley" hypothesis (Mori, 1970), artificial entities bearing a near-identical resemblance to real humans elicit a sense of uneasiness, characterized by lack of familiarity and even feelings of eeriness, even though increasingly realistic artificial

characters in general tend to elicit more positive responses. Although empirical evidence for the pronounced unfamiliarity of near-human entities still appears inconsistent, the bulk of studies support the overall positive association between realism and familiarity (Kätsyri et al., 2015). Abundant exposure to real human faces from an early life could possibly explain the lack of subjective familiarity with virtual faces. In this case, human observers should conversely possess lesser perceptual expertise in processing virtual as compared with real faces. In the present study, we investigate whether participants are indeed impoverished in processing virtual faces—that is, faces that are close yet distinguishable computer-generated approximations of real faces.

Several lines of evidence suggest that the human visual system possesses perceptual expertise with faces that is shaped by exposure (even though social-cognitive motivational factors may also play a role; e.g., Bernstein et al., 2007). One of the earliest and best documented examples is the tendency for perceivers to have more accurate recognition memory for faces from one's own ethnic group in comparison to faces from other ethnicities (Meissner and Brigham, 2001; Young et al., 2012). Algorithmic analysis of three-dimensional head scans has provided support for one prerequisite of this effect, the existence of ethnicity-characteristic facial features (O'Toole et al., 1991; Salah et al., 2008). More accurate recognition of own-ethnicity faces has become widely known as the other-race effect, own-race bias, or cross-race effect. Such terms may be misleading, however, given that this effect is not only biologically determined. For example, individuals from one country who were adopted into another country at an early age showed a reversal of the effect such that they recognized faces originating from their adoption country better than faces originating from their birth country (Sangrigoli et al., 2005). In a similar vein, training has been shown to reduce the recognition disadvantage for other-ethnicity faces (e.g., Hills and Lewis, 2006; Tanaka and Pierce, 2009). Guiding participant's attention to features that are characteristic of other-ethnicity faces can also eliminate the effect (Hills et al., 2013). Such findings both exemplify the malleability of the other-ethnicity effect and argue against its biologically or racially determined origins. Furthermore, the term "race" itself has been called into question both in biology and neuroscience because of its inexact and prejudiced nature (Yudell et al., 2016; Cubelli and Della Sala, 2017). Hence, following Valentine et al. (2016), we refer to this phenomenon as the own-ethnicity bias (OEB). Biases resembling the OEB have been demonstrated also for other variables besides ethnicity—for example, men have better recognition memory for male than female faces, whereas the opposite holds true for women (Wright and Sladden, 2003). This suggests that also the processing of own-gender faces may be fine-tuned by possibly greater exposure to same-gender individuals.

Most studies documenting the OEB effect have used a standard old-new recognition memory paradigm in which participants are first asked to memorize a set of faces and then tested for their ability to discriminate between previously seen (target) and previously unseen (distractor) faces (Meissner and Brigham, 2001). Typical findings show a "mirror effect" in which own-ethnicity faces yield a higher proportion of hits (targets

identified as previously seen) and a lower proportion of false alarms (distractors identified as previously seen) as compared to other-ethnicity faces (e.g., Meissner et al., 2005). Inflated false alarm rate for other-ethnicity faces means that people tend to confuse individuals from other ethnic groups readily with one another—a phenomenon which could be characterized anecdotally with the statement "They all look the same to me" (e.g., Ackerman et al., 2006).

The finding that ethnicity modulates not only the proportion of hit rates but the proportion of false alarms as well has been previously explained in the framework of the face-space coding model of Valentine (Valentine, 1991; Valentine et al., 2016). Generally speaking, this model suggests that faces are represented mentally in a multidimensional space. These dimensions can correspond to any features that serve to discriminate between individuals (e.g., mouth shape or inter-ocular distance); however, they are not explicitly defined by the model. Face-space model posits that these dimensions are selected and scaled to optimize discrimination of frequently encountered faces. Hence, these dimensions are optimized for own-ethnicity faces that are by definition encountered frequently but, assuming infrequent encounters with other ethnic groups, they are less efficient for encoding differences between other-ethnicity faces (cf. Valentine, 1991). As a result, different other-ethnicity faces can share identical values on several dimensions, which means that they end up being clustered more densely in the face-space than own-ethnicity faces. Conversely, encountering an other-ethnicity face activates more exemplars in the face-space, which makes it more difficult to determine whether that face was in fact encountered previously or whether it is merely similar to other previously seen faces. According to the model, this ultimately generates a higher proportion of false alarms for other-ethnicity faces as compared with own-ethnicity faces.

Inversion effect, or the slower and much less accurate recognition of upside-down as compared with upright faces, is considered one of the hallmarks of perceptual expertise with faces or other well-learned objects (Maurer et al., 2002). Allegedly, inversion has a greater effect on configural (or holistic; perceiving relations among features) than featural (or piece-meal; processing individual features) processing of faces. A possible alternative explanation based on the face-space model could be that face inversion, similarly as many other impairments (e.g., blurring, adding noise, or presenting photographic negatives), simply introduces noise to face encoding (Valentine, 1991). Although the interaction between the OEB and inversion is not entirely uncontroversial (for a review, see Young et al., 2012), evidence exists for a greater inversion effect in own-ethnicity than other-ethnicity faces (e.g., Rhodes et al., 1989; Vizioli et al., 2011). Such findings are consistent with the notion that individuals possess more perceptual expertise with own- as compared with other-ethnicity faces. Furthermore, they contradict the notion that inversion would simply add noise to face encoding because if this were the case, inversion should elicit even greater impairment on the already impoverished encoding of other-ethnicity faces. Hence, these findings also suggest that other-ethnicity faces may be processed in a more featural or piece-meal fashion than own-ethnicity faces.

We next turn to the question of whether the processing of virtual faces could be similar to other-ethnicity faces when it comes to face encoding; or more specifically, mirror and inversion effects in face recognition. First, however, we note that contemporary computer-rendering methods do not yet tap face processing expertise fully to the same extent than real human faces. Arguably, FaceGen Modeler (Singular Inversions) is one of the most versatile and most commonly used programs for face perception experiments (e.g., Cook et al., 2012; MacDorman et al., 2013; Balas and Pacella, 2015; Crookes et al., 2015). This program can be used to create both reconstructions of real faces and randomly generated novel faces in a parametric space derived from a large number of three-dimensional face scans (Blanz and Vetter, 1999). Recently, Crookes et al. (2015) contrasted the OEB for real and FaceGen-generated virtual faces using face recognition memory and perceptual discrimination tasks. Their results demonstrated reduced accuracy for virtual faces in both tasks, and an attenuated OEB for virtual as compared with real faces in the recognition memory task. These findings hence show that virtual faces based on FaceGen software are close but not perfect reconstructions of real human faces, and that they elicit a similar but weaker OEB effect than real faces. In a similar recent study, Balas and Pacella (2015) contrasted recognition memory and discrimination accuracy between virtual and real faces, where the former were again generated by FaceGen. Their results demonstrated that participants were less accurate in recognizing virtual faces in comparison to real faces. Similarly, participants were less accurate in matching two faces to an immediately preceding face image in an ABX matching task.

Even though these two studies demonstrate that FaceGen-generated virtual face stimuli perform less efficiently than real human faces, it is questionable whether their results can be generalized to other virtual faces as well. An important distinction between other-ethnicity faces and virtual faces is that whereas other-ethnicity faces may possess genuine ethnicity-characteristic features (cf. O'Toole et al., 1991; Salah et al., 2008), virtual faces are recognized as "virtual" only when they fail to replicate some characteristics of their reference stimuli (real faces). For example, it is possible that FaceGen-generated virtual faces are artifactual or less detailed replications of real faces, or that they differ from real faces in terms of brightness, contrast, or colors. The extent to which such trivial low-level differences could explain previously observed differences between real and virtual faces is presently not known.

An unfortunate characteristic of all virtual faces is that they can in fact have very little in common. This raises the question of whether it is at all justifiable to consider virtual faces as a unified category of research stimuli. Previous studies investigating continua from virtual to real faces have, however, shown that virtual faces are perceived categorically; that is, equally spaced image pairs are discriminated better when they straddle the virtual–real category boundary than when they reside on the same side of it (Looser and Wheatley, 2010; Cheetham et al., 2011). Changes in virtual–real category in sequentially presented faces are also known to elicit fMRI responses in category learning and uncertainty related neural networks (Cheetham et al., 2011). These findings suggest that virtual and real faces

are typically perceived as distinct categories, similarly as faces of different species (Campbell et al., 1997) or faces of different ethnic groups (Levin and Angelone, 2002). Furthermore, exposure may also modulate categorization and evaluation of virtual faces. Burleigh and Schoenherr (2015) demonstrated that more frequent exposure to specific morph levels between two computer-generated faces improves categorization accuracy for these levels. Frequency-based exposure was also found to modulate participants' subjective ratings, albeit at a statistically non-significant level.

In the present investigation, we operationalize virtual faces using FaceGen but also correct them for most obvious artifacts, and match real and virtual faces with respect to specific low-level visual features. The purpose of this procedure is to increase the generalizability of present results beyond that of a specific computer-rendering method. A justifiable concern after such matching procedure, however, is whether real and virtual faces can still be discriminated from each other. Trivially, if computer-generated images were sufficiently similar to real images, the two would be indistinguishable from each other even by experts (cf. Lehmuskallio et al., 2018).

STUDY 1

In this study, we first investigate whether real and virtual face images can be differentiated from each other even after they have been matched for the following low-level visual features: spatial frequency contents (level of details), brightness, contrast, and colors. Most obvious artifacts are also removed from the virtual faces. Importantly, this matching is done for whole images, that is, at global level. It is possible that even after such global-level matching, local features such as the shapes of individual features may serve to differentiate between real and virtual faces. Subtle artifacts may also remain in the local features of virtual faces. Furthermore, it is possible that low-level visual features still vary at the local level after they have been matched globally. For example, it is possible that nose and eye region brightness might differ in two images even though their averages remained the same. Conversely, we predict that real and virtual faces can still be differentiated from each other based on any of such local differences. Hence, we make the following hypothesis for Study 1:

> H1: Real and virtual faces can be differentiated from each other, even after global-level matching for spatial frequency contents, brightness, contrast, and colors.

In practical terms, colors add extra complications to psychophysical experiments given that one has to consider matching three color channels between images instead of only one luminosity channel. Hence, our secondary research question is whether colors truly contribute to differentiating virtual from real faces. Previous studies suggest that real and virtual faces are easier to discriminate from color than grayscale images (Fan et al., 2012, 2014; Farid and Bravo, 2012). However, given that these studies used different image sets for real and virtual faces, it is conceivable that these results would reflect differences between

the employed image samples. Hence, we also aim to test the following secondary hypothesis.

> H2: Real and virtual faces are discriminated better from color than grayscale images.

Methods

Participants

Participants were 48 (29 women) university students whose age ranged from 18 to 30 years ($M = 20.9$ years). All participants identified themselves as Caucasian in ethnic origin. Participants signed to the study anonymously using the SONA system (http://www.sona-systems.com) of Maastricht University, and received course credit in compensation for their participation. All participants gave written informed consent in accordance with the Declaration of Helsinki. The present studies were reviewed and approved by the ethics committee of the Faculty of Psychology and Neuroscience.

Design

The study had a 2 (face type: real, virtual) × 2 (spatial frequency matching: strict, lenient) × 2 (colors: grayscale, color) within-subjects design.

Stimuli

Research stimulus samples are shown in **Figure 1**. Real face stimuli were 12 neutral face images (half female) from Glasgow (Burton et al., 2010) and Radboud (Langner et al., 2010) face image sets. Virtual face stimuli were created using FaceGen Modeler (Singular Inversions; Version 3.13). Real faces (frontal images only) were imported into FaceGen, and an initial alignment was provided using a number of feature points. Reconstructed and original faces were aligned and matched with each other to the extent possible with respect to small variations in head position, gaze direction, and facial expression. Major artifacts (in particular, black line between the lips) were corrected in Photoshop (Adobe; Version CS6). All images were oval-masked to conceal external features (ears and hair), which would otherwise have been clearly unrealistic in the virtual stimuli. Final images were 246 × 326 pixels in size.

All further image manipulations were carried out in Matlab (The Mathworks Inc.; Version R2016a). Grayscale images were produced by weighting original RGB channel values. Inhouse functions based on SHINE toolbox (Willenbockel et al., 2010) were used for standardizing images. Two methods were used for matching energy at different spatial frequencies across the images: matching the whole Fourier spectra ("strict matching") and matching only the rotational average of the Fourier spectra ("lenient matching")—for details, please refer to Willenbockel et al. (2010). We used the latter matching procedure in place of original (non-matched) images, given that leniently matched and original images were practically identical and led to similar results in pilot tests. Prior to spatial frequency matching, image backgrounds were substituted by the average pixel intensity values within the masked face regions to reduce sharp transitions in the images. Mean and standard deviations for the pixel values within the masked

region were standardized across images, and backgrounds in the final images were substituted with a constant gray color. For color images, image matching was carried out separately for each RGB channel (cf. Kobayashi et al., 2012; Railo et al., 2016).

Procedure

This study was carried out as an online evaluation, which was programmed and hosted through Qualtrics platform (http://www.qualtrics.com). Only participants using a laptop or a desktop computer with a sufficiently large display (minimum 12") were included. A total of 96 stimuli (8 conditions × 12 actors) were presented in a pseudo-randomized order. Participants were asked to identify whether each stimulus portrayed a human or a virtual face in a one-interval forced choice task with two response alternatives. Participants were also asked to indicate how confident they were of their choice using a 5-step Likert scale (1—uncertain, 2—somewhat uncertain, 3—somewhat certain, 4—certain, 5—absolutely certain). The questionnaire was self-paced, but participants were instructed to answer each question as quickly and as accurately as possible. Participants were required to carry out the questionnaire in a single session without breaks.

Preprocessing

Hit and false alarm rates for the identification task were transformed into sensitivity index d' and response bias index c, calculated according to signal detection theory using the following standard formulae (Stanislaw and Todorov, 1999; Chapter 2 in Stevens and Pashler, 2002).

$$d' = z(H) - z(F)$$
$$c = -\frac{1}{2}(z(H) + z(F))$$

Here, hit rate (H) refers to the proportion of real faces identified correctly as human, and false alarm rate (F) refers to the proportion of virtual faces identified incorrectly as human. Following the guidelines of Stanislaw and Todorov (1999), H and F were corrected using log-linear method to avoid incalculable values. In the present study, d' reflects the extent to which participants were able to differentiate between real and virtual faces. Theoretically, c can be understood as the difference between participants' response criterion and neutral point where neither response alternative is favored. In the present one-interval task, response criterion can be interpreted in terms of "human" responses. Positive values refer to more conservative response criterion or a tendency to respond "virtual," whereas negative values refer to more liberal response criterion or tendency toward responding "human" for all faces.

Results and Discussion

Results for different conditions are illustrated in **Figure 2**. For testing H1, we first compared d' scores to zero using one-sample T-tests. Test results showed that d' scores were significantly above zero in all conditions, $T_{(47)} > 10.01$, $p < 0.001$, Cohen's $d > 1.44$, which indicates that real and virtual faces were clearly differentiated from each other
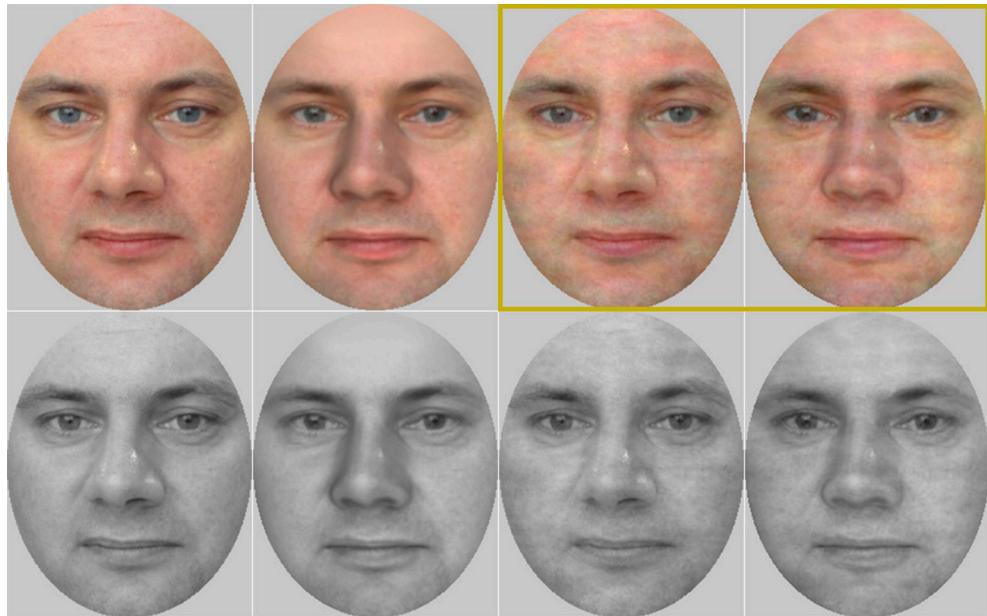
FIGURE 1 | Sample images for one actor extracted from the Glasgow face set (Burton et al., 2010). Upper row: color images, bottom row: grayscale images. From left to right: real face after lenient spatial frequency matching, virtual face after lenient matching, real face after strict matching, and virtual face after strict matching. Only strictly matched color images (surrounded by dark-yellow square on the top-right corner) were used in Study 2.
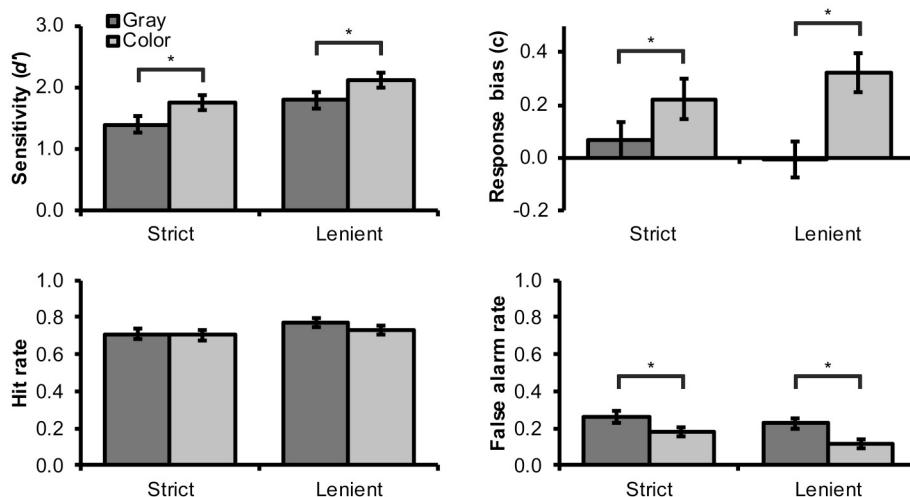
FIGURE 2 | Mean (and SEM) values for d' sensitivity index, c response bias index, hit rate, and false alarm rate for the identification task by color and spatial frequency matching conditions. Statistically significant differences between color and grayscale images are denoted with an asterisk ("*").

in all experimental conditions. Next, a 2×2 within-subjects ANOVA was used to assess the influence of color and spatial frequency matching on d' scores. Significant main effects were observed for spatial frequency matching, $F_{(1, 47)} = 26.08$, $p < 0.001$, $n_p^2 = 0.36$, and color, $F_{(1, 47)} = 25.34$, $p < 0.001$, $n_p^2 = 0.35$. Strict matching elicited lower d' sensitivity scores than lenient matching (**Figure 2**). As predicted by H2, color images elicited higher d' scores than grayscale images.

For completeness, we also analyzed response bias values using similar analysis. We observed a significant main effect for color, $F_{(1, 47)} = 25.17$, $p < 0.001$, $n_p^2 = 0.35$, and a significant interaction between spatial frequency matching and color, $F_{(1, 47)} = 5.25$, $p = 0.027$, $n_p^2 = 0.10$. Specifically, color images elicited higher c values (bias toward responding "virtual") than grayscale images. This effect was weaker for strictly than leniently filtered images (**Figure 2**), which may suggest that it was partly obscured by the strict filtering procedure (however, similar effect was

not observed for false alarm rates; see below). To understand these results better, we next analyzed hit and false alarm rates individually. Color images elicited a lower proportion of false alarms than grayscale images, $F_{(1, 47)} = 30.38$, $p < 0.001$, $n_p^2 = 0.39$. In other words, when shown in color, virtual faces were mistaken less frequently for real faces. There was also a non-significant tendency toward a higher false alarm rate for strictly rather than leniently filtered faces, $F_{(1, 47)} = 3.91$, $p = 0.054$, $n_p^2 = 0.08$. Interaction between spatial frequency matching and color was not significant, $F_{(1, 47)} < 1$, $p = 0.432$, $n_p^2 = 0.01$. No significant effects were observed for hit rates, which suggests that both sensitivity and response bias findings were driven mainly by changes in false alarm rates. We interpret these results to mean that colors are particularly important for recognizing virtual faces as artificial but have a smaller role for the correct recognition of real faces as human.

Confidence ratings were additionally analyzed using a 2 (color) $\times$ 2 (spatial frequency matching) $\times$ 2 (face type) within-subjects ANOVA. The results showed a significant main effect for spatial frequency matching, $F_{(1, 47)} = 75.30$, $p < 0.001$, $\eta_p^2 = 0.62$, and a significant interaction effect for color and face type, $F_{(1, 47)} = 7.19$, $p = 0.010$, $n_p^2 = 0.13$. Strict as compared with lenient spatial frequency matching elicited generally lower confidence ratings regardless of face type ($M = 3.47$ and 3.77, $SD = 0.48$ and 0.49). Simple effect tests showed that confidence ratings for color and grayscale images differed only for virtual faces ($p = 0.003$). Specifically, participants rated higher confidence when categorizing virtual faces from color rather than grayscale images ($M = 3.81$ and 3.56, $SD = 0.49$ and 0.55). This finding further corroborates the importance of colors for recognizing virtual faces.

Not surprisingly, the present results showed that the "strict" spatial frequency matching procedure elicited lower discrimination performance and lower confidence ratings than the "lenient" procedure, which we considered analogous to unmatched stimuli. At the same time, our results confirmed that highly realistic virtual faces (cf. **Figure 1**) could still be differentiated from real human faces relatively easily even after the strict matching procedure. Given that any experimental comparison between unmatched real and virtual faces would be confounded by differences in spatial frequency contents (e.g., overall lack of details in virtual faces), we hence decided to adopt the strict matching procedure for our second experiment. Our other findings replicate the previous finding (Fan et al., 2012, 2014; Farid and Bravo, 2012), with slightly better controlled stimuli, that real and virtual faces are differentiated better and with higher confidence from color as compared with grayscale images. A closer inspection of false alarm rates as well as participants' confidence ratings suggested that colors are particularly important for the correct recognition of virtual faces. Interestingly, visual inspection of **Figure 2** would suggest that color has a roughly similar effect on discrimination accuracy than the present choice of spatial frequency matching. Hence, we conclude that adopting color rather than grayscale images can be used to compensate for the loss of discrimination accuracy caused by strict spatial frequency matching.

STUDY 2

In the second study, we continue to investigate whether our rigorously matched virtual faces tap perceptual expertise similarly as real human faces. We expect to observe a similar mirror pattern as in previous OEB studies in which other-ethnicity faces elicited both a lower proportion of hits and a higher proportion of false alarms than own-ethnicity faces (e.g., Meissner et al., 2005). Following this pattern, aggregate measures based on hits and false alarms have previously indicated lower discrimination accuracy (discrimination between previously seen and novel faces) and lower response bias (overall tendency to respond "previously seen" to all faces) for other-ethnicity faces. We predict analogous effects for virtual faces. That is,

> H1: Virtual faces will elicit lower discrimination accuracy than real faces.
> H2: Virtual faces will elicit lower response bias than real faces.

Previous findings suggest that inflated false alarm rate for other-ethnicity faces—or the "They all look the same to me" phenomenon—is a major factor driving the OEB effect. One explanation for this is that facial encoding dimensions in the face-space model of Valentine (Valentine, 1991; Valentine et al., 2016) are optimized for discriminating frequently seen own-ethnicity faces but that they are suboptimal when it comes to the discrimination of other-ethnicity faces. Assuming that virtual faces contain sufficiently different or distorted features with respect to real human faces, we predict a similar effect for virtual faces as well. That is, we predict that:

> H3: Virtual faces will elicit a higher proportion of false alarms than real faces.

In the present study, we also investigate the effect of inversion on the recognition of virtual faces. In their previous study, Balas and Pacella (2015) observed an equally large inversion effect for virtual and real faces in a perceptual discrimination task. Performance was close to ceiling level for both upright and inverted faces, however, which leaves open the possibility that a more difficult task might be more sensitive to differential inversion effects in real and virtual faces. A diminished inversion effect for virtual faces could be taken as evidence that virtual faces are processed in a more piece-meal and less "face-like" manner than real faces. Here we test the following prediction:

> H4: Real faces will elicit a greater inversion effect as measured with discrimination accuracy than virtual faces.

Previous factor-analytic research on participants' self-reports have demonstrated that the typicality (or distinctiveness) of faces is composed of two orthogonal components: memorability and general or context-free familiarity (Vokey and Read, 1992; Meissner et al., 2005). For the present context, it is interesting that the latter factor combines familiarity with attractiveness and likability. That is, faces resembling frequently encountered faces evoke not only a heightened sense of familiarity, but more

favorable evaluations as well (Vokey and Read, 1992). Even more interestingly, own-ethnicity faces are known to receive higher ratings in terms of these items than other-ethnicity faces (Meissner et al., 2005). One way to interpret this is that familiarity with specific kinds of faces breeds more positive affects, which could also explain why all virtual faces appear more strange and unpleasant—or even eerie—than real human faces (Kätsyri et al., 2015). Another line of research has demonstrated that inversion can eliminate grotesqueness caused by distorted configural features. In particular, this seems to be the case for the so-called Thatcher illusion, in which eyes and mouth are flipped vertically (Stürzel and Spillmann, 2000). If typical human features are distorted in virtual faces, virtual faces should elicit less favorable evaluations than real human faces. Furthermore, if these features are at least partly configural in nature, inversion should reduce their effects. These two hypotheses are stated explicitly below.

H5: Virtual faces receive higher eeriness ratings than real faces.
H6: Inversion decreases the eeriness of virtual as compared with real faces.

Methods

Participants

Participants were 64 (32 men and 32 women) university students or university graduates in the age range 18 to 36 years ($M = 22.6$ years). Participants were recruited via the SONA system of Maastricht University, flyers placed in the campus, and social media. Two original participants who scored high on PI20 prosopagnosia self-report questionnaire (Shah et al., 2015) and additionally received low overall scores in the present recognition memory task were excluded and replaced with new participants. Male and female participants did not differ statistically significantly on PI20 scores ($M = 39.1$ and 40.7, $SD = 8.1$ and 7.7), $T_{(62)} = 0.84$, $p = 0.407$. The majority (89%) of participants reported having played video games with realistic human-like characters at most once per month during the last year. That is, most participants had little experience with realistic virtual characters. All participants identified themselves as Caucasian in ethnic origin. Participants received a 7.5 € voucher in compensation for their participation. All participants gave written informed consent in accordance with the Declaration of Helsinki. The present studies were reviewed and approved by the ethics committee of the Faculty of Psychology and Neuroscience.

Stimuli

Research stimuli were 80 neutral face images (half female) from Glasgow (Burton et al., 2010) and Radboud (Langner et al., 2010) face image sets, replicated both as real and virtual versions. Face images were selected on the basis of distinctiveness preratings (cf. Valentine, 1991; McKone et al., 2007) from a larger set of 100 face images. These initial images were oval-masked and matched for luminance, contrast and colors but not for spatial frequency contents. Twenty-five participants who did not take part in the actual study rated the images for distinctiveness on a 7-step semantic differential scale ranging from "very typical/very difficult to recognize" to "very distinctive/very easy

to recognize." Twenty images were dropped on the basis of individual consideration and the remaining 80 images were divided evenly into eight stimulus sets based on their mean ratings. Finally, the selected images were replicated as real and virtual versions and matched for low-level features similarly as the strictly matched color images in Study 1 (**Figure 1**).

Procedure

The present study design was adapted from two previous OEB studies that included both face ethnicity and inversion as factors (Rhodes et al., 1989; Vizioli et al., 2011). In particular, participants completed standard recognition memory tasks separately for real and virtual faces, with the task order counterbalanced across participants. Recognition memory tasks for real and virtual faces were separated by a 2-min break. Both tasks consisted of a study and a test phase. During the study phase, participants were asked to view and memorize 20 faces presented in a pseudo-randomized order. Each face was presented for 5 s and preceded by a fixation cross for 2 s. All study faces were shown in upright orientation.

In the test phase, the 20 old faces (seen during the study phase) were interleaved with 20 new faces, and all faces were presented in a pseudo-randomized order. Half of the images were shown in upright orientation and the other half in inverted (rotated 180°) orientation. Participants were instructed to answer as quickly and as accurately as possible whether they had seen each face during the study phase or not using response buttons "S" and "L" on the keyboard. The assignment of response buttons was counterbalanced across participants. Each image remained on the screen until a response was received from the participant, and images were separated by 2-s fixation cross trials. Participants saw only real or virtual faces during the same study-test cycle. The eight stimulus sets were counterbalanced with the face type, trial type, and orientation conditions. Male and female participants were assigned evenly into counterbalancing conditions. Prior to the actual recognition memory tasks, participants practiced the study-test procedure with 20 faces which were not included in the actual study.

After the recognition memory tasks, participants were asked to evaluate how human-like and eerie the faces appeared on a 7-step Likert scale ranging from total disagreement to total agreement. Eeriness was defined as "being so mysterious, strange, or unexpected as to send a chill up the spine." Participants rated the same 80 faces they had seen during the memory tasks, each with the same face type (real or virtual) and orientation (upright or inverted). To test whether the order of human-likeness and eeriness ratings would bias the results, participants gave these ratings in either separate blocks beginning from human-likeness (16 participants), in separate blocks beginning from eeriness (16 participants), or simultaneously in the same block (32 participants). In the former two conditions, human-likeness and eeriness were only explained prior to the beginning of their respective blocks. Male and female participants were assigned evenly into these conditions. All tasks were programmed and presented using E-Prime 2.0 (Psychology Software Tools, Pittsburgh, PA), and displayed on a 24" Asus VG248QE monitor.

Preprocessing

Hit and false alarm rates were transformed into d' sensitivity and c response bias indices similarly as in Study 1. In this study sensitivity d' reflects the extent to which participants were able to differentiate between old (seen during the study phase) and new (not seen) faces, whereas response bias c refers to the general tendency to respond "seen" or "not seen." Positive c values refer to more conservative response criterion or a tendency to respond "not seen" for all faces, whereas negative values refer to more liberal response criterion or tendency toward responding "seen" for all faces.

Results

Recognition of Real and Virtual Faces

We used 2×2 within-subjects ANOVAs to analyze the effects of face type and orientation on sensitivity (d') and response bias (c) indices on the one hand, and hit and false alarm rates on the other. We also tested whether any of these indices were influenced by the order of real and virtual face blocks but failed to observe any significant effects for block order or its interaction with face type ($p > 0.270$, $\eta_p^2 < 0.02$). This suggests that block order did not exert substantial generic or face type specific effects in the present study. Given that we had clear *a priori* predictions for our results, we did not adopt multiple-comparison correction in further analyses.

Recognition memory results are illustrated in **Figure 3**. Although visual inspection of this figure suggests that sensitivity scores were slightly higher for real as compared with virtual faces, as predicted by H1, this effect failed to reach statistical significance, $F_{(1, 63)} = 2.75$, $p = 0.102$, $\eta_p^2 = 0.04$.

Hypothesis H2 predicted a more lenient response bias (i.e., lower c scores) for virtual faces. **Figure 3** suggests that response bias may have been less conservative for virtual than for real faces, but only in the upright condition. Given that face inversion exerted a considerable impairment on the processing of faces (see below), stronger response bias effects should in fact have been expected particularly for upright faces. In support, we observed a borderline significant interaction effect between face type and inversion, $F_{(1, 63)} = 3.94$, $p = 0.052$, $\eta_p^2 = 0.059$. Consequently, we decided to test H2 specifically for upright faces. This analysis confirmed a statistically significant and moderately large (Cohen, 1992) effect for face type in upright faces, $F_{(1, 63)} = 4.40$, $p = 0.040$, $\eta_p^2 = 0.650$, but not in inverted faces, $F_{(1, 63)} = 0.78$, $p = 0.381$, $\eta_p^2 = 0.012$.

Following the above logic, we next tested the effect of face type on false alarm rates in upright condition. In support of H3, our results demonstrated a significantly higher false alarm rate with a moderate effect size for virtual rather than real faces in upright orientation (see **Figure 3**), $F_{(1, 63)} = 6.14$, $p = 0.016$, $\eta_p^2 = 0.089$, but not in inverted orientation, $F_{(1, 63)} = 0.00$, $p = 1.000$, $\eta_p^2 = 0.00$. For hit rate, the effect of face type was not significant in either upright orientation, $F_{(1, 63)} = 0.33$, $p = 0.568$, $\eta_p^2 = 0.005$, or in inverted orientation, $F_{(1, 63)} = 1.75$, $p = 0.191$, $\eta_p^2 = 0.027$. These findings suggest that the more lenient response bias for upright virtual faces was driven mainly by false alarm responses, that is, participants' higher tendency

to answer "seen before" to novel virtual faces. The 95% CI for the false alarm rate difference between virtual and real faces was [0.01, 0.10].

Inversion had a statistically significant and large effect on sensitivity, $F_{(1, 63)} = 94.73$, $p < 0.001$, $\eta_p^2 = 0.601$, and response bias, $F_{(1, 63)} = 15.39$, $p < 0.001$, $\eta_p^2 = 0.196$. As can be seen in **Figure 3**, inverted faces received lower sensitivity scores and more liberal response criterion (lower c scores). Looking at this the other way, inverted faces received moderately lower hit rates, $F_{(1, 63)} = 5.58$, $p = 0.021$, $\eta_p^2 = 0.081$, and much higher false alarm rates, $F_{(1, 63)} = 78.67$, $p < 0.001$, $\eta_p^2 = 0.555$, than upright faces (**Figure 3**, lower panels). For inverted and upright faces, the 95% CI for the false alarm rate difference was [0.15, 0.24].

Contrary to H4, the interaction effect between face type and inversion on d' was not statistically significant, $F_{(1, 63)} = 0.15$, $p = 0.700$, $\eta_p^2 = 0.002$. Conversely, simple tests confirmed a significant and large inversion effect for both real, $F_{(1, 63)} = 70.69$, $p < 0.001$, $\eta_p^2 = 0.529$, and virtual faces, $F_{(1, 63)} = 41.63$, $p < 0.001$, $\eta_p^2 = 0.398$.

Self-Report Ratings

We first tested whether rating order (human-likeness first, eeriness first, or both together) had significant main or interaction effects for face type at a lenient significance threshold of $p < 0.100$. Because no significant effects were observed for either human-likeness ($p > 0.128$) or eeriness ($p > 0.440$), this confound variable was dropped from further analyses. Hence, self-report ratings were analyzed using a 2 (face type) $\times$ 2 (inversion) within-subjects ANOVA.

Human-likeness and eeriness ratings are illustrated in **Figure 4**. Real as compared with virtual faces received significantly higher human-likeness ratings with a large effect size, $F_{(1, 63)} = 78.41$, $p < 0.001$, $\eta_p^2 = 0.554$. That is, similarly as in our pretest, participants were clearly able to discriminate virtual from real faces. There was also a significant interaction between face type and inversion such that inversion decreased the human-likeness difference between real and virtual faces (cf. **Figure 4**), $F_{(1, 63)} = 31.25$, $p < 0.001$, $\eta_p^2 = 0.332$. Looking at this the other way, real faces received lower human-likeness ratings when inverted ($p < 0.001$), whereas inversion did not have a statistically significant effect on virtual faces ($p = 0.083$).

In H5, we predicted that virtual faces would receive higher eeriness ratings than real faces. This prediction was confirmed, given that the difference between virtual and real faces was statistically significant and large, $F_{(1, 63)} = 40.34$, $p < 0.001$, $\eta_p^2 = 0.390$. Finally, in H6 we predicted that inversion would reduce or eliminate the eeriness of virtual faces. At first sight, this hypothesis appeared to receive support, given that the interaction between face type and inversion was significant with a moderate effect size, $F_{(1, 63)} = 5.50$, $p = 0.022$, $\eta_p^2 = 0.080$. However, as can be seen in **Figure 4**, inversion in fact increased rather than decreased eeriness for both virtual ($p = 0.009$) and real faces ($p < 0.001$). Apparently, the interaction effect was significant because this increase was greater for real rather than virtual faces and not because inversion decreased the eeriness of virtual faces in particular.
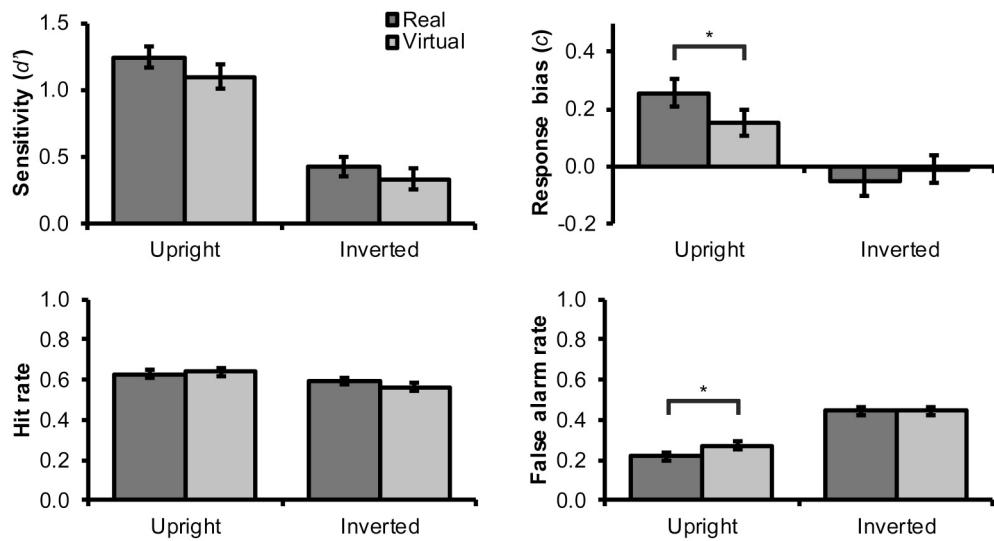
FIGURE 3 | Mean (and SEM) values for d' sensitivity index, c response bias index, hit rate, and false alarm rate for the recognition memory task. Statistically significant differences between virtual and human faces are denoted with an asterisk ("*").
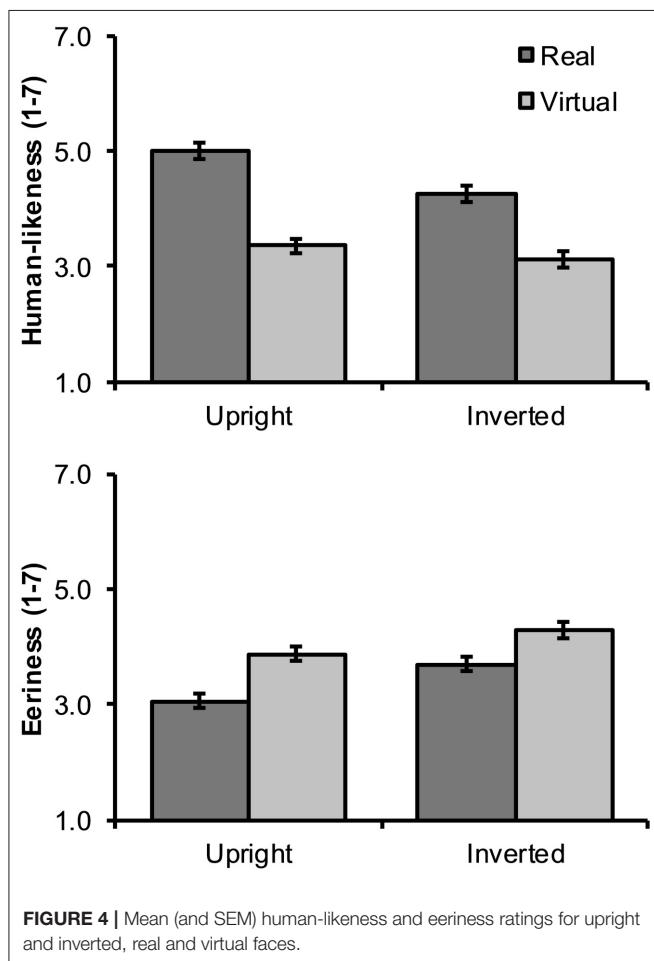
FIGURE 4 | Mean (and SEM) human-likeness and eeriness ratings for upright and inverted, real and virtual faces.

GENERAL DISCUSSION

In the present investigation, we set to find out whether highly-realistic virtual faces tap perceptual expertise similarly as real human faces. Unlike faces of different ethnic groups in humans, virtual, and real faces tend to differ with respect to low-level visual features, which might contribute to differences in perceptual processing. In Study 1, we demonstrated that virtual faces can still be differentiated from real faces even after these two types of faces have been matched for spatial frequency contents, brightness, contrast, and colors. We interpret this to mean that individuals are able to use local features or their configurations to decipher whether a face is real or virtual. In Study 2, we showed that in a recognition memory task, virtual as compared with real faces elicit a less conservative response bias and a higher proportion of false alarms. Virtual and real faces did not differ with respect to discrimination accuracy or the magnitude of inversion effect, however.

The present findings resemble OEB findings in recognition memory studies with real human faces. Such studies have, however, typically identified a mirror pattern in which other-ethnicity faces receive both a lower proportion of hits and a higher proportion of false alarms than own-ethnicity faces (Meissner and Brigham, 2001). This mirror pattern has also been seen as lower discrimination sensitivity in the aggregate index that pits hits against false alarms. In contrast, we observed a difference in response bias but not in discrimination sensitivity. Given that the aggregate response bias measure depends positively on both hits and false alarms, and virtual as compared with human faces elicited a higher proportion of false alarms with a slight tendency toward higher proportion of hits as well (cf. **Figure 3**), this pattern of results is not surprising.

Importantly, the higher proportion of false alarms for virtual faces was predicted on the basis of the highly influential face-space model (Valentine, 1991; Valentine et al., 2016). Here the reasoning was that individuals' hypothetical face-space representation is optimized for real human faces, and that this representation is not necessarily appropriate for encoding virtual faces whose features or feature configurations differ from those of real faces. Similarly as for other- vs. own-ethnicity faces, differences between virtual faces are hence encoded imperfectly, which leads to a denser representation in the face-space. When individuals are making judgments in a recognition memory task, virtual faces then allegedly activate more face exemplars than equivalent real faces, which leads to a false sense of familiarity and a higher proportion of false alarms. The present study hence suggests that, similarly as other- vs. own-ethnicity faces, virtual faces tap perceptual expertise less efficiently than real faces. This effect is particularly evident in false alarm choices. The present study hence makes a contribution to existing research literature by demonstrating this theoretically predicted false alarm effect for virtual faces.

The present investigation is similar to that of Balas and Pacella (2015), given that both they (in their Experiment 1) and we (in Study 2) carried out a recognition memory task for real and virtual faces. The major difference between these studies is that we used stimuli that were matched for low-level features, spatial frequency contents in particular. The present results suggest that such matching eliminates the discrimination advantage for real faces observed by Balas and Pacella. In contrast, their results did not support different response bias or false alarm effects for real and virtual faces. We suggest that this difference originated from other methodological differences. First of all, the present study may have had higher statistical power for detecting a response bias effect because of a higher number of participants (64 against 18) and a within- rather than between-subjects design. Second, the response bias effect may have been more pronounced in the present study because of the less demanding recognition memory task (with 40 instead of 90 faces). The present investigation also differed from the study by Balas and Pacella because we studied inversion effects in a recognition memory task and considered the subjective evaluations of virtual and real faces.

Previous research evidence gives reason to believe that inversion effect is a hallmark of perceptual expertise for faces and other well-learned stimuli (Maurer et al., 2002), and that this effect is stronger for own- as compared with other-ethnicity faces (e.g., Rhodes et al., 1989). Unexpectedly, Balas and Pacella (2015; Experiment 2) demonstrated a similar inversion effect for real and virtual faces in a perceptual discrimination task, possibly due to ceiling effects in their results. The present study replicates this finding in a more difficult and different (recognition memory) task. If inversion effect is a hallmark of perceptual expertise, why did inversion then elicit roughly equal degradation on real and virtual faces? Similarly as Balas and Pacella (2015), we suggest that the human visual system processes virtual faces in a highly face-like manner. This statement is perhaps particularly uncontroversial for such highly realistic virtual stimuli as those used in the present study (cf. **Figure 1**). Inversion had a drastic overall effect on the proportion of false alarms (lower 95% CL for the difference 15 percentage units), which was clearly larger

than the effect of face type in upright faces (upper 95% CL for the difference 10 percentage units). Hence, we suggest that inversion compromised face processing to the extent of concealing the more subtle processing differences between real and virtual faces.

Given that face inversion is thought to influence configural processing more than featural processing, the observed findings do not support the suggestion that virtual faces would be processed in a less configural manner than real faces. However, although this was not a specific aim in the present study, the human-likeness ratings from Study 2 suggest that configural and featural information may have played a different role on the recognition of human-likeness in the case of real and virtual faces. Specifically, our results showed that inversion elicited decreased human-likeness ratings for real faces but had lesser or no influence on virtual faces. This suggests that configural processing, which was impaired by inversion, was important for identifying real faces as human. On the other hand, virtual faces were still recognizable as non-human after inversion, plausibly because this judgement was mainly based on individual features. There is some previous evidence suggesting that eyes could be a particularly important feature for differentiating real from virtual faces (Looser and Wheatley, 2010).

Overall, the present self-report findings from Study 2 confirm the previous observation that virtual faces are always considered more eerie than real faces. The results also demonstrated that this difference is smaller for inverted than for upright faces. At first sight, this seemed to support the prediction that inversion can eliminate the eeriness of virtual faces similarly as with "Thatcherized" faces (Stürzel and Spillmann, 2000). However, a closer inspection of our results showed that inversion elicited increased eeriness for both real and virtual faces but that this increase was larger for real faces. It is plausible that the overall heightened eeriness for inverted faces reflected more effortful processing caused by increased encoding error (Valentine et al., 2016). Furthermore, human-likeness ratings suggested that inversion had a larger effect on the categorization of real as compared with virtual faces. Given that eeriness ratings closely parallel these findings, it is possible that inversion had a differential effect on real and virtual faces simply because inverted real faces were more difficult to recognize as human than upright real faces. Hence, the present findings cannot be taken as support for the prediction that inversion would eliminate the eeriness of virtual faces by reducing configural differences between virtual and real faces.

We want to address some potential limitations of the present investigation and to suggest directions for future research. First, similarly as Balas and Pacella (2015) and Crookes et al. (2015), we used FaceGen software as the basis for our virtual stimuli. However, unlike them, we additionally matched virtual and real faces with respect to various low-level visual features. It could be argued that after this matching, the present virtual stimuli were no longer representative of typical virtual faces. We want to emphasize, however, that the above two studies have already demonstrated the limits of typically used stimuli (e.g., those generated by FaceGen), and that our aim was instead to test whether real and virtual faces are still processed differently after they have been matched for most obvious low-level visual confounds. Hence, the important question is not whether our

stimuli were high in mundane realism (i.e., whether they were similar to modern computer-rendered faces) but whether they were high in psychological realism (i.e., whether they tapped psychological processes relevant for perceiving animacy in faces) (cf. Shadish et al., 2002). This question was addressed in Study 1, which clearly showed that the present stimuli were perceived distinctly as human and non-human stimuli.

Nevertheless, we want to acknowledge other confounds that could still have influenced the present stimuli even after the matching procedure. Because virtual stimuli were generated by replicating real faces in the FaceGen software's parametric space, it is possible that virtual faces or some of their features (e.g., nose shapes) might have been more similar to each other than was the case for original faces. This reduced variability could then trivially explain the inflated false alarm rate for virtual faces. We also note that featural matching was only done at the global level, that is, across whole images. After such global matching, local features might still have had for example varying brightness levels (for example, darker nose region in one image and darker skin region in the other). With more detailed local-level matching, however, maintaining whole-image consistency would have become a practical impossibility. Given these shortcomings, we cannot fully exclude the possibility that the present results were still specific to the present stimuli. We suggest that this problem in fact applies to all studies using virtual stimuli, given the obvious impossibility of creating virtual faces that are visually identical to real faces yet at the same time discriminable from them. Future studies might want to consider using more than one method for producing virtual stimuli to increase the generalizability of their results; however, even this approach does little to solve the fundamental problem related to the lack of unequivocal operationalization of "virtual" or "artificial" stimuli.

An ideal solution to this problem might be to keep the stimuli constant but to present them in varying contexts. We give some suggestions for future research, which at the same time refine the present research questions. First, the effect of perceptual expertise could be tested directly by training participants with either virtual or real faces before the experimental task, for example by adopting a similar training paradigm as Burleigh and Schoenherr (2015). Second, perceptual expertise could also be tested by preselecting participants with high or low exposure to realistic virtual faces in video games and other digital media. Third, future studies could test whether the processing of virtual faces is prone to similar social-cognitive and motivational factors as other-ethnicity and out-group faces (see Young et al., 2012). For example, Bernstein et al. (2007) demonstrated that merely assigning other people as in-group vs. out-group members— for example, members of the same or other universities—elicits

higher discrimination sensitivity in a recognition memory task. Similarly, labeling the same ambiguous real/virtual faces (cf. Cheetham et al., 2011) or even the same human faces as either real or virtual might provoke different processing strategies in individuals. Importantly, all of these hypotheses can be tested by holding the same stimuli constant, which eliminates the influence of visual differences on obtained results.

We would also like to note that performing recognition memory task separately for real and virtual faces could possibly have elicited different processing strategies, which could then have inflated existing response bias differences between them. This effect would in fact resemble the effect of arbitrary labeling as hypothesized above, and it would mean that the present response bias finding was related more to social-cognitive processes than to visual differences between the stimuli. Future studies are required to explore this possibility, however. In particular, the present study could be replicated by interleaving virtual and real faces within the same blocks.

To summarize, the present findings show that virtual faces evoke a higher proportion of false alarms than real faces in a recognition memory task, which suggests that virtual faces do not tap face processing expertise to the same extent than real faces. Furthermore, the present findings suggest that this literal lack of familiarity might then contribute to the uneasiness or even eeriness virtual faces trigger in human observers, which was also observed in the present investigation. The present investigation makes a significant contribution to previous literature by considering low-level visual confounds in the stimuli, by demonstrating that the differential processing of virtual and real faces is particularly evident in false alarm choices, and by linking this result to the qualitative evaluation of virtual faces.

AUTHOR CONTRIBUTIONS

JK designed and implemented the experiment, analyzed the data, and wrote the manuscript.

FUNDING

ACKNOWLEDGMENTS

REFERENCES

Ackerman, J. M., Shapiro, J. R., Neuberg, S. L., Kenrick, D. T., Becker, D. V., Griskevicius, V., et al. (2006). They All Look the Same to Me (Unless They're Angry). *Psychol. Sci.* 17, 836–840. doi: 10.1111/j.1467-9280.2006.01790.x

Balas, B., and Pacella, J. (2015). Artificial faces are harder to remember. *Comput. Hum. Behav.* 52, 331–337. doi: 10.1016/j.chb.2015.06.018

Bernstein, M. J., Young, S. G., and Hugenberg, K. (2007). The cross-category effect mere social categorization is sufficient to elicit an own-group bias

in face recognition. *Psychol. Sci.* 18, 706–712. doi: 10.1111/j.1467-9280.2007.01964.x

Blanz, V., and Vetter, T. (1999). "A morphable model for the synthesis of 3D faces," in *Proceedings of the 26th International Conference on Computer Graphics and Interactive Techniques* (New York, NY: ACM Press/Addison-Wesley Publishing Co.), 187–194.

Burleigh, T. J., and Schoenherr, J. R. (2015). A reappraisal of the uncanny valley: categorical perception or frequency-based sensitization? *Cogn. Sci.* 5:1488. doi: 10.3389/fpsyg.2014.01488

Burton, A. M., White, D., and McNeill, A. (2010). The glasgow face matching test. *Behav. Res. Methods* 42, 286–291. doi: 10.3758/BRM.42.1.286

Campbell, R., Pascalis, O., Coleman, M., Wallace, S. B., and Benson, P. J. (1997). Are faces of different species perceived categorically by human observers? *Proc. R. Soc. Lond. B Biol. Sci.* 264, 1429–1434. doi: 10.1098/rspb.1997.0199

Cheetham, M., Suter, P., and Jäncke, L. (2011). The human likeness dimension of the "uncanny valley hypothesis": behavioral and functional MRI findings. *Front. Hum. Neurosci.* 5:126. doi: 10.3389/fnhum.2011.00126

Cohen, J. (1992). A power primer. *Quant. Methods Psychol.* 112, 155–159. doi: 10.1037/0033-2909.112.1.155

Cook, R., Johnston, A., and Heyes, C. (2012). Self-recognition of avatar motion: How do I know it's me? *Proc. R. Soc. B Biol. Sci.* 279, 669–674. doi: 10.1098/rspb.2011.1264

Crookes, K., Ewing, L., Gildenhuys, J., Kloth, N., Hayward, W. G., Oxner, M., et al. (2015). How well do computer-generated faces tap face expertise? *PLoS ONE* 10:e0141353. doi: 10.1371/journal.pone.0141353

Cubelli, R., and Della Sala, S. (2017). Taking race out of neuroscience too. *Cortex* 99, A1–A2. doi: 10.1016/j.cortex.2017.04.023

Fan, S., Ng, T.-T., Herberg, J. S., Koenig, B. L., and Xin, S. (2012). "Real or fake? Human judgments about photographs and computer-generated images of faces," in *Proceedings of SIGGRAPH Asia 2012 Technical Briefs* (New York, NY: ACM), 17.

Fan, S., Wang, R., Ng, T.-T., Tan, C. Y.-C., Herberg, J. S., and Koenig, B. L. (2014). Human perception of visual realism for photo and computer-generated face images. *ACM Trans. Appl. Percept.* 11, 1–21. doi: 10.1145/2620030

Farid, H., and Bravo, M. J. (2012). Perceptual discrimination of computer generated and photographic faces. *Digit. Investig.* 8, 226–235. doi: 10.1016/j.diin.2011.06.003

Hills, P. J., Cooper, R. E., and Pake, J. M. (2013). Removing the own-race bias in face recognition by attentional shift using fixation crosses to diagnostic features: an eye-tracking study. *Vis. Cogn.* 21, 876–898. doi: 10.1080/13506285.2013.834016

Hills, P. J., and Lewis, M. B. (2006). Reducing the own-race bias in face recognition by shifting attention. *Q. J. Exp. Psychol.* 59, 996–1002. doi: 10.1080/17470210600654750

Kätsyri, J., Förger, K., Mäkäräinen, M., and Takala, T. (2015). A review of empirical evidence on different uncanny valley hypotheses: support for perceptual mismatch as one road to the valley of eeriness. *Front. Psychol.* 6:390. doi: 10.3389/fpsyg.2015.00390

Kobayashi, M., Otsuka, Y., Kanazawa, S., Yamaguchi, M. K., and Kakigi, R. (2012). Size-invariant representation of face in infant brain: an fNIRS-adaptation study. *Neuroreport* 23, 984–988. doi: 10.1097/WNR.0b013e32835a4b86

Langner, O., Dotsch, R., Bijlstra, G., Wigboldus, D. H. J., Hawk, S. T., and van Knippenberg, A. (2010). Presentation and validation of the Radboud Faces Database. *Cogn. Emot.* 24, 1377–1388. doi: 10.1080/02699930903485076

Lehmuskallio, A., Häkkinen, J., and Seppänen, J. (2018). Photorealistic computer-generated images are difficult to distinguish from digital photographs: a case study with professional photographers and photo-editors. *Vis. Commun.* doi: 10.1177/1470357218759809. [Epub ahead of print].

Levin, D. T., and Angelone, B. L. (2002). Categorical perception of race. *Perception* 31, 567–578. doi: 10.1068/p3315

Looser, C. E., and Wheatley, T. (2010). The tipping point of animacy: how, when, and where we perceive life in a face. *Psychol. Sci.* 21, 1854–1862. doi: 10.1177/0956797610388044

MacDorman, K. F., Srinivas, P., and Patel, H. (2013). The uncanny valley does not interfere with level 1 visual perspective taking. *Comput. Hum. Behav.* 29, 1671–1685. doi: 10.1016/j.chb.2013.01.051

Maurer, D., Grand, R. L., and Mondloch, C. J. (2002). The many faces of configural processing. *Trends Cogn. Sci.* 6, 255–260. doi: 10.1016/S1364-6613(02)01903-4

McKone, E., Brewer, J. L., MacPherson, S., Rhodes, G., and Hayward, W. G. (2007). Familiar other-race faces show normal holistic processing and are robust to perceptual stress. *Perception* 36, 224–248. doi: 10.1068/p5499

Meissner, C. A., and Brigham, J. C. (2001). Thirty years of investigating the own-race bias in memory for faces: a meta-analytic review. *Psychol. Public Policy Law* 7, 3–35. doi: 10.1037/1076-8971.7.1.3

Meissner, C. A., Brigham, J. C., and Butz, D. A. (2005). Memory for own- and other-race faces: a dual-process approach. *Appl. Cogn. Psychol.* 19, 545–567. doi: 10.1002/acp.1097

Mori, M. (1970). The Uncanny Valley (K. F. MacDorman & N. Kageki, Trans.). *IEEE Robot. Autom. Mag.* 19, 98–100. doi: 10.1109/MRA.2012.2192811

O'Toole, A. J., Deffenbacher, K., Abdi, H., and Bartlett, J. C. (1991). Simulating the 'other-race effect' as a problem in perceptual learning. *Connect. Sci.* 3, 163–178.

Railo, H., Karhu, V.-M., Mast, J., Pesonen, H., and Koivisto, M. (2016). Rapid and accurate processing of multiple objects in briefly presented scenes. *J. Vis.* 16:8. doi: 10.1167/16.3.8

Rhodes, G., Brake, S., Taylor, K., and Tan, S. (1989). Expertise and configural coding in face recognition. *Br. J. Psychol.* 80, 313–331. doi: 10.1111/j.2044-8295.1989.tb02323.x

Salah, A. A., Alyüz, N., and Akarun, L. (2008). Registration of three-dimensional face scans with average face models. *J. Electron. Imaging* 17:011006. doi: 10.1117/1.2896291

Sangrigoli, S., Pallier, C., Argenti, A. M., Ventureyra, V. A. G., and de Schonen, S. (2005). Reversibility of the other-race effect in face recognition during childhood. *Psychol. Sci.* 16, 440–444. doi: 10.1111/j.0956-7976.2005.01554.x

Shadish, W. R., Cook, T. D., and Campbell, D. T. (2002). *Experimental and Quasi-experimental Designs for Generalized Causal Inference.* Belmont, CA: Wadsworth.

Shah, P., Gaule, A., Sowden, S., Bird, G., and Cook, R. (2015). The 20-item prosopagnosia index (PI20): a self-report instrument for identifying developmental prosopagnosia. *R. Soc. Open Sci.* 2:140343. doi: 10.1098/rsos.140343

Stanislaw, H., and Todorov, N. (1999). Calculation of Signal Detection Theory measures. *Behav. Res. Methods Instrum. Comput.* 31, 137–149. doi: 10.3758/BF03207704

Stevens, S. S., and Pashler, H. E. (2002). *Stevens' Handbook of Experimental Psychology, 3rd Edn.* New York, NY: John Wiley & Sons.

Stürzel, F., and Spillmann, L. (2000). Thatcher Illusion: dependence on Angle of Rotation. *Perception* 29, 937–942. doi: 10.1068/p2888

Tanaka, J. W., and Pierce, L. J. (2009). The neural plasticity of other-race face recognition. *Cogn. Affect. Behav. Neurosci.* 9, 122–131. doi: 10.3758/CABN.9.1.122

Valentine, T. (1991). A unified account of the effects of distinctiveness, inversion, and race in face recognition. *Q. J. Exp. Psychol. Sect. A* 43, 161–204. doi: 10.1080/14640749108400966

Valentine, T., Lewis, M. B., and Hills, P. J. (2016). Face-space: a unifying concept in face recognition research. *Q. J. Exp. Psychol.* 69, 1996–2019. doi: 10.1080/17470218.2014.990392

Vizioli, L., Foreman, K., Rousselet, G. A., and Caldara, R. (2011). Inverting faces elicits sensitivity to race on the N170 component: a cross-cultural study. *J. Vis.* 10, 15–15. doi: 10.1167/10.1.15

Vokey, J. R., and Read, J. D. (1992). Familiarity, memorability, and the effect of typicality on the recognition of faces. *Mem. Cogn.* 20, 291–302. doi: 10.3758/BF03199666

Willenbockel, V., Sadr, J., Fiset, D., Horne, G. O., Gosselin, F., and Tanaka, J. W. (2010). Controlling low-level image properties: the SHINE toolbox. *Behav. Res. Methods* 42, 671–684. doi: 10.3758/BRM.42.3.671

Wright, D. B., and Sladden, B. (2003). An own gender bias and the importance of hair in face recognition. *Acta Psychol.* 114, 101–114. doi: 10.1016/S0001-6918(03)00052-0

Young, S. G., Hugenberg, K., Bernstein, M. J., and Sacco, D. F. (2012). Perception and motivation in face recognition: a critical review of theories of the cross-race effect. *Pers. Soc. Psychol. Rev.* 16, 116–142. doi: 10.1177/1088868311418987

Yudell, M., Roberts, D., DeSalle, R., and Tishkoff, S. (2016). Taking race out of human genetics. *Science* 351, 564–565. doi: 10.1126/science.aac4951

Perceived Ownership of Avatars Influences Visual Perspective Taking

Christian Böffel* and Jochen Müsseler

Institute of Psychology, RWTH Aachen University, Aachen, Germany

Modern computer-based applications often require the user to interact with avatars. Depending on the task at hand, spatial dissociation between the orientations of the user and the avatars might arise. As a consequence, the user has to adopt the avatar's perspective and identify herself/himself with the avatar, possibly changing the user's self-representation in the process. The present study aims to identify the conditions that benefit this change of perspective with objective performance measures and subjective self-estimations by integrating the idea of avatar-ownership into the cognitive phenomenon of spatial compatibility. Two different instructions were used to manipulate a user's perceived ownership of an avatar in otherwise identical situations. Users with the high-ownership instruction reported higher levels of perceived ownership of the avatar and showed larger spatial compatibility effects from the avatar's point of view in comparison to the low ownership instruction. This supports the hypothesis that perceived ownership benefits perspective taking.

Keywords: avatars, ownership, stimulus-response compatibility tasks (SRC), perspective-taking, human computer interaction (HCI)

INTRODUCTION

When we are confronted with avatars in the virtual world, we often have to adopt their perspective in order to complete our task. Sometimes it is necessary to control an avatar; sometimes we merely interact with avatars controlled by others. Avatars are used to represent the user in the digital world and the user is able to interact with the virtual world through the avatar. In both situations, seeing the world through the avatar's eyes can be useful to plan actions or to interpret the actions of others. This process, referred to as visual perspective taking (PT), was observed in various situations and toward a large variety of targets. It occurs toward human confederates (Frischen et al., 2009; Freundlieb et al., 2016, 2017) and even non-human targets like triangles (Zwickel, 2009) or arrows (Santiesteban et al., 2014). In the case of avatars, the distinction between an object and a person is not clear-cut. It is sometimes unclear if someone else controls the avatar or if the avatar is an independent agent controlled by the program itself. While we are generally able to identify objects as non-human, we still tend to attribute human-like agency and mental states to them (Heider and Simmel, 1944). This agency attribution seems to aide visual PT (Zwickel, 2009). In past studies we showed that PT occurs toward avatars, regardless of whether PT is needed to complete the task (Müsseler et al., 2017) or not (Böffel and Müsseler, 2017). In the present study we confronted participants with an avatar that was presented opposite to them on a computer screen in a top–down view. Our goal was to take a closer look at how this avatar is interpreted and how this interpretation can benefit or inhibit PT in a top–down manner.

People have the remarkable ability to incorporate objects — like avatars — into the mental representation of their bodies. The most famous example of this is the rubber hand illusion experiment by Botvinick and Cohen (1998) in which participants were able to feel the touch on a rubber hand. To achieve this illusion, a rubber hand was placed in front of the participants while their real hand was hidden from sight. When the rubber hand and the real hand were brushed synchronously, participants started to "feel" the stimulation on the rubber hand instead of their own and reported that the rubber hand seemed to be part of their body. This sense that an object belongs to the person's own body is referred to as *ownership*. Ma and Hommel (2013) demonstrated that virtual objects — such as virtual hands — could also become part of mental body representations akin to the rubber hand.

Several factors were identified that lead to this sense of ownership. Makin et al. (2008) showed the importance of spatial congruency between the real and the artificial hand for perceived ownership, whereas Shimada et al. (2009) demonstrated the necessity of temporal congruency between the tactile stimulation and the visual perception of this stimulation on the fake hand. Tsakiris (2010) underlined the importance of visual similarity between real and artificial body parts and argued that the artificial body part has to resemble the real body part in order to be embodied. However, this assumption has been called into question by recent experimental findings. Armel and Ramachandran (2003) were able to observe ownership of a table, Ma and Hommel (2015a) demonstrated ownership of objects like balloons and squares and Guterstam et al. (2013) were even able to observe embodiment of empty space. These effects are overall comparable to the perceived ownership of artificial hands even though the objects had no resemblance of real hands or body parts.

A different factor that seems to influence ownership is a sense of connectedness between the actions of the person and the action effects on the side of the object. This controlling aspect — or perceived agency over the object — leads to ownership (Ma and Hommel, 2015a,b) and it can be used to induce a rubber hand-like illusion without the need for tactile stimulation (Kalckert and Ehrsson, 2014). Overall, perceived agency over an object is a promising mechanism for inducing ownership that also influences visual PT. When comparing PT in situations in which a person was actually controlling the arms of an avatar to situations in which this control was merely imagined, only the conditions with actual control were associated with visual PT (Böffel and Müsseler, 2017). Combined with the results of Zwickel (2009), two conflicting characteristics of a situation can be identified that seem to benefit PT: on the one hand perceiving the target as an individual agent seems to aide PT, because it can help us understand someone's actions (Tversky and Hard, 2009), on the other hand, actual control over the target also seems to lead to PT. Both processes seem exclusive, because we cannot see an avatar as an independent agent and attribute intentions to it, if it fully obeys our every command. Our goal is to solve this conflict in the present study.

The results of previously mentioned studies point toward the importance of bottom–up processes in the integration of objects into a person's representation of their action and body. It therefore seems plausible that action representation is a gateway that leads to ownership: if a person's action reliably causes a certain effect, even when this effect is produced through seemingly unconnected mediating objects (e.g., a rubber hand or a balloon), the action effect becomes a relevant part of the action code and the person is able to anticipate this effect as a consequence of her/his action. We believe that ownership of the object that produces the action effect is inferred as a result. We expect that once ownership is acquired, visual PT often follows to facilitate the planning of future actions through the object that is now perceived as part of the person's body.

Because ownership is observed in different situations and of different objects, it seems very likely that ownership toward an avatar is rather easily achievable, especially if the person controls the avatar. Although past studies point to the importance of bottom–up processes for ownership, we aim to demonstrate that it is also subject to top–down modulation. The acquisition of ownership via bottom–up processes seems to be an automatic process and top–down modulation of automatic processes has been demonstrated before (for an overview see Kiefer, 2007). We believe that when confronted with an object, two different explanations of the same situation are able to alter the framework that influences how ownership of an object is acquired. In a high ownership explanation, the participants might shift attention to situational features that support their sense of control, while in the low ownership explanation the opposite is expected. The use of two different instructions that target this sense of control should therefore be able to alter the interpretation of the situation resulting in different levels of perceived ownership of an avatar (measured via self-report questionnaire). Such a result could further our understanding of the nature and dependencies of automatic processes.

When examining visual PT, two different — although similar — tasks have been used in the past: the *own body transformation task* that asks participants to judge on which side of a shown body a certain salient feature is located, and the *avatar-in-scene task* that uses laterality decisions of objects from an avatars point of view. Both tasks share the problem that the results are potentially influenced by stimulus–response compatibility (SRC) effects with unknown consequences (May and Wendt, 2013). SRC refers to the observation that certain mappings of responses to stimuli lead to performance advantages over others (Fitts and Deininger, 1954) and result in faster reaction times and lower error rates. When using stimuli and responses that carry spatial information, SRC is in most cases aligned with spatial correspondence of stimulus and response positions. Conditions in which stimulus and response occur in the same hemifield are generally compatible, whereas conditions with opposing positions are incompatible (for an overview see Proctor and Vu, 2006). A theoretical framework often used to explain SR compatibility are the so called dual-route models (e.g., Kornblum et al., 1990). These models propose that stimulus presentation causes the activation of two routes: the automatic route leads to a direct activation of a response code that spatially corresponds to the stimulus position. A stimulus presented on the left would activate a left response code. A second route

uses the SR mapping, for example given by an instruction, to retrieve the correct response. In the case that both routes activate the same correct response the execution of the response is facilitated, otherwise a conflict occurs that has to be resolved. This conflict leads to slower reaction times and increased error rates. As a consequence of SRC, both, facilitation and interference can be observed (Wallace, 1971). Because SRC effects are often attributed to an overlap of certain features of the task's mental representation, we can use these effects to infer how a certain stimulus is mentally represented. More importantly, it allows us to identify if the stimulus position is coded from the participants own point of view, leading to the typically observed advantages of spatially corresponding stimulus–response parings, or if it produces different compatibility effects indicating that the stimulus position is coded from the avatar's point of view instead. The results of this coding process — referred to as feature codes — are often seen as abstract representations and independent of the modality that was used to create them (Hommel et al., 2001). As consequence, a stimulus coded as "left" from the participant' point of view would form the same feature code as a stimulus coded as "left" from an Avatar's point of view, although they do in fact occupy different locations.

In the present study we use SRC effects to measure visual PT (Böffel and Müsseler, 2017; Müsseler et al., 2017). Previous studies that follow a similar approach point toward a complicated situation. On one hand, studies find that objects are generally coded from the person's own perspective when no agency instruction is used (Gardner and Potts, 2011; Taylor et al., 2016). However, this observation doesn't seem universal. Gardner and Potts (2010) and Taylor et al. (2016) show that under certain circumstances feature codes can be created from the objects perspective instead. Müsseler et al. (2017) demonstrated that SRC effects can arise from an avatar's point of view, rather than the person's own in situations that force the person to take the avatar's perspective in a SRC task and Böffel and Müsseler (2017) showed that these effects can occur even in a Simon task, in which the avatar's orientation is irrelevant. However, these compatibility changes only occur when the dimensional overlap between stimulus and response position from the participant's point of view is low, or the participant's control over the avatar's movements is high. The latter is likely linked to ownership of the avatar that is acquired through bottom–up processes and effect anticipation.

Overall, PT toward avatars is able to influence SRC, which indicates a change in the mental representation of the situation, if the right conditions are met. As a result, PT can lead to an effect of spatial correspondence as seen from the avatar's point of view rather than the person's, effectively reversing the expected effect of spatial compatibility under certain circumstances (Böffel and Müsseler, 2017; Müsseler et al., 2017). Assuming that SRC effects arise based on the mental representation of a task, it is a useful tool to quantify PT because it allows us to infer how the stimulus location is coded (Ottoboni et al., 2005; Hommel, 2011). A tool that we also rely on in the present study.

Based on the described mechanisms, we expect that an increase in perceived ownership aids the incorporation of the avatar and its movements into the person's mental representation of the task. This should lead to an increase of visual PT to facilitate action planning and therefore induce larger compatibility effects from the avatar's perspective. We hope to show what is ultimately more beneficial: high ownership of the avatar or low ownership but higher levels of autonomy on the side of the avatar. We believe that although PT can be an effective mechanism for understanding someone else's actions (Tversky and Hard, 2009) it is even more vital when it helps to plan our own actions through the means of an avatar. Or to put it differently: we think that planning our own actions evokes a stronger need for visual PT than understanding someone else's.

To summarize the goals of this study: first, we want to show that the otherwise automatic acquisition of perceived ownership of an object can be influenced in a top–down manner by a framework provided in the instruction of the task. And second, we want to demonstrate that this change in perceived ownership is associated with changes in visual PT as measured with stimulus compatibility effects. Therefore, we want to pose the following hypotheses.

Hypotheses

We expect that the two different instructions produce quantifiable differences in perceived ownership of the avatar, measured by the avatar-ownership questionnaire with higher self-reported perceived ownership in the high-ownership group in comparison to the low ownership group. We further predict that SRC effects are dependent on perceived ownership. In the high ownership group, we expect a larger benefit of spatially non-corresponding conditions compared to the low ownership group where compatibility drifts toward the participant's perspective rather than the avatar's. This should result in an interaction of spatial correspondence and instruction.

MATERIALS AND METHODS

We used two different instructions for the same task to top–down influence perceived ownership of an avatar: the setup and the avatars used were similar to the ones of Böffel and Müsseler (2017) and Müsseler et al. (2017). The participants were confronted with an avatar that was displayed on a screen and sitting opposite them (**Figure 1**). One instruction described the avatar as fully controlled by the participant, much like a tool (high ownership condition), and the second tried to establish the avatar as an individual agent (low ownership condition). The participants were asked to respond to dark/light blue disks with key presses that resulted in avatar hand movements. In both groups, stimuli, responses, and action effects were identical. The action goal was defined in the same way in both groups: act so that the avatar moves a certain arm. This was done to avoid the influence of different action goals that could otherwise lead to SRC effects related to the location of the action goal rather than response location as described by Hommel (1993a).

Participants

In total 48 students (39 females) from RWTH Aachen University with a mean age of $M = 21.6$ ($SD = 3.9$) participated in this
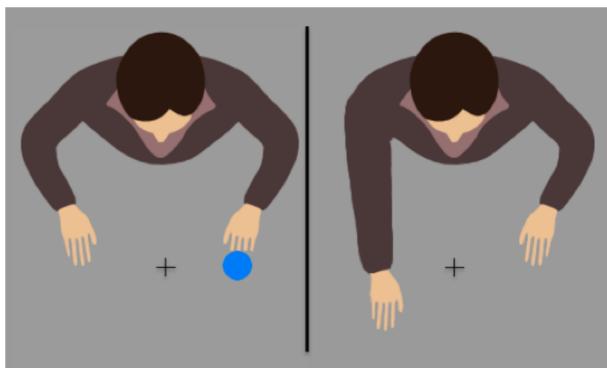
FIGURE 1 | Example condition that demanded a contralateral response, before **(left)** or after **(right)** a right key-press.

experiment for course credit or a monetary compensation of 5 €. All participants had normal or corrected-to-normal vision.

Apparatus and Stimuli

MatLab and the Psychtoolbox Extension v3.0 (Brainard, 1997; Pelli, 1997) were used for stimulus presentation and reaction time measurement. The stimuli were presented on a 22″ CRT monitor (Iiyama Visionmaster Pro 514 with a resolution of 1024 × 768 and 100 Hz refresh rate). The participants were seated 70 cm in front of the monitor and responded with their left and right index fingers on response keys (**Figure 2**). Dark blue (RGB 36 115 254) and light blue circles (RGB 98 193 254), each with a diameter of 50 pixel (1.79°) were used as targets, presented 1.61° to the left or right of a central fixation cross and in front of a gray background (RGB 155 155 155). The avatar had a size of roughly 240 × 200 pixels (8.73° × 8.56°) and was facing the participants with its hands pointing toward the stimuli positions (**Figure 1**).

Procedure

The participants gave written informed consent to the terms of the experiment, including data storage and data usage for publication purposes. After that, half of the participants were instructed to control the avatar's hands by pressing the respective key on the response board: a right key-press moves the right hand and a left key-press moves the left hand. This lead to effector congruency between the participant's hands and the hands of the avatar. The second group was instructed to imagine the avatar as an independent agent that acts according to its own goals and always wants to move both hands. However, the participant can prevent the avatar from moving the ipsilateral hand with a key press. For example, a right key-press stops the avatar's left hand from moving. As a result, the avatar once again only moves its right hand. This means that in both groups the same key press lead to the same observable action effects and the main objective was the same in both instructions: act in such a way that only the contralateral hand is moved if the target is light-blue and only the ipsilateral hand is moved if the target is dark-blue. The mapping of light-and dark-blue targets to ipsi- and contralateral responses

FIGURE 2 | Experimental setup.

was counterbalanced between participants. Because the avatar always showed the same hand movements after a certain key is pressed, regardless of the instruction used, and the goal of the action was always the same only the interpretation of the situation was changed by the instruction. A central fixation cross and the avatar remained visible throughout the experiment. The targets were presented without a time limit until the participants responded. If the response was incorrect, slower than 1,500 ms (lapse) or faster than 100 ms (anticipation) it was labeled as an error and followed by a feedback tone. The waiting period between the response and the beginning of the next trial was 2,250 ms and increased by additional 1,500 ms after an error occurred. Each participant performed 10 blocks, including 8 repetitions of each combination of stimulus position and stimulus color. The first block was a practice block that was excluded from the analysis. The order of trials was randomized within each block. Overall each condition was repeated 80 times over the course of the experiment resulting in a total of 320 trials per participant, excluding practice-trials. The participants needed approximately 25 min to complete the experiment.

After the experiment the participants were asked to fill in a questionnaire that featured the perception of the avatar. The avatar-questionnaire was based on an instrument used by Ma and Hommel (2015b) that targeted the perceived ownership of

virtual hands and is a modified version of the questionnaire Botvinick and Cohen (1998) used to examine the rubber hand illusion. Our modified questionnaire asked the participants to rate 10 statements regarding ownership of the avatar and its hands (e.g., "It felt as if the avatar's hands were part of my body" "The hands of the avatar began to resemble my hands in terms of shape or skin tone") on a seven-step Likert scale ranging from 1 "I strongly disagree" to 7 "I strongly agree." The complete list of items used is shown in **Table 1**. We altered the items used by Ma and Hommel (2015b) to closer resemble the avatar setting while trying to maintain the general objective of the instrument. Three items were omitted because they targeted tactile perceptions, which were not included in our experiment. The instrument was used in a German translation. We calculated an overall perceived ownership score as the sum of the responses for each participant. The possible range of ownership values was therefore 10 to 70 and higher values indicated higher levels of perceived ownership.

Design

The experimental conditions consisted of all possible combination of stimulus position, response position and instruction. Stimulus and response position were used to determine spatial correspondence. The conditions in which stimulus and response positions were both on the participants left or right were labeled as spatially corresponding, others as non-corresponding. This resulted in a 2 × 2 design with the within-subjects factor spatial correspondence (non-corresponding vs. corresponding) and the between-subjects factor instruction (high ownership vs. low ownership).

RESULTS

Reported Ownership

The analysis of the avatar-questionnaire data revealed instruction-based group differences: the high-ownership

TABLE 1 | Items used in the ownership questionnaire.

Q1: It felt as if the avatar's hands were part of my body.

Q2: It seemed that my hand was in the location where the hand of the avatar was.

Q3: I lost the feeling where my hands were located.

Q4: It seemed that my hands were no longer part of my body.

Q5: I had the feeling that I might have additional hands.

Q6: Sometimes I felt as if my hands were turning virtual.

Q7: The hands of the avatar began to resemble my hands in terms of shape or skin tone.

Q8: It appeared (visually) as if the hands of the avatar were drifting toward my hands.

Q9: It seemed like I could have moved the hand on the screen if I wanted, as if it were obeying my will.

Q10: It felt as if my hands took on the same size as the avatar's hands.

The items are based on the instruments used by Botvinick and Cohen (1998) and Ma and Hommel (2015b).

instruction was associated with overall higher levels of self-reported ownership ($M = 22.3$; $SD = 10.9$) compared to the low ownership instruction ($M = 17.8$; $SD = 7.0$). This effect was statistically significant [$t(39.15) = 1.69$; $p_{(one-tailed)} = 0.05$], df were Welch-adjusted to account for differing variances in both groups.

Reaction Times and Percentage Errors

Reaction times longer than 1,500 ms or shorter than 100 ms were regarded as errors and were removed from the RT analyses. A total of 254 trials (1.7%) were excluded this way along with 821 false responses (5.3%) for a total of 1075 errors (7.0%). Mean RTs and percentage errors (PE) were analyzed separately using 2 × 2 mixed design ANOVAs with the within-subjects factor spatial correspondence (corresponding vs. non-corresponding) and the between subject factor instruction (high vs. low ownership). Results are shown in **Figure 3**. The analysis of mean reaction times revealed a significant influence of spatial correspondence $F(1,46) = 5.51$, $p = 0.023$, $\eta_p^2 = 0.11$, overall favoring spatially non-corresponding stimulus–response pairings ($M_{corr.} = 649$ ms vs. $M_{non-corr.} = 630$ ms). This effect was significantly influenced by the factor instruction $F(1,46) = 7.04$, $p = 0.011$, $\eta_p^2 = 0.13$ with a 40 ms advantage of non-corresponding conditions in the high ownership instruction group compared to a 2 ms advantage of spatially corresponding conditions in the low ownership group. Analyzed separately, the 40 ms advantage of non-corresponding conditions in the high ownership condition is statistically significant with [$t(23) = 3.69$, $p = 0.001$, two tailed] while the 2 ms advantage of spatially corresponding conditions in the low ownership group is not [$t(13) = 0.20$, $p = 0.84$, two tailed].

The analysis of percentage errors showed a marginally significant main effect of spatial correspondence $F(1,46) = 3.58$, $p = 0.065$, $\eta_p^2 = 0.07$ that interacted significantly with instruction $F(1,46) = 4.40$, $p = 0.041$, $\eta_p^2 = 0.09$. Spatially non-corresponding SR-mappings were associated with lower error rates compared to corresponding ones in the high-ownership instruction ($M_{corr.} = 8.1\%$ vs. $M_{non-corr.} = 5.6\%$) but not when paired with the low ownership instruction ($M_{corr.} = 7.1\%$ vs. $M_{non-corr.} = 7.2\%$). Similar to the reaction times only the 2.5% points advantage of non-corresponding conditions in the high ownership condition is statistically significant with [$t(23) = 3.57$, $p = 0.002$, two tailed] while the 0.1% points advantage of spatially corresponding conditions in the low ownership group is not [$t(13) = 0.13$, $p = 0.90$, two tailed] when analyzed separately.

DISCUSSION

The analysis of the avatar-questionnaire data revealed instruction-based group differences that were consistent with our expectations. The high ownership instruction resulted in significantly higher values of self-reported ownership compared to the low ownership instruction. We therefore conclude that the manipulation was successful. Overall this supports the idea that top–down processes influence perceived ownership.
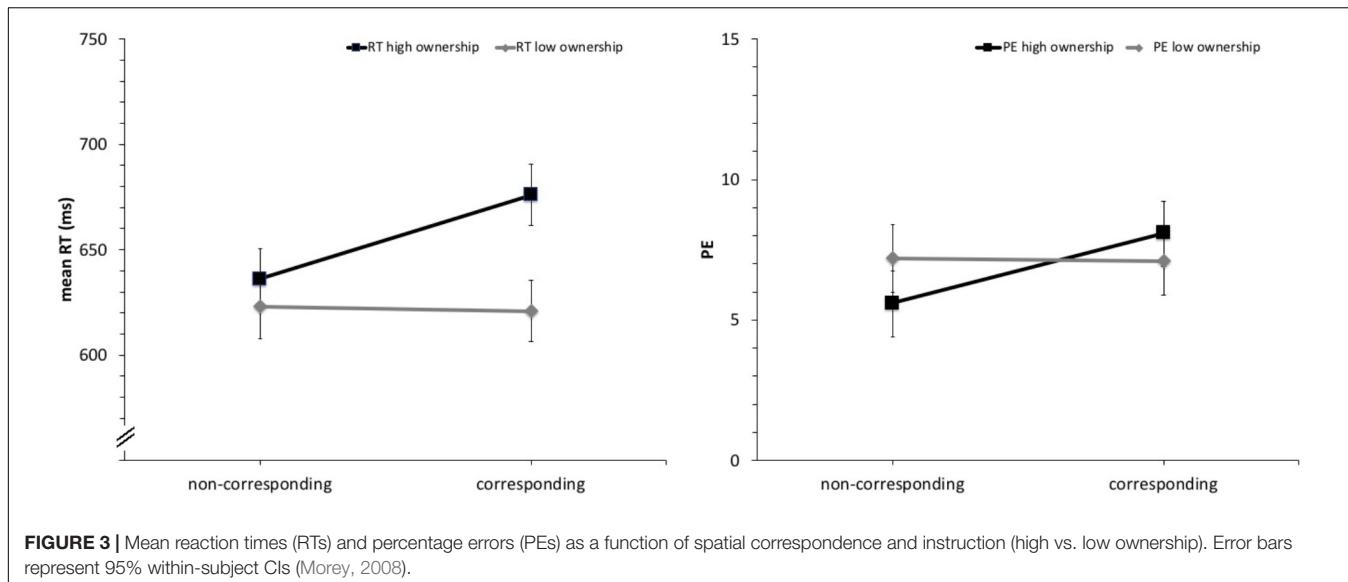
FIGURE 3 | Mean reaction times (RTs) and percentage errors (PEs) as a function of spatial correspondence and instruction (high vs. low ownership). Error bars represent 95% within-subject CIs (Morey, 2008).

The analysis of reaction times and error rates showed that both instruction cause different effects of spatial correspondence in otherwise identical scenarios. While the correspondence effect was negligible in the low ownership condition, it was significantly (more) negative in the high-ownership condition. The high ownership conditions therefore cause compatibility effects that are based on the avatar's point of view instead of the participant's own. This means that the observed effects are similar to the effects we would expect if the participant would actually see the scene from a rotated point of view. We think this is a very strong indicator that the stimuli are coded from the avatar's viewpoint and that the resulting mental representation is the important factor that determines spatial compatibility rather than the actual physical location of the stimuli. As a result, stimuli presented on the left produced compatibility effects as if they were presented on the right and vice versa. A stimulus presented on the left side of the avatar lead to the formation of the same feature code as a stimulus presented on the left of the person, even though their position is in fact different. This is apparently not the case in the low ownership condition which indicates that both conditions lead to different mental representation of the same scene.

The absence of a correspondence effect in the low ownership could point toward the possibility that the task was complicated enough to eliminate the influence of the automatic activation of spatially corresponding responses a phenomenon that can be observed in mixed SRC tasks (Shaffer, 1965). This is most likely a result of a reactive inhibition rather than proactive suppression of the automatic route in complex situations and was described by Proctor and Vu (2010). A similar case could be made for the low ownership condition in our experiment, because the instruction might be sufficiently complex to cause a similar effect. This is apparently not the case in the high ownership condition where an advantage of spatially non-corresponding conditions was observed. The analysis of reaction times showed

higher mean reaction times in the high ownership group. Although this difference is not statistically significant ($p = 0.17$) it seems unlikely that the low-ownership condition is overall more complex.

Why do we still observe a correspondence effect and why is the automatic route not inhibited in the high ownership group? The high-ownership condition has the advantage that the task can be broken down into several steps: step 1: perspective taking, step 2: recoding of the stimulus position, step 3: action. While each step is relatively simple, the completion of all steps combined might cause higher reaction times. The PT in step 1 is also associated with costs that would explain the numerically higher reaction time compared to the low ownership group (Janczyk, 2013). We propose that after PT is completed, the spatial information of the stimulus would be coded within the new frame of reference from the avatar's point of view. At this point the task is identical to a typical SR compatibility task. It leads to the expected effects when accounting for the new mental representation of the stimulus. This mechanism could be similar or identical to the concept of referential coding (Hommel, 1993b) that lays the groundwork for the coding of stimulus features based on different reference frames. The new reference frame provided by the avatar would be rotated by 180° from the participant's point of view and constitutes a rather drastic example of conflicting reference frames. This supports the theory that referential coding of the same situation can either be based on an egocentric or alternative reference frame, based on expectations and knowledge about the situation.

An alternative explanation for the absence of compatibility effects in the low ownership condition might be that both reference frames are activated equally strong, leading to the stimulus position being coded as neutral. The stimulus would cause the formation of both feature codes: "left" and "right." This conflict might result in an overall compatibility effect of zero. Alternatively, one frame of reference may always overwrite the other but both reference frames win this conflict equally

often, resulting in a zero-sum of spatial correspondence effects. The participant might therefore switch between both reference frames, but only one of them would be active at a given time. If the latter is the case, the automatic route of the dual-route model might still be active but its effects are evened-out over the course of the experiment. An alternating activation of the egocentric and allocentric reference frame within the same condition could effectively cause further mixing of compatible and incompatible mappings within those conditions and cause an elimination of SRC effects as described earlier (Shaffer, 1965). It is also possible that the low ownership instruction was interpreted differently by different individuals, leading to PT and reversed spatial correspondence effects in some, but classic correspondence effects in others, again evening out.

Overall this study provides evidence for the influence of top–down processes in perceived ownership, but to conclude that bottom–up processes are not important in the present situation might be a mistake. In this experiment the situation included a reliable congruency between the participant's responses and the movement of the avatar. Such characteristics are expected to invoke a sense of ownership of the avatar (Ma and Hommel, 2015b) that should require no further explanation or instruction. Based on the results of the present study it seems more likely that top–down processes can suppress perceived ownership of an avatar even if the situation would otherwise induce it.

Although we tried to ensure that the final action goal was the same in both instructions, we ultimately cannot rule the influence of sub-goals out. The imagined grabbing of the avatars ipsilateral hand is the most likely example of such a sub-goal. While the final intention is always to produce an avatar movement that is contralateral to the key press, this sub-goal would have an ipsilateral location of intention, in this case the prevention of a movement. A conflict of goal and sub-goal location could be a contributor to the absence of spatial correspondence effects in the low ownership group. How this prevention of an action effect as an intention influences action planning is not entirely clear, although there is some evidence that the location of intention is more important than the location of the actual effect (Hommel, 1993a; Müsseler et al., 2012).

There are some limitations of this study that are noteworthy. First, measuring ownership with a self-report questionnaire might not be ideal. It is unclear whether participants are able to consciously perceive ownership of the same magnitude as it actually influences their actions. On one hand, it might very well be possible that the artificial setting in this study prevents higher degrees of reported ownership, because it seems difficult to agree to the items used in the questionnaire and the process could be largely subconscious. As a result, true ownership effects might be underreported. On the other hand, social desirability bias might have caused an overestimation of ownership. As a result, other means of measuring ownership might be more feasible than self-report. The second limitation lies in the abstract nature of the avatars used in this experiment. The avatars offered no customization and were the same for each participant. While this design-choice ensured constant conditions for all participants, it might have had negative impact on perceived ownership. A higher degree of physical similarity between participant and avatar might result in higher ownership and stronger effects. Customization is a feature that is often, although not always, present in applications that use avatars. Whether the tradeoff of visual constancy for all participants at the cost of higher variability in similarity between participant and avatar is justified is up for debate. From a psychophysics point of view, constancy is crucial whereas from an applied perspective it is often negligible. Since this study relies on a SR compatibility task — a classic experimental paradigm — we chose the first option. Last but not least, it is also likely that the two instructions not only influenced perceived ownership but also the social character of the situation. With high ownership, the situation might be perceived as less social and closer to a tool-use scenario compared to the low ownership instruction that established the avatar as an independent agent. This is particularly interesting, since PT is often seen as a social phenomenon, yet in this study it is only measurable in the situation that is effectively less social.

The result of this study can be applied in the design of human–computer interactions (HCIs). When the user is required to act from the perspective of an avatar in the presence of other avatars, establishing those distracting avatars as independent agents could prove useful to prevent PT toward these distractors. Such unwanted PT could create additional reference frames that are potentially associated with costs and conflicting SRC relations. On the other hand, stressing the control over an avatar might facilitate PT toward this avatar.

DATA AVAILABILITY

The raw data supporting the conclusions of this manuscript will be made available by the authors, without undue reservation, to any qualified researcher.

ETHICS STATEMENT

In accordance with the Declaration of Helsinki (World Medical Association, 2013), all participants gave written informed consent to participate in the study and participation was voluntary. Further, no undue physical or psychological stress by participating in this study was anticipated and the data obtained on individual participants were not used to elucidate properties of the participant but to examine general laws of cognitive information processing. As a result, no ethical concerns were identified in accordance to the ethics guidelines of the DFG (Deutsche Forschungsgemeinschaft [German Research Foundation], 2009).

AUTHOR CONTRIBUTIONS

CB developed the study concept and design, performed the data collection and data analysis under the supervision of JM. All authors have approved this version of the manuscript and its submission.

REFERENCES

Armel, K. C., and Ramachandran, V. S. (2003). Projecting sensations to external objects: evidence from skin conductance response. *Proc. R. Soc. B* 270, 1499–1506. doi: 10.1098/rspb.2003.2364

Böffel, C., and Müsseler, J. (2017). "The simon effect in the presence of avatars," in *Proceedings of the Abstracts of the 3rd International Meeting of the Psychonomix Society*, Amsterdam, 111.

Botvinick, M., and Cohen, J. (1998). Rubber hands "feel" touch that eyes see. *Nature* 391:756. doi: 10.1038/35784

Brainard, D. H. (1997). The psychophysics toolbox. *Spat. Vis.* 10, 433–436. doi: 10.1163/156856897X00357

Deutsche Forschungsgemeinschaft [German Research Foundation] (2009). *FAQ: Informationen für Geistes- und Sozialwissenschaftler/innen.* Available at: http://www.dfg.de/foerderung/faq/geistes_sozialwissenschaften/index.html [accessed March 15, 2018].

Fitts, P. M., and Deininger, R. L. (1954). S-R compatibility: Correspondence among paired elements within stimulus and response codes. *J. Exp. Psychol.* 48, 483–492. doi: 10.1037/h0054967

Freundlieb, M., Kovács, Á. M., and Sebanz, N. (2016). When do humans spontaneously adopt another's visuospatial perspective. *J. Exp. Psychol.* 42, 401–412. doi: 10.1037/xhp0000153

Freundlieb, M., Sebanz, N., and Kovács, Á. M. (2017). Out of your sight, out of my mind: Knowledge about another person's visual access modulates spontaneous visuospatial perspective-taking. *J. Exp. Psychol.* 43, 1065–1072. doi: 10.1037/xhp0000379

Frischen, A., Loach, D., and Tipper, S. P. (2009). Seeing the world through another person's eyes: simulating selective attention via action observation. *Cognition* 111, 212–218. doi: 10.1016/j.cognition.2009.02.003

Gardner, M. R., and Potts, R. (2010). Hand dominance influences the processing of observed bodies. *Brain Cogn.* 73, 35–40. doi: 10.1016/j.bandc.2010.02.002

Gardner, M. R., and Potts, R. (2011). Domain general mechanisms account for imagined transformations of whole body perspective. *Acta Psychol.* 137, 371–381. doi: 10.1016/j.actpsy.2011.04.006

Guterstam, A., Gentile, G., and Ehrsson, H. H. (2013). The invisible hand illusion: multisensory integration leads to the embodiment of a discrete volume of empty space. *J. Cogn. Neurosci.* 25, 1078–1099. doi: 10.1162/jocn_a_00393

Heider, F., and Simmel, M. (1944). An experimental study of apparent behavior. *Am. J. Psychol.* 57, 243–259. doi: 10.2307/1416950

Hommel, B. (1993a). Inverting the Simon effect by intention - Determinants of direction and extent of effects of irrelevant spatial information. *Psychol. Res.* 55, 270–279. doi: 10.1007/BF00419687

Hommel, B. (1993b). The role of attention for the Simon effect. *Psychol. Res.* 55, 208–222.

Hommel, B. (2011). The Simon effect as tool and heuristic. *Acta Psychol.* 136, 189–202. doi: 10.1016/j.actpsy.2010.04.011

Hommel, B., Müsseler, J., Aschersleben, G., and Prinz, W. (2001). The theory of event coding (TEC): a framework for perception and action planning. *Behav. Brain Sci.* 24, 849–878. doi: 10.1017/S0140525X01000103

Janczyk, M. (2013). Level 2 perspective taking entails two processes: Evidence from PRP experiments. *J. Exp. Psychol.* 39, 1878–1887. doi: 10.1037/a0033336

Kalckert, A., and Ehrsson, H. H. (2014). The moving rubber hand illusion revisited: comparing movements and visuotactile stimulation to induce illusory ownership. *Conscious. Cogn.* 26, 117–132. doi: 10.1016/j.concog.2014.02.003

Kiefer, M. (2007). Top-down modulation of unconscious "automatic" processes: a gating framework. *Adv. Cogn. Psychol.* 3, 289–306. doi: 10.2478/v10053-008-0031-2

Kornblum, S., Hasbroucq, T., and Osman, A. (1990). Dimensional overlap: cognitive basis for stimulus-response compatibility–a model and taxonomy. *Psychol. Rev.* 97, 253–270. doi: 10.1037/0033-295X.97.2.253

Ma, K., and Hommel, B. (2013). The virtual-hand illusion: Effects of impact and threat on perceived ownership and affective resonance. *Front. Psychol.* 4:404. doi: 10.3389/fpsyg.2013.00604

Ma, K., and Hommel, B. (2015a). Body-ownership for actively operated non-corporeal objects. *Conscious. Cogn.* 36, 75–86. doi: 10.1016/j.concog.2015.06.003

Ma, K., and Hommel, B. (2015b). The role of agency for perceived ownership in the virtual hand illusion. *Conscious. Cogn.* 36, 277–288. doi: 10.1016/j.concog.2015.07.008

Makin, T. R., Holmes, N. P., and Ehrsson, H. H. (2008). On the other hand: dummy hands and peripersonal space. *Behav. Brain Res.* 91, 1–10. doi: 10.1016/j.bbr.2008.02.041

May, M., and Wendt, M. (2013). Visual perspective taking and laterality decisions: Problems and possible solutions. *Front. Hum. Neurosci.* 7:549. doi: 10.3389/fnhum.2013.00549

Morey, R. D. (2008). Confidence intervals from normalized data: a correction to cousineau (2005). *Tutor. Quant. Methods Psychol.* 4, 61–64. doi: 10.20982/tqmp.04.2.p061

Müsseler, J., Aschersleben, G., Arning, K., and Proctor, R. (2012). Reversed effects of spatial compatibility in natural scenes. *Am. J. Psychol.* 122, 325–336.

Müsseler, J., Ruhland, L., and Böffel, C. (2017). "Reversed effects of spatial compatibility when taking avatar's perspective," in *Proceedings of the Abstracts of the 20th Conference of the European Society for Cognitive Psychology*, Potsdam, 216.

Ottoboni, G., Tessari, A., Cubelli, R., and Umiltà, C. (2005). Is handedness recognition automatic? A study using a Simon-like paradigm. *J. Exp. Psychol.* 31, 778–789. doi: 10.1037/0096-1523.31.4.643b

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: transforming numbers into movies. *Spat. Vis.* 10, 437–442. doi: 10.1163/156856897X00366

Proctor, R. W., and Vu, K.-P. L. (2006). *Stimulus-Response Compatibility Principles: Data, Theory and Application.* Boca Raton, FL: CRC Press. doi: 10.1201/9780203022795

Proctor, R. W., and Vu, K.-P. L. (2010). Stimulus–response compatibility for mixed mappings and tasks with unique responses. *Q. J. Exp. Psychol.* 63, 320–340. doi: 10.1080/17470210902925270

Santiesteban, I., Catmur, C., Hopkins, S. C., Bird, G., and Heyes, C. (2014). Avatars and arrows: implicit mentalizing or domain-general processing? *J. Exp. Psychol.* 40, 929–937. doi: 10.1037/a0035175

Shaffer, L. H. (1965). Choice reaction with variable S-R mapping. *J. Exp. Psychol.* 70, 284–288. doi: 10.1037/h0022207

Shimada, S., Fukuda, K., and Hiraki, K. (2009). Rubber hand illusion under delayed visual feedback. *PLoS One* 4:e6185. doi: 10.1371/journal.pone.0006185

Taylor, A., Flynn, M., Edmonds, C. J., and Gardner, M. R. (2016). Observed bodies generate object-based spatial codes. *Acta Psychol.* 169, 71–78. doi: 10.1016/j.actpsy.2016.05.009

Tsakiris, M. (2010). My body in the brain: a neurocognitive model of body-ownership. *Neuropsychologia* 48, 703–712. doi: 10.1016/j.neuropsychologia.2009.09.034

Tversky, B., and Hard, B. M. (2009). Embodied and disembodied cognition: spatial perspective-taking. *Cognition* 110, 124–129. doi: 10.1016/j.cognition.2008. 10.008

Wallace, R. J. (1971). S-R compatibility and the idea of a response code. *J. Exp. Psychol.* 88, 354–360. doi: 10.1037/h0030892

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

Zwickel, J. (2009). Agency attribution and visuospatial perspective taking. *Psychon. Bull. Rev.* 16, 1089–1093. doi: 10.3758/PBR.16.6.1089

Check for updates

Enacting Space in Virtual Reality: A Comparison Between Money's Road Map Test and Its Virtual Version

Francesca Morganti*

Department of Human and Social Sciences, University of Bergamo, Bergamo, Italy

In the field of spatial cognition research the mutual relationship between perception and action that brings out spatial orientation was lately investigated. Besides, the sameness between creating a cognitive map from the exploration of a not simulated environment, from the use of an allocentric (survey-like) sketched map, and from the interaction with egocentric (route-like) 3D virtual environments, is generally contrived. To understand if different embodied affordances could provide different knowledge organization during wayfinding through the use of distinct spatial simulations, the same group of 61 healthy subjects experienced both the classical version of the Money's Road Map test (M-RMT) and a virtual reality version of the Road Map test (VR-RMT). The M-RMT requires a allocentric to egocentric right/left reasoning to explore a stylized city provided in a survey perspective. The VR-RMT is a 3D version of the same environment through which participants can actively navigate by choosing egocentric-based right/left directions in a route perspective. The results showed that the different embodiments afforded by the two environments and the increasing complexity in turn types provides different spatial outcomes. Results were discussed according to the sensorimotor coupling theory provided from the enactive cognition approach and significances for spatial cognition research were provided.

Keywords: enactive cognition, spatial cognition, virtual reality, Money's Road Map test, egocentric and allocentric coordinates

OPEN ACCESS

INTRODUCTION

The rearmost neuroscientific findings have implied a large overlaying between action and perception inserting the challenge of a spatial cognition research within the enactive approach. This cognitive framework change requires the reshaping of what "interaction" means (Morganti, 2016).

Within the embodied cognition perspective, in fact, is the sensorimotor coupling of the agent's action and of her environmental perception that shapes the possibilities for spatial exploration (Gibson, 1979; Varela et al., 1991; Thompson and Varela, 2001). Thus, spatial cognition derives from the agent's management of an action and from the maintenance of her moment-by-moment sensorimotor schema. This schema "guides" the agent in how to appropriately execute her movements in the specific situation in which she finds herself and what sorts of feedback to expect from the environment (Carassa et al., 2005).

The enactive approach on interaction has some unequivocal implications for spatial cognition research. Orientation, in fact, is a high level cognitive ability that comprises the construction

and use of a spatial representation of the context within which an action is performed. To be effectual it exacts information originated from multiple domains, the perpetual placement of the individual who is acting, combined with the planning of behaviors that are claimed to be ranged with the agent (Gramann et al., 2005). Through catching the opportunities for actions during a new environment exploration, the agent organizes spatial knowledge through egocentred maps (derived from routes traveled in which borders and landmarks can be individuated) and, in the meanwhile, to place herself in the environment by using allocentred maps (based on survey pathway combination) (Brunyeì et al., 2012). Together route and survey viewpoints can be contemplated as "commonplace." Moreover, their reciprocal conversions are an essential procedures backing a productive navigation of intricate environments (Hartley et al., 2003; Ishikawa and Montello, 2006).

Therefore, wayfinding can be conceived as based on the continuous equilibrium between egocentric and allocentric perspectives during the agent's perception–action coupling. Thus, the allocentric perspective supports spatial understandings while the agent is involved in a wayfinding that provides her with egocentric information. Nevertheless, by underwriting what surrounding dynamics are the most befitting among the numerous available at the time, an agent has possibility to plan in advance a path, even in a partly unknown environment, by creating spatial inferences (Morganti et al., 2007). Spatial plans, in fact, can't be considered as pure allocentric action representations (that have to be followed thoughtlessly), but they turn out to be controllers for action to be additionally detailed in the egocentric-based interaction with the surrounding space.

Neuroscience studies support this allocentric/egocentric balance for spatial cognition (Serino et al., 2014), benehating the role of the retrosplenial cortex in the merging of the allocentric data (provided by the Papez circuit) with the egocentric ones received from parietal areas (Burgess, 2006). These neuroscientific evidences evoke how the spatial orientation is inseparable from the embodied perspective and from the specific opportunities for action caught in the explored context (Gunzelmann, 2008).

Thus it is possible to assume that, during a new environment exploration, an agent bodily enacts with a context in a continuous developing process. Accordingly, exploration can be considered as not simply guided by agent goals or motor actions, rather from the everlasting "hook up" of perceptions and actions that creates the agent's way of experiencing the context in which she is included. Moreover, when an agent and a specific environment interact, they are structurally coupled and they co-emerge.

In the last decade, due to the progression of technology, virtual reality simulations were widely introduced in neuroscience and experimental psychology (Morganti, 2004). Together with paper and pencil simulation of environments (such as building plans, city maps, and so on) they were largely used to study spatial cognition. Both these kind of simulations have been generally considered as equivalent to natural place explorations. Moreover, virtual reality by the use of motion devices (such as head- and limb-trackers) can provide a configuration of "natural-like" sensorimotor coupling within the digital environment, providing

the agent with the possibility of actively catch opportunities for action in a computerized three-dimensional space. Even if the spatial knowledge organization derived from virtual environment simulation can be linked to an embodied perception grounded on a situated action, a research question arises here: might the coupling between an agent and the perceptive data provided by the digital environment create a different kind of spatial knowledge representation from the one obtainable to classical map-use? Might it have an impact on actions' choice that an explorer can perform within the environment? Adopting the enactive perspective to spatial simulation-based interaction, in fact, requires reconsidering the definition of the nature of the coupling between the agent and the context and of the possible reciprocal modifications and changes between them (Mellet-d'Huart, 2006). Map-based and virtual reality simulated contexts can provide an agent with specific affordances, and with the possibility to obtain spatial representations from a peculiar coupling with a device-mediated sensorimotor system. This could result in form of agent-environment regularities (e.g., spatial invariants) different in virtual reality simulated and map-based spaces understanding. We consider that as the main issue of our research.

To study how the agent-environment coupling could be in two different spatial simulations, the same group of healthy subjects experienced both the classical paper version of the Money's Road Map test (M-RMT – Money et al., 1967) and a virtual reality version of the Road Map test (VR-RMT – Morganti et al., 2009). As it includes the allocentric egocentric coordination and it is considered an ecologically-like spatial simulation, in the neuropsychological evaluation of spatial ability after brain injury the classical version of M-RMT is generally included. To be solved, in fact, this task requires to egocentrically think about a right/left rotation during the exploration of a sketched city map provided in the allocentric perspective. As the other side of the medal the nowadays exist a virtual VR-RMT that provides participants with an explorable three-dimensional version of the M-RMT in which there is the possibility to actively choose the right/left turns from a egocentric perspective.

The main aim of research is to compare the M-RMT and the VR-RMT in order to understand whether there is any difference between set-out a right/left turn on a body axis (as in the M-RMT) and performing it (as in the VR-RMT) in order to obtain a spatial perspective from the simulated world. Accordingly, our research methodology requires the following steps from participants:

- In the M-RMT condition, participants first look at the map, then delineate how to move on the body axis and finally obtain (and have to keep in mind) the spatial perspective derived from the turn.
- In the VR-RMT condition, participants first look at the map, then in the virtual environment can actively turn right or left on the body axis and obtain the spatial perspective accordingly.

The comparison between M-RMT and VR-RMT proposed here introduces two different spatial simulations that might provide participants with different embodied affordances. They

can, in fact, be considered as tightly linked with different sensory-motor coupling situations. In particular, in the VR-MRT an agent is required to plan in advance a right/left turn and to continuously create relationships between the perspectives obtained in the environment with the result of each turns. While in the M-RMT the agent has to translate information perceived on a map to a possibility of action that can be performed in the environment (but only imagined and taken in mind during exploration). Thus, it is possible to hypothesize that to observe the resulting of a right/left turn in the virtual environment requires unalike cognitive efforts than to ground it on a pure internal cognitive process as in the M-RMT process. These differences in the sensorimotor coupling between the perceptual information and the turn possibilities on the VR-RMT and M-RMT involves a different idea of body (device mediated and not-mediated ones) and it might create different experience for the agent during navigation. Moreover, the invariants of the physical world, obtainable from the active interaction within the virtual environment (the peculiar spatial perspectives faced after a right/left turn in the VR-RMT) might guide the agent's wayfinding in a different manner from the ones provided by the necessity inference on how a spatial perspective can be following a right/left turn in the M-RMT.

Accordingly, it is hypothesized here that the non-identical activities performed in the differently simulated environments will result in distinguishable orientation outcomes. Thus, the main hypothesis is that the peculiar M-RMT and VR-MRT sensorimotor coupling can have role in performing wayfinding and also in facing the increasing complexity of the right/left turns during exploration. Finally, we would like to understand whether, only for the VR-RMT, some individual differences exist in spatial orientation derived from age and computer interaction expertise. We expect that the rotation in VR can be difficult to perform if the participant does not have sufficient expertise in managing computer-based simulations or might present a slight cognitive frailty due to their specific age cohort.

MATERIALS AND METHODS

Materials

The M-RMT (Money et al., 1967) is a test of left–right discrimination. It consists of a stylized city map, depicted in **Figure 1**, in which participants indicate on a 32-step dotted pathway the direction taken at each turn (left or right) in order to follow a designated route. The answers require an allocentric to egocentric based reasoning, because the dotted pathway follows an erratic trace both away from and toward the agent, who is not allowed to turn the map or to make head and body movements to give the correct answer.

The VR-RMT (Morganti et al., 2009), is a virtual reality version of the M-RMT, in which the paper and pencil version is turned into an actively navigable city from an egocentric perspective. No landmarks are depicted as navigation cues, and all the buildings in the virtual simulation have the same texture. The VR-RMT was developed with 3D Game Studio software by which 3D buildings were developed on the basis of buildings' shape and position in

the paper and pencil version of the test. The navigation speed was constant. It was approximately 5 m and 40° per second.

The VR-RMT was administered on an Intel personal computer and was presented on a wall by a video projector that provides a 1,50 m × 1 m image. The participants was seated in a chair approximately 2 m from the virtual environment image depicted on the wall and moved in the virtual environment using a facilitate narrow keyboard (The QueenKey 2.5 × 2.5 narrow keyboard) placed on a small table in front of them.

A snapshot of the M-RMT and VR-RMT was provided in **Figure 1**.

Participants

In this study, we administered both the M-RMT and the VR-RMT to 83 healthy right-handed volunteers aged from 30 to 80 years. Sixty one participants remains enrolled in the study after the assessment of keyboard use and virtual reality familiarity whose mean age was 56.82 and $SD = 15.47$. We divided participants into three groups according to their age. The experimental population presents 19 Young Adults (YA, from 30 to 49 years old), 19 Adults (A, from 49 to 64 years old), and 23 Old Adults (OA, from 65 to 80 years old). In order to avoid confounding variables, such as sex differences in spatial skills, male and female gender was balanced. The participants included 31 females and 30 males with 5 to 19 years of education (Mean = 12,08; $SD = 3,62$). All subjects participated as volunteers and gave informed consent for their data treatment. No participant had a clinical history of neurological and mood disorders such as anxiety/depression.

Procedure

In order to exclude participants with deficits in cognitive domains, the Mini Mental State Examination (MMSE – Folstein et al., 1975) was performed. Participants who had a poor performance (cut off value 24/30) on the MMSE were excluded. After the cognitive evaluation, participants were introduced to the experimental phase.

Using a different virtual environment from the experimental one, a 10-min training session was run to familiarize the participants with the use of a keyboard for navigating in virtual reality. After 10 min, if participants felt comfortable with the keyboard and had satisfactorily demonstrated their ability to guide themselves within the environment, the participants were included in the experimental study. If the participant was not able to navigate the training virtual environment, she was excluded from participation in the experiment. The participants included in the study were also evaluated as slight/average/good in computer interaction by the experimenter, according to the expertise they showed in managing the narrow keyboard to move in the virtual environment. If, according to three expert observers, they were able to quickly move in the keyboard and understanding the correspondence between their finger movements and the effect of them in the virtual environment, they were classified as good. If they require some more training they were classified as average, if they ask for some support from the experimenter they were classified as slight. Nevertheless, all the participant at the end of the training session have to perform the task without experimenter help to be enrolled in the study.
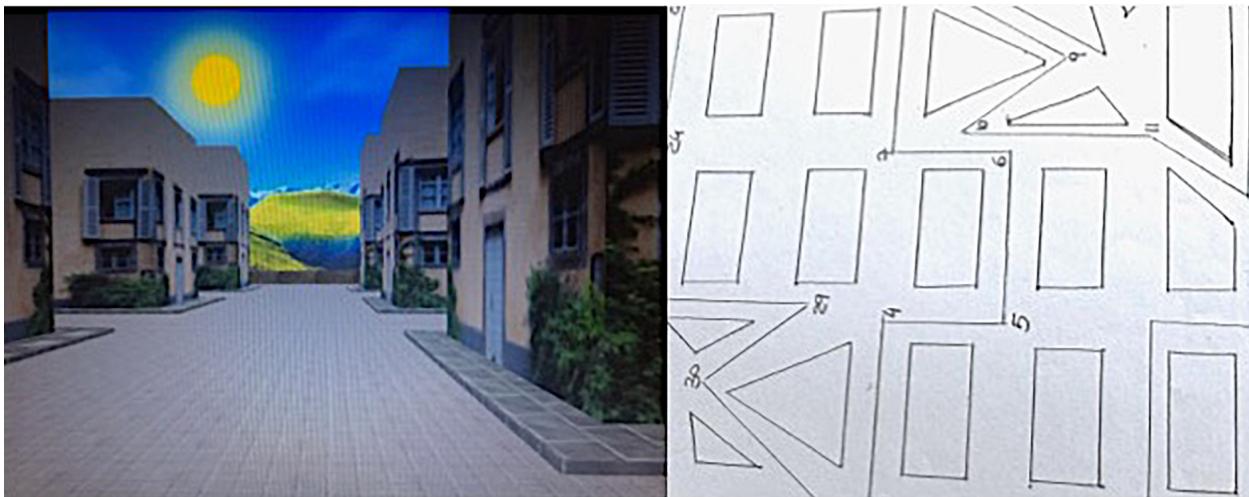
FIGURE 1 | Snapshots of VR-RMT and M-RMT used in the study.

In the experimental phase, the participants were tested individually. They were asked to perform both the M-RMT and the VR-RMT. The two versions of the test were randomly presented to participants. Half the participants performed the M-RMT first and the other half of the participants performed the VR-RMT first. In both the version of the tests, the starting point and the target point were clearly indicated.

In the M-RMT, we asked participants to follow on the sketched map a route taken by a hypothetical traveler. The participant was seated facing the examiner. She was asked to imagine herself moving along a 32-turn (choice points) route indicated by the experimenter on the map. Then, she had to spatially rotate himself to ascertain whether a right or left turn was demanded at each multiple-choice intersection. At each turn point, the participant had to answer the examiner's question: "In order to follow the depicted route, at this point would you be turning right or left?" The map always remained in a fixed position in front of the subjects, who were not allowed to alter their position to facilitate right–left judgments.

In the VR-RMT, the participants viewed virtual environment depicted on the screen with the paper version of the test placed in the table in front of them. While the examiner followed with her finger the route indicated by a dotted line on the paper version of the test, the participant decided which direction she must turn in the virtual environment and turned at each of 32 intersections.

In the M-RMT condition at the top side of the paper the north direction can be easily visible. In the VR-RMT, a sun straight visible from the participant's starting point indicated the corresponding north direction. Before the start of the VR-RMT exploration, the correspondence between the starting position on the paper and in the virtual environment was clearly indicated to participants. Participants could see the paper version of the test during VR-RMT navigation, but they can't rotate the paper in order to follow to the direction taken in the virtual environment. Participants could use the north-sun correspondence to re-orient themselves during the virtual exploration. Each time the participant considered one of the 32 turn points she had reached, she had to orally relate her decision to the experimenter.

In both the M-RMT and VR-RMT, there were equal numbers of right and left turns. A 10-min time limit was imposed for completing the test.

RESULTS
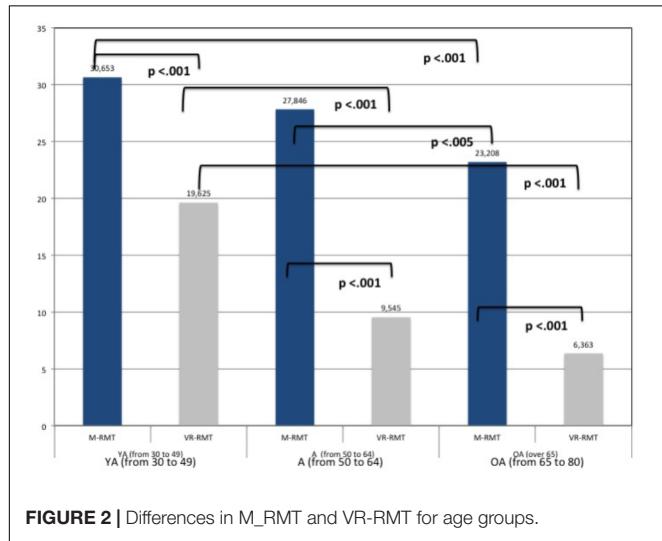
In the first global analyses of performance, both for M-RMT and VR-RMT one point was given for a correct answer—the correct direction (right or left) at each turn—for a maximum of 32 points for each test. In order to test environment consistency first we had a positive correlation between the M-RMT and the VR_MRT (Pearson's $r = 0.58$; $p < 0.001$).

In order to analyze the differences in exploring the two versions of the same environment a repeated measure 2x2x3 ANOVA was conducted. The statistic model includes as within factor Environment (2 levels: M-RMT/VR-MRT) * Presentation Order (2 levels: M-RMT first/VR-MRT first) * Age Group (3 levels: YA/A/OA) as between factors. Descriptive data are depicted in **Table 1**.

Results showed a significant difference [$F(399.21)$, $p < 0.001$] for the factor Environment. Participants better performed the spatial task in the M-RMT (Mean = 27.10; SD = 4.6) than in the VR-MRT (Mean = 11.34; SD = 8.08). Moreover, there is a significant difference in the interaction between Environment and Age Group [$F(8.164)$, $p < 0.001$]. *Post hoc* analysis with Bonferroni adjustment revealed significant differences between YA, A, and OA. When it comes to the M-RMT, there is a better performance by the YA ($p < 0.001$) and the A ($p < 0.001$) compared to the OA; there are no significant differences between the YA and A. As far as the VR-MRT is concerned there is a better performance by the YA compared to the A ($p < 0.001$) and to the OA ($p < 0.001$); there are no significant differences between A and YA.

TABLE 1 | Participant's performances at the M-RMT and VR-RMT according to presentation order and age groups.

	Task order	Age group	Mean	SD	Participants
M_RMT	1 M-RMT first	YA (from 30 to 49)	30,18	2,040	11
		A (from 50 to 64)	28,00	3,521	6
		OA (from 65 to 80)	24,67	4,670	15
		Total	**27,19**	**4,425**	**32**
	2 VR-RMT first	YA (from 30 to 49)	31,13	1,126	8
		A (from 50 to 64)	27,69	3,276	13
		OA (from 65 to 80)	21,75	4,862	8
		Total	**27,00**	**4,877**	**29**
	Total	YA (from 30 to 49)	30,58	1,742	19
		A (from 50 to 64)	27,79	3,259	19
		OA (from 65 to 80)	23,65	4,839	23
		Total	**27,10**	**4,607**	**61**
VR_RMT	1 M-RMT first	YA (from 30 to 49)	20,00	8,050	11
		A (from 50 to 64)	11,17	7,387	6
		OA (from 65 to 80)	6,60	3,269	15
		Total	**12,06**	**8,455**	**32**
	2 VR-RMT first	YA (from 30 to 49)	19,25	9,377	8
		A (from 50 to 64)	7,92	3,499	13
		OA (from 65 to 80)	6,12	2,800	8
		Total	**10,55**	**7,721**	**29**
	Total	YA (from 30 to 49)	19,68	8,387	19
		A (from 50 to 64)	8,95	5,071	19
		OA (from 65 to 80)	6,43	3,057	23
		Total	**11,34**	**8,083**	**61**

FIGURE 2 | Differences in M_RMT and VR-RMT for age groups.

Moreover, pairwise means comparison (t-test) revealed that there are significant differences between the Environments for all the three Age Groups. Data are depicted in **Figure 2**.

Finally there was no significant difference in Presentation Order [$F(1.133)$, $p = 0.292$], nor in Environment * Presentation

Order [$F(0.224)$, $p = 0.638$], nor in the Environment * Presentation Order * Age Group [$F(1.16)$, $p = 0.321$].

From the literature, we know that the spatial task in the M-RMT involves different levels of difficulty, defined by the direction of the virtual traveler on the map as seen from the subject's position (Vingerhoets et al., 1996; Rainville et al., 2002).

In order to account for the fact that left–right discrimination and mental rotation are two different abilities involved in the Road Map spatial task, both for M-RMT and VR-RMT the 32 turns were divided into three types according to the differentiation described by Vingerhoets et al. (1996). As indicated by Vingerhoets and colleagues, we classified the 32 turns of the tests, placing each turn in one of the three following categories:

(a) the correct left–right turn doesn't require mental rotation [no rotation (NR)];
(b) the correct left–right turn requires a 90° mental rotation [half rotation (HR)];
(c) the correct left–right turn requires a > 90°, < 180° mental rotation [full rotation (FR)].

Both the paper and the virtual Road Map present 8 NR, 16 HR, and 8 FR points.

Accordingly, a repeated measure 2x3x3 ANOVA was conducted. The statistic model includes as within factor Environment (2 levels: M-RMT/VR-MRT) * Turn Type (3 levels: NR/HR/FR) * Age Group (3 levels: YA/A/OA) as between factor.

Results showed a significant difference [$F(648.83)$, $p < 0.001$] for Turn Type, for the interaction between Turn Type and Environment [$F(179.53)$, $p < 0.001$] and for the interaction between Turn Type and Age Group. There was no statistical significance [$F(2.34)$, $p = 0.059$] in the interaction between Turn Type, Environment, and Age Group.

With regards to the Environment, pairwise means comparison (t-test) revealed significant differences between M-RMT and VR-MRT for NR [$t(60) = 17.86$; $p < 0.001$], HR [$t(60) = 18.69$; $p < 0.001$], and FR [$t(60) = 16.63$; $p < 0.001$]. Performances in M-RMT revealed higher means compared to performances in VR-MRT.

With regards to the Turn Type, post hoc analysis with Bonferroni adjustment revealed significant differences between NR and HR ($p < 0.001$), HR and FR ($p < 0.001$), but no significant differences between NR and FR ($p = 0.544$) in the M-RMT. There are significant differences between NR and HR ($p < 0.001$), HR and FR ($p < 0.001$), and between NR and FR ($p < 0.001$) in the VR-RMT. Finally, significant differences between M-RMT and VR-RMT are observed for the three turn type(s).

At last, a post hoc analysis with Bonferroni adjustment revealed that Age Group influenced differently Turn Type performances in M-RMT and VR-RMT. For all the three types of rotations there are significant differences between YA and OA (NR $p < 0.001$; HR $p < 0.001$; FR $p < 0.001$), and between A and OA (NR $p < 0.005$; HR $p < 0.003$; FR $p < 0.05$) in M-MRT; while there are significant differences between YA and A (NR $p < 0.001$; HR $p < 0.001$; FR

$p < 0.001$), and between A and OA (NR $p < 0.001$; HR $p < 0.001$; FR $p < 0.001$) in M-MRT. Detailed values are depicted in **Table 2**.

CONCLUSION

Starting from the enactive cognition approach the main research question proposed in this study was about the equivalence between a spatial orientation assessment obtainable from a classical neuropsychological test and the one obtainable from a virtual reality based one. Specifically, as in clinical neuropsychology the classical tests generally provide the patients with an allocentric simulation of space (e.g., a maze or a sketch map) the evaluation of spatial ability might differ from the one derived providing patients with the egocentric perspective possible in the virtual environments. In the classical assessment, in fact, an agent has to translate the allocentric perception in egocentred action, while during the virtual assessment the agent is allowed to move within the environment in the egocentric perspective. As the concept of enaction have introduced the notion of the coevolution of the agent and its environment, the main research question was about if it is possible to create equivalent representations of the surrounding environment in terms of opportunities for action (affordances) and sensorimotor invariants both in allocentred and egocentred spatial simulations.

In exploring a virtual environment an agent took embodied opportunities for action that are granted to the her from the simulation, on the basis of the atypical interaction provided by the computer simulated environment. These kind of affordances are not provided by the environment *per se* but from the interaction between the explorer and the virtual environment. Consequently, it appeared to be necessary to determine if the orientation obtainable from a virtual environment might differ from the spatial orientation obtainable from other

kind of simulations (e.g., an analogical simulation like a sketched map). Thus, the different kind of body–environment coupling was analyzed here in two different forms of the same neuropsychological test.

Even if spatial cognition in virtual environment is comparable to the spatial orientation obtainable from the navigation other simulated spaces, due to the "sense of presence" experienced in it (Carassa et al., 2005; Riva et al., 2011), the present study revealed several significant differences between these two experimental conditions. The VR-RMT appears to be more complex to solve than the M-RMT. This difference between the two tests seems to be directly addressable to the complexity of the turn type in spatial exploration.

Considering nature of the tasks it is possible to observe that in the VR-RMT, the half of participants were asked to use the paper version of the test to perform turns in the virtual environment. It could have been interpreted as a dual task condition, requesting participants to first take a decision about the turns through using the paper-simulated environment and thus to translate the same decision in the virtual-simulated environment. To perform the VR-MRT requires a continuous attention focus change between the two simulations and a perspective switch between the survey of the M-RMT and the route of the VR-MRT. Thus, the finding that the performance was worse in this condition may not be very surprising.

Primarily it is possible to solve the M-RMT by imagining egocentric spatial transformations (Schultz, 1991) whereas in the VR-MRT, participants took decisions for each turn point being in front of the screen and by acting according to the appropriateness of their choices. The M-RMT and the VR-RMT differs in the imagined/perceived perspective taking because in the first task the agent have to set-out how to modify the turn on her body axis and how to derive a new perspective from that turn, whereas in the second task, the agent directly perform the turn on the body axis and directly perceive the point of view modification derived from it. Moreover, the VR-RMT does not require the participant to continuously re-locate herself looking at the map, because the track of each position is done by the experimenter and doesn't require an additional cognitive effort.

Following the second interpretation, we expected a presentation order effect (between the group who experienced the M-RMT or the VR-RMT first) and also a better performance on the VR-RMT. Instead, the participants don't express a presentation order effect and performed worse on the VR-RMT. Thus, independently from the presentation order, the VR-RMT was more complicated to perform than the M-RMT. A possible explanation of this experimental result may be related to the difference between simulation and action: rotating the body on its vertical axis toward the point of reference in virtual reality is more difficult than rotating the body in a mental space. Tversky (2009) underlines that human being continuously experience their own body from inside, influencing the peri-personal space that is independent from the physical environment *per se*. Moreover, it is possible to consider perspective taking and mental rotation as dissociated. When perspective taking, in fact, includes thinking about the changing of the owns egocentric perspective with respect to the surroundings, the mental rotation includes

TABLE 2 | Means and Standard deviations for rotation type, task and age group.

Age group	Environment	Rotation	Mean	SD
YA (form 30 to 49)	M-RMT	No rotation	7,421	0,223
		Half rotation	15,368	0,432
		Full rotation	7,632	0,233
	VR-RMT	No rotation	4,789	0,348
		Half rotation	9,684	0,677
		Full rotation	5,421	0,327
A (from 50 to 64)	M-RMT	No rotation	6,895	0,223
		Half rotation	13,895	0,432
		Full rotation	7,000	0,233
	VR-RMT	No rotation	2,053	0,348
		Half rotation	4,263	0,677
		Full rotation	2,632	0,327
OA (from 65 to 80)	M-RMT	No rotation	5,870	0,203
		Half rotation	11,609	0,393
		Full rotation	6,174	0,212
	VR-RMT	No rotation	1,391	0,316
		Half rotation	2,913	0,616
		Full rotation	2,130	0,297

thinking about the effects of modifying the placements of objects in the surroundings during the maintenance of owns actual perspective in the environment (Hegarty and Waller, 2004).

In addiction, Hintzman et al. (1981) describe spatial knowledge as derived from orientation-specific perspectives, and of relational propositions. Accordingly Kozhevnikov and Hegarty (2001) indicate as the main strategy used in understanding a more than 90° perspective task is to imagine oneself reoriented with respect to the scene. This strategy could have to be used from this study participants. For both the M-RMT and VR-RMT, in order to follow the route participants have to imaginatively anticipate themselves in specific orientation. Generally, an agent is able to move on the gravitational axis while the environment doesn't provide variations. This kind of embodied turn creates an expectation about the spatial perspective (defined by Gibson's affordance theory as "invariants of the physical world") that could have been more efficacious in updating an imaginative world compared to the one of the virtual environment.

These results appear to be partially incongruent with current research in the field. As introduced by Gray and Fu (2004), in interacting with computer-based simulation, individuals were given the option of using the external visualizations to perceive the effect of their actions rather than relying on internal visualization to imagine the effect. In accord with Keehner et al. (2008), it is possible to think that in the VR-MRT task the agents matched the virtual environment snapshots with the right/left turn intentions in looking for the match between the obtained perspectives and the effect of each turns. This continuous reference matching can be considered as tightly coupled with internal cognitive processes. The possibility to externalize representations provided by VR-RMT (by observing the perspective resulting from a right/left turn) may have required more effort than to base it on the embodied imaginative process (as in the M-RMT). This data interpretation is also consistent with the perspective proposed by Di Paolo (2005). Accordingly, here we can suggest that in the VR-MRT, a failure of the sensorimotor coupling between the perceptual information and the turn response on the virtual scenario that doesn't involve the entire body, might have created a meaningless experience for the agent during navigation. Thus the failure of the sensorimotor coupling has been considered as quite useless for spatial orientation.

It is also possible to mention that in the VR-RMT, each mistake in turn taking provides a difference between the agents's expected and taken perspective in space that might influence the next turns affecting the final result more in the VR-RMT than in the M-RMT. This interpretation of the data appears to be supported by the analysis of our results on turn type. Managing HR/FR appears to be easier in the imaginative task than in the virtual one. This is largely observable from the individual differences in the analysis of our data: the results from Age Group comparison showed that our participants were not all equally able to use external visualizations to support spatial orientation in virtual reality. Moreover, the ability to orient them VR-RMT decreases with age. The interaction between Environment and Age Group, in fact, revealed how there is a difference between the younger groups (YA and A) and the older population wayfinding

performed in the M-RMT. It reveals a decline with age in the allocentric to egocentric spatial translation. Whereas in VR-MRT there is a difference of the YA both when compared to A and OA.

This result confirms that the orientation task both in M-RMT and VR-RMT is not equal for all individuals but that it is strictly dependent on the participants' age. Moreover, our data appear to be consistent with the recent findings in age-related decline for wayfinding in complex environments Harris and Wolbers (2014). By using a complex virtual environment for wayfinding ability evaluation in young and old populations they found a wide role of age on the capacity to change from route knowledge to survey one in order to find a target location. Moreover, in their work older participants showed evidence of difficulties in route to survey switching performance, confirming that it can be at least partly explained in terms of prefrontal-noradrenergic network impairment, responsible for egocentric to allocentric coordinating switching behavior.

Finally, the interaction between age group and spatial performances could be also addressed to a computer expertise that can be derived from the age of our participants. We have assumed, in fact, that our age cohorts reflect the possible everyday use of computer or other technological devices in the participants' everyday lives. We had the YA group that could be defined as a "digital native" and were largely exposed to computer-based interactions, the A group that is still a working population and could be quite expert in computer use, and the Old Adult group that is probably retired from work and might not have a large expertise with technologies. These groups appear to be different between M-RMT and VR-RMT. In VR-RMT it appears clear how OA had difficulties in managing turns and that it could be related to the participants' expertise in using computer-based simulations. The data derived from VR-RMT condition are consistent with the evidence that a variability between subjects in spatial task performance is high in virtual reality spaces (Klatzy et al., 1998; Waller et al., 1998). Most of the cognitive abilities involved in understanding space in a virtual simulation seems to be higher cost demanding.

As described above, by considering the sync between both the perspective as essential for spatial navigation and wayfinding, the differences in spatial evaluation obtainable from mainly allocentric or mainly egocentric environment simulations (and from the possibility of interaction they differently provide) have been deeply investigated. Consequently, in order to obtain solid data seems to be necessary to think about an assessment tool specific for virtual environment application (Belingard and Péruch, 2000; Waller, 2000, 2005). Otherwise, within the enactive perspective on cognition, data derived from spatial tasks performed through virtual reality simulations in largely restrictive action possibilities (e.g., neuroimaging studies) could be considered as not completely reliable.

As cognition is the form of embodied action in which cognitive processes arise from recurrent sensorimotor patterns of perception and action (Thompson, 2005), the coupling between organism and environment modulates the construction of a relational domain that is not internally represented in the brain but it is created from the activity and the peculiar coupling with the specific environment. This evidence suggests that

the opportunity of including virtual environments in cognitive evaluation is not exclusively technological, but epistemic. Thus, for spatial cognition evaluation, beyond considering the virtual simulation appropriateness, is equally important to understand the enaction stance that acknowledges orientation as derived from egocentric/allocentric sensorimotor invariance. Data presented here revealed how this sensorimotor invariance differed from the possibility of offloading spatial knowledge, as in the classical and virtual version of M-RMT.

Hence, enactive cognition can be considered ad a privileged point of view in examining virtual reality as more than purely digital place, but as a technical challenge in which an agent is able to find spatial invariants, and to progressively evolve them through the dynamics of the sensory-motor coupling. In this way she understand the environment and the possibilities for action in it.

Thus, the introduction on virtual reality in cognitive science research have to consider how this kind of simulation more than being "realistic" has to technically support the agents' possibility to potentially distinguish the moment-by-moment different paths of encounters with the environment (Di Paolo, 2005, 2009). The peculiar possibilities of sensorimotor coupling, defined for example from the environment characteristics and from the interaction design possibilities provided to the agent can supply explorers with "virtual reality- based" affordances for action and differentiated information feedbacks. Each of these should be deeply considered in order to understand how they could provide distinctive effort for spatial knowledge.

At last, the inclusion of virtual environments within the assessment tools for spatial cognition in neuropsychology may provide an interesting alternative to paper and pencil-based approaches but data derived from this evaluation have to be used with extremely caution. Virtual environments in fact appear here to not involve the same embodied spatial information derived from the navigation performed in other types of environments. Even if it remains a great challenge for enactive cognition research (Varela, 1990).

ETHICS STATEMENT

This study was carried out in accordance with the recommendations of University of Bergamo with written informed consent from all subjects. All subjects gave written informed consent in accordance with the Declaration of Helsinki. The protocol was approved by the University of Bergamo research office.

AUTHOR CONTRIBUTIONS

FM ideated the experiment, collected and analyzed the data, wrote the manuscript.

REFERENCES

Belingard, L., and Péruch, P. (2000). Mental representation and the spatial structure of virtual environments. *Environ. Behav.* 32, 427–442. doi: 10.1177/00139160021972603

Brunyeì, T. T., Gardony, A., Mahoney, C. R., and Taylor, H. A. (2012). Going to town: visualized perspectives and navigation through virtual environments. *Comput. Hum. Behav.* 28, 257–266. doi: 10.1016/j.chb.2011.09.008

Burgess, N. (2006). Spatial memory: how egocentric and allocentric combine. *Trends Cogn. Sci.* 10, 551–557. doi: 10.1016/j.tics.2006.10.005

Carassa, A., Morganti, F., and Tirassa, M. (2005). "A situated cognition perspective on presence," in *Proceedings of the XXVII Annual Conference of the Cognitive Science Society*, eds B. Bara, L. W. Barsalou, and M. Bucciarelli (Alpha, NJ: Sheridan Printing), 384–389.

Di Paolo, E. A. (2005). Autopoiesis, adaptivity, teleology, agency. *Phenomenol. Cogn. Sci.* 4, 429–452. doi: 10.1007/s11097-005-9002-y

Di Paolo, E. A. (2009). Extended life. *Topoi* 1, 9–21. doi: 10.1007/s11245-008-9042-3

Folstein, M. F., Folstein, S. E., and McHugh, P. R. (1975). "Mini-mental state". A practical method for grading the cognitive state of patients for the clinician. *J. Psychiatr. Res.* 12, 189–198. doi: 10.1016/0022-3956(75)90026-6

Gibson, J. J. (1979). *The Ecological Approach to Visual Perception.* Boston, MA: Houghton Mifflin.

Gramann, K., Muller, H. J., Eick, E., and Schonebeck, B. (2005). Evidence of separable spatial representations in a virtual navigation task. *J. Exp. Psychol. Hum. Percept. Perform.* 31, 1199–1223. doi: 10.1037/0096-1523.31.6.1199

Gray, W. D., and Fu, W. T. (2004). Soft constraints in interactive behavior: the case of ignoring perfect knowledge in-the-world for imperfect knowledge in-the-head. *Cognit. Sci.* 28, 359–382.

Gunzelmann, G. (2008). Strategy generalization across orientation tasks: testing a computational cognitive model. *Cognit. Sci.* 32, 835–861. doi: 10.1080/03640210802221957

Harris, M. A., and Wolbers, T. (2014). How age-related strategy switching deficits affect wayfinding in complex environments. *Neurobiol. Aging* 35, 1095–1102. doi: 10.1016/j.neurobiolaging.2013.10.086

Hartley, T., Maguire, E. A., Spiers, H. J., and Burgess, N. (2003). The well-worn route and the path less traveled: distinct neural bases of route following and wayfinding in humans. *Neuron* 37, 877–888. doi: 10.1016/S0896-6273(03)00095-3

Hegarty, M., and Waller, D. (2004). A dissociation between mental rotation and perspective-taking spatial abilities. *Intelligence* 32, 175–191. doi: 10.1016/j.intell.2003.12.001

Hintzman, D. L., O'Dell, C. S., and Arndt, D. R. (1981). Orientation in cognitive maps. *Cognit. Pschol.* 13, 149–206. doi: 10.1016/0010-0285(81)90007-4

Ishikawa, T., and Montello, D. R. (2006). Spatial knowledge acquisition from direct experience in the environment: individual differences in the development of metric knowledge and the integration of separately learned places. *Cognit. Psychol.* 52, 93–129. doi: 10.1016/j.cogpsych.2005.08.003

Keehner, M., Hegarty, M., Cohen, C. A., Khooshabeh, P., and Montello, D. R. (2008). Spatial reasoning with external visualizations: what matters is what you see, not whether you interact. *Cognit. Sci.* 32, 1099–1132. doi: 10.1080/03640210801898177

Klatzy, R. L., Loomis, J. M., Beall, A. C., Chance, S. S., and Golledge, R. G. (1998). Spatial updating of self- position and orientation during real, imagined, and virtual locomotion. *Psychol. Sci.* 9, 293–298. doi: 10.1111/1467-9280.00058

Kozhevnikov, M., and Hegarty, M. (2001). A dissociation between object-manipulation spatial ability and spatial orientation ability. *Mem. Cognit.* 29, 745–756. doi: 10.3758/BF03200477

Mellet-d'Huart, D. (2006). A model of (en)action to approach embodiment: a cornerstone for the design of virtual environment for learning. *Virtual Real.* 10, 253–269. doi: 10.1007/s10055-006-0038-2

Money, J., Alexander, D., and Walker, H. T. (1967). *A Standardized Road Map Test of Direction Sense.* Baltimore: Johns Hopkins Press.

Morganti, F. (2004). Virtual interaction in cognitive neuropsychology. *Stud. Health Technol. Inform.* 99, 55–70.

Morganti, F. (2016). Embodied space in natural and virtual environments: implications for cognitive neuroscience research. *Commun. Comput. Inform. Sci.* 604, 110–119. doi: 10.1007/978-3-319-32270-4_11

Morganti, F., Carassa, A., and Geminiani, G. (2007). Planning optimal paths: a simple assessment of survey spatial knowledge in virtual environments. *Comput. Hum. Behav.* 23, 1982–1996. doi: 10.1016/j.chb.2006.02.006

Morganti, F., Marrakchi, S., Urban, P. P., Iannoccari, G. A., and Riva, G. (2009). A virtual reality based tools for the assessment of "survey to route" spatial organization ability in elderly population: preliminary data. *Cognit. Process.* 10, 257–259. doi: 10.1007/s10339-009-0284-9

Rainville, C., Marchand, N., and Passini, R. (2002). Performances of patients with a dementia of the Alzheimer type in the standardized road-map test of direction sense. *Neuropsychologia* 40, 567–573. doi: 10.1016/S0028-3932(01)00133-6

Riva, G., Waterworth, J. A., Waterworth, E. L., and Mantovani, F. (2011). From intention to action: the role of presence. *New Ideas Psychol.* 29, 24–37. doi: 10.1016/j.newideapsych.2009.11.002

Schultz, K. (1991). The contribution of solution strategy to spatial performance. *Can. J. Psychol.* 45, 474–491. doi: 10.1037/h0084301

Serino, S., Cipresso, P., Morganti, F., and Riva, G. (2014). The role of egocentric and allocentric abilities in Alzheimer's disease: a systematic review. *Ageing Res. Rev.* 16, 32–44. doi: 10.1016/j.arr.2014.04.004

Thompson, E. (2005). Sensorimotor subjectivity and the enactive approach to experience. *Phenomenol. Cognit. Sci.* 4, 407–427. doi: 10.3389/fnint.2013.00015

Thompson, E., and Varela, F. J. (2001). Radical embodiment: neural dynamics and consciousness. *Trends Cognit. Sci.* 5, 418–425. doi: 10.1016/S1364-6613(00)01750-2

Tversky, B. (2009). "Spatial cognition: embodied and situated," in *The Cambridge Handbook of Situated Cognition*, eds P. Robbins and M. Aydede (New York, NY: Cambridge University Press), 201–216.

Varela, F. J. (1990). "Il corpo come macchina ontologica," in *Che Cos'è la Conoscenza*, eds M. Ceruti and F. Preta (Bari: Laterza), 43–53.

Varela, F. J., Thompson, E., and Rosch, E. (1991). *The Embodied Mind: Cognitive Science and Human Experience*. Cambridge, MA: MIT Press.

Vingerhoets, G., Lannoo, E., and Bauwens, S. (1996). Analysis of the money road-map test performance in normal and brain damaged subjects. *Arch. Clin. Neuropsychol.* 11, 1–9. doi: 10.1093/arclin/11.1.1

Waller, D. (2000). Individual differences in spatial learning from computer simulated environment. *J. Exp. Psychol. Appl.* 6, 307–321. doi: 10.1037/1076-898X.6.4.307

Waller, D. (2005). The walkabout: using virtual environments to assess large-scale spatial abilities. *Comput. Hum. Behav.* 21, 243–253. doi: 10.1016/j.chb.2004.02.022

Waller, D., Hunt, E., and Knapp, D. (1998). The transfer of spatial knowledge in virtual environment training. *Presence* 7, 129–143. doi: 10.1162/105474698565631

How Our Cognition Shapes and Is Shaped by Technology: A Common Framework for Understanding Human Tool-Use Interactions in the Past, Present, and Future

François Osiurak[1,2*†], Jordan Navarro[1,2†] and Emanuelle Reynaud[1†]

[1] Laboratoire d'Etude des Mécanismes Cognitifs (EA 3082), Institut de Psychologie, Université de Lyon, Lyon, France, [2] Institut Universitaire de France, Paris, France

Over the evolution, humans have constantly developed and improved their technologies. This evolution began with the use of physical tools, those tools that increase our sensorimotor abilities (e.g., first stone tools, modern knives, hammers, pencils). Although we still use some of these tools, we also employ in daily life more sophisticated tools for which we do not systematically understand the underlying physical principles (e.g., computers, cars). Current research is also turned toward the development of brain–computer interfaces directly linking our brain activity to machines (i.e., symbiotic tools). The ultimate goal of research on this topic is to identify the key cognitive processes involved in these different modes of interaction. As a primary step to fulfill this goal, we offer a first attempt at a common framework, based on the idea that humans shape technologies, which also shape us in return. The framework proposed is organized into three levels, describing how we interact when using physical (Past), sophisticated (Present), and symbiotic (Future) technologies. Here we emphasize the role played by technical reasoning and practical reasoning, two key cognitive processes that could nevertheless be progressively suppressed by the proficient use of sophisticated and symbiotic tools. We hope that this framework will provide a common ground for researchers interested in the cognitive basis of human tool-use interactions, from paleoanthropology to neuroergonomics.

Keywords: tool use, technology, brain–computer interface, automation, technical reasoning

INTRODUCTION

Have you already wondered how researchers living 70 years ago could contact an editor to know whether their manuscript was still under review or not after 5 months? They certainly had to write a mail and wait for a response, perhaps 5 weeks after. Nowadays, we send emails and expect an answer by 2 or 3 days. Perhaps in 1000 years, researchers will just have to think of this and they will receive the answer instantly. These different modes of interaction illustrate the constant modification of our technologies over time, a phenomenon that characterizes our species (Boyd and Richerson, 1985). The ultimate goal of research on this topic is to identify the key cognitive processes involved in these different modes of interaction. As a primary step to fulfill this goal,

we offer a first attempt at a common framework, based on the idea that humans shape technologies, which also shape us in return.

The framework proposed is organized into three levels, describing how we interact when using physical (Past), sophisticated (Present), and symbiotic (Future) technologies[1]. The temporal gradient introduced here implies that, at the species level, physical technologies are anterior to sophisticated ones, which are anterior to symbiotic ones, so that the theoretical proportion of use for each kind of technology is supposed to evolve over time (**Figure 1**). The distinction made between these different kinds of technology is also theorized here at a cognitive level, based on the idea that our modifications on the world are first guided by an intention, needing then the selection of a practical solution (i.e., the practical level), and finally the selection and application of a technical action (i.e., the technical level; **Figure 2**). The thesis defended here is that the technical evolution from physical to sophisticated and symbiotic technologies tends to progressively suppress the technical and practical levels.

Three caveats need to be made at this point. First, there is no overview in the literature about the cognitive processes involved in the different interactions we have with tools and technologies. The major reason for this lack is that this requires a critical, epistemological development as to the way of organizing the field so that researchers from different topics (e.g., stone tools, brain–computer interaction) could communicate within a single and comprehensive framework. The goal of this paper is to fill this gap, by attempting to provide a structured way of organizing the literature based on the evolution of our technology over time. This attempt could be a good starting point for developing such a framework in the future. Second, many cognitive processes are involved in our interactions with tools and technologies. Here we could not address all of them and preferred to concentrate our attention on two key cognitive processes, namely, technical reasoning and practical reasoning. Of course, further theoretical development would be needed to complete our analysis. Third, as with other humans, our interactions with tools and technologies can take different forms according to the role taken by technology (e.g., competition, collaboration). These different levels of interaction that most directly deal with the "social" aspect will be addressed partly in this paper, particularly in the third section. Nevertheless, we acknowledge that a more comprehensive review based on this level of analysis could complete the present review, discussing the potential parallel between our interactions with social (e.g., humans) and non-social (e.g., technologies) agents.

THE PAST: PHYSICAL TOOLS

Physical tools can be defined as those tools that increase our sensorimotor abilities (Virgo et al., 2017). Although we still use a wide variety of physical tools (e.g., hammer, knife), it can be considered that they correspond to the first tools humans have made and used in *pre*-history. At a cognitive level, the use of

all physical tools shares the need for the user to understand physical principles (e.g., percussion, cutting). The characteristics of early stone tools indicate that makers showed evidence of a basic understanding of stone fracture mechanics (Hovers, 2012). The use of physical tools by modern humans also requires this form of physical understanding (Bril et al., 2010).

Some patients can meet difficulties to use everyday tools after left brain damage (Osiurak and Rossetti, 2017). The difficulties concern not only the selection of the appropriate tool, but also the mechanical action performed (e.g., pounding a nail by rubbing it on the nail instead of hammering with it). The same difficulties can be observed when they are asked to solve mechanical problems by using novel tools (Goldenberg and Hagmann, 1998; Jarry et al., 2013). Taken together, these findings indicate that the use of physical tools is grounded on the ability to reason about physical properties of tools and objects based on mechanical knowledge. This is what we call "technical reasoning" (Osiurak et al., 2010; Osiurak and Badets, 2016). This reasoning is critical to form a mental representation of the mechanical action intended. It is also the key process allowing us to generate instances of "technical misusage" (**Figure 2**) also called "function creep," corresponding to the use of a tool in an unusual way (Osiurak et al., 2009). Such instances can be observed relatively early in humans. A 2-years-old child can, for instance, use a tea spoon to hammer a piece of cheese in his mashed carrots, calling the spoon "a hammer." This child knows that the spoon is not a hammer but finds funny to hammer the cheese and handy to use the spoon to do so at that time.

Technical reasoning could be unique to humans (e.g., Penn et al., 2008), explaining a certain number of our specificities such as the use of one tool to create another (e.g., stone knapping) or the use of complex tools that transform our motor energy into different mechanical energies (Osiurak, 2017). Convergent evidence from neuropsychology and cognitive neuroscience indicates that technical reasoning could engage the area PF within the left inferior parietal cortex (Goldenberg and Spatt, 2009; Reynaud et al., 2016), which does not in macaques and other non-human primates (Orban and Caruana, 2014).

Before going on to the next section, one important aspect needs to be considered. Technical reasoning is critical for the making of any technology (physical, sophisticated, symbiotic). For physical technologies, there is no real distance between the maker and the user in that the user needs to mentally make the technology before the use (Osiurak and Heinke, 2017). If you intend to cut a tomato, you are free to select a wide variety of tools. Nevertheless, your selection is based on the physical properties of the tomato, leading you to choose a tool with the appropriate physical properties relatively to the tomato. In a way, you first make your tool mentally (e.g., thinking about something sharp and solid enough) and then you select it *really* accordingly. Things are different for sophisticated technologies, which mainly correspond to interface-based technologies (e.g., computers). A key characteristic of these technologies is that the maker/designer has facilitated the interaction, so that the user has no longer to understand the physical principles underlying the use. In this case, the user does not make mentally the tool before the use but learn the arbitrary relationship between the

---

[1]The terms tool and technology will be hereafter used interchangeably and in a broad sense to refer to any environmental object useful to increase the user's sensorimotor or cognitive capacities (Osiurak et al., 2010).
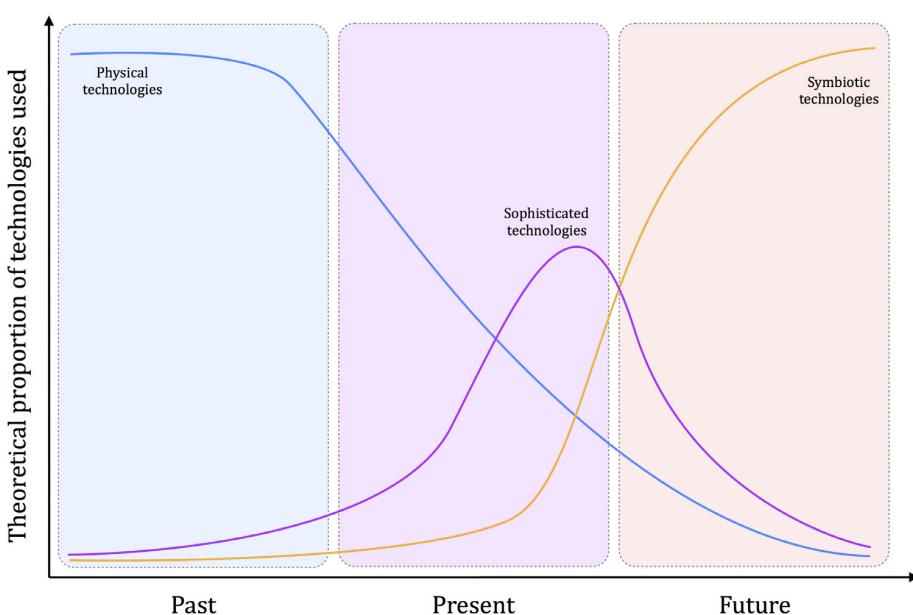
FIGURE 1 | Theoretical proportion of physical, sophisticated, and symbiotic technologies used over time. The core idea is that, at the species level, physical technologies are anterior to sophisticated technologies, which are also anterior to symbiotic technologies. Over time, physical technologies (e.g., stone tools, knifes, hammers) tend to decrease and could be completely absent in a far future. Sophisticated technologies have appeared later and are now a great part of the technologies we use (i.e., interface-based technologies). Again, it can be hypothesized that this kind of technologies will be less and less used. Finally, symbiotic technologies are developing now even if they remain rarely used (e.g., brain–computer interfaces). In a far future, it can be thought that humans will profusely and uniquely use these technologies. The three colored panels correspond to the three time periods (Past, Present, and Future). The color associated to each kind of technologies corresponds to the color of the period where a given technology is dominant (Past: the reign of physical technologies; Present: the reign of sophisticated technologies; Future: the reign of symbiotic technologies).

motor response and its effect. The corollary is that sophisticated technologies may not require, at the technical level (**Figure 2**), technical reasoning skills, but more basic cognitive processes such as associative learning and procedural memory (Osiurak and Heinke, 2017). At least two lines of evidence support this view. First, interface-based technologies (e.g., touchscreens) can be easily used by infants, despite moderate skills to use physical tools (Beck et al., 2011). Likewise, many non-human animals including tool users (e.g., baboons) can use touchscreens very quickly in the absence of any signs of physical tool use (Claidière et al., 2014). Second, patients with damage to the left inferior parietal cortex are impaired to use physical tools, but not interface-based technologies. The opposite pattern can be observed in patients with deficits of procedural memory (e.g., Parkinson's disease), indicating a double dissociation between the ability to use physical versus sophisticated technologies (see Osiurak, 2014, 2017).

THE PRESENT: SOPHISTICATED TOOLS

Stopping the alarm clock after waking up, using tramways, driving a car, interacting with a smartphone, taking the elevator, and so on. With the sophistication of tools and the advent of cognitive tools (e.g., computer spreadsheet) the distance between the making and the use has dramatically increased, so we use many tools we could never build in a lifetime. This does not

change the way we interact with tools: the purpose of a tool is not in the tool itself, but in the user's intentions. A computer screen can be used to stick notes, as a visual barrier, as a mirror, and so forth (i.e., technical misusage). This fact remains whatever the nature of the tool considered, from a very simple stone tool to the most advanced smartphone (e.g., reflecting sunlight). There is a limit, however, in the lack of freedom offered by sophisticated tools to its users at the technical level, because the use of these tools for their usual function needs to master pre-established procedures (see above).

Some sophisticated tools, often referred as automation, do not tend to extend humans but rather to replace them (Young et al., 2007). Those tools that replaces us tend to be poorly accepted by individuals (Navarro et al., 2011). The design of these tools also questions about the human role in our societies, and about what should be automated or not (Hancock, 2014). For instance, a highly automated task completion is often considered as dehumanizing (Coeckelbergh, 2015). People also select an automatic completion of the task only if much more effective than a manual completion (Osiurak et al., 2013; Navarro and Osiurak, 2015, 2017), as if humans tend to avoid the loss of freedom associated to sophisticated tools (**Figure 2**).

Tool use is not neutral for users. Of course, tools are changing the way humans do things, but tools also change humans themselves (Hancock, 2007). All the data available on the Internet provide considerable benefits, yielding information easily. But, it also alters the way people memorize information
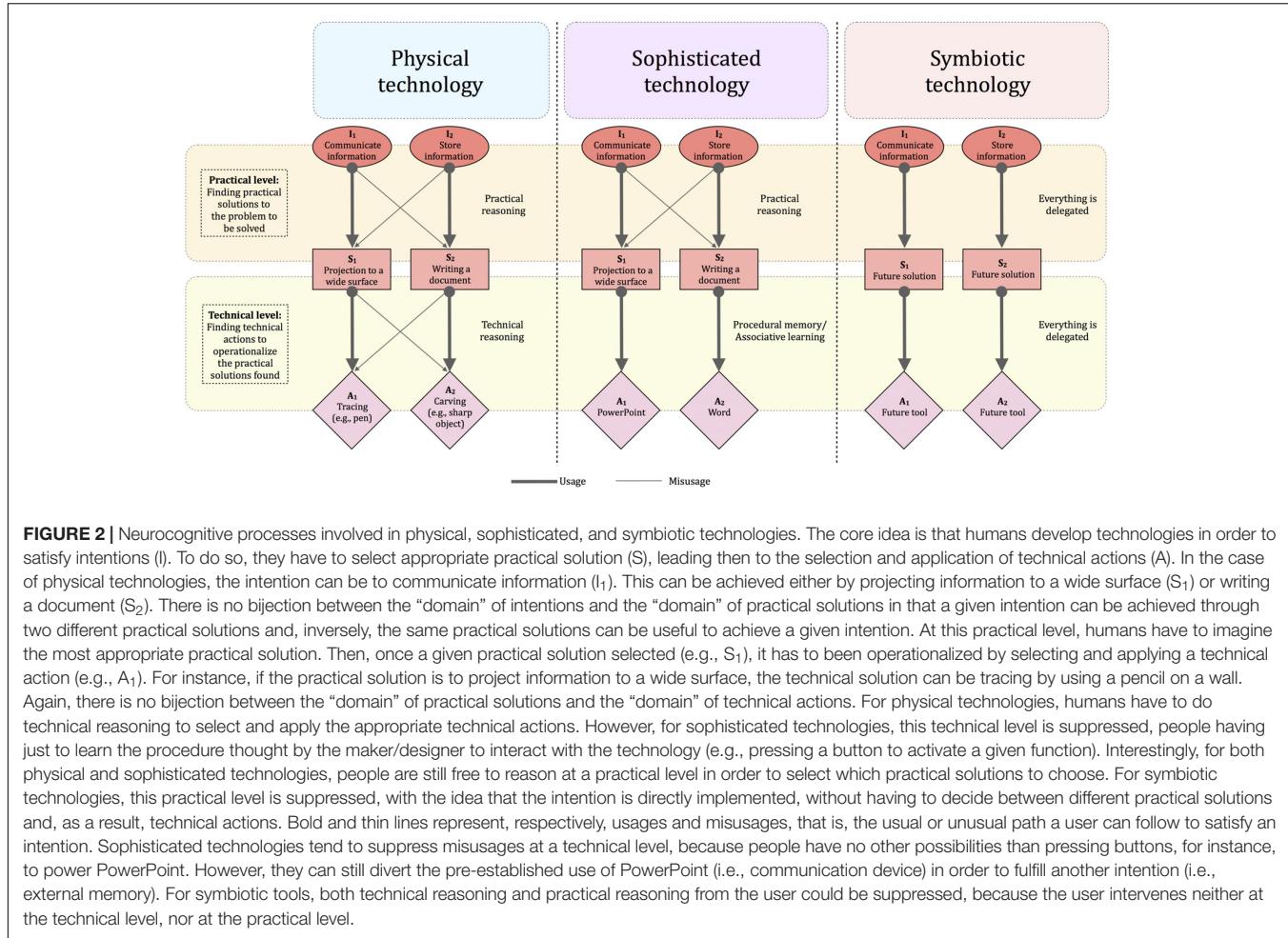
FIGURE 2 | Neurocognitive processes involved in physical, sophisticated, and symbiotic technologies. The core idea is that humans develop technologies in order to satisfy intentions (I). To do so, they have to select appropriate practical solution (S), leading then to the selection and application of technical actions (A). In the case of physical technologies, the intention can be to communicate information (I_1). This can be achieved either by projecting information to a wide surface (S_1) or writing a document (S_2). There is no bijection between the "domain" of intentions and the "domain" of practical solutions in that a given intention can be achieved through two different practical solutions and, inversely, the same practical solutions can be useful to achieve a given intention. At this practical level, humans have to imagine the most appropriate practical solution. Then, once a given practical solution selected (e.g., S_1), it has to been operationalized by selecting and applying a technical action (e.g., A_1). For instance, if the practical solution is to project information to a wide surface, the technical solution can be tracing by using a pencil on a wall. Again, there is no bijection between the "domain" of practical solutions and the "domain" of technical actions. For physical technologies, humans have to do technical reasoning to select and apply the appropriate technical actions. However, for sophisticated technologies, this technical level is suppressed, people having just to learn the procedure thought by the maker/designer to interact with the technology (e.g., pressing a button to activate a given function). Interestingly, for both physical and sophisticated technologies, people are still free to reason at a practical level in order to select which practical solutions to choose. For symbiotic technologies, this practical level is suppressed, with the idea that the intention is directly implemented, without having to decide between different practical solutions and, as a result, technical actions. Bold and thin lines represent, respectively, usages and misusages, that is, the usual or unusual path a user can follow to satisfy an intention. Sophisticated technologies tend to suppress misusages at a technical level, because people have no other possibilities than pressing buttons, for instance, to power PowerPoint. However, they can still divert the pre-established use of PowerPoint (i.e., communication device) in order to fulfill another intention (i.e., external memory). For symbiotic tools, both technical reasoning and practical reasoning from the user could be suppressed, because the user intervenes neither at the technical level, nor at the practical level.

itself in favor of a recall of where to access it Sparrow et al. (2011). Is it for the best or for the worst? This is not a new question, at least in the cognitive ergonomics field. Parasuraman and Riley (1997) stated that automation "*changes the nature of the work that humans do, often in ways unintended and unanticipated by the designers of automation*" (p. 231). Use is described here as the human proneness to activate automation when available. Besides a correct use of automation, misuse (i.e., overreliance on automation) and disuse (i.e., underutilization of automation) have been reported. Thus, the human is reasoning about its interactions with sophisticated tools to adjust his/her behavior according to the context and his/her own objectives (Leplat, 1990). For instance, automation use was found to be related to a balance between trust in automation and user self-confidence (Lee and Moray, 1994). These data can be interpreted as the human nature to keep reasoning based on internal and external assessments (i.e., practical reasoning). This is what we refer to as practical misusage, that is, the ability to divert the pre-established use of a tool (e.g., PowerPoint as a communication device) to fulfill another intention (e.g., storing information; **Figure 2**). A research issue to investigate is the neural bases that support this "practical reasoning." Are there (a) partly the same as those required by technical reasoning?

(b) Rather common to those associated to logical reasoning? Or (c) implying areas known to be engaged in interactions with other humans that would be recycled to reason on human–machine interactions?

Another aspect specific to sophisticated tools is that the perception or inference of tool functions could be sometimes complicated because of the distance between the maker and the user, favoring the occurrence of inappropriate and ineffective use. To counter this phenomenon, a human-centered design has been proposed (Billings, 1991). This design process widely used in a variety of domains (François et al., 2016) is based on the rationale that tool designers should take into account as much as possible users' logic and characteristics during the tool design process. In a way, the consideration of the user in the design process aims at reducing the distance between the maker and the user. Nevertheless, if we assume that humans are keen on practical reasoning, this quest is necessary deceptive as there is no universal reasoning process and, thus, neither universal human–tool interaction, nor natural interaction with sophisticated tools. Inversely, the human–tool interaction is rather artificial because based on an artifice (i.e., a sophisticated tool) for which the user ignores, at least part of, the design philosophy and the working principle.

THE FUTURE: SYMBIOTIC TOOLS

Kid #1: *"You mean you have to use your hands?"*
Kid #2: *"That's like a baby's toy!"*
—*Back to the Future Part II*

Predicting the future of our technology could be a fortune teller's job, had there not been a few mesmerizing anticipation movies and books, featuring great inventions feeding from contemporary science, the society's aspirations, and feeding back companies striving for developing them: inventions such as the Blade Runner flying autonomous cars or the gesture-based user interface from Minority Report prefigure the tools of the future. Some may never be created, some may be part of our everyday lives in 30 years, as the video calls from the first Blade Runner movie are part of our modern lives. This sneak peek into the future shows that all these tools have one thing in common: they seem to be operated seamlessly and conveniently by the user, reducing or abolishing four main constraints: mechanics, space, time, and effort (Osiurak, 2014). Although the depicted vision of our future world is always more technology-oriented, machines never overwhelm the user, who is becoming a part of a human–machine system, as the "commander-in-chief."

Most of the promised futuristic and fantastic tools are operated by thought, voice, or gestures. Because human–machine interaction through devices such as a mouse or keyboard is slow, inefficient, and sometimes not even feasible, the possibility of communicating with machines directly from our thoughts has emerged (Schalk, 2008). The brain–computer interface (BCI) (Wolpaw et al., 2002) field has then rapidly gained interest, first because it could be used in motor rehabilitation programs (Chaudhary et al., 2016), as the aim of BCI is to translate brain activity ("thoughts") into commands understandable by a machine. For achieving this, brain activity is captured by the means of sensors, pre-treated, and assigned to a corresponding action to be performed by the artificial system through an adaptive algorithm that learns to discriminate classes in the brain signals recorded (Mitchell, 1997; Bishop, 2006). A successful BCI interaction very often includes a learning phase attuning the technology to the specificity of the user's cognitive system. The structural inter-individual heterogeneity of the brains themselves, the functional differences, even the intra-individual differences from a time to another, will push the need for the learning algorithms to be highly adapted to a particular individual, if not to his particular mood.

Following this, the tantalizing promises of body-and-mind-operated tools, responding efficiently to the user's intentions, come with the need of individualizing the technology operating the machine. Brain–machine communication needs to be truly adapted to each specific individual for brain patterns to be successfully converted into thoughts. In this ultra-individualized technology, the individual and the tool will then form a system in a tight relationship, depending on each other to "perform" appropriately. The tool is then embodied within the user, and the system they form could be designated as a "symbiotic tool" (Licklider, 1960; Brangier and Hammes-Adélé, 2011). Within this tight interaction, the human has the intention, then the tool operates the technical and practical choices (i.e., suppression of the technical and practical levels; **Figure 2**).

On the journey to a Future in which Technology and Man form a symbiotic system, a few issues remain to be addressed. The first one is the acceptation issue (Davis, 1989). Are we designed to pair with synthetic devices? Can we and shall we accept to be part of a man–machine system? Tools of the Present need the user to accept them. We postulate that the future symbiotic tools will need the user to incorporate them. The second point is to explore the limits of the human cognitive system in terms of BCI performance. To function as smoothly and perfectly as in the Avatar movie for example, many technical issues have to be solved from the maker: the sensors need to be implanted, miniaturized; the algorithms need to be fast and reliable, etc. (Lebedev and Nicolelis, 2006). If the machine-related issues will without any doubt be resolved at some point, only few researches have tackled the man-related issue. Are the neural signals encoding our thoughts specific and reliable enough to be translated into a crystal-clear command? For how long can we maintain a neural state corresponding to a sustained command? Are we (all) designed to be good BCI-commanders, and always? Studies on BCI illiteracy show that 20% of the population cannot produce the brain patterns required for a BCI system to function properly (Vidaurre and Blankertz, 2010). Are their brains faulty, or the techniques immature?

These questions relate to the fundamental enigma of the cognitive system: how can our complex thoughts, dreams, feelings, creativity, instinct, etc. be encoded into less than 10^{15} signals? How can an infinite and unexplored mental world be created by a finite and defined material support? The birth of neuroergonomics (Hancock and Szalma, 2003; Parasuraman, 2003) will certainly help to start answering these issues, and to develop efficient channels of communication with technology.

CONCLUSION

In this review, we depict the different cognitive modes of interaction we have with physical, sophisticated and symbiotic tools. The key idea is that there could be a trend to progressively suppress our involvement at technical and practical levels (**Figure 2**). Interestingly, when considering symbiotic tools, users might be, a day, restricted to produce only intentions and will delegate all remaining efforts and choices to machines. The key issue is whether this restriction has to be viewed as a source of freedom or not? After all, should this scenario be true, what will humans do to occupy their available brain time? We are also aware that this review is biased by our ability to envision future tools, and how technology will evolve in a far future. Perhaps our conception of symbiotic tools is limited, considering only tools that transform our conscious intentions into responses. However, perhaps we will be able to develop technologies that will produce responses based on unconscious thoughts, thereby anticipating our needs even if we are unable to correctly generate them – or even before we generate them (e.g., sending an email to an editor before we intend to do so). In this respect, a critical question for

future research is to determine whether our technological cultural evolution will reach an asymptote as suggested here, or whether other forms of technological interactions will emerge in a far future, again shaping our cognition in return.

AUTHOR CONTRIBUTIONS

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

REFERENCES

Beck, S. R., Apperly, I. A., Chappell, J., Guthrie, C., and Cutting, N. (2011). Making tools isn't child's play. *Cognition* 119, 301–306. doi: 10.1016/j.cognition.2011.01.003

Billings, C. E. (1991). *Human-Centered Aircraft Automation: A Concept and Guideline*. NASA Technology Memorandum No. 103885, Moffett Field, CA: NASA-Ames Research Center.

Bishop, C. M. (2006). Pattern recognition and machine learning. *Pattern Recogn.* 4:738. doi: 10.1117/1.2819119

Boyd, R., and Richerson, P. (1985). *Culture and the Evolutionary Process*. Chicago, IL: University of Chicago Press.

Brangier, É., and Hammes-Adelé, S. (2011). "Beyond the technology acceptance model: elements to validate the human-technology symbiosis model," in *Ergonomics and Health Aspects, HCII 2011, LNCS 6779*, ed. M. M. Robertson (Berlin: Springer-Verlag), 13–21. doi: 10.1007/978-3-642-21716-6_2

Bril, B., Rein, R., Nonaka, T., Wenban-Smith, F., and Dietrich, G. (2010). The role of expertise in tool use: skill differences in functional action adaptations to task constraints. *J. Exp. Psychol. Hum. Percept. Perform.* 36, 825–839. doi: 10.1037/a0018171

Chaudhary, U., Birbaumer, N., and Ramos-murguialday, A. (2016). Brain – computer interfaces for communication and rehabilitation. *Nat. Rev.* 12, 513–525. doi: 10.1038/nrneurol.2016.113

Claidière, N., Smith, K., Kirby, S., and Fagot, J. (2014). Cultural evolution of systematically structured behaviour in a non-human primate. *Proc. R. Soc. Lond. B* 281:20141541. doi: 10.1098/rspb.2014.1541

Coeckelbergh, M. (2015). The tragedy of the master: automation, vulnerability, and distance. *Ethics Inf. Technol.* 17, 219–229. doi: 10.1098/rspb.2014.1541

Davis, F. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Q.* 13, 319–340. doi: 10.2307/249008

François, M., Osiurak, F., Fort, A., Crave, P., and Navarro, J. (2016). Automotive HMI design and participatory user involvement: review and perspectives. *Ergonomics* 139, 1–12. doi: 10.1080/00140139.2016.1188218

Goldenberg, G., and Hagmann, S. (1998). Tool use and mechanical problem solving in apraxia. *Neuropsychologia* 36, 581–589. doi: 10.1016/S0028-3932(97)00165-6

Goldenberg, G., and Spatt, J. (2009). The neural basis of tool use. *Brain* 132, 1645–1655. doi: 10.1093/brain/awp080

Hancock, P. A. (2007). What future for human-machine symbiosis? *Ergonomia* 29, 17–18. doi: 10.1088/1741-2552/aa9817

Hancock, P. A. (2014). Automation: how much is too much? *Ergonomics* 57, 449–454. doi: 10.1080/00140139.2013.816375

Hancock, P. A., and Szalma, J. L. (2003). The future of neuroergonomics. *Theor. Issues Ergon. Sci.* 4, 238–249. doi: 10.1080/1463922021000020927

Hovers, E. (2012). "Invention, reinvention, and innovation: the makings of Oldowan lithic technology," in *Origins of Human Innovation and Creativity*, ed. S. Elias (Amsterdam: Elsevier), 51–68. doi: 10.1016/B978-0-444-53821-5.00005-1

Jarry, C., Osiurak, F., Delafuys, D., Chauviré, V., Etcharry-Bouyx, F., and Le Gall, D. (2013). Apraxia of tool use: more evidence for the technical reasoning hypothesis. *Cortex* 49, 2322–2333. doi: 10.1016/j.cortex.2013.02.011

Lebedev, M. A., and Nicolelis, M. A. L. (2006). Brain-machine interfaces: past, present and future. *Trends Neurosci.* 29, 536–546. doi: 10.1016/j.tins.2006.07.004

Lee, J. D., and Moray, N. (1994). Trust, self-confidence, and operators' adaptation to automation. *Int. J. Hum. Comput. Stud.* 40, 153–184. doi: 10.1006/ijhc.1994.1007

Leplat, J. (1990). Relations between task and activity: elements for elaborating a framework for error analysis. *Ergonomics* 33, 1389–1402. doi: 10.1080/00140139008925340

Licklider, J. C. R. (1960). Man-computer symbiosis. *IRE Trans. Hum. Factors Electron. HFE* 1, 4–11. doi: 10.1109/THFE2.1960.4503259

Mitchell, T. M. (1997). Machine learning. *Annu. Rev. Comput. Sci.* 4:432. doi: 10.1145/242224.242229

Navarro, J., Mars, F., and Young, M. S. (2011). Lateral control assistance in car driving: classification, review and future prospects. *IET Intell. Transp. Syst.* 5, 207–220. doi: 10.1049/iet-its.2010.0087

Navarro, J., and Osiurak, F. (2015). When do use automatic tools rather than doing a task manually? Influence of automatic tool speed. *Am. J. Psychol.* 128, 77–88. doi: 10.5406/amerjpsyc.128.1.0077

Navarro, J., and Osiurak, F. (2017). The more intelligent people are, the more they use tools. *Psychol. Fr.* 62, 85–91. doi: 10.1016/j.psfr.2015.11.002

Orban, G. A., and Caruana, F. (2014). The neural basis of human tool use. *Front. Psychol.* 5:310. doi: 10.3389/fpsyg.2014.00310

Osiurak, F. (2014). What neuropsychology tells us about human tool use? The four constraints theory (4CT): mechanics, space, time, and effort. *Neuropsychol. Rev.* 24, 88–115. doi: 10.1007/s11065-014-9260-y

Osiurak, F. (2017). Cognitive paleoanthropology and technology: toward a parsimonious theory (PATH). *Rev. Gen. Psychol.* 21, 292–307. doi: 10.1037/gpr0000129

Osiurak, F., and Badets, A. (2016). Tool use and affordance: manipulation-based versus reasoning-based approaches. *Psychol. Rev.* 123, 534–568. doi: 10.1037/rev0000027

Osiurak, F., and Heinke, D. (2017). Looking for intoolligence: a unified framework for the cognitive study of human tool use and technology. *Am. Psychol.* doi: 10.1037/amp0000162 [Epub ahead of print].

Osiurak, F., Jarry, C., Allain, P., Aubin, G., Etcharry-Bouyx, F., Richard, I., et al. (2009). Unusual use of objects after unilateral brain damage: the technical reasoning model. *Cortex* 45, 769–783. doi: 10.1016/j.cortex.2008.06.013

Osiurak, F., Jarry, C., and Le Gall, D. (2010). Grasping the affordances, understanding the reasoning: toward a dialectical theory of human tool use. *Psychol. Rev.* 117, 517–540. doi: 10.1037/a0019004

Osiurak, F., and Rossetti, Y. (2017). Definition: limb apraxia. *Cortex* 93:228. doi: 10.1016/j.cortex.2017.03.010

Osiurak, F., Wagner, C., Djerbi, S., and Navarro, J. (2013). To do it or to let an automatic tool do it: the priority of control over effort. *Exp. Psychol.* 60, 453–468. doi: 10.1027/1618-3169/a000219

Parasuraman, R. (2003). Neuroergonomics: research and practice. *Theor. Issues Ergon. Sci.* 4, 5–20. doi: 10.1080/14639220210199753

Parasuraman, R., and Riley, V. (1997). Humans and automation: use, misuse, disuse, abuse. *Hum. Factors J. Hum. Factors Ergon. Soc.* 39, 230–253. doi: 10.1518/001872097778543886

Penn, D. C., Holyoak, K. J., and Povinelli, D. J. (2008). Darwin's mistake: explaining the discontinuity between human and nonhuman minds. *Behav. Brain Sci.* 31, 109–130. doi: 10.1017/S0140525X0800 3543

Reynaud, E., Lesourd, M., Navarro, J., and Osiurak, F. (2016). On the neurocognitive origins of human tool use. a critical review of neuroimaging data. *Neurosci. Biobehav. Rev.* 64, 421–437. doi: 10.1016/j.neubiorev.2016. 03.009

Schalk, G. (2008). Brain–computer symbiosis. *J. Neural Eng.* 5, 1–15. doi: 10.1088/ 1741-2560/5/1/P01

Sparrow, B., Liu, J., and Wegner, D. M. (2011). Google effects on memory: cognitive consequences of having information at our fingertips. *Science* 333, 776–778. doi: 10.1126/science.1207745

Vidaurre, C., and Blankertz, B. (2010). Towards a cure for BCI illiteracy. *Brain Topogr.* 23, 194–198. doi: 10.1007/s10548-009-0121-6

Virgo, J., Pillon, J., Navarro, J., Reynaud, E., and Osiurak, F. (2017). Are you sure you're faster when using a cognitive tool? *Am. J. Psychol.* 130, 493–503. doi: 10.5406/amerjpsyc.130.4.0493

Wolpaw, J., Birbaumer, N., McFarland, D., Pfurtscheller, G., and Vaughan, T. (2002). Brain-computer interfaces for communication and control. *Clin. Neurophysiol.* 113, 767–791. doi: 10.1016/S1388-2457(02)00057-3

Young, M. S., Stanton, N. A., and Harris, D. (2007). Driving automation: learning from aviation about design philosophies. *Int. J. Veh. Des.* 45, 323–338. doi: 10.1504/IJVD.2007.014908

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.

Check for
updates

Exploring Human-Tech Hybridity at the Intersection of Extended Cognition and Distributed Agency: A Focus on Self-Tracking Devices

Rikke Duus[1]*, Mike Cooray[2] and Nadine C. Page[2]

[1] School of Management, University College London, London, United Kingdom, [2] Hult International Business School, London, United Kingdom

In an increasingly technology-textured environment, smart, intelligent and responsive technology has moved onto the body of many individuals. Mobile phones, smart watches, and wearable activity trackers (WATs) are just some of the technologies that are guiding, nudging, monitoring, and reminding individuals in their day-to-day lives. These devices are designed to enhance and support their human users, however, there is a lack of attention to the unintended consequences, the technology non-neutrality and the darker sides of becoming human-tech hybrids. Using the extended mind theory (EMT) and agential intra-action, we aimed at exploring how human-tech hybrids gain collective skills and how these are put to use; how agency is expressed and how this affects the interactions; and what the darker sides are of being a human-tech hybrid. Using a qualitative method, we analyzed the experiences of using a WAT, with a specific focus on how the tracker and the individual solve tasks, share competences, develop new skills, and negotiate for agency and autonomy. We contributed with new insight on human-tech hybridity and presented a concept referred to as the agency pendulum, reflecting the dynamism of agency. Finally, we demonstrated how the EMT and agential intra-action as a combined theoretical lens can be used to explore human-tech hybridity.

Keywords: hybridity, extended mind, agential intra-action, wearable activity trackers, agency

INTRODUCTION

Throughout time, humans have utilized the capabilities and skills derived from interacting with external tools, entities, devices, and artifacts to complement their own cognitive abilities (Thacker, 2003; Dinerstein, 2006; Herbrechter, 2012; Heersmink, 2017). There are many ways that human cognition can be enhanced with external artifacts, some of which are rather mundane, including shopping-lists, books, diaries, recipes, calculators, spreadsheets, and, more recently, mobile phones. The extended mind theory (EMT) is a helpful theoretical apparatus when considering how and in which ways, cognitive processes can become extended across multiple human and nonhuman entities. The EMT reflects Clark and Chalmers (1998) argument that cognitive processes (e.g., memory, information retrieval, and processing) can take place outside of the human mind. Hence, cognitive abilities are a collection, an ensemble (Clark, 2015), of human and other external entities that together perform tasks and solve problems.

The EMT has previously been used as a theoretical lens to explore a number of different contexts, such as the musically extended mind (Krueger, 2014), spirituality and Christian life (Brown and Strawn, 2017), treatment of sex offenders (Ward, 2009), social ant behavior (Bosse et al., 2005), as well as other studies that explore cognitive integration and the extended mind (Menary, 2010), and sense-making (Thompson and Stapleton, 2009). However, studies like these that adopt the EMT are typically conceptual and do not seem to engage directly with research subjects to understand, in practice, how cognitive capabilities become distributed and contribute to the formation of hybrids. In the context of human-tech hybridity, there is a need for further research into the ways in which agency is acquired, expressed and lost as well as the darker sides of these hybrid formations.

We adopt the empirical context of people who use or have recently used wearable activity trackers (WATs) to manage their health and well-being and the trackers that collect, store, and reproduce the health and well-being data. We rely on their accounts of interacting with the WATs and are interested in their lived experiences. WATs are interesting to study in this context, as the technology has gained an on-body position, constantly capturing movements and activities with the potential to influence the person's behaviors, decision-making, and information access. The WATs undertake certain activities with few instructions from the user, e.g., automatically starts monitoring sleep and determines which kind of physical activity the person is performing, while in other situations the person inputs information (e.g., food items), which the WAT transforms into visualizations. We seek to contribute to other recent studies (e.g., Bode and Kristensen, 2016; Etkin, 2016; Fotopoulou and O'Riordan, 2016; Nelson et al., 2016; Rapp and Cena, 2016; Smith and Vonthethoff, 2016) that have used self-monitoring and self-tracking as the empirical context to investigate various areas of human-tech engagement.

We apply the EMT and Barad's (2003) concept of agential intra-action as a combined theoretical lens to explore how cognition is extended to the WATs to solve tasks and provide new insight, while also expressing agency (Clowes, 2018). In this way, the paper builds on the EMT by examining the non-neutrality of the WATs and how they acquire agency in particular situations, which is further conceptualized through our concept of the agency pendulum. As such, this research believes that technology can act as cognitive extensions and at the same time express agency. Hence, we put forward that distributed cognition and distributed agency can be detected when exploring human-tech interactions and that these are important to understand what it means to be a human-tech hybrid.

We posit that hybridity is not a stable condition with predetermined roles and affects. Rather, it is an ongoing process that interweaves the human biological and cognitive with the abilities of other entities. Clark (2007, p. 279) underlines how the human emerges as a "soft self," ready to adapt and be adapted by technological others:

The realization that we are soft selves, wide open to new forms of hybrid cognitive and physical being, should serve to remind us to choose our bio-technological unions very carefully, for in so doing we are choosing who and what we are.

Using the EMT and agential intra-action, we aimed at exploring how human-tech hybrids gain collective skills and how these are put to use; how agency is expressed and how this affects the interactions; and what the darker sides are of becoming a human-tech hybrid.

The literature review starts with a focus on the EMT to explain the core concepts of extended cognition, collective problem solving, coupled systems, and non-neutrality. This is followed by a review of Barad's (2003) agential intra-action concept.

LITERATURE REVIEW

Extended Mind

Situated approaches to understanding cognition, such as embodied, enactive, embedded, and extended, have come to challenge the traditional cognitivist paradigm (Heersmink, 2017). Situated cognition is a form of cognitive extension that can be expressed in a multitude of ways through engagement with a person's external environment. The situated cognition movement has developed primarily since the late 1970s and offers an alternative paradigm for exploring and conceptualizing the mind (Wilson and Clark, 2009). At its core, situated approaches consider human thought as *affected* by the external socio-technological environment (Hutchins, 2014). Hence, when the external environment changes, the individual's cognitive abilities are also impacted. Therefore, when taking a situated approach to cognition, the external socio-technological environment is an important source of influence on human thought.

Clark and Chalmers (1998) concept of extended mind is an attempt to question the locus of cognition as belonging intrinsically and only to the human mind and body. Instead, they advocate that the human mind and cognition is extended across larger systems of different kinds of entities. The focus of the extended mind is how internal and external resources operate together in "driving more-or-less intelligent thought and action" (Sutton et al., 2010, p. 525). Clark (2001, p. 134) explains:

We – more than any other creature on the planet – deploy non-biological elements (instruments, media, notations) to *complement* our basic biological modes of processing, creating extended cognitive systems whose computational and problem-solving profiles are quite different from those of the naked brain.

It is collective problem-solving, involving both a person's internal resources, e.g., the ability to recall items to purchase from the supermarket, combined with the resources afforded by an external entity, e.g., a shopping list with items to purchase. Another example is the internal ability a person may have to find their way around the streets of London to reach a particular destination. This ability is based on the person's prior knowledge and experience of the network of streets, shortcuts and traffic patterns. The task of getting from A

to B, however, is often complemented by the directions and visualizations offered by an external resource; a digital map on the person's smartphone, for example. Hence, to complement the human mind's limited capacity (Norman, 1993), artifacts are created and used as scaffolding to help the person perform certain tasks. Artifacts that offer cognitive scaffolding thereby complement the human information processing capacity by providing information, resources, or capabilities, as and when required, in order to perform the task (Clark, 2015). The human-artifact hybrids gain new capacities (Wilson and Clark, 2009), which in turn affect behaviors, decision-making, and identity formation. Rowlands (2009) also believes that humans use the world around them to extract relevant information which is used to support basic functions such as perception, memory and reasons. These cognitive processes, he believes, take on a hybrid form as they "straddle both internal and external operations" (Rowlands, 2009, p. 2). This need to externalize thought in order to enhance processing capabilities through the use of complementary technologies is an inherent feature of the human experience.

Clark (2015) explains that devices, such as, for example, laptops and smartphones, can be considered bio-external devices that offer resources (e.g., information) for specific tasks, depending on the context and the level of uncertainty. Hence, the context that the human-artifact hybrid is in has an impact on the types of resources needed and the ability of each agent (internal and external) to provide and share the necessary resources. The person and the scaffolding can be so strongly coupled that they become one single cognitive system. Heersmink (2017) explains that the more a person depends on the external information to perform tasks that require cognitive abilities, the deeper the external information, or artifact that provides the information, is integrated with their internal cognitive system. In this way, it is a dynamic relationship and the extent to which the person and the artifact become a single cognitive system, is contingent on factors such as the intensity of information flow, accessibility of the scaffolding artifact, durability of coupling, amount of trust in the scaffold's information, among other (Heersmink, 2015).

To sum up, the concept of extended mind argues for an approach to cognition that is distributed and extended across human and other entities. The surrounding environment is seen as always affecting human thought, memory, decision-making, and actions. Hence, individuals undertake tasks in collaboration with artifacts and can even become a single cognitive system. As such, the EMT considers objects, people, systems, and other external components as part of a larger cognitive system.

In the Section "Coupled Systems," we explore the nature of the human-artifact coupling in further detail through Clark and Chalmers (1998) concept of coupled systems.

Coupled Systems

The concept of coupled systems is the linking of the human organism with an external entity in a two-way interaction, which creates a new cognitive system (Clark and Chalmers, 1998). These human and external entities interact in one system where each plays an active role and acquires collective behavioral

competences. If external parts are de-coupled from the system, the collective competences are reduced or even lost. The external parts, or features, that are embedded in the coupled system have the ability to act and influence the overall system. Clark and Chalmers (1998, p. 51) argue that the external features possess an "ineliminable role," as, if changed, the behavior of the person is likely to change too, even if the internal structure (e.g., the capacity to recall information, plan behavior, etc.) remains the same.

Coupling is one of the more contentious areas of the EMT and is also referred to as the coupling-constitution fallacy (Rowlands, 2009). According to Aizawa (2010), Clark's (2008) argument that a causal dependency between a cognitive process (A) and some other process (B) can make B or A-B constitute a cognitive process is flawed. Clark (2008) has defended this core pillar of EMT by arguing that all couplings are not automatically considered to constitute an extended cognitive process. Rather, focus should be on the *effect* of the coupling and its ability to surface information that is useful within a specific situation of problem-solving. This perspective is further emphasized in Clark and Chalmers (1998) concept of active externalism. Active externalism is grounded in the belief that the external features (e.g., a book, watch, to-do-list, fitness tracker) directly impact the person and the person's behaviors. In this way, the external features play an active role in the creation of the here-and-now and the capabilities of the human-artifact hybrid (Clark and Chalmers, 1998). Often, external features will be taken in use in order to enhance cognitive hybridization by acquiring the ability to process large amounts of information, faster and with a greater level of accuracy (Heersmink, 2017).

There are certain criteria that affect the strength of the human-artifact coupling (Clark and Chalmers, 1998; Heersmink, 2015) and therefore also the extent to which the external features become constitutive of a cognitive process. The coupling needs to be reliable, which means that the external feature or resource needs to be accessible as and when it is required. For example, a shopping list needs to be available when the person needs it in order to solve the particular task, for it to become part of the cognitive resources that the person has available to draw on (Clark and Chalmers, 1998). To create this coupling, it requires a high level of portability, and more importantly, accessibility, to ensure that the coupling is reliable. Clark and Chalmers (1998) argue that occasional decoupling, damage, loss, or malfunction does not put into question this unified cognitive system. They point to how a person's internal cognitive capabilities may also be challenged at times (e.g., from a lack of sleep, illness, or intoxication) and as along as the external features are available when required, then that constitutes a coupling. It is further important that the information that flows from the external source is trusted by the person receiving it. If the information is not trusted, its role as a guide for action will be challenged. Heersmink (2015) adds further detail to the notion of trust. He argues that (dis)trust can be either explicit or implicit. The main difference between explicit and implicit trust is that for explicit trust, the information is consciously evaluated before determining whether it is trustworthy or not.

For implicit trust, the information is assumed either trustworthy or not trustworthy. We tend to trust information implicitly if we have endorsed it in the past, if many people rely on this information to guide their action, or if it is relevant to achieving set goals (Arango-Muñoz, 2013). Tripathi (2010) explains that the more we depend on technologies to carry out or mediate our everyday activities, the more we will need to trust them to do so.

To sum up, a person and an external entity can become a new cognitive system through two-way interaction. However, the external entity needs to possess a high degree of trust, reliance, and accessibility and it must have been endorsed by the person at some point in the past in order to become part of this new system (Clark and Chalmers, 1998).

As we continue to explore facets of EMT as a perspective to understand human-WAT hybridity, it is of great relevance to consider the role the technology plays in shaping intentions and effects. We provide a brief review of the literature on the non-neutrality of technology, which leads us to further explore the agentic expressions of technological entities.

Non-neutrality

The concept of non-neutrality should be seen in the context of human-tech hybridization and as a contrasting perspective to the views that technology is always enhancing (i.e., positive) or that hybridization is neutral (i.e., means to an end). Verbeek (2006) explains that technologies always mediate and shape action. Ihde (1990) and Heersmink (2017) support this argument that human-tech hybridization is not neutral, as technologies do shape intentions and effects. Ihde (2004, p. 120) explains that "To take instruments either for granted or as simply transparent, is to make an implicit assumption that instruments are 'neutral'." Heersmink (2017) identifies three ways to understand the non-neutrality of technology. First, technologies can embody moral and political values. For example, a non-smoking sign encourages smokers not to smoke and instead adopt a behavior that is typically supported by health authorities. Second, technologies can mediate and transform experiences and perspectives on the world. The experience of being-in-the-world is affected by the technologies an individual interacts with, whether that is a motorbike, a heart monitor, or a computer game. The impact a technology has can be detected by paying attention to which aspects of an experience that it amplifies and which it reduces (Tripathi, 2010). For example, the use of whiteboards to mindmap ideas and thoughts amplifies the ability to organize, inter-relate, and prioritize information and, potentially, share it with others. Third, technologies can be seen as having unintended consequences, which are difficult to affect or change. For example, social networking platforms were designed to bring people together, but, unintentionally, have also contributed to issues of social anxiety and isolation for some users. Other unintended consequences can be seen with something as mundane as word processing software and the in-built spell checking feature. Users of this software, may experience their spelling deteriorate due to their misspelling automatically being corrected, sometimes so

quickly that the person may not even notice the correction being made.

As technology becomes more ubiquitous and present in day to day tasks and interactions, there is a risk of over-reliance on the external information that the technologies provide (Carr, 2011). This may lead to a reduction in the cognitive abilities, as tasks, information storing, and problem-solving are outsourced to technologies, and therefore not learned or practiced by the individual.

The non-neutrality of technology underlines the idea that technological entities shape intentions and effects, some of which can be unintended consequences or side effects, which are unexpected. In the final section of this review of the literature, we continue to explore the role of technology in the making of human-tech hybridity.

Agential Intra-action

Like the EMT acts to challenge the traditional cognitivist paradigm, Barad's (2003) work on agential intra-action also takes an oppositional stance. It encourages a re-think and re-view of our relations with the external world by de-centering the dominant human actor and instead focusing on the intertwined nature of humans and other entities (Pickering, 2013).

Agential intra-action considers agency to be emergent and distributed over human and nonhuman forms. Hence, from this perspective, agency is not deterministic, absolute or, indeed, only a human practice. As such, agential intra-action also challenges the dominant human subject (Puig de la Bellacasa, 2009). Nonhuman entities, such as technology, are also seen to express agency, affect, and influence relationships as well as the collective practices, capabilities, and cognitive abilities of human-tech hybrids.

Barad's (2003) focus is on understanding the co-created and co-constructed behaviors, decisions, and experiences, which take place within these hybrid relationships. She explains that "Agency is not an attribute but the ongoing reconfigurings of the world" (Barad, 2003, p. 818). In other words, entities, whether human or other, are seen to express agency (e.g., the ability to influence a situation) as they interact with other entities in the world. This is what Barad (2003, p. 817) refers to as the "ebb and flow of agency." Hence, situations, relations, and identities are ongoing and evolve between actors, who through those intra-actions become and act. To understand how different human and nonhuman entities *become*, it is important to account for both human and nonhuman forms of agency (Barad, 2003). It is also important to emphasize that, in this view, agentic expressions by nonhuman entities are not purely extensions or transferals of human agency. Pickering (2013, p. 25) explains that it is useful *not* to think about agency in terms of "will, intention, calculation, and representation," but rather in terms of performance, doings, actions, impact, and influence. In other words, agency is expressed in places of action, of consequence, of impact – in places where change occurs. Pickering (1995, p. 102) refers to this as the "open-ended dance of agency." Hence, that it is through intra-actions that agency is enacted and that agentic expressions flow back and forth between actors. As such,

human-tech (and other) relationships are always dynamic, always unfolding.

In summary, agential intra-action provides an opportunity to acknowledge nonhuman agentic expressions, doings and actions as the human and technology interact. Adopting this lens in combination with the EMT enables an investigation of how agency becomes distributed across the users and the WATs in situations of extended cognition.

MATERIALS AND METHODS

A qualitative approach to inquiry was adopted to collect and critically evaluate a multitude of perspectives from individuals who have a range of experiences with WATs. The aim of the study was not to provide generalizations across all users of WATs, but rather to explore in-depth the subjective, lived experiences, and human-WAT relationships, which each participant in this study is involved in co-creating. We intend to empirically identify and evaluate a range of experiences, actions, emotive responses, and skills attainment, which can help to illuminate what it is like to be a human-WAT hybrid. This insight, consequently, demonstrates the usefulness of adopting the EMT and agential intra-action as a set of combined theoretical perspectives and drivers to investigate people's relationships with technology. This is a timely inquiry due to technology's fast advancement, interactive nature, and ubiquitous involvement in a multitude of daily life experiences and decisions.

Research Participants and Sampling

The purpose of the empirical data collection and analysis was to understand how human-tracker hybrids gain collective skills and how these are put to use; how agency is expressed and how it affects the interaction; and whether these human-tracker relationships have darker sides. In order to capture insight from users of WATs in relation to these research areas, the research team used stratified purposive sampling (Ritchie et al., 2003) to select eight female participants, living and working in the United Kingdom.

We adopted this sampling approach to ensure that our participants met specific criteria and would be able to contribute with new insight in relation to the main objectives of the study, while also being individually comparable (Bryman and Bell, 2011). This study is part of a larger research inquiry currently focused on women's experiences of using digital devices, such as WATs, to manage health and well-being. Therefore, only female users of WATs were considered for this study, although in future studies this is likely to extend to also include male users.

Beyond gender, prior experience of using and interacting with a WAT to manage own health and well-being was a primary factor for sample selection. All participants had used a WAT for at least 6 months. In line with the stratified purposive sampling approach, we wanted to ensure some diversity of the sample and to take an inclusive approach (Ritchie et al., 2003). Hence, we further sampled according to specific usage levels. To capture a wider range of experiences, it was important not only to include individuals who were highly engaged with their WAT, but also

TABLE 1 | Sampling of participants.

Name	Interest in monitoring day-to-day activities	Level of intensity
Joanna	0	0
Christine	0	0
Mary	1	1
Jane	1	1
Sofia	2	2
Catherine	1	2
Paula	2	2
Maria	1	2

those who had a lower engagement level and those who had become non-users after having used a WAT in the recent past. In terms of defining engagement, we identified this as a combination of two behavioral attributes. First, the extent of interest in monitoring day-to-day activities (e.g., the frequency of checking performance analytics), and second, the level of intensity of the relationship with the WAT in terms of the amount of activities logged. **Table 1** provides an overview of the participant sampling. A "0" reflects current non-engagement (due to no longer using the WAT), while a "2" reflects the highest level of engagement.

There were many other sampling criteria which could have been adopted to select a sample, for example, the reasons for initiating the use of the WAT, types of job/job function (e.g., sedentary versus active), type of WAT (e.g., brand, functional features, position on body), and life stage (e.g., single, couple, family). We chose not to limit the empirical scope of our inquiry beyond engagement level, as we were, primarily, interested in the participants' day-to-day interactions with the WAT, the collective skills acquisition and the, potential, darker sides of the relationship.

We used two of the authors' professional networks to identify and obtain access to participants who fulfilled the sampling criteria. This proved a key strength of our study due to the prior familiarity between the interviewer and interviewee, which allowed us to capture personal insight, stories, and experiences, which the participants felt comfortable sharing (Easterby-Smith et al., 2012).

Table 2 details the participants taking part in the research, including their age group, profession, lifestyle characteristics, activities tracked, and engagement level. To ensure ethical integrity and our participants' anonymity, each individual was given a pseudonym (Ogden, 2008) and details shared, which could enable others to identify them, were also removed or not used explicitly in the study.

Data Collection and Analysis

Participant interviews took place over a period of four months. The interviews were guided by a semi-structured checklist of areas, which were informed by our central research questions. We themed the questions into six categories: About the research participant, About the WAT, Usage, Behavior, Relationships, and Drawbacks/Downsides. Hence, we were interested, not only in the positive aspects of participants' interactions with the WAT, but also those that were perceived to be of a more

TABLE 2 | Participant profiles.

Name	Age	Profession	Lifestyle	Type of tracker	Tracking activities	Engagement level
Joanna	25–30	Business development	Very active and attends the gym several times a week. Highly competitive and driven. Often participated in challenges via the activity tracker. Felt addicted to the tracker. Enjoys team sports. Family orientated.	Fitbit Charge Monitors: Heart rate, Automatic sleep tracking, Steps, Activities, Food and water logging, Distance, Stairs climbed, Caller ID	Used to track: Activities Steps	Used to be a Hyper-engaged user. The WAT needed resetting which has not been done. Has decided not to continue using it.
Christine	25–30	Business development	Very active and attends the gym several times a week. Highly competitive and driven. Often participated in challenges via the activity tracker. Was motivated to challenge others. Felt addicted to the tracker.	Fitbit Charge Monitors: Heart rate, Automatic sleep tracking, Steps, Activities, Food and water logging, Distance, Stairs climbed, Caller ID	Used to track: Steps Sleep Activities	Used to be a Hyper-engaged user. Currently, the WAT is broken and not in use.
Mary	30–35	Human Resources	Very active and attends the gym several times a week. Wants to stay in control and has reduced the number of activities tracked. Wants to be healthy rather than calorie-focused. Does not use the activity tracker app. Accesses data via the tracker itself.	Fitbit Alta Monitors: Automatic sleep tracking, Steps, Smart tracking of activities, Food and water logging, Distance, Move reminders	Steps	Engaged
Jane	55–60	Operations management	Recovering from surgery. Important to do a good number of steps each day. Walking and jazz dance help her to recover and move away from being incapacitated post-surgery.	Fitbit Flex Monitors: Manual sleep tracking, Steps, Activities, Food and water logging, Distance	Steps Water intake Activities	Engaged
Sofia	35–40	Training and development	Highly competitive and enjoys the outdoors. Equestrian at a competitive level. Attends the gym regularly. Walks the dog. More focus on physical activity than gym-based exercise. Always exceeding the targets set.	Jawbone UP Monitors: Heart rate, Automatic sleep tracking, Steps, Activities, Calories, Water intake, Distance	Activities Sleep Steps	Hyper-engaged
Catherine	50–55	Operations management	Very active. Family focused and community orientated. Attends the gym regularly. Is highly aware and interested in fitness and health.	Fitbit Charge Monitors: Heart rate, Automatic sleep tracking, Steps, Activities, Calories, Water intake, Distance, Stairs climbed, Caller ID	Heart rate Steps Sleep Competes against others in work	Engaged
Paula	35–40	Business development	Very active. Recently had knee surgery. Attends team sport. Attends the gym regularly and enjoys weight training. Partner is a personal trainer. Evaluates performance on a weekly basis.	Fitbit Charge Monitors: Heart rate, Automatic sleep tracking, Steps, Activities, Calories, Water intake, Distance, Stairs climbed, Caller ID	Training purposes, specific exercises rather than daily use, heart rate for intensity training Monitors almost everything	Hyper-engaged
Maria	30–35	Program coordinator	Active lifestyle. Often busy on the weekend. Important to always wear the tracker. Highly reliant on tracking sleep. Needs the data to verify daily activity.	Fitbit Charge Monitors: Heart rate, Automatic sleep tracking, Steps, Activities, Calories, Water intake, Distance, Stairs climbed, Caller ID	Sleep Heart rate Steps Activities	Engaged

negative or challenging nature. Specific questions directed at all participants included whether the WAT had made them feel guilty, underperforming and regretful of looking at the analytics.

The semi-structured interview template was used to ensure that similar areas were explored for each participant for comparability and depth of insight (Irvine et al., 2013). It was important to create a dialog and conversation with each participant, enabling them to speak openly about the experiences that reflect their tracker relationship. Therefore, it was important to create a good rapport with each participant by showing interest in their stories, using listening techniques and asking for clarifications and examples at appropriate times during the interviews (King and Horrocks, 2010). To encourage this conversational and open interview format, we, at times, allowed for participants to influence the direction of the interviews (Stern et al., 1998) and describe situations, experiences, and feelings which came to mind (Thompson et al., 1989). In this way, not all interviews were conducted using the same ordering of questions from the interview template, although all areas were covered within each interview.

The interviews were audio recorded and thereafter transcribed verbatim. Each interview lasted between 45 and 60 min. We conducted a thematic analysis through an inductive process (Corbin and Strauss, 2008). The first stage of the data analysis involved a broad coding of themes that was not confined to specific assumptions or directions to allow for a multitude of interpretations and themes. This was a process initiated by the author unfamiliar with the research participants, who shared the initial themes with the research team. Thereafter the other authors read the transcripts, took notes, and identified initial themes. This was followed by several collaborative research meetings where each member of the research team presented their findings and rationales. From this iterative and collaborative process (Spiggle, 1994), the core themes were refined and agreed on. At this stage, we were highly alert to the themes' relevance to the central research questions. We focused on those themes that best reflect our participants' experiences of their interactions with and usage of the WAT and its cognitive resources and how agency becomes distributed across the participants and the WATs.

The validity of the research process can be assessed by the ability of the research team to capture the experiences, actions, emotive responses, and collective skills of the interview participants (Easterby-Smith et al., 2012) and the extent to which the research and analysis methods were effective in addressing the central research questions. The sampling of individuals with varying WAT engagement levels exposed us to a greater breadth of insight, while still focusing on the same core research questions. As all members of the research team were actively involved in the coding, analysis, and theming of the data along with several collaborative research meetings, this enhanced the credibility and validity of the research findings, while also minimizing bias (Denzin, 1989). The main limitation of this approach maybe the reliance on participant self-reporting. In the interviews, participants were required to share their experiences, feelings, and behaviors related to their interactions with the WAT,

rather than, for example, presenting the actual WAT engagement reports. It was important for this study to go beyond and behind performance dashboards to explore and understand participants' subjective experiences and feelings about their interactions with the WAT.

From the iterative and inductive process and guided by our research questions, we identified three themes. With the presentation of these themes, we illuminate how our research participants gain new skills and capacities, ways they become empowered by the WAT and also the darker sides of becoming a human-WAT hybrid.

RESULTS

This empirical investigation has multiple inter-linked purposes. First, we wanted to explore how our participants and the WATs interacted with a particular focus on the role of the WATs in helping the participants to undertake tasks, solve problems, and gain insight. Second, we focused on the distribution of agency within the human-WAT hybrids to illuminate how the varied expressions of agency affect the human-WAT coupling and thereby also the degree of extended cognition. We relied on the human participants' experiences of these events, situations, and daily practices.

Collective and Extended Skills

From the participants' accounts of their experiences, the WATs, as external resources, have played an important educational role in their lives. They have contributed with new insight about their activity levels, calorie burn/in-take, sleep patterns, and other body metrics. The main interfaces for this sharing of knowledge were the WATs and the WATs' mobile and desktop software applications. The performance visualizations acted as a gateway to systemized and categorized records of performance that were automatically logged by the WATs (e.g., sleep data) and some which were manually kept updated by participants (e.g., food/calorie/water intake). Initially, participants had found the WATs' abilities to collate and visually represent the biometric data intriguing and exciting, as they felt they were given access to an "X-ray" of themselves. Catherine, who had been using her WAT for a few months, explained:

> I think it is quite interesting. I watch the activity tracker all the time. How many steps I have taken during the day. I think it gets you to be more active. If I have been racing around, I like to also keep an eye on my heart rate too, just keep an eye on it. And the steps, I am also looking at the steps. I try to do at least 15,000 a day (Catherine).

This was a typical reaction among the participants, who believed it was helpful to use the WATs to keep taps on themselves and track their performance throughout the day. As the WATs were set to track the participants, they were able to provide real-time feedback that gave participants instant insight into how their body was reacting to certain activities (e.g., by tracking heart rate) and their progress toward meeting set goals (e.g., by counting steps). Several participants believed that using the

WAT to capture, store, and visually present the various biometric data had affected how they went about making certain decisions related to their health management, including eating habits, sleep patterns, and activity levels.

The automation of biometric data collection was mentioned as the primary role and responsibility of the WATs. Participants expected the WATs to have these abilities in order to help them fill in the "blank spots" of knowledge related to their health management. It transpired that all participants had the perception that health-related decisions that were informed by data were better decisions than those simply informed by their own opinions, feelings, and experiences. This belief was a major reason for using and interacting with the WATs. It was clearly the participants' view that the WATs contributed with a new set of competencies, which assisted them to better understand and assess calories in food items, create and manage sleep routines, and estimate step counts.

The WATs not only contributed to skills related to capturing, storing, and visualizing performance data, but were also expected to provide certainty and reassurance. For Sofia, the WAT was used specifically as an external resource to evidence to herself and others that she not only meets, but exceeds her daily activity targets and therefore has the right to feel tired in the evening. The WAT provided the data and visualizations, which complemented her own internal resources (e.g., ability to remember and explain to others about her daily routines) to feel reassured and to reassure others of her high levels of activity. The process of creating evidence and reassurance was further enabled as the WAT acted as a diary, containing all past performances and goals exceeded. This is data and insight that would have been challenging for Sofia to memorize accurately without the support of some sort of cognitive scaffolding.

> I keep getting grief because people say I am tired. But when you walk the dogs before work and you look after two horses after work, generally on a good day, I am not normally home until half 8, 9 o'clock. I was tired. So, it is quite interesting to know how much I am doing each day. I love the sleep data! Telling me how much sleep I have had. I love what it tells me! To have an idea about how much I am doing each day. I normally exceed my targets by 140–150%. So, it is also quite interesting to know that when you say you have had a busy day. I love data like that (Sofia).

Participants tended to trust the data that the WATs collected and presented to them. This trust was an important aspect of the interaction between the participants and their WATs and affected their willingness to respond to and interact with the WATs. By not questioning the data, the WATs were relied on for their information, analytics, and input as external resources and seen as important components toward achieving a higher level of daily activity.

It was evident that the WATs were active in nudging and prompting participants to either adopt or avoid a particular behavior or decision. For some participants, the notifications and performance updates provided by the WATs caused them to make time for a walk during their lunch break, review

current decisions on types and amounts of food intake, and systemize movement throughout the day by using alerts. Mary, for example, enabled her WAT to nudge her once an hour to get up and take 250 steps. She felt too sedentary in her job and needed the WAT to remind her to be systematically active. Similarly, Jane had become more conscious of her activity levels after she started using the WAT. She used the WAT to gain a status report of her actively levels at lunch time and if the performance was low, she would make purposive efforts in the afternoon to walk and meet her step targets. In this way, the participants' ways of thinking and evaluating different actions and options related to their health and well-being, were affected by the WATs ability to track, monitor, store, and present their performance data in real-time.

The participants had acquired their WAT for different purposes. Some sought its support as encouragement to achieve a heightened level of exercise, some looked to the WAT to simply document an already active lifestyle and others wanted the WAT specifically to guide them toward a weight loss. Some wore it day and night to track many different activities, while others wore it mainly at the gym or when completing specific types of exercise (e.g., running). Therefore, some participants set their WAT to track and monitor many different kinds of activities, while others were more selective of when they wanted the WAT to monitor and track their performance.

Despite the different reasons for having acquired the WAT, all participants looked to the WAT as an external source of certainty. The data and visualizations of performance complemented participants' internal knowledge about healthy eating habits and exercise. It acted to remove some uncertainty by over-ruling participants' own subjective gut feelings and estimates and provided a perception of an objective truth. The participants' trust in the WATs' data capture and visualizations was heightened by their own inability to gather and process this kind of information with a similar degree of accuracy and speed. For Maria, she used the WAT's sleep tracking feature to monitor the quality of her sleep. She trusted the WAT's ability to provide a truthful representation of whether she had had a good or a poor night's sleep and did not question it.

> I track sleep because then I can justify why I am tired. I can see that I have had a bad night's sleep. I didn't buy the tracker so that I could track my sleep. It is more of a side benefit. But I do like to track sleep to feel justified why I am tired. I feel better when the data tells me that I have had a bad night's sleep. I don't like it when I feel like I have had a very bad night's sleep and my tracker tells me that I have had a good night's sleep. That annoys me! Then I don't have an excuse to feel tired. And I believe it. I trust it. So, I have no excuse to feel tired. I just have to get on with it. I know it's odd! (Maria)

In addition, the overarching purpose for acquiring a WAT was to complement participants' own abilities to estimate and assess their level of activity, calorie intake, and sleep patterns. They acknowledged that these assessments were often inaccurate, faulty, or simply not possible to undertake and keep track of.

This expertise was instead expected of the WATs. For some participants, the WATs helped them to establish new activity routines and give greater insight into calorie consumption, step counts of certain routes and how to achieve better sleep pattern results.

Human Empowerment Through Technology Extension

Participants expressed how the data and visualizations generated by the WATs made them feel empowered. It was seen as a source of simplification of choices and decisions due to the transparency of performance and progress the WATs provided as and when requested by the participants. The WATs and the related mobile phone application kept participants updated on progress throughout the day and this was experienced to increase their confidence in decision-making.

For Sofia, who already had an active lifestyle before she started using the WAT, the role of the WAT was not to encourage her to be more active; rather the opposite. The role of the WAT was to help her manage her need to live up to societal expectations of how much one ought to weigh, exercise, eat, and so on. The analyses created by the WAT became a "pressure releaser" by alerting her of when she had met the set targets and providing the quantitative evidence.

It tells me "It's OK, you're doing enough." You know you read so much, watch so much about how much exercise you are meant to be doing, what the average is, from a fitness perspective, from a weight perspective and that kind of stuff. And I find myself thinking, I don't have much more time in a day. Do I go to the gym? No, I don't have time in a day, unless I maybe take 20 min at lunch to do that. So it is making me go "You do enough Sof." Instead of telling me that I should be doing more, mine is about taking pressure off, instead of telling me that I need to go for a run (Sofia).

This example demonstrates how the WATs can reduce pressure and stress that participants put on themselves to be active. In fact, it can help participants to make decisions *not* to exercise by confirming that their level of activity is already high and they are meeting the set goals. Here, the WATs contribute with objectivity and analytical evidence to support participants' otherwise subjective assessments.

Participants tended to either become more reliant on the WAT in order to make health, fitness, and food related decisions, or grow in confidence to make their own decisions with little input or guidance from the WAT. The more dependent participants needed the data to ascertain their own performance and used it as encouragement to continue. Other participants reported seeing their confidence grow as a result of learning from the tracker and acquiring their own, internalized capacity to manage their health and well-being. This indicates the presence of a knowledge transfer, where some participants learn from the WATs' ability to measure, monitor and track. Participants had gained new knowledge of the length of a particular walking route, the speed at which they walked a mile, the calorie amounts of different foods, and the calories burnt from different types of exercise.

This learnt knowledge had built confidence in them to make more independent decisions related to measuring, monitoring, and evaluating their activity levels, eating habits, and sleep. These were abilities that previously were possessed mainly by the WATs. Joanna explained how she is no longer using her WAT, but that it has helped her to establish an active routine, which she has been able to continue with despite not using the WAT.

Beyond providing the performance data, participants particularly enjoyed receiving the WATs' buzzing vibrations when a goal had been achieved. This provided a tactile interaction between the WATs and the participants that was effective in eliciting a positive emotion in participants, who felt proud and happy to have reached their goals. These vibrations created a physical connection between the WATs and the participants who were able to *feel* the WATs and were alerted to their communication. When the WATs were "silent" (e.g., not vibrating or making a sound), they also became invisible as participants would forget that they were on their wrists. They, however, remained accessible to give updates on performance. Most often, participants would turn to the WATs' mobile phone applications to gain more in-depth performance updates, whereas the WATs on their wrists provided a quick snapshot of progress. The buzzing vibrations and other indicators of reached goals (e.g., flashing lights, sounds) were ways that the WATs directly interacted with the participants and influenced their actions, behaviors, and decisions in real-time. Participants anticipated these interactions with the WATs and were open to be influenced; whether that was to be reassured (e.g., performance is on track) or encouraged (e.g., to heighten activity levels). They trusted the guidance provided by the WATs and used it in real-time to make decisions.

Some participants even wanted the WAT to take on a more proactive and influential role. Paula explained:

I probably want my tracker to be a scary, sort of, army person. In my ideal world, it would be someone who would be like "Come on, get your act together, let's get to the gym." That is the kind of motivation that I like and the kind of motivation that I need as well. I like to be told what to do and "Come on, try a bit harder" (Paula).

The WATs not only used vibrations and sounds to affect participants' decision-making, but also used the color green to induce behavior change. The green color was used in the WATs' mobile applications when goals were met. Participants explained how seeing this color, and knowing this was a sign of success, prompted them to feel happy, self-fulfilled, and positive. The green color signaled goal completion and became synonymous with accomplishment and encouragement.

It is a very clever dashboard in the sense that if you have achieved your goals it is in green. I don't know who has chosen green, but green does make you feel happy, it's like "Yeah go, you have achieved what you needed!." If you have only achieved 75% of your target, then it is amber orange. So, it is like a traffic system almost (Paula).

The WATs also used other ways to communicate with and affect participants such as smileys, badges, and trophies, which they received from completing their goals and taking part in competitions.

The WATs also contributed with more socially enabled capacities to encourage participants to be active. Such capacities included step competitions, which allowed participants to compete against other WAT users. The competitions were mainly daily and weekly step challenges and were effective, for some, in increasing activity levels. Participants often competed against colleagues at work and were keen to keep an eye on everyone's progress. However, not everyone was interested in participating in these challenges. For Joanna and Christine, the competitions had initially provided much fun and excitement, but turned out to be a short-term fad, which did not sustain their interest.

Some WATs were perceived as friendly supporters that mainly offered encouragement, advice, guidance, and data-driven insight. However, for Mary, this was her second WAT. The interactions with her first WAT had been strained, not because there was anything technically wrong with the WAT; it did what it was supposed to do, however, that was exactly the problem. She had felt controlled by the WAT and described it as a "relentless task master." Comparatively, the new WAT was more like an "ally that cheered her on." The interactions with the new WAT reflected a different dynamic. She limited the WAT's influence by carefully managing how she interacted with it and how deeply embedded it was in her daily life. She did this by setting more realistic goals, reducing the number of activities it could track and she requested more infrequent updates from the WAT's mobile phone application. Consequently, she became less obsessed with knowing her weight and less critical of her physical appearance. She started to appreciate her interactions with the WAT and found it helpful when it vibrated and nudged her to be active because she experienced a greater extent of control. This illuminates that within the human-WAT cognitive system, there is an on-going negotiation over influence, which affects decision-making, control, and competencies.

In this theme, it has been evidenced how participants used the WATs and performance data to gain insight, which, for many, was experienced as a form of self-empowerment and extended ability to make better health-related decisions. However, there are also darker sides of relying on WATs and their resources. In the following section, we present further evidence of the complexities of these dynamic human-tech interactions.

Darker Sides of Human-WAT Hybridity

We observed that the hybridity with the on-body, always-accessible WAT also led to some negative experiences for the participants. These experiences reflect the WATs' non-neutrality as well as some of the unintended consequences of an extended mind.

The cognitive abilities that the WATs contributed with (e.g., learning about food calories, calorie burn/in-take, real-time activity levels) made some participants feel insufficient, poor performing, and negative about themselves. Hence, what had started as an exciting experience, turned, for some, into a source of self-loathing and disappointment. Mary explained how she

had taken a break from her first WAT because it was constantly reminding her of how she was *not* meeting her targets, which made her feel guilty about her inability to change her behaviors:

I didn't feel that it [the WAT] was actually working for me and my routine. I think it just kept telling me that I was gaining weight and I got angry and I stopped using it. I think it was the realization of having all that data, it was actually making me realize how unhealthy I was at the time (Mary).

Some participants reported feeling exposed and confronted with what they perceived to be bad habits (e.g., over-eating and a lack of exercise). This led to some emotional distress. By extending monitoring and measuring capabilities to the WATs, it surfaced behavioral traits and habits, which did not match up with some participants' ideal perception of self. The WATs provided quantifiable data, which previously had been ignored, suppressed or simply unknown to the participants.

As the participants expressed a high level of trust in the biometric data, it led to a sense of *bodily disconnect* for some. This disconnect was expressed as a form of alienation between the participant and her own body, fueled by an increased uncertainty about how to best manage and build a strong and healthy body. One participant explained this by saying that she had stopped listening to her internal body and had become reliant on what the WAT told her. The perceived superiority of the WAT to provide a more truthful and accurate assessment affected especially those participants who had become reliant on the WAT's abilities to capture, store, and analyze their data. Some participants were reliant on the data to confirm that they had indeed completed the particular activity. Hence, when the WAT was physically absent, which also led to an absence of the data, they experienced a reduction in the ability to monitor, measure, and assess their activity levels. In this way, participants did not feel capable of completing the tasks that the WATs could undertake.

If I forget to charge it or forget to put it on then I get very annoyed. Because then I have done steps but there is no data to prove that. I don't like that. It is a very bad habit. One day I did a lot of activities and I had forgotten to put it on and I felt very disappointed, even though it is just me who looks at the data. I want to have the data. If you have done the steps and you haven't recorded it, have you really done it? (Maria).

Some participants had developed an intensive dependency relationship with the data, feeling obsessed with checking it. This included tracking progress, analyzing performance, and responding to the data. It offered the capability not only to track performance in real-time, but also to store this data for later comparisons, evaluation, and analysis. This meant that participants' interactions with the WATs were not just real-time, but also involved understanding longer term performance trends. Joanna reflected on how the cognitive capabilities provided by the WAT had led to a sense of obsession:

It was obsessive. It would be all you thought about. You would be constantly checking your steps. How many have I done?

You'd walk to the toilet and then check how many steps that was. Always refreshing the app to see if anyone else had done more steps. I'd be like "Oh no she is ahead of me I need to go for a walk." It definitely disturbed me during the day. I'd have my phone on my desk, so I'd be like "Let's have a little look" when its after lunch, because people would have gone to the gym and I'd be like "Oh I need to go for a walk." So yeah, it probably didn't help productivity in my work life (Joanna).

The extended capabilities of the WATs were not always a positive influence on participants, but led to feelings of stress, self-blame, and the need to improve the performance data even if it was inconvenient or unwanted by the participant. Jane explained how the presence of the data made her determined to meet her targets and when she did not achieve these, she felt angry with herself. She put additional pressure on herself to catch up with the "lost steps" the following day.

I feel cross with myself. All I had to do was to go around the block to meet my target. I should have gone around and done that. I was a bit cross with myself. I should have done that and today I'll try and do 12,000 steps to make up for yesterday (Jane).

The need to reach targets was intensified if targets were part of a competition with other WAT users. Christine and Paula explained how the new capability of tracking and quantifying their exercise and being part of competitions had become a dominant influence on their lives:

It is ridiculously addictive, actually got a bit stupid because it started to interfere with my life and if I hadn't reached a certain amount it would be like 10 o'clock in the evening and I would be like, "I'm going for a run." I'm going out just so that I can get my steps up, just to beat the people that I was in a competition with. Your life had become a constant challenge (Christine).

It is addictive. So, at the very beginning, reaching the goal of the steps was quite important. If someone invites you into a challenge then it gets quite competitive and you do try and beat the other people, especially when it is a narrow margin (Paula).

Some participants took action to change the capabilities of the WAT. They did not stop using the WAT as an external feature, but they purposively reduced the WAT's ability to influence them. Participants explained a number of moves taken to change the outcomes of the WAT's biometric data processing. Some participants inputted a lower food calorie amount for the items eaten to keep their daily calorie intake below the maximum target set on the WAT. Some kept the WAT's step count target purposely low to ensure that they would meet, or even exceed, the target. This would then trigger the WAT to congratulate them on their achievements, consequently inducing positive emotions. Others were strategic about when to request the WAT to sync their performance data from the WAT wristband to the mobile application. In particular, when competing in challenges/competitions, participants requested this sync of data only on the final day of the challenge/competition so not to reveal their progress to the others in the competition. Others were selective about the types of activities they requested the WAT to track, de-selecting some low-performing categories (e.g., sleep and high-intensity activities).

The influence of using the WAT as external scaffolding also became apparent in how the WAT was able to make visible specific performance targets. The mere presence of these targets created the expectation that these needed to be met. When they were not met, it often elicited feelings of frustration and disappointment for participants. Feeling that these performance targets needed to be met meant that participants, at times, adopted behaviors and actions which they would have rather not. Mary, for example, explained how the ability of her first WAT to monitor and assess her eating and drinking habits made her feel disempowered. As she inputted her food and drinks consumption, her habits became quantified, stored, and visually presented to her in the WAT's mobile application. These were habits that had not previously been a concern to her, but the trend data presented by the WAT made her feel embarrassed. It came to a point when she could no longer continue with her normal eating and drinking habits without a high amount of self-doubt. Consequently, the WAT became a source of self-loathing.

It came to the point where I was thinking "Do I have those two glasses of red wine or not?" I was too embarrassed, my Fitbit embarrassed me every day (Mary).

In summary, the investigation revealed that participants' experiences of using WATs to undertake specific monitoring, tracking, and analytical tasks were not without challenges. Darker sides of attaining extended cognitive abilities transpired in the form of self-doubt, bodily alienation, and a mindset fixated on goal/target completion. As a consequence, all participants made moves to limit the WATs' influence and/or the accuracy of its analytics by adjusting the biometric or food input. This was a way of managing the new situation, where cognitive processes related to health management were supported by the abilities of external devices.

DISCUSSION

The study demonstrates that our relationship with technology is complex, even when the technology is relatively mundane and simple, such as WATs. In this study, the EMT has been shown to be an effective analytical apparatus to explore how cognition can become extended beyond the human and the impact this has on decision-making, behaviors, and experiences. The EMT is best adopted as a theoretical lens when the distribution of cognition occurs in relation to specific tasks for which the human needs the support and resources of external entities.

This study builds on the EMT by examining the non-neutrality of the WATs and how they acquire agency in particular situations. Agential intra-action supports the view that external entities are not neutral or non-expressive, but can, in their intra-action with others, for example humans, become expressive, shape intent, and action (Barad, 2003). This dimension of human-tech

relationships is important alongside the EMT as it helps us to illuminate, not only distributed cognition, but also distributed action and effects. With our empirical investigation, we have contributed with insight on human-tech hybridity, and also demonstrated how the EMT and agential intra-action can be used in conjunction to drive knowledge in empirical studies.

It has become evident that those who use a WAT interact with it with the purpose of generating personal data and insight that can lead to certainty and which would be challenging to attain without the support of a WAT. This is similar to findings in the work of Schroeder et al. (2018) who investigated how people use self-tracking technology to better manage migraines. It was found that many people track symptoms and use app alerts to try to predict when a migraine is likely to come on, hence empowering the person to reduce the factors leading to a migraine or better prepare for it. In this way, the self-tracking technology, much like the WATs, is an external resource that can enhance people's capabilities to predict and make decisions.

The WATs become influential external resources in an environment where individuals are focused on identifying the best ways to manage their health and well-being and struggle to trust their internal instinct and gut feel to do so. Hence, the WATs are used as scaffolding (Norman, 1993) to reduce perceived uncertainty and gain support in decision-making.

Table 3 provides an overview of the types of tasks where the WAT offer its resources to complement the human cognitive abilities through enabling extended memory, data capture, and analytical capabilities. The WATs supplement the humans with additional resources, which make the humans better able to judge and take decisions about how to attain a certain calorie intake, maintain a certain intensity of activities (e.g., steps, calorie burn, movement notifications), and review and assess past performance trends. While an individual could capture much of the information related to exercise and food/drinks consumption using a manual logging method, the WATs provide efficiency, automation, objectivity (unless the human manipulates what is logged), consistency, and real-time analytical capabilities based on large and diverse sets of data. Moreover, was the human to undertake these tasks manually, the tools used to capture the information, e.g., notebooks and pens, would also become external scaffolding.

Our study found that initially the WATs are exciting external entities that enable access to biometric information, considered to give new and enhanced decision-making abilities. It is an example of a bio-external device (Clark, 2015) that is flexible and individualized in the way it can share its resources (e.g., data trends, visualizations, notifications, goals) with the individual who draws on its resources and capabilities. In this way, the interactions between the individual and the WAT are dynamic, as the more the person uses it and also manually logs activity, the more data the WAT has to capture, store, analyze, and make available to the person [see also Rapp and Tirassa's (2017) work on a new theory of the self and related guidelines for the design of personal informatics technology].

In most situations, the analyzed and visualized outputs are considered to be trustworthy, which is based on the individuals' implicit trust in the data (Arango-Muñoz, 2013). The individuals do not question what the WATs tell them. Although in some cases, an individual may not like the outputs that the WAT presents them with, for example, a display of poor sleep patterns, and chooses for the WAT to stop monitoring this activity. This, however, is typically not because the data is mistrusted, but often because it is disliked. In addition to trustworthy, the data outputs are also considered to be reliable and, due to the WATs' on-body position, accessible at almost all times. These factors have an impact on the strength of the coupling (Heersmink, 2017) between the person and the WAT.

Coupled systems of heterogeneous entities can create new cognitive systems, where each entity takes on an active role (Clark and Chalmers, 1998). In this study, it is evident that both the WATs and the individuals co-constitute new hybrids that have emerged as a consequence of the human-WAT intra-actions (Barad, 2003). The WATs affect, and for some transform, the experience of being-in-the-world, as they provide a new layer of quantified "life data." At the same time, the individuals also contribute to the coupling by giving the WATs access to trackable and quantifiable behaviors and allowing the devices to be present throughout the day and, for some, the night. However, Rapp and Tirabeni (2018) point to a potential loss of agency and control a person may feel over own body when engaging in self-tracking activities. In their study on mechanisms of externalization of the body among amateur and elite athletes, it appears that elite athletes are better at regulating their interactions with the tracker and know when to trust their subjective sensations, whereas amateur athletes are more reliant on the data to assess their performance. Hence, it seems that the human-WAT relationships may also be affected by the human's existing "practice in relation to her body" (Rapp and Tirabeni, 2018, p. 14).

In this study, it was observed that the WATs have attained an ineliminable role in most human-WAT hybrid relationships (Clark and Chalmers, 1998). This is seen in how the collective competences acquired by the human-WAT hybrids are affected when the WATs are absent. This absence leads to an inability to accurately estimate the number of steps taken, calories burnt, and other data, which the WATs normally collate. This loss in ability is not something the individual believes he or she can restore through their own cognitive abilities. The absence-induced reduction in cognitive information gathering and processing further highlights what it is that the WATs impact and contribute with (Tripathi, 2010); namely, the ability to record, store, analyze, and visually present, at the individual's request, insight into the individual's health and activity performance and progress. Not all people, however, feel unable to make estimates about their activities when the WAT is absent. Some have internalized new behaviors as a consequence of interacting with the WAT and have, as a result, heightened their confidence in making health-related decisions without the input from the WAT. Therefore, when the WAT is absent, they become less affected and are able to make unassisted estimates related

Task	WAT resources	Human cognitive processes
To capture, store, analyze, systemize and categorize real-time biometric data	The ability to record, store, analyze and visually present performance data. The ability to undertake various analyses of the data based on the human's request.	The ability to operate the WAT, manage settings and enable/disable the WAT's biometric data collection abilities. The cognitive ability to set up a sync between the WAT, the WAT's mobile phone app and the WAT's desktop interface. The cognitive ability to set health and activity goals.
To ascertain times when an increase in movement/exercise is recommended	The ability to manage large biometric datasets, undertake pattern recognition analysis, compare performance with set goals, and provide notifications/nudges.	The ability to judge whether to respond to the notification/nudge, which may depend on other situational factors.
To monitor and track calorie intake	The ability to store, categorize and visually present consumed calories over time and according to food/drinks categories.	The ability to identify the best match between a selected food/drinks item and food/drinks items available on the WAT's mobile app or desktop interface.
To gain insight into sleep patterns and rhythms	The ability to automatically detect when sleep is initiated (or respond to sleep mode being activated) and record hours slept, sleep patterns and present a daily analysis of sleep quality.	The ability to manage the sleep mode setting.
To monitor heart rate in real-time	The ability to monitor and record heart rate in real-time, quantify heart rate performance, and provide notifications/nudges when deemed above an advisable level.	The ability to judge whether a change in behavior is required to either increase or reduce the heart rate.
To gain certainty about the activities required to attain a healthy and active lifestyle	The ability to generate real-time and quantitative information that reflects in-the-moment performance as well as analyses that are based on longitudinal performances.	The ability to process the information and analyses produced to complement own information processing capacity.

to calorie intake, distance covered, and the quality of their sleep.

From the participants' accounts, the access to activity and performance data helped to create a sense of heightened control and reduced fear of making poor decisions (Schulz, 2011) related to their health. Schüll (2016) emphasizes that self-tracking devices are both sources of responsibility and delegation. People who self-track wish to make informed decisions and take responsibility for their actions and behaviors, but at the same time, they also delegate part of the responsibility for this to an external device. Specifically, what is delegated to the external technology is the responsibility to "calculate and act upon itself." This is seen in how the WATs are given the responsibility for calculations and assisting the individual's decisions through nudges and notifications. However, the darker sides of the human-WAT hybrid also reveal that relying on the resources of a WAT is not uncomplicated.

Drawing on the views of Ihde (1990); Pickering (2013) and Heersmink (2017) in terms of the non-neutrality of technology and the belief that technology, like humans, can also express agency, we contribute with the concept of the *agency pendulum*. The agency pendulum draws inspiration from Barad's (2003) rendition of agency; that it is enactments which emerge from intra-actions between all sorts of entities including those that are not human. Hence, agency is not a constant and it is not an attribute assigned to an entity (Ewalt, 2016). Instead, it is played out, expressed, and seen through effects and collective capabilities (Pickering, 1995).

The agency pendulum swings between the human and the WAT, which means that, at times and in specific situations, the human is enabled to affect and create change; in other situations, it is the WAT that influences and impacts decisions and behaviors (**Table 4**). The agency pendulum does not reflect a symmetrical division of agency. Its movements are individualized to the specific human-WAT hybrid and are affected by, for example, how much the human cares about, listens to, and becomes affected by the expressions of the WAT as well as the expressive abilities of the WAT. The agency pendulum acts as a metaphor for distributed agency and attempts to highlight the non-neutrality of the WATs as they intra-act with the humans.

The human enacts agency in situations when the individual limits or extends the WAT's presence and influence by increasing or decreasing the collection of particular biometric data. Nafus

Human
- Change the settings to increase or decrease the number and types of metrics being tracked
- Prevent the WAT from synching with the mobile phone app or WAT's desktop interface
- Alter the WAT input data to generate particular results from the WAT's analysis

WAT
- Send real-time notifications and nudges that lead to behavior change or affect decision-making
- Share health data that alter attitudes toward food and drink consumption
- Gain trust as an objective and truthful reflection of health performance

and Sherman (2014) describe these kinds of practices as forms of "soft resistance" to the automatic collection of personal informatics. The soft resistance becomes visible when people purposely choose when automatic data collection is a helpful external exercise and when they would rather use internalized decision-making criteria. In our study, human agency is also expressed when the WAT is prevented from synching the captured data with the mobile phone app or desktop interface. This prevents the WAT from undertaking the analysis and visualization of the data, which is a central task. The human is also seen to express agency when the data inputted is altered to achieve certain outputs. This could be food calorie data that is lowered or a step count target that is set purposely low. In these situations, the agency pendulum swings toward the human, as the human carries out acts that affect him/herself, the intra-action with the WAT and the influence that the WAT can have. These acts challenge the "reflexive monitoring self," which is often described as a "rational, motivated, and data-centric" individual (Lupton, 2016, p. 115). Rather, at times, the human acts to reduce or alter the pattern of tracking and consequent self-surveillance.

When the agency pendulum swings toward the WAT, it is often when the WAT makes expressions, which end up affecting the human. These acts are outcomes of the WAT's design, but nonetheless, when they are expressed, they have an influence on human behavior and decision-making. When the WAT sends notifications and nudges the individual to adopt or avoid a particular behavior or make a particular decision, then that is an assertion of the WAT's agency. At times, this leads participants to feel a loss of control over own actions, a wish to reclaim control and an experience that the decisions they are making are not truly their own. Similarly, when the WAT presents food and drinks consumption trend analytics, then this is a situation that can impact the individual's decision-making or feeling about self. The WAT also enacts agency when it acquires a position of trust, i.e., the human considers it trustworthy and is willing to trust the information it communicates based on its information processing and analytics. The WAT is provided the autonomy to conduct these calculations and visualizations and communicate them to the human, taking on the position of an autonomous system (Ohlin and Olsson, 2015).

The agency pendulum acts as a helpful metaphor when exploring this intersection of distributed cognition and distributed agency. It enables an illumination of the impacts and influences of the external features beyond their roles as offering cognitive scaffolding. It can add a further dimension of insight to understand the dynamics of human-tech hybridity by including a focus on distributed action and impact. Considering the EMT and agential intra-action as part of a connected exploration, enables an investigation that examines both how cognition is extended to other entities to solve tasks and also how these external entities, in their intra-actions with the human, come to enact agency, i.e., become expressive, shape intent, and action. Hence, we put forward the notion that distributed cognition and distributed agency are not capacities that are mutually exclusive, but are in fact closely tied together.

CONCLUSION

The EMT and the concept of agential intra-action can be used as a combined theoretical apparatus to explore the distribution of cognition from the human to other external entities and the dynamism of agentic enactments, as both the human and the external entities attain the ability to influence, impact, and create change. The theoretical lenses provide specific concepts that can be applied to surface how abilities, competences and capacities can become co-constituted in human-tech hybrid relationships and how these affect the ability to make decisions and solve problems. The research identified specific ways in which the WATs support the humans' cognitive abilities by contributing with an extended memory, data capture and analysis capabilities.

The research also highlighted that, while technologies, such as WATs, can contribute with new abilities and insight, there are unintended implications of this engagement between the human and the WAT. Attaining an extended mind and interacting with an external entity that has the ability to influence, guide, and illuminate health-related behaviors through constant monitoring and tracking, can lead to some experiences of stress, disappointment, and self-blame. Hence, as entities and new technologies are developed to support human cognition, it is important to also consider the side effects and non-neutrality of the technology and how that impacts the human experience.

To further capture and conceptualize the dynamism of distributed agency, we presented the concept of the agency pendulum. The agency pendulum swings between the human and the external entity, which means that, at times and in specific situations, the human is enabled to affect and create change; in other situations, it is the external entity that influences and impacts decisions and behaviors.

We posit that it is particular useful to consider distributed agency as an additional layer of theoretical exploration when considering the EMT in order to also capture the non-neutrality of the external entities that support human cognitive abilities.

ETHICS STATEMENT

This study was carried out in accordance with the recommendations of the Ethics Committee at Ashridge Executive Education at Hult International Business School with written informed consent in accordance with the Declaration of Helsinki.

AUTHOR CONTRIBUTIONS

This study is part of on-going research into wearable and other digital technologies by RD and MC. RD and MC developed the theoretical framing, undertook the empirical data collection, and drove the analysis and presentation of themes and contributions. NP took part in the data analysis and initial identification of themes.

REFERENCES

Aizawa, K. (2010). The coupling-constitution fallacy revisited. *Cogn. Syst. Res.* 11, 332–342. doi: 10.1016/j.cogsys.2010.07.001

Arango-Muñoz, S. (2013). Scaffolded memory and metacognitive feelings. *Rev. Philos. Psychol.* 4, 135–152. doi: 10.1007/s13164-012-0124-1

Barad, K. (2003). Posthumanist performativity: toward an understanding of how matter comes to matter. *Signs* 28, 801–831. doi: 10.1086/345321

Bode, M., and Kristensen, D. B. (2016). "The digital doppelgänger within. a study on self-tracking and the quantified self movement," in *Assembling Consumption: Researching Actors, Networks and Markets*, eds R. Canniford and B. Domen (New York, NY: Routledge), 119–135.

Bosse, T., Jonker, C. M., Schut, M. C., and Treur, J. (2005). Simulation and analysis of a shared extended mind. *Simulation* 81, 719–732. doi: 10.1177/0037549705060260

Brown, W. S., and Strawn, B. D. (2017). Beyond the isolated self: extended mind and spirituality. *Theol. Sci.* 4, 411–423. doi: 10.1080/14746700.2017.1369755

Bryman, A., and Bell, E. (2011). *Business Research Methods.* Oxford: Oxford University Press.

Carr, N. (2011). *The Shallows: What the Internet is Doing to Our Brains.* New York, NY: W.W. Norton & Company.

Clark, A. (2001). Reasons, robots and the extended mind. *Mind Lang.* 16, 121–145. doi: 10.1111/1468-0017.00162

Clark, A. (2007). Re-inventing ourselves: the plasticity of embodiment, sensing, and mind. *J. Med. Philos.* 32, 263–282. doi: 10.1080/03605310701397024

Clark, A. (2008). *Supersizing the Mind: Embodiment, Action, and Cognitive Extension.* New York, NY: Oxford University Press.

Clark, A. (2015). What "Extended Me" knows. *Synthese* 92, 3757–3775. doi: 10.1007/s11229-015-0719-z

Clark, A., and Chalmers, D. (1998). The extended mind. *Analysis* 58, 10–23. doi: 10.1093/analys/58.1.7

Clowes, R. W. (2018). Immaterial engagement: human agency and the cognitive ecology of the internet. *Phenomenol. Cogn. Sci.* 39, 1–21. doi: 10.1007/s11097-018-9560-4

Corbin, J., and Strauss, A. (2008). *Basics of Qualitative Research: Techniques and Procedures for Developing Grounded Theory*, 3rd Edn. Thousand Oaks, CA: Sage Publications, Inc. doi: 10.4135/9781452230153

Denzin, N. K. (1989). *Interpretive Interactionism.* Newbury Park, CA: Sage.

Dinerstein, J. (2006). Technology and Its discontents: on the verge of the posthuman. *Am. Q.* 58, 569–595. doi: 10.1353/aq.2006.0056

Easterby-Smith, M., Thorpe, R., and Jackson, P. (2012). *Management Research*, 4th Edn. London: Sage.

Etkin, J. (2016). The hidden cost of personal quantification. *J. Consum. Res.* 42, 967–984. doi: 10.1093/jcr/ucv095

Ewalt, J. P. (2016). The Agency of the spatial. *Womens Stud. Commun.* 39, 137–140. doi: 10.1080/07491409.2016.1176788

Fotopoulou, A., and O'Riordan, K. (2016). Training to self-care: fitness tracking, biopedagogy and the healthy consumer. *Health Sociol. Rev.* 26, 54–58. doi: 10.1080/14461242.2016.1184582

Heersmink, R. (2015). Dimensions of integration in embedded and extended cognitive systems. *Phenomenol. Cogn. Sci.* 14, 577–598. doi: 10.1007/s11097-014-9355-1

Heersmink, R. (2017). Extended mind and cognitive enhancement: moral aspects of cognitive artifacts. *Phenomenol. Cogn. Sci.* 16, 17–32. doi: 10.1007/s11097-015-9448-5

Herbrechter, S. (2012). Posthumanism, subjectivity, autobiography. *Subjectivity.* 5, 327–347. doi: 10.1057/sub.2012.13

Hutchins, E. (2014). The cultural ecosystem of human cognition. *Philos. Psychol.* 27, 34–49. doi: 10.1080/09515089.2013.830548

Ihde, D. (1990). *Technology and the Lifeworld: from Garden to Earth.* Bloomington, IN: Indiana University Press.

Ihde, D. (2004). Has the philosophy of technology arrived? a state-of-the-art review. *Philos. Sci.* 71, 117–131. doi: 10.1086/381417

Irvine, A., Drew, P., and Sainsbury, R. (2013). 'Am I not answering your questions properly?' clarification, adequacy and responsiveness in semi-structured telephone and face-to-face interviews. *Qual. Res.* 13, 87–106. doi: 10.1097/QAD.0b013e328350995c

King, N., and Horrocks, C. (2010). *Interviews in Qualitative Research.* Thousand Oaks, CA: Sage Publishing.

Krueger, J. (2014). Affordances and the musically extended mind. *Front. Psychol.* 4:1003. doi: 10.3389/fpsyg.2013.01003

Lupton, D. (2016). The diverse domains of quantified selves: self-tracking modes and dataveillance. *Econ. Soc.* 45, 101–122. doi: 10.1080/03085147.2016.1143726

Menary, R. (2010). "Cognitive integration and the extended mind," in *The Extended Mind*, ed. R. Menary (Cambridge,MA: MIT Press), 227–244.

Nafus, D., and Sherman, J. (2014). This one does not go up to 11: the quantified self movement as an alternative big data practice. *Int. J. Commun.* 8, 1784–1794.

Nelson, E. C., Verhagen, T., and Noordzij, M. L. (2016). Health empowerment through activity trackers: an empirical smart wristband study. *Comput. Hum. Behav.* 62, 364–374. doi: 10.1016/j.chb.2016.03.065

Norman, D. (1993). *Things That Make US Smart: Defending Human Attributes in the Age of the Machine.* New York, NY: Basic Books.

Ogden, R. (2008). "Pseudonym," in *The SAGE Encyclopedia of Qualitative Research Methods*, ed. M. Lisa (Thousand Oaks, CA: Sage Publications).

Ohlin, F., and Olsson, C. M. (2015). "Intelligent computing in personal informatics: key design considerations," in *Proceedings of the 20th International Conference on Intelligent User Interfaces (IUI '15)*, New York, NY: Association for Computing Machinery, 263–274. doi: 10.1145/2678025.2701378

Pickering, A. (1995). *The Mangle of Practice: Time, Agency, and Science.* Chicago, IL: University of Chicago Press. doi: 10.7208/chicago/9780226668253.001.0001

Pickering, A. (2013). "Living in the Material World," in *Materiality and Space. Technology, Work and Globalization*, eds F. X. de Vaujany and N. Mitev (London: Palgrave Macmillan).

Puig de la Bellacasa, M. (2009). Touching technologies, touching visions. The reclaiming of sensorial experience and the politics of speculative thinking. *Subjectivity* 28, 297–315. doi: 10.1057/sub.2009.17

Rapp, A., and Cena, F. (2016). Personal informatics for everyday life: how users without prior self- tracking experience engage with personal data. *Int. J. Hum. Comput. Stud.* 94, 1–17. doi: 10.1016/j.ijhcs.2016.05.006

Rapp, A., and Tirabeni, L. (2018). Personal informatics for sport: meaning, body, and social relations in amateur and elite athletes. *ACM Trans. Comput. Hum. Interact.* 25, 1–30. doi: 10.1145/3196829

Rapp, A., and Tirassa, M. (2017). Know thyself: a theory of the self for personal informatics. *Hum. Comput. Interact.* 32, 335–380. doi: 10.1080/07370024.2017.1285704

Ritchie, J., Lewis, J., and Elam, G. (2003). "Designing and selecting samples," in *Qualitative Research Practice,* eds J. Ritchie and J. Lewis (Thousand Oaks, CA: Sage Publications), 77–108.

Rowlands, M. (2009). Extended cognition and the mark of the cognitive. *Philos. Psychol.* 22, 1–19. doi: 10.1080/09515080802703620

Schroeder, J., Chung, C. F., Epstein, D. A., Karkar, R., Parsons, A., Murinova, N., et al. (2018). "Examining self-tracking by people with migraine: goals, needs, and opportunities in a chronic health condition," in *Proceedings of the 2018 Designing Interactive Systems Conference (DIS '18)*, (New York, NY: ACM), 135–148. doi: 10.1145/3196709.3196738

Schüll, N. D. (2016). Data for life: wearable technology and the design of self-care. *Biosocieties* 11, 317–333. doi: 10.1057/biosoc.2015.47

Schulz, K. (2011). *Being Wrong: Adventures in the Margin of Error.* New York: HarperCollins.

Smith, G. J. D., and Vonthethoff, B. (2016). Health by numbers? Exploring the practice and experience of datafied health. *Health Sociol. Rev.* 26, 6–21. doi: 10.1080/14461242.2016.1196600

Spiggle, S. (1994). Analysis and interpretation of qualitative data in consumer research. *J. Consum. Res.* 21, 491–503. doi: 10.1086/209413

Stern, B. B., Thompson, C. J., and Arnould, E. J. (1998). Narrative analysis of a marketing relationship: the consumer's perspectiv. *Psychol. Mark.* 15, 195–214. doi: 10.1002/(SICI)1520-6793(199805)15:3<195::AID-MAR1>3.0.CO;2-5

Sutton, J., Harris, C. B., Keil, P., and Barnier, A. J. (2010). The psychology of memory, extended cognition, and socially distributed remembering. *Phenomenol. Cogn. Sci.* 9, 521–560. doi: 10.1007/s11097-010-9182-y

Thacker, E. (2003). Data made flesh: biotechnology and the discourse of the posthuman. *Cult. Crit.* 53, 72–97. doi: 10.1353/cul.2003.0029

Thompson, C. J., Locander, W. B., and Pollio, H. R. (1989). Putting consumer experience back into consumer research: the philosophy and method of existential-phenomenology. *J. Consum. Res.* 19, 133–147. doi: 10.1086/209203

Thompson, E., and Stapleton, M. (2009). Making sense of sense-making: reflections on enactive and extended mind theories. *Topoi* 28, 23–30. doi: 10.1007/s11245-008-9043-2

Tripathi, A. K. (2010). Ethics and aesthetics of technologies. *AI Soc.* 25, 5–9. doi: 10.1007/s00146-010-0265-7

Verbeek, P. P. (2006). Materializing morality: design ethics and technological mediation. *Sci. Technol. Hum. Values* 31, 361–380. doi: 10.1177/0162243905285847

Ward, T. (2009). The extended mind theory of cognitive distortions in sex offenders. *J. Sex. Aggress.* 15, 247–259. doi: 10.1080/13552600903263087

Wilson, R., and Clark, A. (2009). "How to situate cognition: letting nature take its course," in *The Cambridge Handbook of Situated Cognition*, eds P. Robbins and M. Aydede (New York, NY: Cambridge University Press), 55–77.

frontiers
in Psychology

Designing Smart Objects to Support Affording Situations: Exploiting Affordance Through an Understanding of Forms of Engagement

Chris Baber*

School of Engineering, University of Birmingham, Birmingham, United Kingdom

In this paper I consider how the concept of "affordance" has been adapted from the original writings of Gibson and applied to interaction design. I argue that a clear understanding of affordance shifts the goal of interaction design from one of solely focusing on either the physical object or the capabilities of the person, toward an understanding of interactivity. To do this, I develop the concept of Forms of Engagement, originally proposed to account for tool use. Finally, I extend this concept to interacting with modified tangible user interfaces, or "animate objects." These animate objects not only sense how they are being used, but also communicate with each other to develop a shared intent, and provide prompts and cues to encourage specific actions. In this way, the human-object-environment system creates affording situations in pursuit of shared intentions and goals. In order to determine when to provide prompts and cues, the objects need to have a model of how they ought to be used and what intention they are being used to achieve. Consequently, affordances become not only the means by which actions are encouraged but also the manner in which intentions are identified and agreed.

Keywords: affordance, smart objects, animate objects, interactivity, forms of engagement

INTRODUCTION

This paper is motivated by three simple questions: (i) how do people know how to use smart objects (i.e., how do people respond to the form and function of smart objects in order to achieve goals)? (ii) how do objects make sense of the manner in which they are being used (i.e., can objects recognize different ways in which a person interacts with them)? (iii) how should designers design smart objects to enable people to use these appropriately (i.e., is it possible to better inform design practice so that we can predict the successes and challenges of interacting with smart objects)? Unpacking this a little, a "smart object" (Kortuem et al., 2010) is some artifact with which a person can interact, but which is capable of sensing that it is being interacted with, capable of making inferences from these sensor data, capable of communicating these inferences with other artifacts, and capable of guiding the person to perform further actions. Knowing how to use an object *could* involve problem-solving in which features of the object are associated with functions, and these functions associated with a plan to act. But often, there is little overt, conscious awareness in performing the action.

By way of a motivating example, imagine that you are reaching to pick up a cup containing a hot drink. The handle of the cup could be grasped in a particular way (say, two fingers through the handle and the thumb resting on the top), or the body of the cup could be held in your palm with fingers and thumb wrapping around it. Which grasp you select depends on, among other things, the heat of the contents of the cup, whether the cup is full to the brim, whether the handle is on one side or the other. However, it is unlikely that your selection arises from conscious deliberation: you simply pick up the cup. As Wittgenstein noted, "The aspects of things that are most important for us are hidden because of their simplicity and familiarity." (Wittgenstein, 1958, p.50). The concept of affordance helps frame this activity and explain how it can be performed without conscious intervention. In other words, very often, we simply "know" what to do. From the perspective of cognitive psychology, this knowledge has been termed "procedural" (Anderson, 1981), tacit (Polyani, 1966), "implicit" (Berry and Broadbent, 1988), or "automatic" (Schneider and Shiffrin, 1977; Shiffrin and Schneider, 1977). It is from these different traditions that one can appreciate what "affordance" involves. Relating this to design, one could also suggest that understanding such implicit, subliminal interaction could align neatly with some formulations of the concept of nudging, in which people might be encouraged to perform actions on certain ways and where such encouragement would be at the edge of conscious awareness (Thaler and Sunstein, 2008). Nudging could, for example, involve the cup prompting the user to pick it up (perhaps to encourage the person to drink more water) or it could encourage picking the cup up with one hand rather than the other (perhaps as part of rehabilitation) or it could encourage the person not to pick it up (perhaps to discourage the person from drinking coffee after a certain time of day). In these instances, the cup takes on the role of a smart (possibly irritating, possibly helpful) partner in performing an action. For me, the question is whether this partnering could be both beneficial and performed without conscious awareness. So, could interaction with a smart object be described in terms of affordance. I begin this paper with a short account of how the concept of affordance has developed, with particular reference to interaction design.

A Brief History of "Affordance"

For many people working in the field of Human-Computer Interaction (HCI), their first encounter with the concept of "affordance" probably came from Norman's (1988) *The Psychology of Everyday Things* [he later, in 2002, rewrote this as *The Design of Everyday Things*]. In this book, Norman presents "affordance" as an act of interpretation, in which the form of an object is seen in relation to a specific action. So, the flat plate on a door "affords" pushing. From this perspective, the "affordance" is a visual clue, provided by the object, as to its intended functioning: "Plates are for pushing. Knobs are for turning. Slots are for inserting things into. Balls are for throwing or bouncing." (p. 9). What is deceptively attractive about this notion, for design at least, is the implication that the physical form of the object corresponds with a conceptual model that the user of the object brings to the interaction. In other

words, Norman's (1988) definition, while it looks to be based on perception, is really about interpreting the object's functions in terms of specific features, and linking this interpretation to a goal that one wishes to achieve. Returning to our example of picking up a cup, this implies that one needs to selectively determine which features of the cup (and its contents) are most salient to the goal of drinking from it (under certain constraints, like not spilling the contents or scalding one's hand). In other words, there is an implication that, prior to performing an action, one engages in a sort of problem-solving which allows salient features to be elicited and interpreted. Later, Norman (1999) distinguished "perceived affordances" from what he defined as Gibsonian or "real affordances." It is worth noting at this point that there are extreme differences between "perceived" and "real" affordances. For one thing, Gibson's (1977, 1979) claim is that we have a perceptual system which is tuned (through evolution, experience, learning) to the environment. This means that there is no requirement for any form of interpretation of information; we just "see" (or hear or otherwise perceive) a pattern to which we can respond. To repeat our example, a cup full of steaming hot coffee is "seen" as a different object (supporting different actions) to a half-full cup of cold milk. When Norman uses the word "perceive," this is not in the same manner that Gibson uses it; Norman seems to suggest that perception is an active process of extracting features and assigning meaning, whereas for Gibson, perception is the capability of being sensitive to information. Later still, Norman (2008) separated "real affordances" from signifiers, i.e., perceptual information about objects.

It is worth tracking the term "affordance" back further. Gibson taught a course on the phenomenological philosophy of Merleau-Ponty, and the Gestalt psychologist Koffka was a colleague of Gibson's in the 1930s (Kaufer and Chemero, 2015). Key to Merleau-Ponty's (1945) Phenomenology is the notion of *intentionality*, which is concerned with how we "see" an object in terms of how we will interact with it (rather than as a collection of features). That is, we see the *intentional object* in relation to our intended action. One way of appreciating this, is through the concept of "Gestalt" (with which Merleau-Ponty was familiar), which is not some property of the object but rather the combination of the sensory stimulation evoked by an object in a given context. In Norman's (1988) glossing of "affordance," the object becomes imbued with meaning in a way that Gibson (and Gestalt psychologists, and Merleau-Ponty) resisted. This means not only that the "Gestalt" is more than the sum of its parts, but also that the object can be interacted with differently under different conditions. This reiterates our distinction between cups of hot coffee and cold milk. In order to interact with an object, the individual must have the ability to act upon or with that object; and so, the individual can be considered in terms of effectivities (Turvey and Shaw, 1979). In this respect, environmental constraints (in terms of properties of objects) are responded to in terms of bodily constraints (in terms of effectivities). Stoffregen (2003) and (Chemero, 2003; Chemero et al., 2003) dispute the implication that "affordance" arises because the object elicits a dispositional response in the user, and they propose that this should not be regarded in

terms of dispositions (that is, consistent responses to objects) but rather in terms of abilities (that is, flexible and adaptive styles of interaction). Furthermore, as Osiurak et al. (2017) point out, the notion of effectivity conflates two kinds of action possibilities—those offered by the body and those provided by objects.

In terms of Gestalt psychology, Lewin (1936) developed the concept of *Aufforderungscharaktere* (translated as "demand character," "invitation-character" or "prompt-character") indicating the properties of an object which call for a certain behavior. This describes interaction with an object in a context, in terms of "valences" (which are a function of the person's (motivational) state and the properties of the environment in which they are acting). This implies (I feel) that the relationship between object and action would vary according to person and environment (much as Merleau-Ponty, 1945 suggests). In contrast, Gibson (1979) claims that "affordances" are invariant and quotes his colleague Koffka as saying, "*Each things say what it is… a fruit says 'Eat me', water says 'Drink me', thunder says 'Fear me'…*" (Koffka, 1955, p.7). My problem with this claim is that it seems to return us to the idea that an "affordance" is a property of the object and is independent of the viewer. In contrast, in order to perceive an object's affordance, one needs to have prior experience of using objects of this type and a set of beliefs as to how such objects ought to be used. This gives a strong cultural and experiential basis to the response to affordance in ways that Gibson was seeking to avoid through his insistence that perception of affordance was a direct response to the visual appearance of an object. Gaver (1991) suggested that one could separate affordance from perceptual information, and introduced terms such as "false affordance" (in which the form of object implies a possible action, say a decal on a product that looks like a button) and "hidden affordance" (in which perceptual information is obscured). Although the notions of "false" and "hidden" affordance are useful, this relies on the conflation of "affordance" with function. This creates further confusion in the application of affordance to design—should we be concerned with designing visual signifiers that cue an action (which is, surely, much the same as stating that the form of an object signifies its functions, which designers know anyway) or does affordance provide another perspective on design?

From his interpretation of Gibson's various proposals about "affordance," Chemero (2009) suggests that, "Affordances are neither properties of the animal alone nor properties of the environment alone. Instead, they are relations between the abilities of an animal and some feature of a situation." (p.191). This observation is significant to the current paper for three reasons. First, it recognizes that affordances arise through relations in animal-object-environment systems (rather than existing as properties of any constituent component). This raises questions about what the designer is designing in order to support affordance. My answer to this is that design, in this context is less about the fashioning of objects (although, of course, these are important) and more about choreographing situations in which people interact with objects. Second, the idea that affordances are relations implies that people rarely attend

to the specific features of the context in which these relations occur. In their discussion of affordance, Still and Dark (2013) suggest that people respond to affordances "automatically," i.e., with little or no conscious awareness or need for attentional control. Similarly, the use of highly familiar objects would involve minimal attentional demand, but when confronted with a novel or unfamiliar object, there would be a need to construct a plan of how to interact with it (Humphreys, 2001; Humphreys et al., 2010). Furthermore, if affordances guide action then this could only be for someone able to perceive the relevant "information," able to perform the relevant action, and able to relate the action to a desirable goal (Roux and Bril, 2005; Fairlie and Barham, 2016). As Kirsh (2013) has it, "goals make perception enactive" (Kirsh, 2013, p. 10). To illustrate this, he gives the example of a stonemason (or bricklayer) who "…will look at bricks for places to apply cement; when looking at an odd brick he will 'see' the particular trowel shape that is needed." (Kirsh, 2013, p. 9). For someone without the experience of bricklaying, there is less likely to be distinctions between bricks and less likely for these distinctions to result in changes in action.

Formally Describing Affordance

Lewin (1936), who we have already noted as a providing a precursor definition of what became known as "affordance," developed a simple equation (Equation 1) to model behavior (*B*) as a function *f* of Person (*P*) and environment (*E*).

$$B = f(P, E) \qquad (1)$$

This simply states that behavior of a person is directly connected to the environment in which they act. In order to address some of the issues surrounding the debate over what "affordance" might be, Turvey (1992) proposed a formal definition (Equation 2 which one can see is inheriting Lewin's idea). This can be expressed as:

$$W_{p, q} = j(X_p, Z_q) \text{ possesses } r \qquad (2)$$

In other words an Environment or World, *W*, has properties p and q which can be defined as the joining, j, of an object *X* (with property *p*) and an animal *Z* (with effectivity *q*) in order to produce an affordance relationship, *r*. In this account, the animal has a set of dispositions, characterized in terms of effectivity, which enable it to respond to object properties. So, an adult human hand can grasp the handle of a full cup and lift it in a way that a child's smaller hand might not be able to: from Equation (2), the cup_handle (for the adult) affords grasping (because its property, *p*, defined by its size and shape, matches the disposition, *q*, of the person, defined by hand-size), and the full_cup affords lifting because of the adult's strength. As noted previously, Stoffregen (2003) questioned Turvey's (1992) claim that effectivities are dispositions. He suggested that it makes more sense to regard these as abilities that can be called upon in a given situation. This is useful because it means that the response that a

designer might expect to elicit using a given form could be correct in terms of effectivity but not ability, and so, affordance is about matching ability <u>not</u> disposition.

For Stoffregen (2003), affordance emerges from the World-Object-Animal system and is not a property of any one of these in isolation. Thus, Stoffregen (2003) offered Equation (3).

$$W_{p,q} = (X_p, Z_q) \text{ possesses } h \qquad (3)$$

What the formal descriptions struggle to present is the discretion with which such responses are made. In other words, is it possible to not respond to an object's "solicitation" of a response? Certainly, this is not easy to see from Turvey's (1992) account. For Stoffregen (2003), the post-hoc description of an affordance as something that has occurred in a system, rather blurs this problem.

To consider the problem more concretely, the notion of Stimulus-Response Compatibility has been a staple part of Ergonomics design thinking for the past half century. To illustrate this idea, imagine that you have a row of 4 lights in front of you (labeled 1–4), and between you and the lights is a row of 4 buttons (labeled A–D). The buttons and lights are arranged so that 1 and A are adjacent, etc. When one of the lights turns on, you must press one of the buttons to turn off this light as quickly as possible. In the adjacent (or congruent) arrangement, when light1 turns on, you press button A. In an incongruent arrangement, when light 1 turns on, you have to press, say, button C. Not surprisingly, the congruent arrangement leads to much faster performance. Early accounts of the SRC suggested that the performance differences were due to "translation" (Fitts and Seeger, 1953; Fitts and Deininger, 1954; Welford, 1976). People prefer arrangements in which the elements (light and button) are congruent, and this is termed a Population Stereotype (there is some work to suggest that different cultures might have slightly different Population Stereotypes). Furthermore, most people produce faster responses with fewer errors in Sets of stimulus-response pairings which have this preferred arrangement, and this defines Stimulus-Response Compatibility (SRC). A popular explanation of SRC relates to the ability to extract salient features and pair these with an appropriate response. This is the "dimensional overlap" model (Kornbulm et al., 1990) and broadly contrasts the overlap of dimensions (elements) in a set (i.e., the congruence of arrangements) with the relevance of elements within a set (i.e., how the features of a stimulus relate to a response). For example, button presses could conceivably be made in response to proper names. In this case, there is no overlap between the layout of the buttons and the nature of the stimulus, and there is no relevance of stimulus content to response. On the other hand, button presses might be to the lights (which might be labeled with proper names). In this case, there is no relevance of the names, but there might be overlap between the position of the light and the position of the button. Finally, the congruent condition (arranging buttons and lights as described earlier) has both overlap and relevance.

The relevance of SRC to HCI has been recently reviewed in a paper by Proctor and Vu (2016), and they suggest that

it continues to provide useful guidelines for design. There is much to be said for the empirical evidence from SRC. From the perspective of affordance, it could be argued that SRC arises when information from environment (stimulus) relates to ability (response). In other words, there is potential argument that removes the need to appeal to a "translation" or a "dimensional overlap" to explain this. In their paper, Proctor and Vu (2016) argue against "affordance" and suggest that it merely describes a particular form of spatial compatibility. I felt that they misrepresented the basic ideas of affordance and agree with Stins and Michaels (1997), who argued that, in SRC studies, the "information" could include more than just the position of the response buttons (as SRC tends to assume). Crossing one's hands in SRC experiments leads to an increase in reaction time, even when the position of stimulus and response objects remain constant, and this does not seem to be the result of a simple biomechanical constraint; reactions using crossed hands cannot be explained solely by conflict management, as proposed by the dimensional overlap model. This suggests that the relationship between response and stimulus involves more than the simple mappings that SRC assumes. Further, SRC studies often fail to control properly for the different compatibility effects that could arise from the use of different response actions that are required. Finally, SRC studies do not seem to be able to account for how changes in ability can lead to changes in performance. Having said that, the formal approaches to affordance outlined, above do not account for this either. If we refer back to the formalisms outlined in Equations (2) and (3), it is difficult to see how these could account for the differences in SRC. In both congruent and incongruent conditions, X_p would be "light on," and Z_q would be "press button." So, perhaps, we need to elaborate the X_p description to include X_{p1} "light on" + X_{p2} "light adjacent to button" (in the congruent condition), and to elaborate Z_{q1} "associate light label with button label" + Z_{q2} "press button" in the incongruent condition.

While the formal descriptions of Turvey (1992) and Stoffregen (2003) are directed at the immediate relationship between an object and its user, this does not fully capture the situation in which the relationship arises. For Kirlik (2004), a problem with Stoffregen's (2003) equation is that there does not appear to any constraint on how to define the parameters. Abbate and Bass (2017) develop a variation of Stoffregen's (2003) formalism that works with *a priori* constraints (Equation 4):

$$Possesses(affordance_i)(X_p, Z_q) \qquad (4)$$

This relationship becomes expandable with specific values of the elements of X that are relevant to a given "goal" and with specific values that define the ability of Z required to respond to these features. As an example, Abbate and Bass (2017) propose that an aircraft cabin door is plugged into its fitting under high external pressure, and that (on the ground) the door can be opened by pulling out a lever and then turning it. So, in this case, there are two affordances of interest, i.e., *leverLiftable,* and *doorOpenable*. These can be defined as follows:

possesses(leverLiftable)(X_p, Z_q) = true if:
$X_p.Airspace.Aircraft.Cabin.Door.Lever_{p1}[Slot][bottom\_of]=overlapping \wedge$
$X_p.Airspace.Aircraft.Cabin.Door._{p1} \text{ x } (X_p.Airspace.Aircraft.Cabin._{p1} - X_p.Airspace_{p1}) + X_p.Airspace.Aircraft.Cabin.Door.Lever_{p2} \leq$
$Z_q.Airspace.Aircraft.Cabin.Door.Lever_{q1}[position\_up]$

possesses(doorOpenable)(X_p, Z_q) = true if:
$X_p.Airspace.Aircraft.Cabin.Door.Lever_{p1}[Slot][top\_of]=overlapping \wedge$
$X_p.Airspace.Aircraft.Cabin.Door._{p2}[Cabin][left\_of] = contained\_within \wedge$
$Z_q.Airspace.Aircraft.Cabin.Door._{q1}[position\_back]= true \wedge Z_q.Airspace.Aircraft.Cabin.Door._{q1}[translate\_left]=true$

This formal description elaborates the context under which the lever "affords" lifting and the door "affords" opening (in terms of external air pressure and the position of the lever, and in terms of the action performed by the person). In order for the person to perform the action, they need to apply the appropriate force to the lever—so this is intended to reflect ability rather than disposition. However, there is something missing from these formal accounts, and that is the rationale for performing the action in the first place. One way of considering this is to turn to suggestions that "affordance" is hierarchical and can be described in terms of different levels.

Levels of Affordance

Although Abbate and Bass (2017) relate values for X and Z to an affordance relationship, they do not say how the affordance itself relates to a particular "goal," such as lift_lever or open_door. McGrenere and Ho (2000) use the term "possibility for action" to indicate that there might be levels of affordance. One way of thinking of this is in terms of "sequential affordance" (Gaver, 1996). For example, grasping a lever handle "affords" lifting, which then releases the door and, so "affords," opening the door. In this sequence, affordances are "nested," i.e., the lever's "graspability" is nested in the door's "openability." I am not convinced that it makes sense to call this a "sequence of affordances," so much as a sequence of actions, but can see how one could apply the formal descriptions outlined above to each "state" in the ongoing sequence of interactions between person and object. What is interesting about this perspective is that the "door_handle" contributes to several "affording situations." Consider, for example, turning the door handle when you were carrying a pile of books or a cup of coffee, as opposed to turning it with an unencumbered hand.

The notion that affordances could have multiple instances was also discussed by Hartson (2003) who suggested that affordances could be: cognitive, physical, sensory, with each of these helping users to perform cognitive, physical or sensory action. This seems to me to conflate different notions of "affordance" in ways that are not helpful. For instance, while affordance describes the relationship between the form of an object and the person's action, it is not obvious how this relates to cognitive and sensory actions. Similarly, Turner (2005) contrasted what he termed "simple affordance" (which draws on Gibson's definition) with "complex affordance" (which involves interpretation and response to an object's form in terms of the user's culture, history, praxis). However, applying the term "affordance" to such different behaviors can only serve to increase confusion. To this end, I proposed a different terminology to describe these different levels.

Forms of Engagement

In order to explore the concept of affordance further, and to make use of the suggestion that there are different levels of "affordance" that provide constraints of the ways in which we interact with objects, I developed the idea of forms of engagement (Baber, 2003, 2006). In this, the focus is on the ways in which we engage with objects and how different forms can serve to support and constrain each other. The most recent version of this concept is illustrated by **Figure 1**. The arrows are intended to indicate the relation "constrains." Note that, at the center of **Figure 1** is a dotted box which is labeled "affordance." This describes a relationship between the ability to recognize salient features in an object (Environmental Engagement) and the ability to act using that object (Motor Engagement).

Figure 1 separates the effectivity of the person, in terms of *Morphological Engagement*, from ability, in terms of motor engagement. There are two reasons for this: first, morphology is partly dispositional, e.g., in terms of the size of the hand; and second, hand shaping will be influenced by subsequent actions, e.g., when reaching to grasp an object, hand shape is modified in anticipation of the type of grip required to respond to properties of the object, such as weight, fullness, slipperiness etc. (Wing et al., 1986), and this will also be influenced by *Motor Engagement*, i.e., Rosenbaum et al. (1992) notion of "end-state comfort" explains why people might adopt an uncomfortable grip at the beginning of an action, in order to end an action with a comfortable grip. For example, if a wine glass is upside down on the table, you will probably twist the hand awkwardly to pick it up in order to turn it right-way up. So, there are a limited set of ways in which an object can be grasped by the human hand and the selection of grasp combines object properties with intended movements. That is, a hand of a given size will have limits of how it can grasp objects, but *how* the grasp is performed reflects the ability and intentions of the person, which will vary according to a host of situational factors, as well as prior experience.

In order to act on an object, there is a need to respond to the "information" that it conveys. I am using the word information in a Gibsonian sense, and apply the term *Environmental Engagement* to reflect this. Consequently, an affordance arises as the result of the relationship between *Environmental* and *Motor Engagement*. For example, people can make rapid judgements about whether to turn their body to fit through narrow apertures as they approach these (Warren and Whang, 1987) and can make such judgements even when their bodies have been modified to an unfamiliar size, e.g., when wearing "pregnancy packs" on the front (Franchak and Adolph, 2014), or when wearing rugby shoulder pads (Higuchi et al., 2011). Furthermore, increasing the weight of the body, e.g., by wearing a heavy rucksack, can
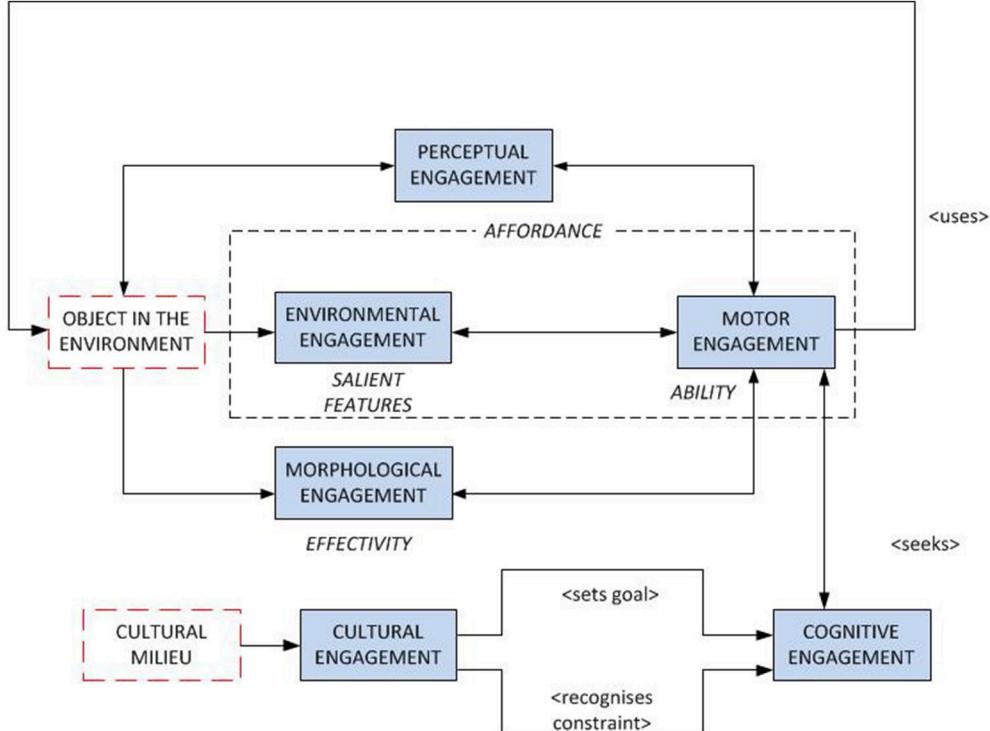
FIGURE 1 | Forms of Engagement (2017 version).

alter judgements of the steepness of a hill (Proffitt, 2006). The implication is that there is a "body-scaled" perception of some features of the environment that can guide some actions (Warren, 1984; Fajen, 2007). In other words, people are able to "see" aspects of the object, or the environment, in terms of an action that they both want, and are able, to perform. We can directly relate this proposal to Equation (2), e.g., imagine we are interested in stair-climbing, and the property of the world, Xp, is the height of a stair riser, and the property of the person, Zq, is their leg length.

This could, of course, be termed "perception-action coupling" (which is a common expression of Gibson's ideas and a reasonable explanation of affordance from the perspective adopted in this paper). So, I retain the term "affordance" for the specific relationship between object and action—and regard this as an emerging property of the world-object-person system. However, this relationship is bounded by the other forms of engagement. The suggestion that *Motor Engagement* is directed toward subsequent action implies an intention, but I argue that there is equal scope that the "intention" can be defined in response to the *Motor Engagement* (opportunistic or situated action). At the very least, there is a two-way exchange between the action-as-performed and the goal-state of that action. The role of *Cognitive Engagement* is to provide this high-level management on ongoing actions. Across the various forms of engagement, *Perceptual Engagement* relates salient features to changing state of the object-person system. Finally, the notion of an "acceptable" goal could relate to the culture in which one is acting. This *Cultural Engagement* relates to the idea of "complex affordance"

(Turner, 2005). It could also relate to the concept of "cultural affordances" developed by Ramstead et al. (2016).

The basic concept of Forms of Engagement is intended to retain "affordance" as a simple relationship between the actions a person performs to the features of the object that they are using. The connections between the different forms represent the constraints that shape and respond to this relationship. I claim that this provides a useful way of conceptualizing interaction, and use this to explore ways in which one can design animate objects.

Animate Objects

Having proposed that interaction comprises a number of Forms of Engagement, one can relate these to the possible inferences that animate objects could make as they are being interacted with. At the most basic level, sensors on the object could provide data to characterize the motion, orientation, position, etc. of the object. However, what would be most useful is not just identifying that a movement has been made but also to identify how well that movement has been made, e.g., has it been performed smoothly, hesitantly, with tremor etc. In this way, the object would be able to make inferences about the user's Motor Engagement and abilities. Additional sensing capability could be added to monitor hand shape and movement as it approaches the object, in order to make inferences concerning Morphological Engagement. This could be used to determine the type of action that the person might be intending to make, even before picking up or handling the object. Previously I have contrasted these as epistemic or ergotic

gestures, to reflect the fact that such actions could be treated as "gestures" which have the intention of altering the state of the user's environment (Baber, 2014).

The object, assuming that it can modify its appearance, could encourage Environmental Engagement through changes that emphasize specific features. So, when a handle rises on the side of a cup, people are more likely to use the hand on that side of the cup to pick it up (Baber et al., 2017). Having some knowledge of where the object is being used could also influence the definition of appropriate actions, through Cultural Engagement. Combining inferences drawn from Motor and Morphological Engagement, the object could infer the most likely intention of the person, and use this inference to provide additional cues and guidance (Jean-Baptiste et al., 2015).

Let us assume that the "smart object" looks like something familiar, say a cup, which has been fitted with sensors (Gellersen et al., 1999; Baber et al., 2017). On the one hand, this is an object that we "know" how to use, but on the other hand, this is an alien object that is capable of doing things that we do not, necessarily, fully understand. The cup could, for example, be part of a system that monitors our daily liquid intake and the system could have a "goal" of ensuring that we drink a specified quantity of liquid, or it might be part of a system that has the "goal" of reducing our caffeine intake. One way in which such "goals" could be communicated to the user would for the artifacts themselves (through lights, sounds, movement etc.) to provide feedback and prompts to the person. In this way, the form of the objects could display their function. I am interested in this relationship between form and function (both in terms of "normal" and "smart" objects), and how the "function" of an object corresponds to the action in which it is used. There are many instances in which the "action" is quite different from the designed "function," e.g., a laptop could be used to prop the leg of a wobbly desk, or as a tray to carry several coffee cups, or as a weapon.

Implications for Design

I close this paper with some observations on how the concept of Forms of Engagement could apply to broader areas of HCI design. There seems to me to be a division between those practitioners who are interested in usability and those interested in user experience (Baber, 2015). The "usability" focus tends to emphasize performance (although, of course, International Standards Organization definitions of usability include efficiency, effectiveness *and* experience), while "user experience" tends to focus on the emotional response (from pleasure to frustration) that users get from their interactions with technology. Broadly, I would suggest that usability takes as its "context of use," the region in **Figure 1** that is defined by Environmental, Motor, Morphological, Perceptual and Cognitive Engagement, while "user experience" takes as its focus the region in **Figure 1** that is defined by Cultural, Cognitive and Perceptual Engagement. Of course, I am not claiming that there is not overlap between these regions, but it seems to me that the differences in practice relate to the different levels of analysis that practitioners emphasize. It would, one hopes, be profitable and useful to merge these practices of evaluation of HCI.

A final point for this paper is that I do not believe that it is possible to "design affordance" into an object. This is the fundamental argument made in this paper. However, I do believe that it is possible to create affording situations—and that this is what good design has always sought to achieve. Knowing how a person with given ability would interact with an object to achieve a given goal in a given context is central to ISO definitions of Human-Centred Design. What I have offered in this paper is a conceptual framework that illustrates this goal, and relates it to an unambiguous interpretation of the concept of "affordance."

AUTHOR CONTRIBUTIONS

The author confirms being the sole contributor of this work and approved it for publication.

REFERENCES

Abbate, A. J., and Bass, E. J. (2017). "Modeling affordance using formal methods," in *Proceedings of the Humans Factors and Ergonomics Society 2017 Annual Meeting* (Santa Monica, CA: HFES), 723–727. doi: 10.1177/1541931213601666

Anderson, J. R. (1981). *Cognitive Skills and their Acquisition,* Hillsdale, NJ: LEA.

Baber, C., Khattab, A., Russell, M., Hermsdörfer, J., and Wing, A. (2017). Creating affording situations: coaching through animate objects. *Sensors* 17, 2308. doi: 10.3390/s17102308

Baber, C. (2003). *Cognition and Tool Use*. London: Taylor and Francis.

Baber, C. (2006). Cognitive aspects of tool use. *Appl. Ergon.* 37, 3–15. doi: 10.1016/j.apergo.2005.06.004

Baber, C. (2014). "Objects as agents: how ergotic and epistemic gestures could benefit gesture-based interaction," in CHI'14 Workshop on Gesture-based Interaction Design: communication and cognition (Toronto).

Baber, C. (2015). "Evaluating human computer interaction," in Evaluating Human Work, eds J. R. Wilson and S. Sharples (Boca Raton, FL: CRC Press), 359–381.

Berry, D. C., and Broadbent, D. E. (1988). Interactive tasks and the implicit-explicit distinction, *Br. J. Psychol.* 79, 251–272.

Chemero, A., Klein, C., and Cordeiro, W. (2003). Events as changes in the layout of affordances. *Ecol. Psychol.* 15, 19–28. doi: 10.1207/S15326969ECO1501_02

Chemero, A. (2003). An outline for a theory of affordances. *Ecol. Psychol.* 15, 181–195. doi: 10.1207/S15326969ECO1502_5

Chemero, A. (2009). *Radical Embodied Cognitive Science*. Cambridge, MA: MIT Press.

Fairlie, J. E., and Barham, L. S. (2016). From Chaine-Operatoire to Observational Analysis: a pilot study of a new methodology for analysing changes in cognitive task-structuring strategies across different Hominin tool-making events, *Cambridge Archaeol. J.* 26, 643–664. doi: 10.1017/S0959774316000469

Fajen, B. R. (2007). Affordance-based control of visually guided action. *Ecol. Psychol.* 19, 383–410. doi: 10.1080/10407410701557877

Fitts, P. M., and Deininger, R. L. (1954). S-R compatibility: Correspondence among paired elements within stimulus and response codes. *J. Exp. Psychol.* 48, 483–492. doi: 10.1037/h0054967

Fitts, P. M., and Seeger, C. M. (1953). S-R compatibility: Spatial characteristics of stimulus and response codes. *J. Exp. Psychol.* 46, 199–210. doi: 10.1037/h0062827

Franchak, J. M., and Adolph, K. E. (2014). Gut estimates: pregnant women perceive possibilities for walking through doorways, *Attention Percept. Psychophys.* 76, 460–472. doi: 10.3758/s13414-013-0578-y

Gaver, W. (1991). *Technology Affordances, CHI'91*, NewYork, NY: ACM.

Gaver, W. (1996). Affordances for interaction: the social is material for design. *Ecol. Psychol.* 8, 111–129. doi: 10.1207/s15326969eco0802_2

Gellersen, H.-W., Beigl, M., and Krull, H. (1999). "The MediaCup: Awareness technology embedded in an everyday object," in *Handheld and Ubiquitous Computing 1st International Symposium HUC'99* (Berlin: Springer).

Gibson, J. J. (1977). "The theory of affordance," in *Perceiving, Acting, and Knowing: Toward an Ecological Psychology,* eds R. E. Shaw and J. Bransford (Hillsdale, NJ: LEA), 62–82.

Gibson, J. J. (1979). *The Ecological Approach to Visual Perception*, Hillsdale, NJ: LEA.

Hartson, R. (2003). Cognitive, physical, sensory, and functional affordances in interaction design. *Behav. Inform. Technol.* 22, 315–338. doi: 10.1080/01449290310001592587

Higuchi, T., Murai, G., Kijima, A., Seya, Y., Wagman, J. B., and Imanaka, K. (2011). Athletic experience influences shoulder rotations when running through apertures. *Hum. Mov. Sci.* 30, 534–539. doi: 10.1016/j.humov.2010.08.003

Humphreys, G. W., Yoon, E. Y., Kumar, S., Lestou, V., Kitadono, K., and Roberts, K. L. (2010). The interaction of attention and action: from seeing action to acting on Perception. *Br. J. Psychol.* 101, 185–206. doi: 10.1348/000712609X458927

Humphreys, G. W. (2001). Objects, affordances, action. *Psychologist* 14, 408–412.

Jean-Baptiste, E. M., Rotshtein, P., and Russell, M. (2015). "POMDP based action planning and human error detection," in *IFIP International Conference on Artificial Intelligence Applications and Innovations* (Berlin: Springer International Publishing), 250–265. doi: 10.1007/978-3-319-23868-5_18

Kaufer, S., and Chemero, A. (2015). *Phenomenology: An Introduction.* New York, NY: Wiley.

Kirlik, A. (2004). On Stoffregen's definition of affordance. *Ecol. Psychol.* 16, 73–77. doi: 10.1207/s15326969eco1601_10

Kirsh, D. (2013). Embodied cognition and the magical future of interaction design. *ACM Trans. Comput. Hum. Interact.* 20, 1–30. doi: 10.1145/2442106.2442109

Koffka, K. (1955). *Principles of Gestalt Psychology.* London: Routledge and Kegan Paul.

Kornbulm, S., Hasbrouq, T., and Osman, A. (1990). Dimensional overlap: cognitive basis for stimulus-response compatibility – a model and taxonomy. *Psychol. Rev.* 97, 253–270. doi: 10.1037/0033-295X.97.2.253

Kortuem, G., Kawsar, F., Fitton, D., and Sundramoorthy, V. (2010). *Smart objects as building blocks for the Internet of Things, IEEE Internet Computing.* Washington, DC: IEEE Computer Society.

Lewin, K. (1936). *Principles of Topological Psychology,* New York, NY: McGraw-Hill.

McGrenere, J., and Ho, W. (2000). *Affordances: clarifying and evolving a concept.* New York, NY: ACM.

Merleau-Ponty, M. (1945). *Phenomenology of Perception.* London: Routledge.

Norman, D. A. (1988). *The Psychology of Everyday Things.* New York, NY: Basic Books.

Norman, D. A. (1999). Affordance, conventions, and design. *Interactions* 6, 38–43. doi: 10.1145/301153.301168

Norman, D. (2008). Signifiers, not affordances. *Interactions* 15, 18–19. doi: 10.1145/1409040.1409044

Osiurak, F., Rossetti, Y., and Badets, A. (2017). What is an affordance? 40 years later. *Neurosci. Biobehav. Rev.* 77, 403–417. doi: 10.1016/j.neubiorev.2017.04.014

Polyani, M. (1966). *The Tacit Dimension,* Chicago, IL: University of Chicago Press.

Proctor, R. W., and Vu, K. P. L. (2016). Principles for designing interfaces compatible with human information processing. *Int. J. Hum. Comput. Inter.* 32, 2–22. doi: 10.1080/10447318.2016.1105009

Profitt, D. R. (2006). Embodied perception and the economy of action. *Perspect. Psychol. Sci.* 1, 110–122. doi: 10.1111/j.1745-6916.2006.00008.x

Ramstead, M. J. D., Veissiere, S. P. L., and Kirmayer, L. J. (2016). Cultural affordances: scaffolding local worlds through shared intentionality and regimes of attention. *Front. Psychol.* 7:1090. doi: 10.3389/fpsyg.2016.01090

Rosenbaum, D. A., Vaughan, J., Barnes, H. J., and Jorgensen, M. J. (1992). Time course of movement planning: selection of handgrips for object manipulation. *J. Exp. Psychol. Learn. Mem. Cogn.* 18: 1058. doi: 10.1037/0278-7393.18.5.1058

Roux, V., and Bril, B. (2005). General introduction: a dynamic systems framework for studying a uniquely Hominin behaviour. *J. Archaeol. Sci. Rep.* 5, 225–239.

Schneider, W., and Shiffrin, R. M. (1977). Controlled and automatic human information processing: I. Detection search, and attention. *Psychol. Rev.* 84, 1–66. doi: 10.1037/0033-295X.84.1.1

Shiffrin, R., and Schneider, W. (1977). Controlled and automatic human information processing: II. Perceptual learning, automatic attending, and a general theory. *Psychol. Rev.* 84, 127–190. doi: 10.1037/0033-295X.84.2.127

Still, J. D., and Dark, V. J. (2013). Cognitively describing and designing affordances. *Design Stud.* 34, 285–301. doi: 10.1016/j.destud.2012.11.005

Stins, J. F., and Michaels, C. F. (1997). Stimulus-response compatibility is information-action compatibility. *Ecol. Psychol.* 9, 25–45. doi: 10.1207/s15326 969eco0901_2

Stoffregen, T. A. (2003). Affordances as properties of the animal-environment system. *Ecol. Psychol.* 15, 115–134. doi: 10.1207/S15326969ECO1502_2

Thaler, R. H., and Sunstein, C. R. (2008). *Nudge: Improving Decisions about Health, Wealth and Happiness.* London: Penguin Books.

Turner, P. (2005). Affordance as context. *Interact. Comput.* 17, 787–800. doi: 10.1016/j.intcom.2005.04.003

Turvey, M. T., and Shaw, R. (1979). "The primacy of perceiving: an ecological reformulation of perception for understanding memory," in *Perspectives on Memory Research: Essays in Honor of Uppsala University's 500th Anniversary*, ed L.-G. Nilsson (Hillsdale, NJ: Lawrence Erlbaum Associates), 167–222.

Turvey, M. T. (1992). Affordances and prospective control: an outline of the ontology. *Ecol. Psychol.* 4, 173–187. doi: 10.1207/s15326969eco0403_3

Warren, W. H., and Whang, S. (1987). Visual guidance of walking through apertures: body-scaled information for affordances. *J. Exp. Psychol. Hum. Percept. Perform.* 13, 371–383. doi: 10.1037/0096-1523.13.3.371

Warren, W. H. (1984). Perceiving affordances: visual guidance of stair climbing. *J. Exp. Psychol. Hum. Percept. Perform.* 10, 683–703. doi: 10.1037/0096-1523.10.5.683

Welford, A. T. (1976). *Skilled Performance. Perceptual and Motor Skills.* Glenview, IL: Foreman.

Wing, A. M., Turton, A., and Fraser, C. (1986). Grasp size and accuracy of approach in reaching. *J. Motor Behav.* 18, 245–260. doi: 10.1080/00222895.1986.10735380

Wittgenstein, L. (1958). *Philosophical Investigations,* New York: Basil Blackwell.

Check for updates

Understanding and Resolving Failures in Human-Robot Interaction: Literature Review and Model Development

Shanee Honig* and Tal Oron-Gilad

Mobile Robotics Laboratory and HRI Laboratory, Department of Industrial Engineering and Management, Ben-Gurion University of the Negev, Beer Sheva, Israel

While substantial effort has been invested in making robots more reliable, experience demonstrates that robots operating in unstructured environments are often challenged by frequent failures. Despite this, robots have not yet reached a level of design that allows effective management of faulty or unexpected behavior by untrained users. To understand why this may be the case, an in-depth literature review was done to explore when people perceive and resolve robot failures, how robots communicate failure, how failures influence people's perceptions and feelings toward robots, and how these effects can be mitigated. Fifty-two studies were identified relating to communicating failures and their causes, the influence of failures on human-robot interaction (HRI), and mitigating failures. Since little research has been done on these topics within the HRI community, insights from the fields of human computer interaction (HCI), human factors engineering, cognitive engineering and experimental psychology are presented and discussed. Based on the literature, we developed a model of information processing for robotic failures (Robot Failure Human Information Processing, RF-HIP), that guides the discussion of our findings. The model describes the way people perceive, process, and act on failures in human robot interaction. The model includes three main parts: (1) communicating failures, (2) perception and comprehension of failures, and (3) solving failures. Each part contains several stages, all influenced by contextual considerations and mitigation strategies. Several gaps in the literature have become evident as a result of this evaluation. More focus has been given to technical failures than interaction failures. Few studies focused on human errors, on communicating failures, or the cognitive, psychological, and social determinants that impact the design of mitigation strategies. By providing the stages of human information processing, RF-HIP can be used as a tool to promote the development of user-centered failure-handling strategies for HRIs.

Keywords: human-robot interaction, failure, user-centered, information processing, context

INTRODUCTION

While substantial effort has been invested in making robots more reliable, experience demonstrates that robots are often challenged by frequent failures. The Mean Time Between Failure (MTBF) for robots in field environments is often within a few hours (Tsarouhas and Fourlas, 2016). Despite this, mobile robots have not yet reached a level of design that allow effective management of faulty or unexpected behavior. In fact, research suggests that the relationship between symptoms and cause of failure is often not clear even to trained roboticists (Steinbauer, 2013). Having to rely on a professional to understand and resolve a robot's faulty behavior is a barrier to acceptance amongst untrained users. Customer support also becomes costly when users are unable to differentiate between technical errors (software bugs or hardware failures) and problems resulting from improper use (misuse; Parasuraman and Riley, 1997) or unrealistic expectations. Moreover, how a robot manages failure influences willingness to use the robot again (Lee et al., 2010), the degree of deterioration in task performance (Ragni et al., 2016), user trust in the robot (Hamacher et al., 2016), and people's perceptions of the robot (Gompei and Umemuro, 2015), suggesting that failure handling may have substantial commercial and economic benefits. Yet, little is known about how to create failure management tools for robots that are appropriate for untrained users. We shed light on this topic, with the goal of developing design tools and design guidelines that facilitate development of robot interactions that enable untrained users to quickly and easily identify and act on failures, while maintaining a positive user experience.

To tackle the challenging problem of failure handling for untrained users, it is first necessary to review the cognitive considerations that critically influence naive users' ability to detect and solve robot failures, and evaluate whether these considerations have been properly addressed in the existing Human-Robot Interaction (HRI) literature. This paper presents a detailed look at the literature in HRI regarding when people perceive and resolve robot failures, how robots communicate failure, how failures influence people's perceptions and feelings toward robots, and how these effects can be mitigated. Since little research has been done on these topics within the HRI community, insights from the fields of Human Computer Interaction (HCI), human factors engineering, cognitive engineering and experimental psychology are presented and discussed. To the best of our knowledge, a thorough review of robotic failure handling from a user-centered perspective has not yet been conducted. Based on the literature, we developed a model of information processing for robotic failures (the Robot Failure Human Information Processing Model, RF-HIP) that guides the discussion of our findings. As robots become more present in day-to-day life, especially for elderly users who are inexperienced with robotic applications (Beer and Takayama, 2011), we anticipate that such reviews and models will become increasingly useful. Researchers could use them to better understand what influences failure handling in HRIs, to identify possible knowledge gaps and to promote future research directions. Roboticists, engineers, and designers could use them to guide design choices that will increase user acceptance and decrease customer support costs. Policy makers could use them to decide on standards for the necessary failure-handling techniques required to make robots safe for general use.

The paper is organized as follows: first, the types of failures that may occur during HRIs are discussed. Second, search criteria and an overview of the relevant HRI literature that matched these criteria is presented. Third, cognitive determinants that are likely to influence a person's ability to perceive and resolve failures are combined with current research in robotic user-centered failure handling to create a model of information processing. Finally, gaps in the HRI literature are presented and discussed.

DEFINING AND CLASSIFYING ERRORS

Various definitions exist for the terms "failure," "error," and "fault." In line with (Laprie, 1995; Carlson and Murphy, 2005; Steinbauer, 2013; Brooks, 2017), we adopted terminology in which *failure* refers to "a degraded state of ability which causes the behavior or service being performed by the system to deviate from the ideal, normal, or correct functionality" (Brooks, 2017). This definition includes both perceived failures, unexpected behavior and actual failures, which is consistent with findings that suggest that intentional yet unexpected or incoherent behaviors are sometimes interpreted as erroneous (Short et al., 2010; Lemaignan et al., 2015). Failures result from one or more *errors*, which refer to system states (electrical, logical, or mechanical) that can lead to a failure. Errors result from one or more *faults*, which refer to anything that causes the system to enter an error state. For example, a robot may experience a *failure* resulting from an *error* in face-recognition, caused by poor illumination (*fault*).

It is improbable to identify all possible types of robotic failures since mobile robots operate in unstructured changing environments with a wide variety of possible interactions. Yet, several taxonomies for classifying errors and failures have been proposed. Laprie (1995) classified failures according to severity, defining *benign failures* (failures whose consequences are comparable to the benefits of the service they are preventing) and *catastrophic failures* (failures with a higher cost by one or more orders of magnitude than the service). Ross (Ross et al., 2004) categorized system errors according to failure recoverability, defining *anticipated errors* (when the agent backtracks through the plan to achieve the same goal through an alternate course of action), *exceptional errors* (when the current plan cannot cope with the failure, and re-planning can be done to formulate a strategy to achieve the original goal), *unrecoverable errors* (when the current plan cannot cope with the error and re-planning cannot be done), and *socially recoverable errors* (when the agent can continue on with the original plan with appropriate assistance from other agents within its environment). Giuliani et al. (2015) classified failures according to their type, defining *technical failures* (caused by technical shortcomings of the robot) and *social norm violations* (when

the robot deviates from the social script or uses inappropriate social signals, e.g., looking away from a person while talking to them).

Carlson and Murphy (2005) devised an extensive error classification taxonomy by analyzing how Unmanned Ground Vehicles (UGVs) failed in the field using studies from urban search and rescue and military field applications. The classification, based on Laprie (1995) and Norman (2002) categorized errors according to the source of failure (the fault), and included two main categories: (1) *physical failures*, which are failures caused by physical errors in the system's effectors, sensors, control system, power sources, or communications, and (2) *human failures*, which are caused by human-made errors. They further classified physical failures according to *severity* (*terminal failure*—terminates the system's current mission; *nonterminal failures*—degrades its ability to perform its mission) and *repairability* (*field repairable*—repairable with tools that accompany the system in the field; *nonfield repairable*—cannot be repaired with tools that accompany the system in the field), and human failures according to *design failures* (errors introduced during design, construction, or post-production modifications, e.g., programmed to greet people with "goodbye") and *interaction failures* (errors introduced by unintended violations of operating procedures). Interaction failures included *mistakes* (performing an action that is wrong), and *slips* [attempting to do the right thing unsuccessfully, e.g., accidentally pressing the wrong button (Barakova et al., 2015)].

While the (Carlson and Murphy, 2005) taxonomy is extensive, there are additional interaction failures that were not accounted for. For example, it did not consider other types of human errors, such as *lapses*, which occur as a result of lapses of memory and/or attention (e.g., forgetting to turn the robot off), and *deliberate violations*, which are intentional illegitimate actions (e.g., directing the robot to run into a wall) (Reason, 1990). Three main taxonomies of human errors are frequently cited in the literature (Stanton and Salmon, 2009): (1) Norman's error categorization (Norman, 1981), which divides human errors into those that result from misinterpretations of the situation, those that result from faulty activation of schemas (knowledge structures) due to similar trigger conditions, and those that result from activating schemas too early, too late, or not at all; (2) Rasmussen's error categorization (Rasmussen, 1982), which divides human errors by the level of cognitive control within which they occur (skill-, rule-, or knowledge-based), and (3) Reason's categorization (Reason, 1990), which builds on Rassmussen's ideas and divides human errors into slips, lapses, mistakes and violations (described above). Moreover, the (Carlson and Murphy, 2005) taxonomy doesn't consider uncertainties in the interaction that result from varying environments and other agents. (Sutcliffe and Rugg, 1998) described 10 environmental and social factors that may increase the likelihood of errors, and classified them into *group level judgement*, *working environment*, and *organizational flaws*.

Steinbauer (2013) collected information regarding failures that occurred to teams in RoboCup competitions, and classified them into four categories: *Interaction* (problems that arise from uncertainties in the interaction with the environment, other agents, and humans), *algorithms* (problems in methods and algorithms), *software* (design and implementation faults of software systems), and *hardware* (physical faults of the robotic equipment). They used several attributes to classify faults and their properties, including the fault's relevance to different robotic systems (*relevance*), the context in which the fault occurred (*condition*), indicators used to identify the failure (*symptoms*), how the failure impacted the mission (*impact*: non-critical, repairable, and terminal), and the frequency of the occurrence of a fault (*frequency*: never, sporadic, regularly, frequently).

Brooks (2017), based on Lutz and Woodhouse (1999), identified two main types of failure: communication failures and processing failures. Communication failures are related to data being passed between modules, including *missing data* (incomplete messages or dropped packets), *incorrect data* (data generated incorrectly or distorted during transmission), *bad timing* (data sent too early, before the receiver is ready to handle it, or too late, causing delays in reaction), and *extra data* (data sent multiple times but only expected once, or sending larger messages than expected). Processing failures include *abnormal terminations*, that could happen due to unhandled exceptions, segmentation fault, or dead-lock; *missing events*, that could happen when a conditional statement is not triggered or a callback or interrupt never fires; *incorrect logic* due to bad assumptions or unforeseen conditions; and *timing or ordering*, where events take place in a different order than expected or a waiting period times-out before information arrives.

We propose an inclusive human-robot failure taxonomy that combines the above system and human oriented classifications (**Figure 1**). According to this taxonomy, the main distinction is between two types of failures: *technical failures* and *interaction failures*. *Technical failures* are caused either by hardware errors or problems in the robot's software system. Software errors are further classified into design failures, communication failures, and processing failures. Following Steinbauer's categorization (Steinbauer, 2013), *interaction failures* refer to problems that arise from uncertainties in the interaction with the environment, other agents, and humans. These include social norm violations and various types of human errors as noted in Reason (1990). Each failure event, regardless of its source, can be categorized by the following attributes:

- *Functional Severity:* criticality of the failure to the robot's functioning (non-critical, recoverable, terminal).
- *Social Severity:* criticality of the failure to future acceptance of the robot's services (non-critical, recoverable, unrecoverable).
- *Relevance:* relevance of the fault to different robot systems, which can be high (relevant to almost all robotic systems), medium (relevant only to some robotic systems), or low (highly specialized failures).
- *Frequency:* how often the failure occurs (never, sporadic, regularly, frequently).
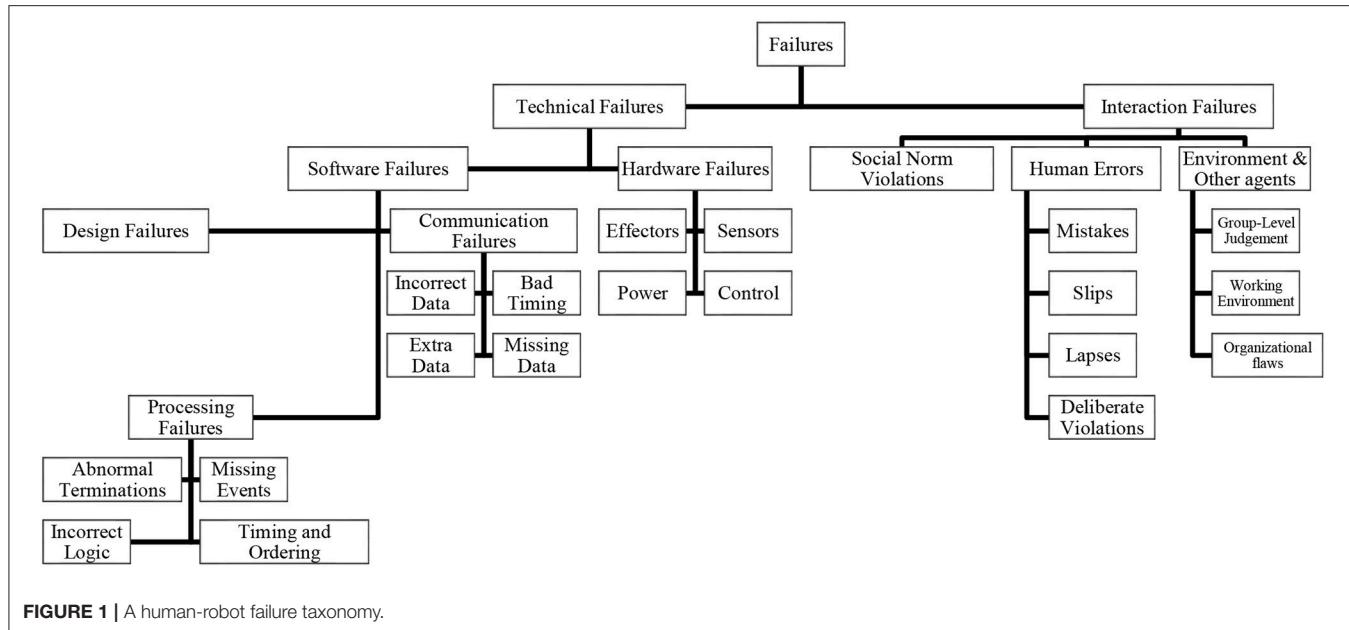
FIGURE 1 | A human-robot failure taxonomy.

- *Condition:* the context in which the fault and failure occurred.
- *Symptoms*: indicators used to identify the failure.

LITERATURE REVIEW ON USER-CENTERED FAILURE HANDLING

Various search engines were used to conduct the online literature search on human-centered failure handling in robots, including Google Scholar, IEEE, ACM, Science Direct, Springer, Sage Journals, Taylor & Francis Online, and Cambridge Core. Robotics conferences and journals covered in this search include ICRA, IROS, RO-MAN, SMC, Robotics and Autonomous Systems, Human Machine Systems, HRI, International Journal of Social Robotics, Autonomous Robots, International Journal of Robotics Research, Robotica, Intelligent Robots and Systems, and Advanced Robotics, amongst others. Keywords used were: robot, error, failure, recovery, reliability. Included in the review are articles that address robotic failure-handling from the perspective of the human operator, user or bystander, rather than from a systems perspective. That is, we focused on studies that evaluated some aspect of the bilateral relationship between end-user's needs, wants and limitations and robotic failure. Articles that dealt with errors without addressing the user or the interaction were not included in the review. Given the vast amount of research on technical considerations of robot reliability and error handling, we cannot claim our search to be exhaustive, however given the large number of resources surveyed, we do believe it is indicative of current trends.

Figure 2 shows the result of the literature search of HRI articles that evaluated some aspect of user-centered failure handling. Altogether, 52 relevant papers were identified, where 40 of them were published in conference proceedings, 8 in academic journals, 1 doctoral dissertation, 2 theses, and 1 technical report.

Papers were classified into three main topics: (a) communicating failures and their causes, i.e., how should a robot communicate to its user and bystanders that an error has occurred; (b) the influence of failures on HRI, i.e., how do failures influence user perceptions of the robot and user behavior; and (c) mitigating failures, i.e., approaches on how to mitigate the negative effects of failure on HRIs. The following sections provide an overview of methodologies used in the literature, including the types of errors and symptoms studied, evaluation methods and metrics, the types of robotic systems used, and experimental environments.

Errors and Symptoms Studied

Almost all errors researched in the literature exemplified technical failures (e.g., Gieselmann, 2006; Kim and Hinds, 2006; Gieselmann and Ostendorf, 2007; Spexard et al., 2008; Kim et al., 2009; Groom et al., 2010; Lee et al., 2010; Takayama et al., 2011; Desai et al., 2012, 2013; Kahn et al., 2012; Rosenthal et al., 2012; Shiomi et al., 2013; Yasuda and Matsumoto, 2013; Kaniarasu and Steinfeld, 2014; Lohan et al., 2014; Cha et al., 2015; Gehle et al., 2015; Giuliani et al., 2015; Gompei and Umemuro, 2015; Hamacher, 2015; Knepper et al., 2015; Mirnig et al., 2015, 2017; Mubin and Bartneck, 2015; Salem et al., 2015; Bajones et al., 2016; Brooks et al., 2016; Hamacher et al., 2016; Hayes et al., 2016; Ragni et al., 2016; Robinette et al., 2016; Engelhardt and Hansson, 2017; Law et al., 2017; Sarkar et al., 2017; van der Woerdt and Haselager, 2017; Kwon et al., 2018). Only a few evaluated the impact of social norm violations (e.g., Short et al., 2010; Salem et al., 2013; Giuliani et al., 2015; Mirnig et al., 2015, 2017; van der Woerdt and Haselager, 2017), and none focused on human errors. Some articles did not specify the type of error used (e.g., Ross et al., 2004; Cassenti, 2007).

A robot's failure symptoms in the literature include the robot not completing a given task (e.g., Takayama et al., 2011; Rosenthal et al., 2012; Brooks et al., 2016; Robinette et al., 2016; Mirnig et al.,

Distribution of user-centered failure handling by topic (Top)

Communicating failures and their causes: Cha et al., 2015; Brooks 2017; Kwon et al., 2018

The influence of failures on human-robot interaction: Giuliani et al., 2015; Mirnig et al., 2015; Mirnig et al., 2017; Brooks et al., 2016; Brooks 2017; Gompei & Umemuro; Hayes et al., 2016; Lee et al., 2010; Gieselmann 2006; Ragni et al., 2016; Salem et al., 2013; Salem et al., 2015; Short et al., 2010; Desai et al., 2012; Desai et al., 2013; Kim et al., 2009; Kim et al., 2017; Schütte et al., 2017; De Visser & Parasuraman; Bajones et al., 2016; Sarkar et al., 2017; Kahn et al., 2012; Mubin & Bartneck; Hamacher 2015; Hamacher et al., 2016; Adubor et al., 2017; Robinette et al., 2016; Lucas et al., 2018; van der Woerdt & Haselager; Yasuda & Matsumoto; Law et al., 2017; Rossi et al., 2017a; Rossi et al., 2017b; Lemaignan et al., 2015; Lucas et al., 2017; Takayama et al., 2011; Gehle et al., 2015

Mitigating failures: Cassenti 2007; Spexard et al., 2008; Brooks et al., 2016; Brooks 2017; Desai et al., 2013; Rosenthal et al., 2012; Groom et al., 2010; Knepper et al., 2015; Kaniarasu & Steinfeld; Yasuda & Matsumoto; Bajones et al., 2016; Shiomi et al., 2013; Hamacher 2015; Hamacher et al., 2016; Lohan et al., 2014; Engelhardt & Hansson; Kaniarasu et al., 2013; Schütte et al., 2017; Kim & Hinds 2006; Drury et al., 2003; Lucas et al., 2017; Lucas et al., 2018; Lee et al., 2010; Gieselmann & Ostendorf

Distribution by publication year (Bottom) — boxes per year (bottom to top):

- 2003: Drury et al.
- 2006: Gieselmann; Kim & Hinds
- 2007: Gieselmann & Ostendorf; Cassenti
- 2008: Spexard et al.
- 2009: Kim et al.
- 2010: Lee et al.; Short et al.; Groom et al.
- 2011: Takayama et al.; De Visser & Parasuraman
- 2012: Rosenthal et al.; Desai et al.; Kahn et al.
- 2013: Salem et al.; Yasuda & Matsumoto; Desai et al.; Shiomi et al.; Kaniarasu et al.
- 2014: Kaniarasu & Steinfeld; Lohan et al.
- 2015: Giuliani et al.; Mirnig et al.; Gehle et al.; Gompei & Umemuro; Knepper et al.; Salem et al.; Hamacher; Mubin & Bartneck; Cha et al.; Lemaignan et al.
- 2016: Brooks et al.; Hayes et al.; Ragni et al.; Robinette .et al; Bajones et al.; Hamacher et al.
- 2017: Mirnig et al.; Engelhardt & Hansson; van der Woerdt & Haselager; Adubor et al.; Brooks; Schütte et al.; Law et al.; Rossi et al (b); Rossi et al (a); Lucas et al.; Kim et al.; Sarkar et al.
- 2018: Kwon et al.; Lucas et al.

FIGURE 2 | Distribution of user-centered failure handling by topic (**Top**) and by publication year (**Bottom**).

2017; Kwon et al., 2018), running into obstacles (e.g., Brooks et al., 2016), performing the wrong action (e.g., Kim et al., 2009; Lee et al., 2010; Desai et al., 2012, 2013; Yasuda and Matsumoto, 2013; Kaniarasu and Steinfeld, 2014; Gehle et al., 2015; Mubin and Bartneck, 2015; Salem et al., 2015; Brooks et al., 2016; Hayes et al., 2016; Robinette et al., 2016; Mirnig et al., 2017; Sarkar et al., 2017; van der Woerdt and Haselager, 2017), performing the right action incorrectly or incompletely (e.g., Takayama et al., 2011; Shiomi et al., 2013; Cha et al., 2015; Hamacher, 2015; Brooks et al., 2016; Hamacher et al., 2016; Adubor et al., 2017; Sarkar et al., 2017; van der Woerdt and Haselager, 2017; Kwon et al., 2018), producing no action or speech (irresponsiveness) (e.g., Gieselmann, 2006; Lohan et al., 2014; Bajones et al., 2016; Robinette et al., 2016; Lucas et al., 2017, 2018), timing speech improperly (e.g., Mirnig et al., 2017), failing to produce speech (e.g., Gieselmann and Ostendorf, 2007; Mirnig et al., 2017), producing inappropriate speech or erroneous instruction (e.g., Gieselmann, 2006; Gieselmann and Ostendorf, 2007; Short et al., 2010; Gehle et al., 2015; Gompei and Umemuro, 2015; Lucas et al., 2017, 2018; Mirnig et al., 2017; Sarkar et al., 2017), repeating statements or body movements (e.g., Gieselmann and Ostendorf, 2007; Spexard et al., 2008; Lucas et al., 2017; Kwon et al., 2018), producing unexpected or erratic behavior (e.g., Kim and Hinds, 2006; Spexard et al., 2008; Short et al., 2010; Desai et al., 2012; Salem et al., 2013, 2015; Lemaignan et al., 2015; Robinette et al., 2016; van der Woerdt and Haselager, 2017), making knowledge-based mistakes (e.g., Groom et al., 2010; Short et al., 2010; Kahn et al., 2012; Rosenthal et al., 2012; Salem et al., 2015; Hayes et al., 2016; Ragni et al., 2016; Engelhardt and Hansson, 2017; Law et al., 2017), overtly stating there is a problem (e.g., Spexard et al., 2008; Bajones et al., 2016; Lucas et al., 2018), asking for help (e.g., Ross et al., 2004; Hüttenrauch and Severinson-Eklundh, 2006; Spexard et al., 2008; Rosenthal et al., 2012; Yasuda and Matsumoto, 2013; Knepper et al., 2015; Bajones et al., 2016; Srinivasan and Takayama, 2016), producing body language associated with failure (e.g., Takayama et al., 2011), and questioning for additional information (e.g., Gieselmann, 2006; Lucas et al., 2018).

Evaluation Methods and Metrics

Error recovery strategies and reactions to errors have been evaluated using surveys (e.g., Lee et al., 2010; Takayama et al., 2011; Cha et al., 2015; Brooks et al., 2016; Adubor et al., 2017; Kim et al., 2017; Rossi et al., 2017b; van der Woerdt and Haselager, 2017; Kwon et al., 2018), video analysis of HRIs (e.g., Giuliani et al., 2015; Mirnig et al., 2015), and unstructured observational

studies (e.g., Gieselmann, 2006; Gehle et al., 2015), however most studies used controlled user experiments (e.g., Spexard et al., 2008; Short et al., 2010; Desai et al., 2013; Salem et al., 2013, 2015; Gompei and Umemuro, 2015; Knepper et al., 2015; Hayes et al., 2016; Ragni et al., 2016; Robinette et al., 2016; Mirnig et al., 2017; Lucas et al., 2018). One study introduced an idea on how to improve situation awareness (SA; see Comprehension and Memory section) in erroneous situations without any formal evaluation (Cassenti, 2007).

User perceptions of the robot that have been evaluated in erroneous situations include the robot's perceived agency (Lemaignan et al., 2015; van der Woerdt and Haselager, 2017), predictability (van der Woerdt and Haselager, 2017), apologeticness (Shiomi et al., 2013), moral accountability (Kahn et al., 2012), friendliness (Groom et al., 2010; Shiomi et al., 2013; Kim et al., 2017), propensity to damage (van der Woerdt and Haselager, 2017), trustworthiness (Gompei and Umemuro, 2015; Brooks et al., 2016; Hamacher et al., 2016; Rossi et al., 2017a; Sarkar et al., 2017; van der Woerdt and Haselager, 2017; Kwon et al., 2018), likeability (Groom et al., 2010; Salem et al., 2013; Bajones et al., 2016; Engelhardt and Hansson, 2017; Mirnig et al., 2017; Sarkar et al., 2017), reliability (Short et al., 2010; Salem et al., 2015), familiarity (Gompei and Umemuro, 2015), anthropomorphism (Lee et al., 2010; Salem et al., 2013, 2015; Lemaignan et al., 2015; Mubin and Bartneck, 2015; Mirnig et al., 2017; Sarkar et al., 2017), animacy (Engelhardt and Hansson, 2017; Sarkar et al., 2017), technical competence (Groom et al., 2010; Short et al., 2010; Desai et al., 2013; Salem et al., 2015; Brooks et al., 2016; Engelhardt and Hansson, 2017; Sarkar et al., 2017), dependability (Brooks et al., 2016), intelligence (Mubin and Bartneck, 2015; Salem et al., 2015; Bajones et al., 2016; Engelhardt and Hansson, 2017; Mirnig et al., 2017; Sarkar et al., 2017), belligerence (Groom et al., 2010) and safety (Salem et al., 2015; Adubor et al., 2017; Sarkar et al., 2017). Studies have also evaluated the effects of errors on engagement (Lemaignan et al., 2015; Law et al., 2017), future contact intensions with the robot (Short et al., 2010; Salem et al., 2013, 2015; Brooks et al., 2016; Robinette et al., 2016; Kwon et al., 2018), the robot being a good teammate (Kwon et al., 2018), psychological closeness with the robot (Salem et al., 2015; Sarkar et al., 2017), rapport and persuasion (Lucas et al., 2018), creating a shared reality (Salem et al., 2013), compliance (Rosenthal et al., 2012; Salem et al., 2015; Robinette et al., 2016; Mirnig et al., 2017), attitudes toward robots (Salem et al., 2013; Gompei and Umemuro, 2015; Kim et al., 2017; Sarkar et al., 2017), and participant's emotional state (e.g., comfortable, safe, relaxed, confused) (Groom et al., 2010; Yasuda and Matsumoto, 2013; Hamacher, 2015; Robinette et al., 2016).

The quality of error recovery and communication strategies have been evaluated using various performance metrics, including whether users managed to resolve the problems (Spexard et al., 2008), attribution of blame (Kim and Hinds, 2006), the frequency of use of recovery feature (Spexard et al., 2008), the number of error-free user interactions (Gieselmann and Ostendorf, 2007; Knepper et al., 2015), time per repair (Rosenthal et al., 2012; Knepper et al., 2015; van der Woerdt and Haselager, 2017), time until task completion (De Visser and Parasuraman, 2011; Rosenthal et al., 2012; Schütte et al.,

2017), user comfort (Engelhardt and Hansson, 2017), user satisfaction (Gieselmann and Ostendorf, 2007; Shiomi et al., 2013), task performance and completion (Gieselmann and Ostendorf, 2007; De Visser and Parasuraman, 2011; Desai et al., 2013; Salem et al., 2013; Knepper et al., 2015; Brooks, 2017; Schütte et al., 2017), workload (Brooks, 2017), confidence (De Visser and Parasuraman, 2011; Brooks, 2017), comprehension of information (Brooks, 2017; Kwon et al., 2018), the number of times participant had to stop their primary task to handle the robot (Brooks, 2017), trust in robot (De Visser and Parasuraman, 2011; Rosenthal et al., 2012; Hamacher et al., 2016), the participant's emotional state (Groom et al., 2010) and their influence on user impressions of the robot (Groom et al., 2010; Shiomi et al., 2013; Bajones et al., 2016; Engelhardt and Hansson, 2017; Kwon et al., 2018). Brooks (2017) devised a measurement scale of people's reaction to failure called the REACTION scale, which claims to compare different failure situations based on the severity of the failures, the context risk involved, and effectiveness of recovery strategy. Rossi et al. (2017b) found that people, regardless of age or gender, are fairly consistent in how they rate the severity of robot errors.

The method of measuring each criterion varied; to assess the quality of interaction, research teams mainly used custom made questionnaires with Likert scales and unstructured interviews with a large variety of different questions (e.g., Kim and Hinds, 2006; Short et al., 2010; Rosenthal et al., 2012; Desai et al., 2013; Knepper et al., 2015; Hayes et al., 2016; Robinette et al., 2016; Kwon et al., 2018; Lucas et al., 2018). The most common structured and validated questionnaires used include the Godspeed questionnaire (used in Salem et al., 2015; Bajones et al., 2016; Engelhardt and Hansson, 2017; Mirnig et al., 2017; Sarkar et al., 2017) and NASA TLX (used in Desai et al., 2012, 2013; Hamacher, 2015; Hamacher et al., 2016; Brooks, 2017). Some evaluations were done using video-analysis (Kahn et al., 2012; Hamacher et al., 2016; Sarkar et al., 2017); looking at behavioral data (Kahn et al., 2012; Bajones et al., 2016; Hamacher et al., 2016; Sarkar et al., 2017), verbal statements made during the experiment (Kahn et al., 2012; Bajones et al., 2016; Hamacher et al., 2016), and the number and type of errors made (Bajones et al., 2016). About half of the experimental studies were performed using the Wizard-of-Oz technique (Riek, 2012) (e.g., Gieselmann, 2006; Groom et al., 2010; Short et al., 2010; Kahn et al., 2012; Rosenthal et al., 2012; Yasuda and Matsumoto, 2013; Mubin and Bartneck, 2015; Lucas et al., 2018), and half programmed erroneous behavior to be performed automatically (e.g., Gehle et al., 2015; Gompei and Umemuro, 2015; Hamacher, 2015; Hayes et al., 2016). Only a few studied unplanned failures (e.g., Giuliani et al., 2015; Knepper et al., 2015; Mirnig et al., 2015).

The number of participants used in each study varied, however with the exception of Gieselmann (2006), all had more than 10, which is arguably sufficient to obtain meaningful results through user studies (Nielson, 2000). Most experiments were done on Americans (21) and Europeans (18). Few studies involved non-Western participants (Shiomi et al., 2013; Yasuda and Matsumoto, 2013; Gompei and Umemuro, 2015; Kim et al., 2017), and only one evaluated cross-cultural differences

(Rossi et al., 2017a). Participants varied in age, however most studies were primarily implemented on younger adults. One study evaluated children (Lemaignan et al., 2015); none focused on elderly participants above the age of 75. With the exception of seven studies, the distribution between male and female participants was relatively equal (more equal than 60–40%). Sixteen (31%) of the studies evaluated participants with little experience with robots, 2 (3.8%) studies evaluated experienced participants, and 30 (58%) studies did not state the participants' level of experience with robots. Only four studies (7.7%) evaluated both experienced and inexperienced participants (Hamacher, 2015; Hamacher et al., 2016; Rossi et al., 2017a; Lucas et al., 2018).

Robotic Systems

A wide variety of robotic systems are used to study human centered failure handling. NAO was by far the most commonly used robot (Gehle et al., 2015; Giuliani et al., 2015; Gompei and Umemuro, 2015; Mirnig et al., 2015, 2017; Engelhardt and Hansson, 2017; van der Woerdt and Haselager, 2017; Lucas et al., 2018), however several other off-the-shelf solutions were used, including BIRON (Spexard et al., 2008), Kuka youBots (Knepper et al., 2015), iRobot ATRV-JR (Desai et al., 2012), Robovie-mR2 (Shiomi et al., 2013), Snackbot (Lee et al., 2010), and Baxter (Adubor et al., 2017; Sarkar et al., 2017). Several systems were custom made for the purpose of the research (Yasuda and Matsumoto, 2013; Lohan et al., 2014; Lemaignan et al., 2015; Mubin and Bartneck, 2015). About half of the 52 studies used humanoid robots [robots that possess some human-like features (Walters et al., 2008)], and half used mechanoid robots [robots that are machine-like in appearance (Walters et al., 2008)].

Environment

Experimental evaluations were mostly done indoors, with single-persons (86%). Only one study evaluated robotic failures in outdoor environments (Giuliani et al., 2015), and five of the studies evaluated robotic failures indoors when more than one person was present (Kim and Hinds, 2006; Rosenthal et al., 2012; Gehle et al., 2015; Lemaignan et al., 2015; Bajones et al., 2016). With the exception of Cassenti (2007), which proposed a strategy for helping users recover from errors after prolonged time in which no interaction with the robot was made, all of the studies focused on errors that occurred during interaction with the robot.

A UNIFIED INFORMATION PROCESSING MODEL FOR USER CENTERED FAILURE HANDLING

In order to develop interactions that enable untrained users to easily identify and solve failures, it is critical to consider cognitive factors that influence the ability to perceive and act upon a robotic failure. Interacting with a robot in a moment of failure is inherently an information-processing task—the user must perceive information from the robot and environment, process it to identify if an error has occurred, recall what can be done to fix it or enter a command to obtain additional information, select

and then execute responses based on that information. Thus, for failure-handling management tools to be easy to use, the human-robot interface must be designed to meet the information processing capabilities of users.

There are many theories regarding how people process information (e.g., McClelland, 1979; Card et al., 1983, 1986; Miller, 1988; Kieras and Meyer, 1997). One information-processing model that seems particular relevant is the Communication-Human Information Processing (C-HIP) Model (Wogalter, 2006a), which describes the way people process warnings. In situations of failure, indicators from the robot, user and environment can be viewed as warnings of the robot's degraded state of ability. The model includes three main parts: (1) sending the warning, (2) processing it by the receiver, and (3) acting. The parts are described using nine stages that must be completed for people to be compliant with a warning. A bottleneck at any given stage can impede on processing at subsequent stages, and feedback from later stages and additional sources (such as environmental and personal attributes of the receiver) can affect processing in earlier stages.

After reviewing the cognitive considerations that influence people's ability to detect and solve robot failures, as well as the current literature in failure handling in HRIs, we developed an information processing model called the Robot Failure Human Information Processing (RF-HIP) Model, modeled after C-HIP (Wogalter, 2006a), to describe the way people perceive, process, and act on failures in human robot interactions (**Figure 3**). By providing the stages of information processing and factors that influence them, RF-HIP can be used as a tool to systematize the assessment process involved in determining why a particular approach of handling failure is successful or unsuccessful in order to facilitate better design. The model, which will be used to guide the presentation of the relevant literature, includes three main parts: (1) communicating failures, (2) perception and comprehension of failures, and (3) solving failures. Each part contains several stages, all heavily influenced by contextual considerations (the source, task, receiver, environment and other agents) and mitigation strategies. The model differs from C-HIP in three primary ways: (1) there is a separate stage for decision making, (2) it accounts for unplanned failure indicators (symptoms) and for subconscious behavior, and (3) it highlights the bilateral relationship between all stages of information processing, contextual factors and mitigation strategies. The components of the model are discussed in the following sections.

Source

The source is the transmitter of symptoms indicative of a failure. The source of failure is typically the robot, however it could also be the user or other humans in the environment (e.g., in case of human error or when a person produces behavioral responses to robot failure). In situations where a symptom is identified by the source, the source must determine whether it can handle it on its own by ignoring or eliminating the problem, or whether it needs to produce a warning of the symptom to others. If the failure is technical, there are several automatic methods that can be used to detect the error (e.g., Murphy and Hershberger, 1999; Canham et al., 2003) and automatically
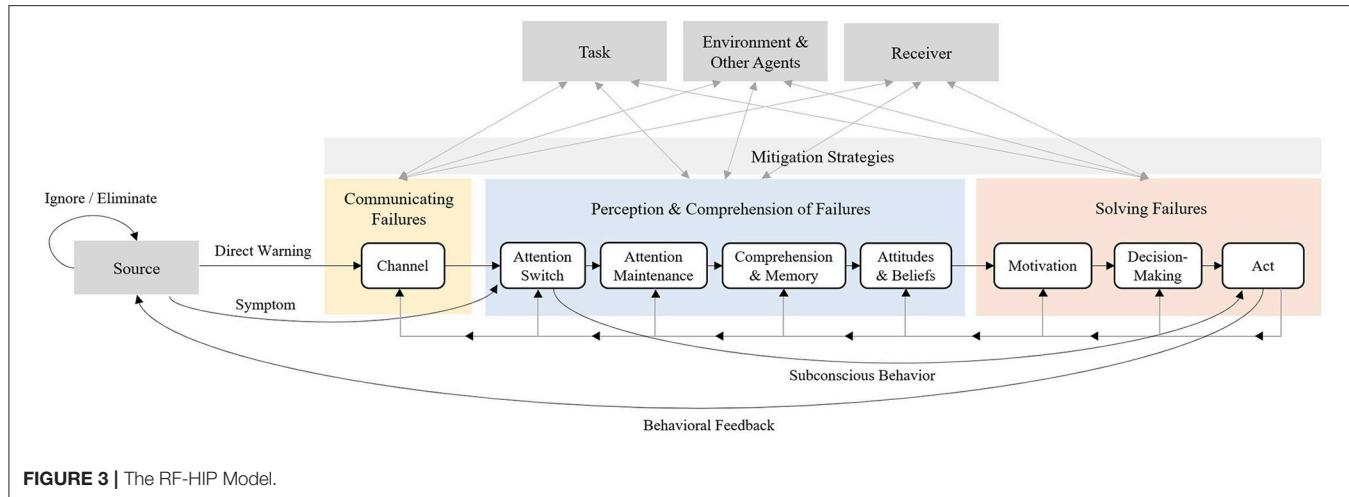
FIGURE 3 | The RF-HIP Model.

determine the appropriate recovery method, without involving human assistance (e.g., Murphy and Hershberger, 1999; Mendoza et al., 2015). Several methods also exist to predict and resolve human error in HCI that could be applied to robots (e.g., Embrey, 1986; Baber and Stanton, 1994). Sometimes the symptom is itself a type of warning that is outwardly projected (e.g., the robot's wheel falling off), so the receiver perceives it without the source actively deciding on how to communicate the failure. In such cases, the source may not always be aware of the symptom (e.g., a robot may not be aware when it deviates from social norms).

Warnings can be direct or indirect: a direct warning occurs when the person is directly exposed to the symptom or to a warning from the source, whereas an indirect warning is received in other ways (e.g., learning about the problem from a family member). Various characteristics of the source influence perceived beliefs, credibility, and relevance of symptoms and warnings (Wogalter, 2006a).

Communicating Failures
Channel
The channel is the medium and modality which the source uses to transmit information regarding a failure to receivers. While some robot failures can be detected through changes in the robot's behavior or posture (e.g., Takayama et al., 2011; Kwon et al., 2018), changes in the robot's physicality (e.g., a wheel falling off), or changes in the user's behavior (see section Act), other issues (e.g., missing data) produce no obvious symptoms. Moreover, overt changes in robotic behavior may remain undetected by users as a result of poor situation awareness, inexperience with the robot, or lack of supervision (Brooks, 2017). Consequently, various methods have been suggested to intentionally communicate failures and their causes to users and bystanders of robotic systems when possible. If the source identifies a need for a direct warning, it must determine how the relevant agents should be warned. Depending on the source, different channels of communication and delivery methods will be possible.

Visual indicators on robot
Brooks (2017) investigated the use of standardized icons displayed on the body of a robot as a method of conveying information about an autonomous robot's internal system state. Specifically, they attempted to convey information about whether the robot is safe to be around and whether it is working properly using five target messages (ok, help, off, safe, and dangerous). Results indicated that icons are a viable method for communicating system state information to untrained bystanders.

Other types of on-robot visual indicators have also been used to indicate robotic errors. One approach is using light (or lack of it)—the Neato robotic vacuum cleaners display an amber light around the main button when it cannot start cleaning[1]; Baraka et al. (2016) used flashing red lights to indicate path obstructions; and Robinette et al. (2016) turned off the robot's lights to indicate inoperability. Another common method is using on-robot screen displays. In Sarkar et al. (2017), Baxter's screen showed a sad smiley face with explanatory text whenever an error was made. Similarly, Jibo[2] (a personal assistant robot) shows an error code and message on its screen whenever there is an issue[3].

The primary advantage of using visual indicators on the robot to display failure states is that their placement allows the message to be communicated not only to the robot operator but also to bystanders without any mediating artifacts. Another advantage is that insights and design principles from human factors and HCI literature (e.g., Nielsen, 2001; Wogalter and, 2006c; Egelman et al., 2009; Bauer et al., 2013) could be used as inspiration. There are, however, disadvantages to using visual indicators on the robot. For one, visual indicators on the robot can only influence people who are actively looking and paying attention to the robot.

---

[1]"Status Lights," *Neato Robotics* (2017). Available online at: https://support.neatorobotics.com/hc/en-us/articles/225370027-Status-Lights (Accessed December 14, 2017).

[2]"Hey! I'm Jibo." Available online at: https://www.jibo.com/. [Accessed: 14-Dec-2017].

[3]"Jibo - Error messages." Available online at: https://support.jibo.com/jibo/articles/en_US/FAQ/error-messages. [Accessed: 14-Dec-2017].

Remote operators and people performing multiple tasks may not notice the indicators in time to act upon them, which is particularly important in failure situations. Second, the message could at times be occluded, depending on the robot's speed and posture relative to the human observer. Third, icons and status lights can effectively convey only simple messaging that represent distinct alternative states of the robot. Screens on the robot can communicate more complex information, however it requires the user to physically come close to the robot, which may not always be safe for certain types of failure. Lastly, the public nature of such indicators may not always be socially appropriate—people may feel uncomfortable having others know about certain errors taking place. For example, a robot unable to track the users' legs because they are too wide or narrow relative to its expectations may cause embarrassment.

Secondary screens

Another method of communicating a robot's failure state is by using a secondary screen (such as a smartphone) to provide additional information about the robot. This strategy is one of the most popular in today's commercial robots (e.g., Kuri[4], iRobot Home Robots[5], Neato Robotics[6]) and has several advantages: (1) it enables users to interact with the robot using familiar methods of interaction, (2) complex information can be more easily conveyed on-screen, and (3) status information can be accessed remotely and covertly. The main disadvantage of this method is that it inherently shifts the user's eyes and attention away from the robot and from the tasks they are performing, which hinders situation-awareness and could be dangerous in threatening situations. Cassenti (2007) proposed presenting a video replay strategy using a secondary screen to quickly provide situation awareness after prolonged times of robot neglect.

Audio and speech

Our ability to localize acoustic sources and apply selective attention to one acoustic stream out of many, even at a distance, makes the audio modality popular for communicating failures. As such, many mobile robots use audio and speech to communicate robotic failures. Some use simple audio tones to gather user attention (e.g., Brooks, 2017), whereas others communicate failure using more complicated speech, such as Jibo[2] and the robot in Schütte et al. (2017). Cha et al. (2015) found that people perceived robots speaking conversationally as more capable than those that could only maintain a functional level of speech. However, this changed when the robot made an error—after an error, robots with conversational speech were perceived as less capable than those with functional speech. This effect is similar to equivalent research in HCI (Weinstock et al., 2012) that found that when a visually aesthetic user interface errors, the error lowers perceptions of satisfaction, human automation cooperation and trust more than when a non-visually aesthetic interface errors. Several researchers suggest to

use verbal communication cautiously since dialogue can lead to biased perceptions of the robot's capabilities (Fong et al., 2003; Cha et al., 2015). Simpler audio signals can be used to signal the existance of a problem, however, they cannot effectively explain the cause of error.

Modality comparisons

Very few studies assessed the benefits of different modalities for communicating failures in HRIs. Cha et al. (2016) evaluated a robot which utilized both light and sound of varying levels of urgency to request help from bystanders when it experienced difficulty. Results indicated that participants interpreted light and sound signals differently: sound alerted the user that the robot needed help and the light indicated the level of urgency of the help request. Moreover, participants preferred a more attention-grabbing signal when the urgency of the request was high, and when the urgency of the request was lower, they preferred the robot to take into account the participant's level of availability by utilizing greetings and being more polite.

Brooks (2017) compared between a designated smartphone application and a light-and-button based interface in their ability to help inexperienced users better detect and solve failures while performing a secondary task. Unlike the previous example, which used an indicator to help users detect robot requests, this example focused also on its ability to help users solve errors. Results indicated that participants were able to obtain information about the robots, identify solutions to problems and allocate their time more appropriately using the app.

Further studies from the warning literature provide insight regarding how to create comprehensible warnings. Warnings presented in more than one modality generally facilitate better comprehension than those presented in a single modality (Wogalter, 2006a). While there is conflicting evidence of whether written text or speech are better for comprehending language-based warnings (Mayer, 2002; Wogalter, 2006b), reading language allows people to review the material and tends to be faster, so it may be more appropriate for long or complex messages. In contrast, shorter, less complex messages have a greater impact when presented auditorily than visually, and are generally better for switching attention (Wogalter, 2006a). A short auditory warning that directs the users' attention to more detailed information could be used to capture attention while facilitating the processing of more complex information (Wogalter, 2006a).

Perception and Comprehension of Failures
Attention Switch and Maintenance

For a failure event to influence user behavior, attention must be switched to it for the user to perceive the information (Wogalter, 2006a). Moreover, attention must be maintained by users to perform desired behaviors properly and avoid certain types of human errors, such as slips (Reason, 1990). The conditions under which a person shifts their attention can be used to guide the design of robotic failure indicators. Sudden changes in the environment [e.g., change in luminance (Theeuwes, 1995), motion onset (Abrams and Christ, 2003), and abrupt appearance or disappearance of stimuli (Pratt and McAuliffe, 2001)] or the

---

[4]Life with Kuri (2017). Available online at: https://www.heykuri.com/living-with-a-personal-robot (Accessed December 13, 2017)

[5]iRobot Home Robots (2018). Available online at: http://www.irobot.com/ (Accessed January 06, 2018).

[6]Neato Robotics. Available online at: https://www.neatorobotics.com/ (Accessed January 01, 2018).

robot's behavior (Okada et al., 2003; Sato et al., 2007) could be used to quickly and involuntarily shift people's attention to urgent failure situations or to cue users to attend to information elsewhere. These involuntary shifts of attention tend to be brief (Buschman and Miller, 2010), and are dependent upon users' expectations (Posner et al., 1978; Folk et al., 1992). In contrast, long term exposure to a warning could make it unable to attract attention at later times ("inhibition of return"; Posner and Cohen, 1984; Klein, 2000), so the use of permanent cues must be considered carefully.

Voluntary shifts of attention can be sustained for longer periods of time (Welsh et al., 2009) and can result from a wider variety of stimuli (Sears and Jacko, 2009), allowing more freedom in the design of failure indicators. Various factors affect people's ability to identify and attend to a specific stimulus, including the degree of similarity to other items in the environment (von Grünau et al., 1994; Gorbunova, 2017), interest (Renninger and Wozniak, 1985), temporal and physical location of warnings (Frantz and Rhoades, 1993; Wogalter et al., 1995), the task (Welsh et al., 2009), age (Yamaguchi et al., 1995), and practice (Feinstein et al., 1994). This emphasizes the importance of taking contextual factors into consideration when designing warnings for failure. Fischer et al. (2014) found that verbal greetings attracted attention better than simpler audio signals, but they did not improve the likelihood of the person to perform the robot's request.

The design of a warning should be guided by the response required from the user (stimuli-response compatibility; Sears and Jacko, 2009). For example, reaction time is lower when people are asked to respond vocally to an auditory stimulus or with motion to a spatial attribute (Wang and Proctor, 1996). Spatial correspondence (Fitts and Seeger, 1953; Fitts and Deininger, 1954; Reeve and Proctor, 1990), similarity (Kornblum et al., 1990), and logical relations (rules) (Duncan, 1978) between the stimulus and response sets have all been shown to improve stimulus-response compatibility. Since it is not always clear in which circumstances compatibility effects are going to occur (Proctor and Vu, 2009), designers need to repeatedly test warnings on users, particularly for urgent failures.

A robot's warning can be noticed yet fail to maintain attention long enough for the user to extract meaning from it (Wogalter, 2006a). The required duration of attention maintenance has been shown to rely on the channel of communication as well as on the complexity and form of the content (Wogalter, 2006a). Generally speaking, if a warning contains too much information, is too hard to read, or the relevance of the information is low or unclear, people may decide it is too much effort, lose interest and direct their attention elsewhere (Wogalter, 2006a). Moreover, as felt involvement with product information increases, consumers have been shown to spend more time attending to the information (Celsi and Olson, 1988). Combining pictures with written or spoken text has been shown to increase attention to information in comparison to text alone (Houts et al., 2006). Visual warnings with organized information groupings and generous white space are more likely to hold attention than a single block of text (Wogalter and Vigilante, 2006). The use of humor has also been shown as an effective way to gain

and maintain attention (Weinberger and Gulas, 1992). These strategies could be used in the design of warnings to promote compliance.

Comprehension and Memory

Users must be able to understand the meaning of a failing robot's symptoms or the warning it provides to understand what the failure is and how to react. During the comprehension process, incoming perceptual inputs that have passed attentional filters are connected to past experiences or knowledge to construct an understanding of the event (Harris et al., 2006). This continuing interaction of comprehension and memory is important to understanding what may influence a person's ability to relate erroneous behavior to "normal" robotic behavior, to comprehend the meaning of a failure indicator and to resolve robotic failures.

Characteristics of memory have several implications for robotic failure situations. While people can remember large amounts of information over their lifetime, only a small portion is available to them at any given time for processing (Bettman, 1979; Lang, 2000). As a result, memories and knowledge may not become available without an external cue (Wogalter, 2006a), and those that are readily available may quickly become unavailable due to interference or decay (Proctor and Vu, 2009). This emphasizes the importance of considering external factors, such as user tasks and bystanders, and of providing informative cues to help the user recall and resolve a failure.

In failure-handling situations, recall and comprehension of relevant information (warnings, robotic commands, and possible solutions) could be made easier by exploring influential factors. Studies indicate that it is easier to recall information that is visual (Paivio and Csapo, 1973), concrete (Butter, 1970; Sheehan and Antrobus, 1972), repeated (Kintsch et al., 1975), specific (Mani and Johnson-Laird, 1982), personal (Van Lancker, 1991), novel (Kishiyama and Yonelinas, 2003), typical (Reeve and Aggleton, 1998), humorous (Schmidt, 1994; Summerfelt et al., 2010; Carlson, 2011) and self-generated (Wheeler and Gabbert, 2017). The likelihood a retrieval cue leads to recollection depends on the similarity between the features encoded initially and those provided by the retrieval cue, distinguishability from other cues, and association with the newly learned information (Wheeler and Gabbert, 2017). Storing information to memory seems to depend on deep processing of the meaning of new material, determined by the degree to which one understands the information to form meaningful associations and elaborations with existing knowledge (Bower, 2000), as well as on arousal (Butter, 1970) and individual differences (Verhaeghen and Marcoen, 1996) [e.g., age (Anderson et al., 2000), mood (Bower et al., 1978)]. Various techniques have been developed to improve recall and storage from and to memory (e.g., Bower, 1970a,b; Ritchie and Karge, 1996; Gobet et al., 2001). Such techniques could be used by robot designers to help select appropriate cues that help users recall information that is relevant to the failure.

Comprehension has been shown to be influenced by background knowledge (Tannenbaum et al., 2006), wording (Kintsch et al., 1975), typographic design (Frase and Schwartz, 1979), personality (Sadeghi et al., 2012), felt involvement (Celsi and Olson, 1988), motivation (Sideridis et al., 2006), expectations

(Haberlandt, 1982), training (Dewitz et al., 1987), experience (Macias, 2003), level of automation (Carmody and Gluckman, 1993), interface design (Canham and Hegarty, 2010), workload (Perry et al., 2008) and stress level (Perry et al., 2008). One common way to classify a person's level of comprehension is by evaluating their Situation Awareness (SA) (Endsley, 1988). Drury et al. (2003) defined components of situation awareness that are relevant to HRI: (1) awareness of the locations, identities, activities, states, and surroundings of the robot and fellow human collaborators, (2) awareness of the robot's knowledge of the human's commands and any human constraints, (3) awareness of the knowledge that the robots have of the activities and plans of other robots, and (4) awareness of the overall goals of the joint human-robot activities and progress toward the goal. They then related these types of awareness to critical incidents at an urban search and rescue competition in which the operator or robot encountered a problem, and found that all critical incidents resulted from awareness violations (Drury et al., 2003). Techniques that improve situation awareness could be used by robot designers to help prevent various types of failures.

Beliefs and Attitudes

At this stage of processing, the comprehended information merges with existing beliefs and attitudes. A mental model can be a useful concept for understanding this process. As the user interacts with the robot, they receive feedback from the system and the environment that allows them to develop a representation (a mental model) of how they believe the system behaves for a given task. These representations lead to expectations, which in turn direct perception and behavior (Stanton, 2009). Studies in the field of HCI found that users infer models that are consistent with their experiences, even when there is lack of evidence that supports their assumptions (Payne, 2009). Moreover, instead of developing unified models, they develop separate beliefs about parts of the system, processes, or behaviors that are not necessarily complementary (Payne, 1991). While incorrect mental models can lead to difficulties in problem solving, the use of appropriate mental models can help people learn, remember and execute procedures faster (Kieras and Bovair, 1984). Mental models can also explain human errors: if action is directed by mental models, then the selection of inappropriate models or erroneous activation of appropriate models will lead to errors (Norman, 1981). Designers can increase the usability of a robotic interface for handling failures using metaphors that promote the use of applicable mental models and by correcting inappropriate mental models through feedback.

In the HRI literature, mistakes made by robots influence how the robot is perceived. Failures reduce robots' perceived sincerity (Gompei and Umemuro, 2015), competence (Cha et al., 2015; Salem et al., 2015; Ragni et al., 2016), reliability (Salem et al., 2015; Ragni et al., 2016), understandability (Salem et al., 2015), trustworthiness (De Visser and Parasuraman, 2011; Desai et al., 2013; Salem et al., 2015; Law et al., 2017), intelligence (Takayama et al., 2011; Bajones et al., 2016; Ragni et al., 2016), and likeability (Bajones et al., 2016; Mirnig et al., 2017), and increase perceived familiarity (Gompei and

Umemuro, 2015). In Kahn et al. (2012), participants who interacted with a humanoid robot that incorrectly assessed their performance perceived the robot as having emotional and social attributes. Research is inconclusive regarding the effect of failures on the robot's perceived anthropomorphism. Salem et al. (2013) found that errors made robots seem more human, whereas Salem et al. (2015) found that it made robots seem less human. Mirnig et al. (2017), in contrast, did not find differences in people's ratings of the robot's anthropomorphism and perceived intelligence. These differences may be a result of the different robots used, or the different interaction contexts (task, environment).

User perceptions of the robot in a failure situation seem to be influenced by a number of factors. In contrast to Salem et al. (2015), which found that failure reduced perceived reliability, technical competence, understandability, and trustworthiness of a home-care assistant robot, the manufacturing robot in Sarkar et al. (2017) was perceived in a similar manner regardless whether it was faulty or not. According to Sarkar et al. (2017), these differences may stem from the type of failures (Sarkar et al., 2017 involved subtle interaction failures, whereas Salem et al., 2015 produced physical failures with potentially irreversible consequences), or the nature of the experimental task (the industrial context in Sarkar et al., 2017 compared to a more "social" setting in Salem et al., 2015). Rossi et al. (2017a) found that errors with severe consequences lead to greater loss of trust in the robot. Furthermore, user perceptions of the robot in a failure situation may depend on attribution of the cause of failure—in an online survey (van der Woerdt and Haselager, 2017), participants were shown a video portraying a NAO robot failing a task either due to lack of ability or lack of effort. In case of failure, participants attributed more agency to the robot that displayed lack of effort compared to videos in which it displayed lack of ability. The timing of failure also seems to influence how the failure affects perceptions of the robot. Gompei and Umemuro (2015) investigated the effect of a failure's timing: when the robot made speech errors on the first day of contact, the robot's familiarity score did not change; when the robot made its first speech error on the second day of contact, the robots' familiarity score moderately improved as a result of the error. Similarly, Lucas et al. (2017, 2018) found that errors that occur later in a robot's dialogue, particularly after a period of good performance, reduce the robot's persuasiveness.

While robotic failures have been shown to reduce the perceived trustworthiness of robots (De Visser and Parasuraman, 2011; Hancock et al., 2011; Desai et al., 2013; Salem et al., 2015; Law et al., 2017), users' compliance with robot instructions may not be affected. Robinette et al. (2016, 2017) evaluated whether people will trust and follow the directions of a faulty robot in emergency evacuee scenarios. Results showed that the vast majority of participants followed the instructions of the robot despite erraneous behaviors. In line with this finding, Salem et al. (2015) found that while the robot's erratic behavior affected its perceived reliability and trustworthiness, it did not impact participants' willingness to comply with its instructions, even when the requests were unusual. Severity of the outcome affected

compliance with robot requests (Salem et al., 2015). Similar effects were found by Tokushige et al. (2017) as a result of unexpected recommendations.

While there are some indicators that people may prefer predictable behavior in robots (Mubin and Bartneck, 2015), others suggest that people feel more engaged by unpredictable behavior (Short et al., 2010; Fink et al., 2012; Lemaignan et al., 2015; Law et al., 2017). Various studies seem to suggest that failures can be a source of pleasurable interaction with robots (Bainbridge et al., 2008; Yasuda and Matsumoto, 2013; Gompei and Umemuro, 2015; Ragni et al., 2016; Mirnig et al., 2017). In a study by Ragni et al. (2016) despite the faulty robot being rated worse than the error-free robot, participants reported greater enjoyment when the robot made errors. Similarly, Mirnig et al. (2017) found that participants liked faulty robots better than robots that interacted flawlessly. Annotations of video data showed that gaze shifts, smiling and laughter are typical reactions to unexpected robot behavior. While these studies provide insight regarding reactions to robotic failures, the non-criticality of the errors coupled with low personal relevance to the participants may have impacted results.

Desai et al. (2013) investigated the influence of varying reliability on real-time trust and found that periods of low reliability earlier during the interaction have a more negative impact on overall trust than periods of low reliability later in the interaction. In contrast, a preliminary study by Desai et al. (2012) found that people trust a robot less when reliability drops occurred late or in the middle of runs. Within the broader human-automation literature there is certain agreement that trust depends on the timing, consequence, and expectations associated with failures of the automation (Lee and See, 2004).

Solving Failures
Motivation
Solving a robotic failure requires the user to be motivated to solve the problem. Even if the users are not capable of solving the failure themselves, they need to be motivated enough to inform other agents of the problem (such as a caregiver or a technician) in order for it to be addressed. While some problems may significantly impact users, motivating them implicitly, other failures may not be sufficient to motivate them enough to solve the problem, particularly if the interface is hard to understand or operate. Thus, creating successful failure-handling solutions requires skills in motivating and persuading people. Captology, the study of persuasive technologies is a relatively new endeavor in HRI (see Siegel, 2008; Ham and Spahn, 2015). Research has explored effect of a robot's physical presence (Kidd and Breazeal, 2004; Shinozawa et al., 2005; Bainbridge et al., 2008), touch and gesture (Shiomi et al., 2010; Ham et al., 2011; Nakagawa et al., 2011; Chidambaram et al., 2012; Baroni et al., 2014), gazing (Ham et al., 2011), robot and user gender (Siegel, 2008; Nakagawa et al., 2011), vocal cues (Chidambaram et al., 2012; Baroni et al., 2014), interpersonal distance (Siegel, 2008), reciprocity (Lee and Liang, 2016), conversational errors (Lucas et al., 2018), agency (Ham and Midden, 2011), and perceived autonomy (Siegel, 2008)

on persuasive effects. However, none of these studies focused specifically on the influence of motivation in solving robotic failures.

Robots are sometimes viewed as tools, and other times viewed more as social actors (Breazeal, 2004). According to Fogg et al. (2009), there is a difference in how computers can be used to persuade, depending on whether they are viewed as a tool or social actor. Computers as tools can persuade by providing tailored information, triggering decision making, increasing self-efficacy, and guiding people through a process. In contrast, computers as social actors can persuade people by providing social support via praise or criticism, modeling behaviors or attitudes, and leveraging social rules (e.g., turn taking, politeness norms, praise and reciprocity).

Decision-Making
Once individuals have perceived the failure symptoms and/or warnings, comprehended them, formed beliefs and attitudes regarding the situation, and gained enough motivation to solve the issues, they must decide what can be done to solve the failure. Most problems are well beyond the capacity of comprehension to be solved optimally. Reaction time typically increases with the number of stimulus-response alternatives (the Hick-Hyman law; Hick, 1952; Hyman, 1953). Consequently, for problem solving to be effective in a robotic failure situation, search must be constrained to a limited number of possible solutions or approaches (Proctor and Vu, 2009).

A common way novice users constrain search in situations of uncertainty is to use heuristics (Tversky and Kahneman, 1974). Research demonstrates that our judgements are based on the subset of relevant information most accessible in memory, and that we rarely retrieve all relevant information (Bodenhausen and Wyer, 1987; Schwarz, 1998). One particularly common strategy is "satisficing" (Simon, 1956), which refers to searching through available alternatives and choosing the first that meets some minimum acceptable threshold. Some other examples include (but are not limited to) representativeness (Tversky and Kahneman, 1973), availability (Tversky and Kahneman, 1973), and adjustment (Epley and Gilovich, 2006) heuristics. The problem with using heuristics is that they often lead to cognitive biases, which influence the quality of the decision. Many biases in human decision making have been discovered (Croskerry, 2003) [e.g., the framing effect (Tversky and Kahneman, 1981), confirmation bias (Nickerson, 1998), and overconfidence effect (Dunning et al., 1990)]. Consequently, people generally make nonoptimal decisions.

Various efforts have been made to improve and debias decision making, which could be implemented to better support users during robotic failure situations. Three general approaches have been suggested and shown to produce positive results (Morewedge et al., 2015): (1) recalibrating incentives to reward healthy behavior, (2) optimizing how choice options are presented and obtained, and (3) debiasing training interventions. Small changes in presentation and elicitation of choices are particularly effective, cheap and easy to implement, taking many forms such as information framing (Levin and Gaeth, 1988; Larrick and Soll, 2008)

and default selection (Johnson and Goldstein, 2003; Chapman et al., 2010). These recommendations, alongside additional strategies (e.g., Croskerry, 2003), could be used to help facilitate the design of failure-management interfaces for robots to improve the problem-solving abilities of untrained users.

Act

This stage of processing refers to both the execution of the person's decision regarding how to respond to the robotic failure, as well as automatic behaviors that are triggered without maintaining attention. People seem to have various predictable behavioral responses to robotic failures that can be used by robots to identify when a failure has occurred. Failure has been shown to influence users' gaze patterns (Gehle et al., 2015; Hayes et al., 2016; Mirnig et al., 2017), facial expressions (Hayes et al., 2016; Mirnig et al., 2017), head movements (Hayes et al., 2016; Mirnig et al., 2017; Trung et al., 2017), body movements (Mirnig et al., 2017; Trung et al., 2017), and verbal communication (Gieselmann, 2006; Giuliani et al., 2015). Gieselmann (2006) found that indicators for errors in human-robot conversation included sudden changes of the current dialogue topic, indicating non-understanding by asking unspecific questions, asking for additional information and repeating the previous question. Additional indicators used to detect errors in spoken human-robot dialogues include people being silent, asking for help, repeating central elements or asking the robot repeatedly for the same information, saying things that are inconsistent with the current discourse or with the robot's expectations, trying to correct a preceding utterance, hyperarticulating speech, or asking for something they know the robot cannot do, such as making coffee (Gieselmann and Ostendorf, 2007).

Giuliani et al. (2015) and Mirnig et al. (2015) analyzed video data showing social HRIs in which the robot unintentionally made an error. Results indicated that in erraneous situations, participants often used head movements, smiled, raised eyebrows, and looked back and forth between the robot and experimenter or a group member if present. Moreover, the type of error (social norm violation or technical failure) as well as the presence of other people seemed to impact people's reactions to the failure. More specifically, during social norm violations, participants spoke more, were more likely to look back and forth between the robot and objects in front of them and say task-related sentences to the robot than during technical failures. When no experimenter or person was visible, participants used fewer non-verbal social signals (e.g., smiling, nodding, and head shaking), and more often shifted their gaze between the robot's hand, the robot's head, and other objects in front of them than when the experimenter was visible, or when interacting in groups with the robot. The presence and response speed of these social signals were dependent on the type of error made and the type of task the robot was performing.

There is also reason to believe that the modality of the failure influences people's reactions. Short et al. (2010) investigated people's reactions to playing rock–paper–scissors with a humanoid robot that either played fair, cheated through action by changing the selected hand gesture or cheated verbally

by declaring a different hand gesture than the one used. Results indicated that participants showed more verbal social signals to the robot that cheated. Interestingly, verbal cheating was perceived as malfunctions, often leading to reactions of confusion, whereas cheating through action was perceived as deliberate cheating, leading to more exaggerated reactions, showing surprise, amusement, and occasionally anger.

Contextual Factors

Receiver

The receiver is the person(s) or target audience whom witness the warning or symptom, typically the user. Personal attributes of robot users have been shown to affect all stages of information processing, and in turn, the stage of information processing influences the users' experiences and behaviors. Contributing factors surveyed include the user's attitudes and beliefs, interest, practice and training, experience, background knowledge, workload, stress level, situation awareness, mental model, and gender.

Environment and Other Agents

External stimuli from the environment compete for the receiver's limited attention and comprehension resources, limiting information processing. For instance, a friend saying "Hi" when the robot is trying to indicate that the motors stopped working could prevent the user from attending to a visual warning. A noisy environment may cause the user not to hear the robot's low battery beep, or not to be able to concentrate enough to lead it back to its charger. In some cases, this could be an advantage: social norm violations, for instance, could be missed and therefore not negatively influence the interaction. The individual may act on the environment and change it, so there is a bilateral relationship between the environment and the stages of information processing. In situations where the user does not have the know-how, ability or the tools to fix the problem, the involvement of other agents may be necessary to solve the failure.

Task

Task refers to attributes of either the robot's task, the person's task, or a joint task to be completed together. From the literature, it is evident that the task a person is performing can compete for their limited attention and comprehension resources and by doing so, impact the stages of information processing. In turn, cognitive resources devoted to the failure have an impact on the task: an increase in automation during failure condition reduces operator performance (the "lumberjack analogy"; Sebok and Wickens, 2017). Several studies seem to indicate that task performance is significantly influenced by robotic failures. In Ragni et al. (2016), participants competed against a robot in reasoning and memory tasks where the robot either performed with or without errors. Results indicated that task performance was significantly lower in the faulty robot condition. Similarly, in Desai et al. (2012), drops in reliability were shown to affect participants' self-assessments of performance. Salem et al. (2013) evaluated whether participants who were presented with incongruent multimodal instructions by the robot performed worse at their task than those who were presented with unimodal

or congruent multimodal information by the robot and found that incongruent coverbal gesturing reduced task performance. One contrasting account is the manufacturing scenario described in Sarkar et al. (2017), where a physical object was assembled and then disassembled under regular and error conditions. In this scenario, faults did not affect the successful completion of a manufacturing task. The authors proposed that these results may be because the types of failures they implemented (missing an action and/or giving the wrong instructions) did not impede the possibility of a successful manufacturing outcome.

Mitigation Strategies

Various mitigation strategies can be attempted both by the user and robot in order to prevent and handle the negative influences of failure. Mitigation strategies could be applied in any stage of information processing. The stage of processing, in turn, affects the effectiveness of the mitigation strategy applied. The following sections discuss the various strategies that have been implemented to mitigate the negative effects of failure in HRI.

Setting Expectations

Giving the user advance notice regarding potential failures influences how they respond to subsequent failures. This is consistent with studies that found that robotic errors have a stronger negative effect after a period of good performance (Lucas et al., 2018). One online study by Lee et al. (2010) found that setting expectations by forewarning participants of the abilities of the robot improved evaluations of the robot and judgments of the quality of the service. Providing options helped increase people's willingness to use the robotic service again after failure, however was not particularly effective in improving perceptions of the robot (Lee et al., 2010). Additional studies found that providing confidence feedback on the robot's performance encourages better control allocation without affecting user trust (Desai et al., 2013; Kaniarasu et al., 2013).

Communicating Properly

Several researchers have evaluated the impact of politeness strategies, such as apologizing (Lee et al., 2010; Peltason and Wrede, 2011) or expressing regret (Hamacher, 2015), on human-robot error interactions. When robots employ these strategies, perceptions of robots and responses to disagreement are improved (Takayama et al., 2009; Torrey, 2009). In Hamacher et al. (2016) apologizing, expressing regret and expressing reparation lead to similar trust ratings as a non-failing robot.

Various repair strategies have been used to help robots gracefully recover from verbal misunderstandings and speech errors (Gieselmann, 2006). *Achievement strategies* involve explaining the meaning of an utterance, e.g., paraphrasing, restructuring the sentence, repetition, and asking for help. *Functional reduction strategies* involve replacing the original intention by a different, simpler one, for instance, telling the robot to go to the kitchen instead of telling it to pick up the cup in the kitchen. *Formal reduction strategies* involve simplifying the grammar or the vocabulary used, and *ratification* involves confirming or repeating the last utterance made (e.g., "yes, I asked you to press the green button"). Gieselmann (2006) evaluated the

use of these strategies in a domestic HRI scenario, and found that the most common error recovery strategies were achievement strategies and functional reduction strategies.

There is little research evaluating what information should be communicated to help users cope with robotic failure situations. One research study (Cameron et al., 2016a) proposed a method to evaluate whether a robot should respond to an error with (1) simple instructions for the user to follow (e.g., "Follow me back to the lift"); (2) competency-oriented statements that emphasize the robot's abilities, the current situation, and goal (e.g., "That sign said we are on C floor and we need to go to B floor. Follow me back to the lift"); (3) inclusion of apology-oriented statements that emphasize attempts to relate to users but do not indicate competency (e.g., "Sorry about the error; we all make mistakes sometimes. Follow me back to the lift"); or (4) inclusion of both the competency- and apology-oriented statements. However, to the best of our knowledge, the results of this experiment have yet to be published. Other studies proposed communicating the cause of error with varying degrees of success. One experiment found that having the robot place blame for a failure reduced user trust (Kaniarasu and Steinfeld, 2014). Another study found that attributing blame to the user led people to feel less comfortable with the robot, perceiving it as less friendly and competent, even when the person was likely aware that they were the source of problem (Groom et al., 2010). Kim and Hinds (2006) found that providing the cause of failure could facilitate more accurate blame-attribution as long as the robots' explanation correlated to the background knowledge of participants. If not, providing the cause decreased people's perceived understanding of the system. Kwon et al. (2018) proposed expressing physical limitations through motions that communicate what the robot attempted to accomplish and why it was unable to accomplish it. The use of these motions was found to increase positive evaluations of the robot and willingness to collaborate.

It also seems to be important for the robot to produce appropriate verbal and non-verbal responses to an error. One study evaluated how a robot's gaze behavior (no gaze, looking at the other, looking down, and looking away) during mistakes change people's impressions (Shiomi et al., 2013). Experimental results showed that "looking at the other" outperformed different gaze behaviors, communicating degrees of perceived apologetics and friendliness and providing more reflection. Takayama et al. (2011) found that showing a goal-oriented reaction to a task outcome (i.e., disappointment in response to failure and happiness in response to success) made the robot appear smarter than when it did not react, regardless of whether the robot succeeded or failed in the task. Hamacher et al. (2016) found that demonstrating appropriate emotions and awareness of error (e.g., regret or enthusiasm) significantly tempers dissatisfaction with a robot's erroneous behavior and improves trust. Gieselmann (2006) evaluated user reactions to different robot error indicators and found that people preferred the robot asking a specific question to obtain additional information when it didn't understand their utterance. Indicating non-understanding with unspecific questions left users confused, since they did not know what the

robot did not understand, hindering their ability to solve the error.

Asking for Help

Several researchers proposed having robots request help from a human partner when they encounter an error (Ross et al., 2004; Hüttenrauch and Severinson-Eklundh, 2006; Rosenthal et al., 2012; Yasuda and Matsumoto, 2013; Knepper et al., 2015; Bajones et al., 2016). This strategy is computationally less expensive than re-planning, however it is not always applicable (e.g., when the people around do not have the ability or knowledge to help the robot solve the problem). In situations where it is applicable, asking for help can lead to negative experiences (e.g., Mutlu and Forlizzi, 2008) and can be very expensive in terms of monitoring time and cognitive load (Rosenthal et al., 2012). In such cases, it seems the way the robot asks for help matters. Knepper et al. (2015) developed a system that allows a robot to specify the kind of help that is needed in a way that removes as much ambiguity as possible. Users reported that they felt the system was more effective at communicating needs than other tested methods; preferring the precise requests over general phrasings. Moreover, the system improved the subjective evaluation of the robot and the speed and accuracy of human intervention when the robot experienced a problem. Maintaining polite communication also seems to matter: Yasuda and Matsumoto (2013) experimented with a robotic trashcan that spilled garbage, asked a person to pick up the trash for it and then "bowed" in appreciation. Most people found the experience to be positive, despite the spilled garbage and request for help. Another study found that participants who saw the robot stating its limitations before asking for help reported liking the robot more than those who saw control statements (Cameron et al., 2016b).

Rosenthal et al. (2012) sought to understand the willingness and availability of occupants to help a service robot. In their study, a robot visited different offices at different times of day, with different types of requests, and recorded willingness to provide help and the duration of that help. Participants were equally willing to help with all types of requests. Interestingly, willingness to help was not affected by the length of time the question took to answer nor the incentives the occupants received. In a related study, Srinivasan and Takayama (2016) evaluated factors that influence people's behavioral willingness to help a robot, finding that it depends on the robot's social role (peer or assistant), familiarity (new vs. 10 years experience), level of autonomy (autonomous or teleoperated), politeness strategy (direct request, positive politeness, negative politeness, or indirect request), and size of request (small or large). More specifically, people were more willing to help a peer robot that made smaller requests (i.e. that require less effort to fulfill), was more familiar, and used a positive politeness strategy (attended to the listener's wants, conveyed liking, and made the listener feel good about themselves). Moreover, Participants were nearly 50% quicker to help the robot when they believed that it was behaving autonomously rather than being teleoperated by a person.

The aforementioned work largely deals with preventing failures related to limited capabilities or missing information by proactively requesting help. However, some failures cannot be foreseen in advance and may not be included in the robot's planner (i.e., Black Swans; Sebok and Wickens, 2017). Bajones et al. (2016) performed a multi-user Wizard-of-Oz experiment in which they asked participants to help a malfunctioning robot restore the interaction flow after an error occurred. Results indicated that all 38 participants were willing to help the robot with repeated failure situations, regardless of the role they were given in the interaction ("director" or "builder"). Moreover, they found that the person who gave the last command was more likely to help, followed by the person who was closer. Malfunctions that could be actively fixed by the participants did not negatively impact perceived intelligence and likability ratings of the robot.

Mix and Match

Researchers have combined mitigation strategies in order to increase their effect. Spexard et al. (2008) implemented a model that decided on the best strategy based on the initiative taker and the solution provider of an error. Hardware defects caused the robot to inform the user of the reason why it could not move and ask for help, mode confusion or the robot behaving unexpectedly caused it to prompt the user to reset the system, software failures caused the robot to inform the user about the break-down, asking them to contact a technician. Using these help strategies, all participants successfully coped with the problem without external help.

There is very little work on comparing different failure recovery strategies. One exception is Lee et al. (2010), which investigated people's reactions to different recovery methods (apologies, compensation, and options for the user) in an online survey. All the recovery strategies increased positive ratings of the robot's politeness, however, only the apology strategy was effective in making the robot seem more competent, and in making the participants feel closer to the robot and liking it more. The compensation strategy was most effective in increasing perception of satisfaction with the service, but less effective than the apology and option strategies in increasing their perceived willingness to use the service again. The results also suggest that tailoring the recovery strategy to people's orientation to services is important—people with a relational orientation responded particularly well to an apology whereas those with a more utilitarian orientation responded better to compensation. Moreover, apologies were shown to be better for people who treated the robot more like an agent, while compensation was better for people who treating it like a tool. Another study that investigated different failure recovery strategies is Engelhardt and Hansson (2017), which compared between: "ignore" (the robot ignores that a failure has occurred and moves on with the task), "apology" (the robot apologizes for failing and moves on) and "problem solving" (the robot tries to solve the problem with the help of the human). Results showed that the apology strategy scored the lowest on likeability and perceived intelligence, and that the ignore strategy lead to better perceptions of perceived intelligence and animacy. Problem-solving clearly minimized

the negative effects of failure better than apologizing, but the "ignore" condition often scored at least as well as problem-solving.

Several theories have been suggested to explain successful mitigation strategies. According to Booth (1991), whether system errors are helpful or disruptive depends on (i) the ease with which the user can recover from an error; and (ii) the extent to which the system provides cues or features that productively direct the user toward a more appropriate understanding. In line with this theory, Brooks et al. (2016) argued that providing *human support* (providing information that supports or improves the user's situation awareness with respect to the failure and the status of the task being performed) or *task support* (helping the user complete the task they wanted to accomplish) will mitigate negative effects caused by failure; and that combining the two techniques should minimize problems without negative side effects. Moreover, they hypothesized that recovery strategies which reduce the negative effects of a failure will also increase the likelihood of users wanting to use the system again. To test these hypotheses, they conducted two between-subjects survey studies (Brooks et al., 2016). Results indicated that human support was better correlated to whether the information conveyed could be used by the person to affect the outcome of the situation. Task support, as well as a combination of task support and human support, significantly improved people's reaction to failure in all but one scenario. Recovery strategies that reduced the negative effects of a failure were shown to increase the likelihood of users to want to use the system again.

DISCUSSION

The majority of published works on robotic failures focus on technical aspects of making the robots more reliable. Few studies have actively worked toward making failure-handling user friendly, however the growing number of publications on the topic seems to indicate an increase in interest. Successful failure-handling strategies that enable untrained users to quickly and easily identify and solve failures require a holistic approach to design and development. The technical knowledge of hardware and software must be integrated with cognitive aspects of information processing, psychological knowledge of interaction dynamics and domain-specific knowledge of the user, the robot, the target application, and the environment. To achieve this, additional research is essential. By combining insights from a large variety of fields into a single framework, RF-HIP can be used to guide these discussions, and provides an initial hypothesis regarding how people might process symptoms and warnings in situations of robotic failure. In a similar manner to how C-HIP supports the design of new warnings and alerts, the stages of processing could be used to help determine why a particular approach of handling failure is successful while another is unsuccessful; leading to informed design tools and guidelines that facilitate the development of robot interactions that enable untrained users to quickly and easily identify and act upon failures.

Several gaps in the literature have become evident as a result of this analysis. First, it seems that most efforts have been focused on how failures influence user perceptions of the robot and user behavior, looking primarily at cause and effect. Little work has been done on evaluating how a robot should communicate that an error has occurred. Almost no work has been done to understand the underlying cognitive, psychological, and social determinants behind these relationships and how they may impact selection of mitigation strategies. Second, there seems to be a great asymmetry in the types of failures being studied and subsequent failure-handling strategies proposed: while there is a lot of emphasis on recovery strategies to cope with technical failures, there aren't any strategies to cope with recovery from human errors—equivalent to cancel or undo in HCI. Moreover, social-environmental considerations such as the work environment, group-level judgement, and organizational flaws have not been taken into consideration. Third, the importance of motivation to how people perceive, comprehend and solve robotic failures seems to be lost in the literature—studies typically evaluate people in unnatural settings, using tasks that are low in personal relevance. As a result, the ecological validity of most of the studies is low. It would be interesting to evaluate how motivation might influence responses in a more natural setting, when participants have a real stake in whether the robot will succeed or fail. Fourth, the failure attributes identified (functional severity, social severity, relevance, frequency, condition and symptoms) have not received almost any consideration in the HRI literature in terms of how they influence the way in which the failure should be communicated, the HRI, and the selection of mitigation strategies. For the most part, these attributes are unexplored territory and require targeted assessment. Lastly, since most studies used indoor, single-person environments, the effects of various aspects of the environment (e.g., other agents, weather, lighting, size of space) on perceptions of failures and preferences of communication and mitigation strategies remain unknown.

Another challenge the robotics community is facing in failure-handling is benchmarking and comparability. The wide variety of robotic implementations, evaluation environments and measures, coupled with lack of consistency on which implementation and evaluation details are reported in scientific publications, make it difficult and nearly impossible to compare subjective and objective performance metrics from different failure-handling studies. We are unaware of any frameworks that specify how all the contextual considerations identified in this paper should affect robot behavior in order to produce a pleasurable experience. Development of such frameworks are likely going to come from comparing and combining different implementation methods with insights from a wide variety of user studies. A common benchmark must be crafted for a set of robots, tasks, environments, and conditions. Consistent subjective measures and batteries of questionnaires along with clear quantitative evaluation measures must also be defined.

From the literature survey it is evident that many aspects remain to be studied in the field of user-centered failure handling, making this an exciting time to be active in the field. The

importance of studying cognitive considerations that critically influence naive users' ability to detect and solve robot failures is evident. While the current paper proposes how failure warnings and symptoms may be perceived by people, the specifics of the proposed framework must be thoroughly tested and verified. Moreover, whether the RF-HIP model can be used to predict the impact of various forms of robot design on a users' ability to handle failures is still to be determined. Hopefully, this review provides a good starting point for discussing what needs to be done in order to develop robot interactions that enable untrained users to quickly and easily identify and solve failures.

AUTHOR CONTRIBUTIONS

SH is the first author of this publication and main contributor. TO-G is her Ph.D. advisor.

FUNDING

REFERENCES

Abrams, R. A., and Christ, S. E. (2003). Motion onset captures attention. *Psychol. Sci.* 14, 427–432. doi: 10.1111/1467-9280.01458

Adubor, O., St. John, R., and Steinfeld, A. (2017). "Personal safety is more important than cost of damage during robot failure," in *Proceedings of the Companion of the 2017 ACM/IEEE International Conference on Human-Robot Interaction - HRI '17* (Vienna), 403–403.

Anderson, S. J., Cohen, G., and Taylor, S. (2000). Rewriting the past: Some factors affecting the variability of personal memories. *Appl. Cogn. Psychol.* 14, 435–454. doi: 10.1002/1099-0720(200009)14:5<435::AID-ACP662>3.0.CO;2-B

Baber, C., and Stanton, N. A. (1994). Task analysis for error identification: a methodology for designing error-tolerant consumer products. *Ergonomics* 37, 1923–1941.

Bainbridge, W. A., Hart, J., Kim, E. S., and Scassellati, B. (2008). "The effect of presence on human-robot interaction," in *Robot and Human Interactive Communication, 2008. RO-MAN 2008. The 17th IEEE International Symposium on* (Munich), 701–706.

Bajones, M., Weiss, A., and Vincze, M. (2016). "Help, anyone? a user study for modeling robotic behavior to mitigate malfunctions with the help of the user," in *5th International Symposium on New Frontiers in Human-Robot Interaction 2016* (Sheffield, UK).

Baraka, K., Rosenthal, S., and Veloso, M. (2016). "Enhancing human understanding of a mobile robot's state and actions using expressive lights," in *2016 25th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)* (New York, NY), 652–657.

Barakova, E. I., Bajracharya, P., Willemsen, M., Lourens, T., and Huskens, B. (2015). Long-term LEGO therapy with humanoid robot for children with ASD. *Expert Syst.* 32, 698–709. doi: 10.1111/exsy.12098

Baroni, I., Nalin, M., Zelati, M. C., Oleari, E., and Sanna, A. (2014). "Designing motivational robot: how robots might motivate children to eat fruits and vegetables," in *The 23rd IEEE International Symposium on Robot and Human Interactive Communication 2014 RO-MAN* (Edinburgh, UK), 796–801.

Bauer, L., Bravo-Lillo, C., Cranor, L. F., and Fragkaki, E. (2013). *Warning Design Guidelines*. Pittsburgh, PA. Available online at: http://repository.cmu.edu/cylab/113/ (Access date: 14-Dec-2017).

Beer, J. M., and Takayama, L. (2011). "Mobile remote presence systems for older adults : acceptance, benefits, and concerns," in *Proceedings of the 6th International Conference on Human-Robot Interaction (HRI)* (Lausanne), 19–26.

Bettman, J. R. (1979). Memory factors in consumer choice: a review. *J. Mark.* 43, 37–53.

Bodenhausen, G. V, and Wyer, R. S. (1987). "Social cognition and social reality: information acquisition and use in the laboratory and the real world," in *Social Information Processing and Survey Methodology*, eds H. Hippler, N. Schwarz, and S. Sudman (New York, NY: Springer), 6–41.

Booth, P. A. (1991). Errors and theory in human-computer interaction. *Acta Psychol.* 78, 69–96.

Bower, G. H. (1970a). Organizational factors in memory. *Cogn. Psychol.* 1, 18–46.

Bower, G. H. (1970b). Analysis of a mnemonic device: modern psychology uncovers the powerful components of an ancient system for improving memory. *Am. Sci.* 58, 496–510.

Bower, G. H. (2000). "A brief history of memory research," in *The Oxford handbook of memory*, eds E. Tulving and F. Craik (New York, NY: Oxford University Press), 3–32.

Bower, G. H., Monteiro, K. P., and Gilligan, S. G. (1978). Emotional mood as a context for learning and recall. *J. Verbal Learn. Verbal Behav.* 17, 573–585.

Breazeal, C. (2004). Social interactions in HRI: the robot view. *IEEE Trans. Syst. Man Cybern. Part C Appl. Rev.* 34, 181–186. doi: 10.1109/TSMCC.2004.826268

Brooks, D. J. (2017). *"A Human-Centric Approach to Autonomous Robot Failures,"* Ph.D. dissertation, Department of Computer Science, University. Lowell, MA.

Brooks, D. J., Begum, M., and Yanco, H. A. (2016). "Analysis of reactions towards failures and recovery strategies for autonomous robots," in *25th IEEE International Symposium on Robot and Human Interactive Communication, RO-MAN, 2016* (New York, NY), 487–492.

Buschman, T. J., and Miller, E. K. (2010). Shifting the spotlight of attention: evidence for discrete computations in cognition. *Front. Hum. Neurosci.* 4:194. doi: 10.3389/fnhum.2010.00194

Butter, M. J. (1970). Differential recall of paired associates as a function of arousal and concreteness-imagery levels. *J. Exp. Psychol.* 84, 252–256.

Cameron, D., Collins, E., Cheung, H., Chua, A., Aitken, J. M., and Law, J. (2016a). "Don't worry, we'll get there: Developing robot personalities to maintain user interaction after robot error," in *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, Vol. 9793 (Cham; Edinburgh, UK: Springer), 409–412.

Cameron, D., Loh, E. J., Chua, A., Collins, E., Aitken, J. M., and Law, J. (2016b). "Robot-stated limitations but not intentions promote user assistance," in *5th International Symposium on New Frontiers in Human-Robot Interaction* (Sheffield, UK), 1–6.

Canham, M., and Hegarty, M. (2010). Effects of knowledge and display design on comprehension of complex graphics. *Learn. Instr.* 20, 155–166. doi: 10.1016/j.learninstruc.2009.02.014

Canham, R., Jackson, A. H., and Tyrrell, A. (2003). "Robot error detection using an artificial immune system," in *Proceedings of the NASA/DoD Conference on Evolvable Hardware, 2003* (Chicago, IL), 199–207.

Card, S. K., Moran, T. P., and Newell, A. (1986). "The model human processor: an engineering model of human performance," in *Handbook of Perception and Human Performance, Vol. 2. Cognitive Processes and Performance*, eds K. R. Boff, L. Kaufman, and J. P. Thomas (Oxford, UK: John Wiley & Sons), 1–35.

Card, S. K., Newell, A., and Moran, T. P. (1983). *The Psychology of Human-Computer Interaction*. Hillsdale, NJ: L. Erlbaum Associates Inc.

Carlson, J., and Murphy, R. R. (2005). How UGVs physically fail in the field. *IEEE Trans. Robot.* 21, 423–437. doi: 10.1109/TRO.2004.838027

Carlson, K. A. (2011). The impact of humor on memory: is the humor effect about humor? *Humor Int. J. Humor Res.* 24, 21–41. doi: 10.1515/humr.2011.002

Carmody, M. A., and Gluckman, J. P. (1993). "Task specific effects of automation and automation failure on performance, workload and situational awareness," in *Proceedings of the Seventh International Symposium on Aviation Psychology* (Columbus, OH), 167–171.

Cassenti, D. N. (2007). Recovery from automation error after robot neglect. *Proc. Hum. Factors Ergon. Soc. Annu. Meet* 51, 1096–1100. doi: 10.1177/154193120705101809

Celsi, R. L., and Olson, J. C. (1988). The role of involvement in attention and comprehension processes. *J. Consum. Res.* 15, 210–224.

Cha, E., Dragan, A. D., and Srinivasa, S. S. (2015). "Perceived robot capability," in *2015 24th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)* (Kobe), 541–548.

Cha, E., Mataric, M., and Fong, T. (2016). "Nonverbal signaling for non-humanoid robots during human-robot collaboration," in *2016 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI), 2016* (Christchurch), 601–602.

Chapman, G. B., Li, M., Colby, H., and Yoon, H. (2010). Opting in vs opting out of influenza vaccination. *JAMA* 304, 43–44. doi: 10.1001/jama.2010.892

Chidambaram, V., Chiang, Y.-H., and Mutlu, B. (2012). "Designing persuasive robots: how robots might persuade people using vocal and nonverbal cues," in *Proceedings of the Seventh Annual ACM/IEEE International Conference on Human-Robot Interaction* (Boston, MA), 293–300.

Croskerry, P. (2003). The importance of cognitive errors in diagnosis and strategies to minimize them. *Acad. Med.* 78, 775–780. doi: 10.1097/00001888-200308000-00003

De Visser, E., and Parasuraman, R. (2011). Adaptive aiding of human-robot teaming: effects of imperfect automation on performance, trust, and workload. *J. Cogn. Eng. Decis. Mak.* 5, 209–231. doi: 10.1177/1555343411410160

Desai, M., Kaniarasu, P., Medvedev, M., Steinfeld, A., and Yanco, H. (2013). "Impact of robot failures and feedback on real-time trust," in *2013 8th ACM/IEEE International Conference on Human-Robot Interaction (HRI)* (Tokyo), 251–258.

Desai, M., Medvedev, M., Vázquez, M., McSheehy, S., Gadea-Omelchenko, S., Bruggeman, C., et al. (2012). *"Effects of changing reliability on trust of robot systems,"* in *Proceedings of the Seventh Annual ACM/IEEE International Conference on Human-Robot Interaction - HRI '12* (Boston, MA), 73.

Dewitz, P., Carr, E. M., and Patberg, J. P. (1987). Effects of inference training on comprehension and comprehension monitoring. *Read. Res. Q.* 22, 99–121. doi: 10.2307/747723

Drury, J. L., Scholtz, J., and Yanco, H., A. (2003). "Awareness in human-robot interactions," in *IEEE International Conference on Systems, Man and Cybernetics* (Washington, DC), 912–918.

Duncan, J. (1978). Response selection in spatial choice reaction: Further evidence against associative models. *Q. J. Exp. Psychol.* 30, 429–440. doi: 10.1080/00335557843000034

Dunning, D., Griffin, D. W., Milojkovic, J. D., and Ross, L. (1990). The overconfidence effect in social prediction. *J. Pers. Soc. Psychol.* 58, 568–581.

Egelman, S., Cranor, L. F., Hong, J. I., Herbsleb, J. D., and Bellovin, S. M. (2009). *Trust Me: Design Patterns for Constructing Trustworthy Trust Indicators.* Ph.D. dissertation, Department of Computer Science, Carnegie Mellon University, Pittsburgh, PA.

Embrey, D. E. (1986). "SHERPA: a systematic human error reduction and prediction approach," in *Proceedings of the International Topical Meeting on Advances in Human Factors in Nuclear Power Systems* (Knoxville, TN), 184–193.

Endsley, M. R. (1988). "Situation awareness global assessment technique (SAGAT)," in *Aerospace and Electronics Conference, NAECON 1988., Proceedings of the IEEE 1988 National* (Dayton, OH), 789–795.

Engelhardt, S., and Hansson, E. (2017). *"A Comparison of Three Robot Recovery Strategies to Minimize the Negative Impact of Failure in Social HRI,"* BS thesis, Department of Computer Science and Communication, KTH Royal Institute of Technology, Stockholm.

Epley, N., and Gilovich, T. (2006). The anchoring-and-adjustment heuristic: why the adjustments are insufficient. *Psychol. Sci.* 17, 311–318. doi: 10.1111/j.1467-9280.2006.01704.x

Feinstein, A., Brown, R., and Ron, M. (1994). Effects of practice of serial tests of attention in healthy subjects. *J. Clin. Exp. Neuropsychol.* 16, 436–447.

Fink, J., Mubin, O., Kaplan, F., and Dillenbourg, P. (2012). "Anthropomorphic language in online forums about Roomba, AIBO and the iPad," in *Advanced Robotics and its Social Impacts (ARSO), IEEE Workshop on 2012* (Munich), 54–59.

Fischer, K., Soto, B., Pantofaru, C., and Takayama, L. (2014). "Initiating interactions in order to get help: effects of social framing on people's responses to robots' requests for assistance," in *The 23rd IEEE International Symposium on Robot and Human Interactive Communication, RO-MAN* (Edinburgh, UK), 999–1005.

Fitts, P. M., and Deininger, R. L. (1954). SR compatibility: correspondence among paired elements within stimulus and response codes. *J. Exp. Psychol.* 48, 483–492.

Fitts, P. M., and Seeger, C. M. (1953). SR compatibility: spatial characteristics of stimulus and response codes. *J. Exp. Psychol.* 46, 199–210.

Fogg, B. J., Cuellar, G., and Danielson, D. (2009). "Motivating, influencing, and persuading users: an introduction to captology," in *Human Computer Interaction Fundamentals, 1st Edn*, eds A. Sears and J. A. Jacko (New York, NY: CRC press), 109–122.

Folk, C. L., Remington, R. W., and Johnston, J. C. (1992). Involuntary covert orienting is contingent on attentional control settings. *J. Exp. Psychol. Hum. Percept. Perform.* 18:1030.

Fong, T., Nourbakhsh, I., and Dautenhahn, K. (2003). A survey of socially interactive robots. *Rob. Auton. Syst.* 42, 143–166. doi: 10.1016/S0921-8890(02)00372-X

Frantz, J. P., and Rhoades, T. P. (1993). A task-analytic approach to the temporal and spatial placement of product warnings *Hum. Factors* 35, 719–730.

Frase, L. T., and Schwartz, B. J. (1979). Typographical cues that facilitate comprehension. *J. Educ. Psychol.* 71, 197–206.

Gehle, R., Pitsch, K., Dankert, T., and Wrede, S. (2015). "Trouble-based group dynamics in real-world HRI - Reactions on unexpected next moves of a museum guide robot," in *24th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN), 2015* (Kobe), 407–412.

Gieselmann, P. (2006). "Comparing error-handling strategies in human-human and human-robot dialogues," in *Proceedings of the 8th Conference National Language Process (KONVENS)* (Konstanz), 24–31.

Gieselmann, P., and Ostendorf, M. (2007). "Problem-sensitive response generation in human-robot Dialogs," in *Proceedings of the 8th SIGDial Workshop on Discourse and Dialogue* (Antwerp), 219–222.

Giuliani, M., Mirnig, N., Stollnberger, G., Stadler, S., Buchner, R., and Tscheligi, M. (2015). Systematic analysis of video data from different human–robot interaction studies: a categorization of social signals during error situations. *Front. Psychol.* 6:931. doi: 10.3389/fpsyg.2015.00931

Gobet, F., Lane, P. C., Croker, S., Cheng, P. C., Jones, G., Oliver, I., et al. (2001). *Chunking mechanisms in human learning. Trends Cogn. Sci.* 5, 236–243. doi: 10.1016/S1364-6613(00)01662-4

Gompei, T., and Umemuro, H. (2015). "A robot's slip of the tongue: Effect of speech error on the familiarity of a humanoid robot," in *24th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)* (Kobe), 331–336.

Gorbunova, E. S. (2017). Perceptual similarity in visual search for multiple targets. *Acta Psychol.* 173, 46–54. doi: 10.1016/j.actpsy.2016.11.010

Groom, V., Chen, J., Johnson, T., Kara, F., A., and Nass, C. (2010). "Critic, compatriot, or chump?: Responses to robot blame attribution," in *2010 5th ACM/IEEE International Conference on Human-Robot Interaction (HRI)* (Osaka), 211–217.

Haberlandt, K. (1982). Reader expectations in text comprehension. *Adv. Psychol.* 9, 239–249.

Ham, J. R. C., and Midden, C. J. H. (2011). "Persuasive robots can help save energy: the influence of the agency of a robot on its power to influence sustainable behavior," in *The 9th International Conference On Environmental Psychology* (Eindhoven).

Ham, J., and Spahn, A. (2015). "Shall i show you some other shirts too? The psychology and ethics of persuasive robots," in *A Construction Manual for Robots' Ethical Systems*, ed R. Trappl (Cham: Springer), 63–81.

Ham, J., Bokhorst, R., Cuijpers, R., van der Pol, D., and Cabibihan, J. J. (2011). "Making robots persuasive: the influence of combining persuasive strategies (gazing and gestures) by a storytelling robot on its persuasive power," in *Social Robotics. ICSR 2011. Lecture Notes in Computer Science*, Vol 7072, eds B. Mutlu, C. Bartneck, J. Ham, V. Evers, and T. Kanda (Berlin; Heidelberg; Amsterdam: Springer), 71–83. doi: 10.1007/978-3-642-25504-5_8

Hamacher, A. (2015). *"Believing in BERT: Making Good on Bad Robot Behavior,"* M.S. thesis, Department of Brain Science, University College London.

Hamacher, A., Bianchi-Berthouze, N., Pipe, A., G., and Eder, K. (2016). "Believing in BERT: using expressive communication to enhance trust and counteract operational error in physical Human-robot interaction," in *25th IEEE International Symposium on Robot and Human Interactive Communication, RO-MAN 2016* (New York, NY), 493–500.

Hancock, P. A., Billings, D. R., Schaefer, K. E., Chen, J. Y., de Visser, E. J., and Parasuraman, R. (2011). A meta-analysis of factors affecting trust in human-robot interaction. *Hum. Factors* 53, 517–527. doi: 10.1177/0018720811417254

Harris, R. J., Cady, E. T., and Tran, T. Q. (2006). "Comprehension and memory," in *Psychology of entertainment*, eds B. Jennings and P. Vorderer (Mahwah, NJ: Lawrence Erlbaum Associates), 71–84.

Hayes, C. J., Moosaei, M., and Riek, L., D. (2016). "Exploring implicit human responses to robot mistakes in a learning from demonstration task," in *25th IEEE International Symposium on Robot and Human Interactive Communication, RO-MAN, 2016* (New York, NY), 246–252.

Hick, W. E. (1952). On the rate of gain of information. *Q. J. Exp. Psychol.* 4, 11–26.

Houts, P. S., Doak, C. C., Doak, L. G., and Loscalzo, M. J. (2006). The role of pictures in improving health communication: a review of research on attention, comprehension, recall, and adherence. *Patient Educ. Couns.* 61, 173–190. doi: 10.1016/j.pec.2005.05.004

Hüttenrauch, H., and Severinson-Eklundh, K. (2006). To help or not to help a service robot: Bystander intervention as a resource in human-robot collaboration. *Interact. Stud.* 7, 455–477. doi: 10.1075/is.7.3.15hut

Hyman, R. (1953). Stimulus information as a determinant of reaction time. *J. Exp. Psychol.* 45, 188–196. doi: 10.1037/h0056940

Johnson, E. J., and Goldstein, D. (2003). Do defaults save lives? *Science* 302, 1338–1339. doi: 10.1126/science.1091721

Kahn, P. H., Kanda, T., Ishiguro, H., Gill, B. T., Ruckert, J. H., Shen, S., et al. (2012). "Do people hold a humanoid robot morally accountable for the harm it causes?" in *Proceedings of the Seventh Annual ACM/IEEE International Conference on Human-Robot Interaction - HRI '12* (New York, NY; Boston, MA: ACM), 33–40.

Kaniarasu, P., and Steinfeld, A. M. (2014). "Effects of blame on trust in human robot interaction," in *IEEE RO-MAN - 23rd IEEE International Symposium on Robot and Human Interactive Communication: Human-Robot Co-Existence: Adaptive Interfaces and Systems for Daily Life, Therapy, Assistance and Socially Engaging Interactions, 2014* (Edinburgh, UK), 850–855.

Kaniarasu, P., Steinfeld, A., Desai, M., and Yanco, H. (2013). "Robot confidence and trust alignment," in *2013 8th ACM/IEEE International Conference on Human-Robot Interaction (HRI)* (Tokyo; New York, NY: ACM), 155–156.

Kidd, C. D., and Breazeal, C. (2004). "Effect of a robot on user perceptions," in *Proceedings 2004 IEEE/RSJ International Conference on Intelligent Robots and Systems 2004 (IROS 2004)* (Sendai), 3559–3564.

Kieras, D. E., and Bovair, S. (1984). The role of a mental model in learning to operate a device. *Cogn. Sci.* 8, 255–273.

Kieras, D. E., and Meyer, D. E. (1997). An overview of the EPIC architecture for cognition and performance with application to human-computer interaction. *Hum. Comput. Interact.* 12, 391–438.

Kim, E., S., Leyzberg, D., Tsui, K., M., and Scassellati, B. (2009). "How people talk when teaching a robot," in *Proceedings of the 4th ACM/IEEE Conference on Human robot Interaction - HRI '09* (La Jolla, CA; New York, NY: ACM), 23.

Kim, H. Y., Kim, B., Jun, S., and Kim, J. (2017). "An imperfectly perfect robot: discovering interaction design strategy for learning companion," in *Proceedings of the Companion of the 2017 ACM/IEEE International Conference on Human-Robot Interaction - HRI '17* (Vienna, New York, NY: ACM), 165–166.

Kim, T., and Hinds, P. (2006). "Who should i blame? effects of autonomy and transparency on attributions in human-robot interaction," in *ROMAN 2006 - The 15th IEEE International Symposium on Robot and Human Interactive Communication*, 80–85.

Kintsch, W., Kozminsky, E., Streby, W. J., McKoon, G., and Keenan, J. M. (1975). Comprehension and recall of text as a function of content variables. *J. Verbal Learn. Verbal Behav.* 14, 196–214.

Kishiyama, M. M., and Yonelinas, A. P. (2003). Novelty effects on recollection and familiarity in recognition memory. *Mem. Cogn.* 31, 1045–1051. doi: 10.3758/BF03196125

Klein, R. M. (2000). Inhibition of return. *Trends Cogn. Sci.* 4, 138–147. doi: 10.1016/S1364-6613(00)01452-2

Knepper, R. A., Tellex, S., Li, A., Roy, N., and Rus, D. (2015). Recovering from failure by asking for help. *Auton. Robots* 39, 347–362. doi: 10.1007/s10514-015-9460-1

Kornblum, S., Hasbroucq, T., and Osman, A. (1990). Dimensional overlap: cognitive basis for stimulus-response compatibility-a model and taxonomy. *Psychol. Rev.* 97, 253–270.

Kwon, M., Huang, S. H., and Dragan, A. D. (2018). "Expressing Robot Incapability," in *Proceedings of the Thirteenth Annual ACM/IEEE International Conference on Human-Robot Interaction - HRI '18* (Chicago, IL; New York, NY: ACM), 87–95.

Lang, A. (2000). The limited capacity model of mediated message processing. *J. Commun.* 50, 46–70. doi: 10.1111/j.1460-2466.2000.tb02833.x

Laprie, J.-C. (1995). "Dependable computing and fault tolerance: concepts and terminology," in *25th International Symposium on Fault-Tolerant Computing, "Highlights from Twenty-Five Years"* (Pasadena, CA) 2–11.

Larrick, R. P., and Soll, J. B. (2008). The MPG illusion. *Science* 320, 1593–1594. doi: 10.1126/science.1154983

Law, E., Cai, Vicky., Liu, Q. F., Sasy, S., Goh, J., Blidaru, A., et al. (2017). "A Wizard-of-Oz study of curiosity in human-robot interaction," in *26th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN 2017)* (Lisbon), 607–614.

Lee, J. D., and See, K. A. (2004). Trust in automation: designing for appropriate reliance. *Hum. Factors.* 46, 50–80.doi: 10.1518/hfes.46.1.50_30392

Lee, M. K., Kiesler, S., Forlizzi, J., Srinivasa, S., and Rybski, P. (2010). "Gracefully mitigating breakdowns in robotic services," in *2010 5th ACM/IEEE International Conference on Human-Robot Interaction (HRI)* (Osaka), 203–210.

Lee, S. A., and Liang, Y. J. (2016). The role of reciprocity in verbally persuasive robots. *Cyberpsychol. Behav. Soc. Netw.* 19, 524–527. doi: 10.1089/cyber.2016.0124

Lemaignan, S., Fink, J., Mondada, F., and Dillenbourg, P. (2015). "You're doing it wrong! Studying unexpected behaviors in child-robot interaction," in *Social Robotics. ICSR 2015. Lecture Notes in Computer Science 9388 LNCS* (Cham; Paris: Springer), 390–400. doi: 10.1007/978-3-319-25554-5_39

Levin, I. P., and Gaeth, G. J. (1988). How consumers are affected by the framing of attribute information before and after consuming the product. *J. Consum. Res.* 15, 374–378.

Lohan, K. S., Deshmukh, A., and Aylett, R. (2014). "How can a robot signal its incapability to perform a certain task to humans in an acceptable manner?" in *The 23rd IEEE International Symposium on Robot and Human Interactive Communication* (Edinburgh, UK), 814–819.

Lucas, G. M., Boberg, J., Traum, D., Artstein, R., Gratch, Jonathan., Gainer, Alesia., et al. (2018). "Getting to know each other: the role of social dialogue in recovery from errors in social robots," in *Human-Robot Interaction '18* (New York, NY; Chicago, IL: ACM), 344–351.

Lucas, G. M., Nagai, Y., Komatsu, T., Hanheide, M., Natale L. (2017). "The role of social dialogue and errors in robots," in *HAI 2017 - Proceedings of the 5th International Conference on Human Agent Interaction* (New York, NY; Bielefeld: ACM), 431–433.

Lutz, R. R., and Woodhouse, R. M. (1999). "Bi-directional analysis for certification of safety-critical software," in *1st International Software Assurance Certification Conference (ISACC'99)* (Chantilly, VA).

Macias, W. (2003). A beginning look at the effects of interactivity, product involvement and web experience on comprehension: brand web sites as interactive advertising. *J. Curr. Issues Res. Advert.* 25, 31–44. doi: 10.1080/10641734.2003.10505147

Mani, K., and Johnson-Laird, P. N. (1982). The mental representation of spatial descriptions. *Mem. Cogn.* 10, 181–187.

Mayer, R. E. (2002). "Multimedia learning," in *Psychology of Learning and Motivation*, Vol. 41, ed B. H. Ross (Elsevier), 85–139. doi: 10.1016/S0079-7421(02)80005-6

McClelland, J. L. (1979). On the time relations of mental processes: an examination of systems of processes in cascade. *Psychol. Rev.* 86:287.

Mendoza, J. P., Veloso, M., and Simmons, R. (2015). "Plan execution monitoring through detection of unmet expectations about action outcomes," in *2015 IEEE International Conference on Robotics and Automation (ICRA)* (Seattle, WA), 3247–3252.

Miller, J. (1988). Discrete and continuous models of human information processing: theoretical distinctions and empirical results. *Acta Psychol.* 67, 191–257.

Mirnig, N., Giuliani, M., Stollnberger, G., Stadler, S., Buchner, R., and Tscheligi, M. (2015). "Impact of robot actions on social signals and reaction times in HRI error situations," in *Lecture Notes in Computer Science*, Vol. 9388, eds A. Tapus, E. André, J. C. Martin, F. Ferland, and M. Ammi (Cham), 461–471.

Mirnig, N., Stollnberger, G., Miksch, M., Stadler, S., Giuliani, M., and Tscheligi, M. (2017). To err is robot: how humans assess and act toward an erroneous social robot. *Front. Robot. AI* 4:21. doi: 10.3389/frobt.2017.00021

Morewedge, C. K., Yoon, H., Scopelliti, I., Symborski, C. W., Korris, J. H., and Kassam, K. S. (2015). Debiasing decisions: improved decision making with a single training intervention. *Pol. Insights Behav. Brain Sci.* 2, 129–140. doi: 10.1177/2372732215600886

Mubin, O., and Bartneck, C. (2015). "Do as I say: Exploring human response to a predictable and unpredictable robot," in *Proceedings of the 2015 British HCI Conference on - British HCI '15* (Lincoln; New York, NY: ACM) 110–116.

Murphy, R. R., and Hershberger, D. (1999). Handling sensing failures in autonomous mobile robots. *Int. J. Rob. Res.* 18, 382–400.

Mutlu, B., and Forlizzi, J. (2008). "Robots in organizations: the role of workflow, social, and environmental factors in human-robot interaction," in *Proceedings of the 3rd International Conference on Human Robot Interaction - HRI '08* (Amsterdam; New York, NY: ACM), 287.

Nakagawa, K., Shiomi, M., Shinozawa, K., Matsumura, R., Ishiguro, H., and Hagita, N. (2011). "Effect of robot's active touch on people's motivation," in *Proceedings of the 6th International Conference on Human-robot Interaction* (Lausanne: New York, NY: ACM), 465–472.

Nickerson, R. S. (1998). Confirmation bias: a ubiquitous phenomenon in many guises. *Rev. Gen. Psychol.* 2, 175–220.

Nielsen, J. (2001). "Error Message Guidelines," *Nielsen Norman Group*. Available online at: https://www.nngroup.com/articles/error-message-guidelines/ (Accessed January 01, 2018).

Nielson, J. (2000). "Why you only need to test with 5 users," *Nielsen Norman Group*. Available online at: https://www.nngroup.com/articles/why-you-only-need-to-test-with-5-users/. [Accessed: January 10, 2018].

Norman, D. (2002). *The Design of Everyday Things*. Cambridge, MA: The MIT Press.

Norman, D. A. (1981). Categorization of action slips. *Psychol. Rev.* 88, 1–15.

Okada, T., Sato, W., Murai, T., Kubota, Y., and Toichi, M. (2003). Eye gaze triggers visuospatial attentional shift in individuals with autism. *Psychologia* 46, 246–254. doi: 10.2117/psysoc.2003.246

Paivio, A., and Csapo, K. (1973). Picture superiority in free recall: imagery or dual coding? *Cogn. Psychol.* 5, 176–206.

Parasuraman, R., and Riley, V. (1997). Humans and automation: use, misuse, disuse, abuse. *Hum. Factors J. Hum. Factors Ergon. Soc.* 39, 230–253.

Payne, S. J. (1991). A descriptive study of mental models. *Behav. Inf. Technol.* 10, 3–21.

Payne, S. J. (2009). "Mental models in human-computer interaction," in *Human-Computer Interaction Fundamentals, 1st Edn*, eds A. Sears, J. A. Jacko (New York, NY: CRC press), 40–50.

Peltason, J., and Wrede, B. (2011). The curious robot as a case-study for comparing dialog systems. *AI Mag. 32*, 85–99.doi: 10.1609/aimag.v32i4.2382

Perry, C. M., Sheik-Nainar, M. A., Segall, N., Ma, R., and Kaber, D. B. (2008). Effects of physical workload on cognitive task performance and situation awareness. *Theor. Issues Ergon. Sci.* 9, 95–113. doi: 10.1080/14639220600959237

Posner, M. I., and Cohen, Y. (1984). Components of visual orienting. *Atten. Perform. X Control Lang. Process.* 32, 531–556.

Posner, M. I., Nissen, M. J., and Ogden, W. C. (1978). Attended and unattended processing modes: the role of set for spatial location. *Modes Perceiving Process. Inform.* 137:158.

Pratt, J., and McAuliffe, J. (2001). The effects of onsets and offsets on visual attention. *Psychol. Res.* 65, 185–191. doi: 10.1007/s004260100058

Proctor, R., and Vu, K.-P. (2009). "Human information processing: an overview for human-computer interaction," in *Human-Computer Interaction Fundamentals, 1st Edn*, eds A. Sears and J. A. Jacko (New York, NY: CRC press), 20–38.

Ragni, M., Rudenko, A., Kuhnert, B., and Arras, K., O. (2016). "Errare humanum est: Erroneous robots in human-robot interaction," in *25th IEEE International Symposium on Robot and Human Interactive Communication, RO-MAN 2016* (New York, NY), 501–506.

Rasmussen, J. (1982). Human errors. A taxonomy for describing human malfunction in industrial installations. *J. Occup. Accid.* 4, 311–333.

Reason, J. (1990). *Human Error*. Cambridge: Cambridge University Press.

Reeve, D. K., and Aggleton, J. P. (1998). On the specificity of expert knowledge about a soap opera: an everyday story of farming folk. *Appl. Cogn. Psychol.* 12, 35–42.

Reeve, T. G., and Proctor, R. W. (1990). The salient-features coding principle for spatial-and symbolic-compatibility effects. *Adv. Psychol.* 65, 163–180.

Renninger, K., and Wozniak, R. H. (1985). Effect of interest on attentional shift, recognition, and recall in young children. *Dev. Psychol.* 21, 624–632.

Riek, L. (2012). Wizard of Oz Studies in HRI: a systematic review and new reporting guidelines. *J. Human Robot Interact.* 1, 119–136. doi: 10.5898/JHRI.1.1.Riek

Ritchie, D., and Karge, B. D. (1996). Making information memorable: enhanced knowledge retention and recall through the elaboration process. *Prev. Sch. Fail. Altern. Educ. Child. Youth* 41, 28–33.

Robinette, P., Howard, A., and Wagner, A., R. (2017). "Conceptualizing overtrust in robots: why do people trust a robot that previously failed?" in *Autonomy and Artificial Intelligence: A Threat or Savior?*, eds W. Lawless, R. Mittu, D. Sofge, and S. Russell (Cham: Springer), 129–155.

Robinette, P., Li, W., Allen, R., Howard, A. M., and Wagner, A. R. (2016). "Overtrust of robots in emergency evacuation scenarios," in *2016 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI)* (Christchurch), 101–108.

Rosenthal, S., Veloso, M., and Dey, A. K. (2012). Is someone in this office available to help me? : Proactively seeking help from spatially-situated humans. *J. Intell. Robot. Syst. Theory Appl.* 66, 205–221. doi: 10.1007/s10846-011-9610-4

Ross, R., Collier, R., and O'Hare, G. M. P. (2004). "Demonstrating Social Error Recovery with AgentFactory," in *3rd International Joint Conference on Autonomous Agents and Multi-agent Systems (AAMAS04)* (New York, NY), 1424–1425.

Rossi, A., Dautenhahn, K., Koay, K. L., and Walters, M. L. (2017a). "How the timing and magnitude of robot errors influence peoples' trust of robots in an emergency scenario," in *Lecture Notes in Computer Science*, eds A. Kheddar, E. Yoshida, S. S. Ge, K. Suzuki, J.-J. Cabibihan, F. Eyssel, and H. He (Cham: Springer International Publishing), 42–52.

Rossi, A., Dautenhahn, K., Koay, K., L., and Walters, M., L. (2017b). "Human perceptions of the severity of domestic robot errors," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, Vol. 10652 LNAI, eds A. Kheddar, E. Yoshida, S. S. Ge, K. Suzuki, J.-J. Cabibihan, F. Eyssel, and H. He (Cham: Springer), 647–656.

Sadeghi, N., Mohd Kasim, Z., Hoon Tan, B., and Sathi Abdullah, F. (2012). Learning styles, personality types and reading comprehension performance. *Eng. Lang. Teach.* 5, 116–123. doi: 10.5539/elt.v5n4p116

Salem, M., Eyssel, F., Rohlfing, K., Kopp, S., and Joublin, F. (2013). To err is human(-like): effects of robot gesture on perceived anthropomorphism and likability. *Int. J. Soc. Robot.* 5, 313–323. doi: 10.1007/s12369-013-0196-9

Salem, M., Lakatos, G., Amirabdollahian, F., and Dautenhahn, K. (2015). "Would you trust a (Faulty) robot? effects of error, task type and personality on human-robot cooperation and trust," in *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction - HRI '15* (Portland, OR; New York, NY: ACM), 141–148.

Sarkar, S., Araiza-Illan, D., and Eder, K. (2017). *Effects of Faults, Experience, and Personality on Trust in a Robot Co-Worker*. Available Online at: https://arxiv.org/abs/1703.02335 (Submitted on: 7-Mar-2017).

Sato, W., Okada, T., and Toichi, M. (2007). Attentional shift by gaze is triggered without awareness. *Exp. Brain Res.* 183, 87–94. doi: 10.1007/s00221-007-1025-x

Schmidt, S. R. (1994). Effects of humor on sentence memory. *J. Exp. Psychol. Learn. Mem. Cogn.* 20, 953–867.

Schütte, N., Mac Namee, B., and Kelleher, J. (2017). Robot perception errors and human resolution strategies in situated human–robot dialogue. *Adv. Robot.* 31, 243–257. doi: 10.1080/01691864.2016.1268973

Schwarz, N. (1998). Accessible content and accessibility experiences: the interplay of declarative and experiential information in judgment. *Personal. Soc. Psychol. Rev.* 2, 87–99.

Sears, A., and Jacko, J. A. (eds.). (2009). *Human-Computer Interaction Fundamentals, 1st Edn*. New York, NY: CRC press.

Sebok, A., and Wickens, C. D. (2017). Implementing lumberjacks and black swans into model-based tools to support human-automation interaction. *Hum. Factors* 59, 189–203. doi: 10.1177/0018720816665201

Sheehan, P. W., and Antrobus, J. S. (1972). *The Function and Nature of Imagery*. Oxford, UK: Academic Press.

Shinozawa, K., Naya, F., Yamato, J., and Kogure, K. (2005). Differences in effect of robot and screen agent recommendations on human decision-making. *Int. J. Hum. Comput. Stud.* 62, 267–279. doi: 10.1016/j.ijhcs.2004.11.003

Shiomi, M., Nakagawa, K., Matsumura, R., Shinozawa, K., Ishiguro, H., and Hagita, N. (2010). "Could i have a word?': effects of robot's whisper," in *2010 IEEE/RSJ*

International Conference on Intelligent Robots and Systems (IROS) (Taipei), 3899–3904.

Shiomi, M., Nakawaga, K., and Hagita, N. (2013). Design of a gaze behavior at a small mistake moment for a robot. *Interact. Stud.* 14, 317–328. doi: 10.1075/is.14.3.01shi

Short, E., Hart, J., Vu, M., and Scassellati, B. (2010). "No fair!! An interaction with a cheating robot," in *2010 5th ACM/IEEE International Conference on Human-Robot Interaction (HRI)* (Osaka), 219–226.

Sideridis, G. D., Mouzaki, A., Simos, P., and Protopapas, A. (2006). Classification of students with reading comprehension difficulties: the roles of motivation, affect, and psychopathology. *Learn. Disabil. Q.* 29, 159–180. doi: 10.2307/30035505

Siegel, M. S. (2008). *Persuasive Robotics: How Robots Change our Minds.* M.S. thesis, Media Arts and Science, Massachusetts Institute of Technology, Cambridge, MA.

Simon, H. A. (1956). Rational choice and the structure of the environment. *Psychol. Rev.* 63, 129–138.

Spexard, T. P., Hanheide, M., Li, S., and Wrede, B. (2008). "Oops, something is wrong error detection and recovery for advanced human-robot-interaction," in *Proceedings of the ICRA workshop on Social Interaction with Intelligent Indoor Robots* (Pasadena, CA).

Srinivasan, V., and Takayama, L. (2016). "Help me please: robot politeness strategies for soliciting help," in *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems - CHI '16* (San Jose, CA), 4945–4955.

Stanton, N. A., and Salmon, P. M. (2009). Human error taxonomies applied to driving: a generic driver error taxonomy and its implications for intelligent transport systems. *Saf. Sci.* 47, 227–237. doi: 10.1016/j.ssci.2008.03.006

Stanton, N., A. (2009). "Human-error identification in human-computer interaction," in *Human-Computer Interaction Fundamentals, 1st Edn,* eds A. Sears, and J. A. Jacko (*New York, NY: CRC press*), 123–134.

Steinbauer, G. (2013). "A Survey about Faults of Robots Used in RoboCup," in *Lecture Notes in Computer Science,* Vol. 7500, eds Z. T. van der, X. Chen, P. Stone, L. E. Sucar (Berlin; Heidelberg: Springer), 344–355.

Summerfelt, H., Lippman, L., and Hyman I. E. Jr. (2010). The effect of humor on memory: constrained by the pun. *J. Gen. Psychol.* 137, 376–394. doi: 10.1080/00221309.2010.499398

Sutcliffe, A. and Rugg, G. (1998). A taxonomy of error types for failure analysis and risk assessment. *Int. J. Hum. Comput. Interact.* 10, 381–405.

Takayama, L., Dooley, D., and Ju, W. (2011). "Expressing thought: improving robot readability with animation principles," in *Proceedings of the 6th ACM/IEEE International Conference on Human-Robot Interaction (HRI)* (Lausanne), 69–76.

Takayama, L., Groom, V., and Nass, C. (2009). "I'm sorry, dave: i'm afraid i won't do that: social aspects of human-agent conflict," in *Proceedings of the 27th International Conference on Human Factors in Computing Systems - CHI 09* (Boston, MA), 2099.

Tannenbaum, K. R., Torgesen, J. K., and Wagner, R. K. (2006). Relationships between word knowledge and reading comprehension in third-grade children. *Sci. Stud. Read.* 10, 381–398. doi: 10.1207/s1532799xssr1004_3

Theeuwes, J. (1995). Abrupt luminance change pops out; abrupt color change does not. *Attent. Percep. Psychophys.* 57, 637–644.

Tokushige, H., Narumi, T., Ono, S., Fuwamoto, Y., Tanikawa, T., and Hirose, M. (2017). "Trust lengthens decision time on unexpected recommendations in human-agent interaction," in *Proceedings of the 5th International Conference on Human Agent Interaction - HAI '17* (Bielefeld), 245–252.

Torrey, C. (2009). *How Robots Can Help: Communication Strategies that Improve Social Outcomes. Ph.D.* dissertation, Computer Science, Carnegie Mellon University, Pittsburgh, PA.

Trung, P., Giuliani, M., Miksch, M., Stollnberger, G., Stadler, S., Mirnig, N., et al. (2017). "Head and shoulders: automatic error detection in human-robot interaction," in *Proceedings of the 19th ACM International Conference on Multimodal Interaction - ICMI 2017* (Glasgow, UK), 181–188.

Tsarouhas, P. H., and Fourlas, G. K. (2016). Mission reliability estimation of mobile robot system. *Int. J. Syst. Assur. Eng. Manag.* 7:220. doi: 10.1007/s13198-015-0408-9

Tversky, A., and Kahneman, D. (1973). Availability: a heuristic for judging frequency and probability. *Cogn. Psychol.* 5, 207–232.

Tversky, A., and Kahneman, D. (1974). Judgment under uncertainty: heuristics and biases. *Science* 185, 1124–1131.

Tversky, A., and Kahneman, D. (1981). The framing of decisions and the psychology of choice. *Science* 211, 453–458.

van der Woerdt, S., and Haselager, P. (2017). "Lack of Effort or Lack of Ability? Robot Failures and Human Perception of Agency and Responsibility," in *BNAIC 2016: Artificial Intelligence. Communications in Computer and Information Science*, Vol. 765, eds B. Bredeweg and T. Bosse (Cham: Springer), 155–168.

Van Lancker, D. (1991). Personal relevance and the human right hemisphere. *Brain Cogn.* 17, 64–92.

Verhaeghen, P., and Marcoen, A. (1996). On the mechanisms of plasticity in young and older adults after instruction in the method of loci: evidence for an amplification model. *Psychol. Aging* 11, 164–178. doi: 10.1037//0882-7974.11.1.164

von Grünau, M., Dubé, S., and Galera, C. (1994). Local and global factors of similarity in visual search. *Percept. Psychophys.* 55, 575–592.

Walters, M. L., Syrdal, D. S., Dautenhahn, K., Te Boekhorst, R., and Koay, K. L. (2008). Avoiding the uncanny valley: robot appearance, personality and consistency of behavior in an attention-seeking home scenario for a robot companion. *Auton. Robots* 24, 159–178. doi: 10.1007/s10514-007-9058-3

Wang, H., and Proctor, R. W. (1996). Stimulus–response compatibility as a function of stimulus code and response modality. *J. Exp. Psychol. Hum. Percept. Perform.* 22, 1201–1217.

Weinberger, M. G., and Gulas, C. S. (1992). The impact of humor in advertising: a review. *J. Advert.* 21, 35–59.

Weinstock, A., Oron-Gilad, T., and Parmet, Y. (2012). The effect of system aesthetics on trust, cooperation, satisfaction and annoyance in an imperfect automated system. *Work* 41 (Suppl. 1), 258–265. doi: 10.3233/WOR-2012-0166-258

Welsh, T. N., Chua, R., Weeks, D., and Goodman, D. (2009). "Perceptual-motor interaction: some implications for HCI," in *Human-Computer Interaction Fundamentals,* eds A. Sears and J. A Jacko (New York, NY: CRC Press), 7.

Wheeler, R. L., and Gabbert, F. (2017). Using self-generated cues to facilitate recall: a narrative review. *Front. Psychol* 8:1830. doi: 10.3389/fpsyg.2017.01830

Wogalter, M. S. (2006a). "Communication-human information processing (C-HIP) model," in *Handbook of Warnings,* ed M. S. Wogalter (Mahwah, NJ: Lawrence Erlbaum Associates), 51–61.

Wogalter, M. S. Ed. (2006b). "Warning channel: modality and media," in *Handbook of warnings,* ed M. S. Wogalter (Mahwah, NJ: Lawrence Erlbaum Associates), 841.

Wogalter, M. S., and Vigilante, W. J. (2006). "Attention switch and maintenance," in *Handbook of Warnings,* ed M. S Wogalter (Mahwah, NJ: Lawrence Erlbaum Associates).

Wogalter, M. S., Barlow, T., and Murphy, S. A. (1995). Compliance to owner's manual warnings: influence of familiarity and the placement of a supplemental directive. *Ergonomics* 38, 1081–1091.

Wogalter, M., S. (2006c). "Purposes and scope of warnings," in *Handbook of warnings*, eds M. S. Wogalter (Mahwah, NJ: Lawrence Erlbaum Associates), 3–9.

Yamaguchi, S., Tsuchiya, H., and Kobayashi, S. (1995). Electrophysiologic correlates of age effects on visuospatial attention shift. *Cogn. Brain Res.* 3, 41–49.

Yasuda, H., and Matsumoto, M. (2013). "Psychological impact on human when a robot makes mistakes," in *Proceedings of the 2013 IEEE/SICE International Symposium on System Integration* (Kobe), 335–339.

The Internet Is Not a Tool: Reappraising the Model for Internet-Addiction Disorder Based on the Constraints and Opportunities of the Digital Environment

Alessandro Musetti* and Paola Corsano

Department of Humanities, Social Sciences and Cultural Industries, Università Degli Studi di Parma, Parma, Italy

HISTORICAL OVERVIEW

The Internet was born in the United States in the second half of the twentieth century; it was initially used for military purposes but has since become a powerful instrument for nonmilitary use, including the exchange of information all over the world, thanks to the introduction of tools such as the web browser. *From the start*, the World Wide Web assumed several functions (e.g., recreation, education, and business) but preserved a private dimension. To connect, people needed access to an Internet-connected computer, which represented a separation from real life, or a virtual reality. A video-terminal device helped these people to immerse themselves in salient but virtual images and sounds; this immersion could induce symptoms such as dissociation (Schimmenti and Caretti, 2010). In the 1990s, scientists developed a conceptualization of the misuse of the Internet and of Internet-addiction disorder (IAD) that was coherent with their conception of the Internet as virtual reality. The strongest criterion for distinguishing healthy Internet use from misuse was connection time; this criterion was supported by several empirical studies regarding its relationship with psychopathological symptoms (Young, 1998; Quayle and Taylor, 2003; Musetti et al., 2016, 2017).

However, over the last two decades, Internet use has given rise to global sociocultural changes and has had important implications for the functioning of people's minds (Clowes, 2015). Today, digital and connectable tools such as smartphones are powerful, very small, portable, and (thanks to WiFi and cloud technology) able to store a great deal of salient information about people's lives. These tools thus assume the function of an e-memory (electronic memory) by expanding cognitive memory (Clowes, 2015). Virtual reality is no longer synonymous with the Internet, so there is a need to reformulate the conceptualization of the Internet by taking into account its evolution. The extent of digital information in every sphere of people's lives has caused the integration of the Internet into the cognitive tasks people perform in their daily routines, leading to the consideration of the Internet as part of an extended concept of cognition (Smart et al., 2017). The concept of the Internet as a tool to connect to a virtual reality that is separate from the real world is no longer current, so a new concept of the Internet that takes its environmental features into account is needed. This concept is in line with Floridi's (2014) idea of an infosphere that shapes people's reality. The conceptualization of the Internet as an environment rather than as a tool leads to the reformulation of IAD theory. If the Internet is not just a tool to be utilized, the theoretical model of IAD cannot be based on behavior connected to its overuse, misuse, or abuse.

Based on this opinion, we present arguments in favor of reconsidering the Internet as an environment rather than as a tool. In the following section, we explore the Internet's role in cognitive ecology, as well as the inadequacy of treating the Internet as a tool and thus of the current Internet-addiction model.

THE INTERNET AND COGNITIVE ECOLOGY

One conceptualization that could help explain the idea that the Internet is a superstructure within which people operate is that of cognitive ecology (Smart, 2017), which has been defined as "the multidimensional contexts in which we remember, feel, think, sense, communicate, imagine, and act, often collaboratively, on the fly, and in rich ongoing interaction with our environments" (Tribble and Sutton, 2011, p. 94). Today's society is digital (Lupton, 2015), and the Internet represents the main part of its cognitive ecology. In the theory of situated cognition (Robbins and Aydede, 2009), cognition is embodied (Gallagher, 2005), embedded (Rupert, 2004), extended, and distributed or collective (Smart et al., 2017). These theories reconceptualize cognition; instead of the classical, individualistic and intra-brain conception of cognition, these theories take into account the relationships among the brain, the body, and the environment to determine the functional products of the mind (Smart et al., 2017). Thanks to the Internet's development (in terms of devices, apps, and social platforms), it can be seen as the principal structure of embodied, embedded, extended, and distributed cognition. Proponents of the embodied-cognition thesis claim that extra-neural bodily factors shape the course of cognitive processing (Anderson, 2003; Shapiro, 2007, 2011). Mobile or wearable devices such as smartphones are today part of people's daily engagements, and they allow continuous online access, which shapes the course of their daily activities and interactions (Smart et al., 2017). By contrast, proponents of the embedded-cognition thesis claim that the extra-organismic environment plays a role (although not a constitutive one) in cognitive states and processes (Rupert, 2004), thus reallocating cognition to within biological boundaries (Smart et al., 2017). The Internet can be inserted within this vision of cognition. For example, augmented reality devices (Smart et al., 2017) such as Google Glass can enrich the sensory experience and have repercussions on cognitive processes. Advocates for the extended-cognition thesis claim that cognitive processes supervene on the relation between a cognitive agent and the social environment in which that agent is situated (Smart et al., 2017). Internal (biological) structures and external devices work in a pair relationship in which biological structures can perform the same operations as external factors (see Clark and Chalmers, 1998) or in a complementary relationship in which external devices can perform operations that biological structures cannot, and vice-versa (see Sutton, 2010; Heersmink, 2015, 2016). The debate regarding the parity or complementary of the Internet and the brain has not yet been resolved (Smart et al., 2017), and it is not our aim to discuss that issue here. What is important in this context is that

Internet devices are so widespread in the social environment that they are the principal external factor through which people's brains relate to and structure external representations; these devices have thus become integrated in people's cognitive architectures (Halpin et al., 2010). Consider the examples of how the use of GPS has modified people's spatial navigation, including its important impact on the neural mechanisms of spatial cognition (Maguire et al., 2000), or considering how Facebook use shapes the representation of the self, including an important impact on the self-concept. This effect is not merely about the interaction between a cognitive agent and environmental devices or about the scaffolding function that external factors have within the mind. The Internet is more than just a scaffold that guides and integrates the mind as it performs functions that the mind cannot accomplish alone (Sterelny, 2010). Rather, people created the Internet to meet people's needs, and the Internet's functions, such as that of e-memory, have changed the ways in which people remember and behave in the world (i.e., a person can recover remote information without having to store every piece of information from day to day). The Internet has changed people's brain structures, which have in turn evolved in such a way as to change how the Internet meets new needs (Clowes, 2013). This view requires consideration of the Internet as an extended function of the mind, including its actual effects on the development of the brain's circuits. In a similar vein, the advent of cooked food changed not only people's tastes but also their digestive functions and the structures of their jaws and teeth; it thus had repercussions on environmental adaptation and species conservation (Wrangham, 2009; Sterelny, 2010). The last thesis regarding the Internet's crucial role is that of distributed cognition. This thesis relates to the cognitive processes (e.g., focusing, reasoning, remembering, and problem-solving) that a collection of individuals share. Again, the Internet has allowed people to take advantage of a huge network of geographically distributed individuals who process cognitive operations at the same time and on the same issue. This opportunity boosts collaboration, information exchange, and the coordination of collective efforts and collective decision-making (Chi et al., 2008; Chi, 2009; Smart et al., 2017). These theories of cognition are today a matter of debate. Some authors have preferred one vision over others; others have considered the theories to not be mutually exclusive and to instead by various integrated aspects of cognition. In *the article*, we want to underline that, irrespective of the vision that one embraces, the Internet represents a fundamental part of cognitive processing. It not only boosts cerebral operations but also shapes, modulates, and changes neurobiological structures, functioning, and development; the Internet is also, in turn, shaped and developed in a process that resembles a spiral of mutual influence toward ever-higher steps of development.

In this sense, a view of the Internet as a mere tool to be utilized functionally or dysfunctionally, as in the model of Internet addiction, is reductive in this era. Thus, considering the Internet as a digital environment that encloses and characterizes cognitive processes is more useful for understanding the phenomenon that we are studying.

THE INTERNET AS MORE THAN A TOOL

Consider the people of the nineteenth century, who began to deal with great technological changes (due to the Second Industrial Revolution). The invention of the train, for example, represented a substantial change in the connection between long distances and/or in the amount of people or material carried. People also had to learn to use trains by acquiring new behaviors such as buying tickets and waiting for the departure time; these behaviors could be functional or dysfunctional (examples of the latter include buying an expensive ticket or getting on the wrong train). Although the train was intended as an instrument for traveling to a destination, its growth into a global network and its various functions (industrial, civil, and military) fostered the sociocultural revolution of the 1800s. The train changed the way people thought about industry; thus, in the nineteenth century, the bourgeoisie affirmed its power, and science and literature became more liberal. In other words, what began as a mere instrument evolved into an environmental change that people had to adapt to.

The example of the train concretely describes the difference between a tool and a sociocultural environment. The dynamics of the person–tool interaction have been thoroughly studied and represent the basis for the strong Vygotskian psychological tradition (Luria and Vygotsky, 1992). According to this tradition, children organize their behavior by learning to use tools or through external stimuli (Vygotsky, 1997). For example, a child might pay attention to a tool and then name the tool; the name of the tool thus becomes a word in the child's internal speech, thus inducing a new step in the child's reasoning and language functions (Bodrova et al., 2011). This explains how the development of higher brain functions is mediated by the utilization of tools, a view that fits well with the thesis of embodied cognition, according to which external tools shape the course of cognitive processing. It also fits with the thesis of scaffolding cognition, according to which external tools drive cognitive functioning. Within the latter conceptualization, the Internet can be seen as a tool through which people interact and whose use shapes the course of their cognitive processing. However, this view is reductive because it does not take into account the extra-brain operations that the Internet can provide but that the brain cannot. For instance, in the scaffolding view, people can interact with a social platform that reminds them of a salient episode that occurred in their past, thus shaping their emotional reactions and/or thoughts. However, in this view, social-platform interaction does not allow for the improvement of memory systems to provide a better ability to remember salient episodes from the past. Rather, the social platform is seen as a context inside which a limited memory system can take advantage of externally stored information, thus optimizing its work and allowing cognitive resources to be delivered to other processes. In other words, although the Internet—at least in its embryonic form, when recreation was the main online activity—was once considered a tool that shaped and mediated cognition and behavior, today, it is considered an environment that characterizes the people of today. To return to the example of the train, at the beginning, it was considered to be a tool for enhancing travel, but after a few decades, it began to shape the environment that characterized people in the industrial era. Interestingly, Floridi (2014) explained how tools, in addition to being utilized to boost behaviors, have also changed the sociocultural fabrics of various eras, thereby marking the evolution of humanity. The use of bronze (starting in 3000 BC) changed the prehistoric world into the Bronze Age. Similarly, today, people are part of an information society (also known as the infosphere) and can access whatever information they lack (e.g., facts about laws, politics, or science), meaning that there are no boundaries between their online and offline lives—a state known as "onlife" (Floridi, 2014).

As the reader may have noted, the arguments in favor of considering the Internet as an environment have multiplied and advanced. It is important to underline this vision here because the classical model and the resulting research into IAD are based on an obsolete conceptualization of the Internet as a tool.

REAPPRAISING INTERNET-ADDICTION DISORDER

Over the last three decades, the literature on this phenomenon has been abundant, but scholars have not reached an agreement on which criteria must be focused on when determining the dividing line between pathological or nonpathological Internet use (Musetti et al., 2016). The main models of Internet-related pathologies retrace those of other addictions (Young, 1998). If the theorists of IAD do not consider the Internet to constitute the current information society, they risk pathologizing a normal behavior, similarly to what happened for new addictions (as with new terms such as "shopaholic" or "workaholic"; see, e.g., Billieux et al., 2015). Without the environmental framework of the Internet, the theorization of pathological Internet use is limited to a reductive list of potentially problematic behaviors (Schimmenti, 2017), such as using the Internet for pornography or gambling. It is noteworthy that the *DSM-5* does not resolve this impasse, as it does not mention IAD; the only related disorder, online gaming disorder, is inserted in a section regarding diagnoses that require further study (American Psychiatric Association, 2013). The seven symptoms of IAD in the classical model are withdrawal; tolerance; concern over Internet use; heavier or more frequent Internet use than intended; centralized activities to obtain more from the Internet; loss of interest in other social, occupational, and recreational activities; and disregard for the physical or psychological consequences of Internet use (Young, 1998). These criteria must be present for at least 1 year. Clearly, these criteria are not applicable to the vision of the Internet as an environment. If the Internet constitutes the social fabric, it becomes impossible to withdraw from it, making it impossible to be concerned over Internet use; it likewise becomes impossible to focus on obtaining the Internet. In particular, the criterion of "heavier or more frequent use of the Internet than intended" lacks a comparative parameter in the environmental view of the Internet. How much Internet use is normal if the Internet is ingrained in every part of people's lives and also extends their cognition?

In the environmental view, considering the amount of time spent online to be a pathological criterion would mean seeing the entire information society as pathological. Moreover, and paradoxically, a rehabilitation treatment based on this criterion would be centered on reduced Internet access, thus limiting the use of extended and collective cognition (Smart et al., 2017), which could have important repercussions with regard to social adaptation that, in turn, would favor an increase in other pathological criteria, such as withdrawal from social occupation or recreation.

THE INTERNET AS A SOCIAL ENVIRONMENT

Our position is that the classical IAD model should be reformulated to match the vision of the Internet as a social environment. First, researchers must determine whether it is actually possible to be addicted to the Internet. In other words, can people become addicted to their social fabrics? Perhaps it is possible for a person to manifest difficulties or abnormalities when adapting to a social environment. In a similar vein, new models should ignore utilization-related criteria and instead focus on the symptoms that indicate social maladaptation, which may resemble manifestations of known symptoms such as dissociation, depression, anxiety, and personality disorder (Musetti et al., 2018). If this new focus were applied, a question would need be raised about what preexisting pathological conditions would predispose a person to have difficulty adapting to an environment (Caplan, 2002). Considering the Internet as the current socio-cognitive environment, a person's preexisting intra-brain features could favor the success or failure of the adaptation process. In an interesting model, scholars have suggested that maladaptive cognitions precede the symptomatology of IAD (Davis, 2001; Taymur et al., 2016),

thus underlining the comorbidity of IAD with heterogeneous psychopathological diagnoses (Orsal et al., 2013). A child presenting with an attention disorder will have some difficulty adapting to a school environment and to a social network of peers, and this difficulty will often impair the development of the child's intellectual and other cognitive functions. Similarly, a person who is cognitively poorly equipped could fail to take advantage of the Internet's contextual affordances (Ryding and Kaye, 2017). This could result in the unsuccessful extension and/or distribution of cognition processes, with repercussions for the person's cognitive development and risks of pathological adaptation to the digitized environment. A similar view could be used in studies on the appropriate treatments for cognitively predisposing features and to help explain the adaptation processes.

CONCLUSION

We are in favor of treating the Internet as a social environment in which a cognitive agent exists. Our proposal is that Internet use should not be seen as a mere instrumental action to achieve a goal (and which could be functional or dysfunctional); rather, we propose treating Internet use as an action situated in the digital context, as part of a system with a proper structure and rules. Considering the concept of the Internet as a social environment, the classical IAD model should be reformulated, as its implications are obsolete and misleading when applied to studies on the pathological population or on potential treatments.

AUTHOR CONTRIBUTIONS

AM: devised and structured the paper; PC: contribute to development and deep revision of the work, with literature analysis and agreement for final approval of the paper.

REFERENCES

American Psychiatric Association (2013). *Diagnostic and Statistical Manual of Mental Disorders, 5th Edn.* Washington, DC: American Psychiatric Association.

Anderson, M. L. (2003). Embodied cognition: a field guide. *Artif. Intell.* 149, 91–130. doi: 10.1016/S0004-3702(03)00054-7

Billieux, J., Thorens, G., Khazaal, Y., Zullino, D., Achab, S., and Van der Linden, M. (2015). Problematic involvement in online games: a cluster analytic approach. *Comput. Human Behav.* 43, 242–250. doi: 10.1016/j.chb.2014.10.055

Bodrova, E., Leong, D. J., and Akhutina, T. V. (2011). When everything new is well-forgotten old: Vygotsky/Luria insights in the development of executive functions. *New Dir. Child Adolesc. Dev.* 2011, 11–28. doi: 10.1002/cd.301

Caplan, S. E. (2002). Problematic Internet use and psychosocial well-being: development of a theory-based cognitive–behavioral measurement instrument. *Comput. Hum. Behav.* 18, 553–575. doi: 10.1016/S0747-5632(02)00004-3

Chi, E. H. (2009). "Augmented social cognition: using social web technology to enhance the ability of groups to remember, think, and reason," in *35th SIGMOD International Conference on Management of Data* (New York, NY: ACM), 973–984.

Chi, E. H., Pirolli, P., Suh, B., Kittur, A., Pendleton, B. A., and Mytkowicz, T. (2008). "Augmented social cognition," in *AAAI Spring Symposium on Social Information Processing* (Stanford, CA), 11–17.

Clark, A., and Chalmers, D. (1998). The extended mind. *Analysis* 58, 7–19. doi: 10.1093/analys/58.1.7

Clowes, R. W. (2013). The cognitive integration of E-memory. *Rev. Philos. Psychol.* 4, 107–133. doi: 10.1007/s13164-013-0130-y

Clowes, R. W. (2015). Thinking in the cloud: the cognitive incorporation of cloud-based technology. *Philos. Technol.* 28, 261–296. doi: 10.1007/s13347-014-0153-z

Davis, R. A. (2001). A cognitive-behavioral model of pathological Internet use. *Comput. Human Behav.* 17, 187–195. doi: 10.1016/S0747-5632(00)00041-8

Floridi, L. (2014). *The 4th Revolution: How the Infosphere Is Reshaping Human Reality.* Oxford: Oxford University Press.

Gallagher, S. (2005). *How the Body Shapes the Mind.* Oxford: Oxford University Press.

Halpin, H., Clark, A., and Wheeler, M. (2010). "Towards a philosophy of the web: representation, enaction, collective intelligence," in *Web Science Conference* (Raleigh, NC), 1–5.

Heersmink, R. (2015). Dimensions of integration in embedded and extended cognitive systems. *Phenomenol. Cogn. Sci.* 14, 577–598. doi: 10.1007/s11097-014-9355-1

Heersmink, R. (2016). The internet, cognitive enhancement, and the values of cognition. *Minds Machines* 26, 389–407. doi: 10.1007/s11023-016-9404-3

Lupton, D. (2015). *Digital Sociology.* Abingdon: Routledge.

Luria, A. R., and Vygotsky, L. S. (1992). *Ape, Primitive Man, and Child Essays in the History of Behavior*. New York, NY: Harvester Wheatsheaf.

Maguire, E. A., Gadian, D. G., Johnsrude, I. S., Good, C. D., Ashburner, J., Frackowiak, R. S., et al. (2000). Navigation-related structural change in the hippocampi of taxi drivers. *Proc. Natl. Acad. Sci. U.S.A.* 97, 4398–4403. doi: 10.1073/pnas.070039597

Musetti, A., Cattivelli, R., Giacobbi, M., Zuglian, P., Ceccarini, M., Capelli, F., et al. (2016). Challenges in internet addiction disorder: is a diagnosis feasible or not? *Front. Psychol.* 7:842. doi: 10.3389/fpsyg.2016.00842

Musetti, A., Cattivelli, R., Zuglian, P., Terrone, G., Pozzoli, S., Capelli, F., et al. (2017). Internet addiction disorder o internet related psychopathology? [Internet Addiction disorder or Internet Related Psychopathology?]. *Giornale Ital. Psicol.* 44, 359–382. doi: 10.1421/87345

Musetti, A., Terrone, G., and Schimmenti, A. (2018). An exploratory study on problematic Internet use predictors: which role for attachment and dissociation? *Clin. Neuropsychiatry* 15, 35–41.

Orsal, O., Unsal, A., and Ozalp, S. S. (2013). Evaluation of Internet addiction and depression among university students. *Proc. Soc. Behav. Sci.* 82, 445–454. doi: 10.1016/j.sbspro.2013.06.291

Quayle, E., and Taylor, M. (2003). Model of problematic internet use in people with a sexual interest in children. *CyberPsychol. Behav.* 6, 93–106. doi: 10.1089/109493103321168009

Robbins, P., and Aydede, M. (2009). "A short primer on situated cognition," in *The Cambridge Handbook of Situated Cognition,* eds P. Robbins and M. Aydede (Cambridge: Cambridge University Press), 3–10.

Rupert, R. (2004). Challenges to the hypothesis of extended cognition. *J. Philos.* 101, 389–428. doi: 10.5840/jphil2004101826

Ryding, F. C., and Kaye, L. K. (2017). "Internet Addiction": a conceptual minefield. *Int. J. Ment. Health Addict.* 16:225. doi: 10.1007/s11469-017-9811-6

Schimmenti, A. (2017). Psicopatologia e uso problematico di internet. commento a ≪internet addiction disorder o internet related psychopathology?≫ [psychopathology and problematic internet use. commentary on ≪internet addiction disorder or internet related psychopathology?≫] *Giornale Ital. Psicol.* 44, 1007–1013. doi: 10.1421/88779

Schimmenti, A., and Caretti, V. (2010). Psychic retreats or psychic pits?: Unbearable states of mind and technological addiction. *Psychoanal. Psychol.* 27:115. doi: 10.1037/a0019414

Shapiro, L. A. (2007). The embodied cognition research programme. *Philos. Compass* 2, 338–346. doi: 10.1111/j.1747-9991.2007.00064.x

Shapiro, L. A. (2011). *Embodied Cognition*. Abingdon: Routledge.

Smart, P. (2017). Situating machine intelligence within the cognitive ecology of the internet. *Minds Machines.* 27, 357–380. doi: 10.1007/s11023-016-9416-z

Smart, P., Heersmink, R., and Clowes, R. W. (2017). "The cognitive ecology of the Internet," in *Cognition Beyond the Brain*, eds S. J. Cowley and F. Vallée-Tourangeau (ChamSpringer), 251–282.

Sterelny, K. (2010). Minds: extended or scaffolded?. *Phenomenol. Cogn. Sci.* 9, 465–481. doi: 10.1007/s11097-010-9174-y

Sutton, J. (2010). "Exograms and interdisciplinarity: history, the extended mind, and the civilizing process," in *The Extended Mind*, ed R. Menary (Cambridge, MA: MIT Press), 189–225.

Taymur, I., Budak, E., Demirci, H., Akdag, H. A., Güngör, B. B., and Özdel, K. (2016). A study of the relationship between internet addiction, psychopathology and dysfunctional beliefs. *Comput. Human Behav.* 61, 532–536. doi: 10.1016/j.chb.2016.03.043

Tribble, E., and Sutton, J. (2011). Cognitive ecology as a framework for Shakespearean studies. *Shakespeare Stud.* 39, 94–103.

Vygotsky, L. S. (1997). "The history of the development of higher mental functions," in *The Collected Works of L. S. Vygotsky*: Vol. 4, eds R. Rieber and A. Carton (New York, NY: Plenum Press), 1–26.

Wrangham, R. (2009). *Catching fire: How Cooking Made us Human*. London: Profile.

Young, K. S. (1998). Internet addiction: the emergence of a new clinical disorder. *Cyberpsychol. Behav.* 1, 237–244. doi: 10.1089/cpb.1998.1.237

Hypernatural Monitoring: A Social Rehearsal Account of Smartphone Addiction

Samuel P. L. Veissière[1,2,3,4]* and Moriah Stendel[1,3,4]

[1] Department of Psychiatry, McGill University, Montreal, QC, Canada, [2] Department of Anthropology, McGill University, Montreal, QC, Canada, [3] Raz Lab in Cognitive Neuroscience, McGill University, Montreal, QC, Canada, [4] Culture, Mind, and Brain Program, McGill University, Montreal, QC, Canada

We present a deflationary account of smartphone addiction by situating this purportedly antisocial phenomenon within the fundamentally *social* dispositions of our species. While we agree with contemporary critics that the hyper-connectedness and unpredictable rewards of mobile technology can modulate negative affect, we propose to place the locus of addiction on an evolutionarily older mechanism: the human need to monitor and be monitored by others. Drawing from key findings in evolutionary anthropology and the cognitive science of religion, we articulate a *hypernatural monitoring* model of smartphone addiction grounded in a general *social rehearsal* theory of human cognition. Building on recent predictive-processing views of perception and addiction in cognitive neuroscience, we describe the role of social reward anticipation and prediction errors in mediating dysfunctional smartphone use. We conclude with insights from contemplative philosophies and harm-reduction models on finding the right rituals for honoring social connections and setting intentional protocols for the consumption of social information.

Keywords: smartphone addiction, social neuroscience, evolutionary anthropology, predictive-processing, cultural affordances, social rehearsal, hungry ghosts

INTRODUCTION

As this paper was undergoing final review, a new wave of editorials about the noxious effects of smartphone use was sweeping the news. Major Apple shareholders, backed by petitions from customers, were now demanding that the tech giant address the growing problem of smartphone addiction and its impact on children's development (Kawa, 2018). As cognitive scientists who have studied the impact of the internet on human behavior (Veissière, 2016a,b), our aim is to present a nuanced view of the relationship between mobile information technology and human well-being. While we agree that excessive smartphone use can be detrimental to mental health, we aim to recast current understandings of the mechanisms involved in these addictive patterns in a broader evolutionary focus.

In this paper, we offer the provocative claim that current moral panics over smartphone addiction overlook a factor of fundamental importance: there is nothing inherently addictive about mobile technology. We suggest, rather, that it is the *social* expectations and rewards of connecting with other people and seeking to learn from others that induce and sustain addictive relationships with smartphones. Much has been said about Internet addiction and the new medias and technologies that connect us and make us lonely at the same time, leading to adverse mental health consequences (Twenge, 2017).The deeply prosocial

nature of these mechanisms, however, is often understated. Compulsive smartphone use, we claim, is not so much antisocial as fundamentally social. Specifically, we argue that mobile technology addiction is driven by the human urge to connect with people, and the related necessity to be seen, heard, thought about, guided, and monitored by others, that reaches deep in our social brains and far in our evolutionary past.

Smartphones, we claim, provide a potentially unhealthy platform for another healthy impulse. As we will see, they can also enable us to remember and celebrate the role of other people in making us who we are, and help us treasure the bonds that make us a uniquely social species.

In fleshing out the social roots of smartphone addiction – and by extension, of human behavior and well-being – we do not intend to produce a general meta-theory that dismisses other, non-social forms of excessive smartphone use. The hyper-sociality of smart-device addiction, rather, may likely occur on a continuum from the directly social to the indirectly social.

Playing video-games, outsourcing difficult tasks like memorizing schedules or spatial orientation, and having instant access to news and information are among of battery of everyday smartphone functions that are known to be highly addictive (Alter, 2017). At a glance, these domains are not readily apparent as social. From an evolutionary perspective, however, the human capacity to function optimally in any environment (and indeed human intelligence itself) is predicated on having access to a large, cumulative repertoire of contextually relevant cultural information devised by others, and that no single individual could invent on her own, or recreate alone in her own lifetime (Henrich, 2016; Mercier and Sperber, 2017). Seeking news and information, to put it simply, are ways to *learn from others,* and to stay updated on *culturally relevant* events and people. Video-gaming is similarly underpinned by social dimensions that may not be readily visible to users and critics alike. While many video-games involve explicit social rewards from playing online with other users (Snodgrass et al., 2016) other uniquely addictive smartphone games like Candy crush do not. The unpredictable rewards derived from so-called "ludic loops" of increased difficulty (Alter, 2017), as we expand in the Section "Predictive-Processing and Smartphones," typically activate neurobiological systems that increase reward-seeking behavior and addictions in other domains (West et al., 2015). In the next section, we present findings supporting the hypothesis that most smartphone notifications, from email and texting to social media, modulate addictive behavior through the anticipation of *social rewards.* The rewards derived from playing games, however, are social in more indirect ways. The human drive for gaming and competition, indeed, is also rooted in social evolutionary mechanisms, in which intra- and inter-group competition have helped drive the iterative spread of skill, knowledge, and technology from generation to generation (Bell et al., 2009; Richerson et al., 2016). In seeking to excel at a difficult game, we are rehearsing excellence in particular domains of skill, but also in the domain of social competition itself. Smartphones, as we will argue, provide a hyper-efficient extension of deep evolutionary urges for connection with others, learning from others, but also comparing ourselves to and competing with others.

The Sociality of Smartphone Use

When it comes to smartphone use, current scientific literature and intuitive wisdom are overwhelmingly pessimistic, warning us of the dangers these new technologies enable. According to current research, smartphone use is associated with depression (Steers et al., 2014; Andreassen et al., 2016), materialism (Lee et al., 2014; Twenge, 2017), and social anxiety (Billieux et al., 2015; Emanuel et al., 2015; Hussain et al., 2017), spawning a generation of anti-social, chronically anxious, self-obsessed 'zombies' (Lu and Lo, 2017). While these findings raise important concerns about the 'dark side' of smartphone use, they tend to focus on new technologies as the sole locus of addiction and pathology. We propose to bring this problem into a broader evolutionary focus, and will go on to argue that the current 'smartphone obsession' is neither grounded in, nor indicative of a paradigmatic shift in the psychosocial context in which human experience is invariably framed. Popular accounts, we argue, miss the mark on a crucially important factor: it is not so much smartphones themselves that are addictive, but rather the *sociality* that they afford. We insist that this drive for sociality is a fundamental feature of human evolution that predates smartphones by hundreds of thousands – by some accounts several millions – of years (Hrdy, 2007). Simply put, smartphone addiction is hyper-social, not anti-social.

There is ample evidence to support the claim that smartphone use is inherently prosocial, and by extension, that this prosociality is a core locus of smartphone addiction. First, the majority of smartphone use is spent on social activities such as social networking, text messaging, and phone calls (Li and Chung, 2006; Lopez-Fernandez et al., 2014). Even less interactive smartphone use, like information seeking or surfing the web, has now become implicitly social: 'likes', views, and comments are social indices of prestige and collective attention. Second, individuals who use their devices for primarily social purposes are quicker to develop habitual smartphone use (Van Deursen et al., 2015). These findings suggest that it is not just the smartphone itself that is addictive but rather the—direct or indirect—social interaction it enables.

Gendered dimensions of smartphone addiction provide further clues into its inherent sociality. Current findings in evolutionary psychology and social neuroscience indicate that women are on average more proficient at social cognition and tend to display more prosocial behavior than men (Eckel and Grossman, 1998; Andreoni and Vesterlund, 2001; Meier, 2007; Laasch and Conaway, 2009; Rand et al., 2016; Soutschek et al., 2017; see Espinosa and Kovářík, 2015 for alternate explanations). This gender discrepancy is maintained in smartphone use, with numerous studies showing that women use their phones for social purposes significantly more than men do (Tufekci, 2008; Van Deursen et al., 2015). According to our hypothesis, the prosocial nature of female smartphone use would render females more susceptible to addiction. Recent estimates confirm this view: females are more likely to develop addictive smartphone behaviors, experience more anxiety if they cannot use their smartphones, and feel less in control over checking their phones (Thompson and Lougheed, 2012; Van Deursen et al., 2015).

Imagined Other Minds Guide Our Expectations

Despite minor gendered differences in social cognition, it is not controversial that humans as a whole are a prosocial species. Beyond amply documented findings in developmental psychology attesting to the intrinsic co-evolutionary links between cognition and sociality (Moll and Tomasello, 2007; Tomasello, 2009; Tomasello et al., 2012), recent research on mind-wandering has shown that a large part of our spontaneous mental lives is devoted to rehearsing social scenarios. A recent large-scale investigation using experience-sampling, for example, demonstrated that nearly half of waking time is spent in mind-wandering episodes unrelated to the task at hand (Killingsworth and Gilbert, 2010). Although science on daydreaming often describes the consequences of a wandering mind (e.g., Mrazek et al., 2013), it is likely premature to believe that a cognitive function that occupies such a large percentage of mental life does not confer some adaptive benefit. To explain the ubiquity of mind-wandering, Poerio and Smallwood (2016) have proposed that the phenomenon is evolutionarily adaptive, serving as a platform for offline social cognition. Supporting this view, research shows that all but a small fraction of daydreaming involves social scenarios (Mar et al., 2012; Song and Wang, 2012). Moreover, mind-wandering and social cognition rely on shared neural activation, whereby the neural activity that occurs during daydreaming significantly overlaps with that of core social processes like mentalizing and perspective taking – the very processes that enable an individual to socially flourish (Poerio and Smallwood, 2016). Recent models on the evolution of depression help confirm this social hypothesis for the mechanisms of ordinary cognition. In a series of influential papers, Paul Andrews and colleagues have argued that 'depression' (a disorder characterized by cognitive rumination) confers specific *social* advantages to help keep social problems in mental focus. Again, it is of note that women (who are demonstrably more proficient than men at social cognition) experience depression at much higher rates than men. Andrews and colleagues see this as further evidence that a significant part of mental life is dedicated to rehearsing social scenarios (Andrews and Thomson, 2009; Andrews et al., 2012, 2015). All in all, a growing consensus between developmental psychology, cognitive neuroscience, and phenomenology strongly suggests that humans are almost always thinking about and *through* other people (Frith, 2002; Tomasello, 2009; Mar et al., 2012; Ramstead et al., 2016). The time is ripe, then, to elaborate a generalized social rehearsal theory of cognition. In the following sections, we expand on this theory and apply it to smartphone use.

Social Media and Internet Notifications as Hyper-Natural Monitoring

In a series of recent papers, Ramstead et al. (2016; see also Ramstead et al., 2017; Veissière, 2017) have described symbolically enriched human worlds as organized landscapes of "cultural affordances" grounded in mutual, recursively nested expectations about shared standards of behavior. 'Culture', on this view can be conceptualized as patterned allocations of attention; that is, the practice of selectively paying attention, ascribing meaning, and guiding behavior to certain features of the world according to what we expect others to also expect and pay attention to. While what is made salient through collectively shaped attentional preferences acquires different values and affords different experiences from group to group, the capacity for shared attention extrapolated to large groups of generalized 'like me' others is a species-wide disposition – the very disposition, mediated by joint-intentionality, that gives rise to cultural forms of life among Homo Sapiens (Ramstead et al., 2016; Veissière, 2017).

On this view, over the course of normal cognitive and social development, humans learn to see the world through the perspective of other people and intuitively imagine context-relevant agents (usually imbued with prestige) to guide them in their actions (Veissière, 2017). From context to context and moment to moment, we outsource a large part of our thinking, feelings, and decision-making to sometimes explicit, most often implicit scenarios of the "what would so-and-so think, feel, or expect me to do" variety.

This reassuring feeling of being watched and guided by imaginary others has been hypothesized to play an important role in the evolution of cooperation, morality, organized religion, and large-scale social life (Whitehouse, 2004; Boyer, 2008; Norenzayan and Shariff, 2008; Atran and Henrich, 2010; Norenzayan et al., 2013). According to this view, often called the *super-natural monitoring hypothesis,* we fashioned our Gods and Spirits to better flesh out the imaginary agents that guide our ordinary cognition, consciousness, action, and moral attitudes.

Instant text messaging, email, and social media provide a platform for our hungry need to be connected, but also for our need to watch and monitor others, and better still, for our need to be seen, heard from, thought about, monitored, judged, and appraised by others. We might call this the *hyper-natural monitoring hypothesis.*

The prevailing – and hyperbolic – view on smartphone use is that it is a sly weapon, responsible for pandemic-like waves of mass loneliness, anxiety, insecurity, materialism, and narcissism among today's youth – particularly the so-called 'digital natives' born after 1994 (Roberts et al., 2015; Weiser, 2015; Pearson and Hussain, 2015; Twenge, 2017). As Jean Twenge has pointed out in her recent book on digital natives (Twenge, 2017), the advent of electronically mediated childhoods in the West was also concurrent with a general shift in parenting culture, and the rise of so-called 'helicopter parenting'[1] in particular. Drawing on extensive survey research, she points out that children and youth born after 1994 spent considerably less unsupervised time socializing with their peers than their forebears, and significantly more time on electronic devices. While precise causality behind

---

[1] "Helicopter parenting" is used as a derogatory term to describe obsessive parental supervision in most dimensions of children's lives. Although the phrase first appeared in the l960′s (Ginott, 1965/2009), it is often said to characterize the post-1980s childrearing culture of "hovering around" one's child. "Lawnmower parenting" (where one paves the way for children in all aspects of their lives), is sometimes used to describe more extreme forms of helicopter parenting. In November 2017, the Economist reported that parents in the United States and nine European countries (except for France), now spent 50% more time with their children than in 1965 (The Economist, 2017).

these two correlated factors cannot be ascertained, we can only note that youth who otherwise do not interact with their peers "in real life" (*irl* in internet lingo) seek to do so with the means available to their generation. Online-mediated life, more to the point, is always, already real life, and as such, it is inherently social.

What current moral panics about digital media often fail to consider, thus, is that *the desire to see and be seen*, and *judge and be judged* is precisely *about other people*. There is nothing abnormal, as such, about seeking self-worth through other people's point of view. We propose, thus, to think of this urge as fundamentally normal, and anchored in core mechanisms of social cognition that are distinct to our species. On our social rehearsal and monitoring view, smartphones simply equip us with a novel medium to channel innate human sociality. Their proclivity to induce addiction, in turn, simply points to how much others matter to us and how we want to matter to them.

PREDICTIVE-PROCESSING AND SMARTPHONES

If the primary motivation of smartphone use is prosocial, why can this technology lead to such negative outcomes? We turn to the science of addiction to describe how mobile technology in particular has sent us into a vortex of anxiety-inducing, hyper-excited, hyper-monitoring.

A Brief Venture into the Neuroscience of Addiction

The exact nature and neurochemical correlates of smartphone addiction are currently unknown (Elhai et al., 2017). Key insights from the neuroscience of learning and addiction, however, can offer important insights into our attachment to the strange flickering and buzzing bricks that seem to regulate our lives.

As we have seen, smartphone use is at once constitutive of and constituted by a complex landscape of sociality. This landscape, however, is also modulated by notifications from dozens of applications that deliver beeps and buzzes, mostly to alert us that another human has interacted with us. We should now consider where and how 'addiction' fits in this picture. Social interaction (digital or not) activates the dopaminergic reward circuits in the basal ganglia (See Krach et al., 2010 for a review). It is important to note that these same circuits are implicated in addictive drug use (Belin et al., 2009), compulsive video-gaming, and reward-seeking in general (West et al., 2015). These are circuits that are also responsible for associative learning: the process by which an individual learns to associate two stimuli (Hebb, 1976; Seger, 2006; Yin and Knowlton, 2006). For associative learning to occur, an initial exposure to a new stimulus must occur alongside a reflex-eliciting stimulus. With a smartphone, nearly all notifications that the user encounters elicit a social value and thus activate the dopaminergic reward circuit, leading the user to anticipate and seek these rewarding notifications. With each occurrence this link grows stronger, and the user will anticipate and seek these rewarding notifications, paving the road for habitual behavior.
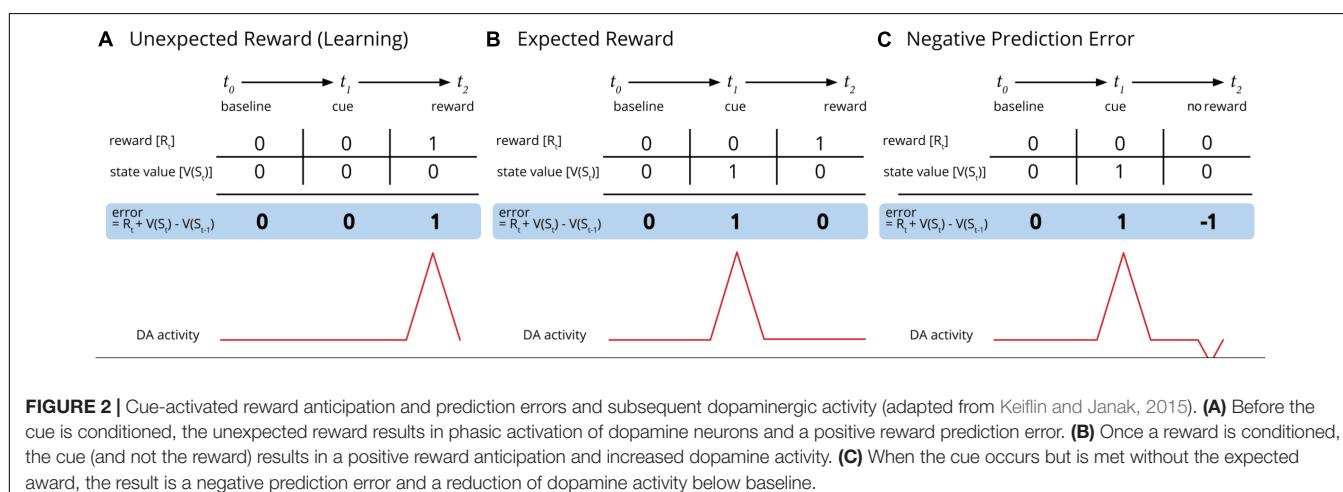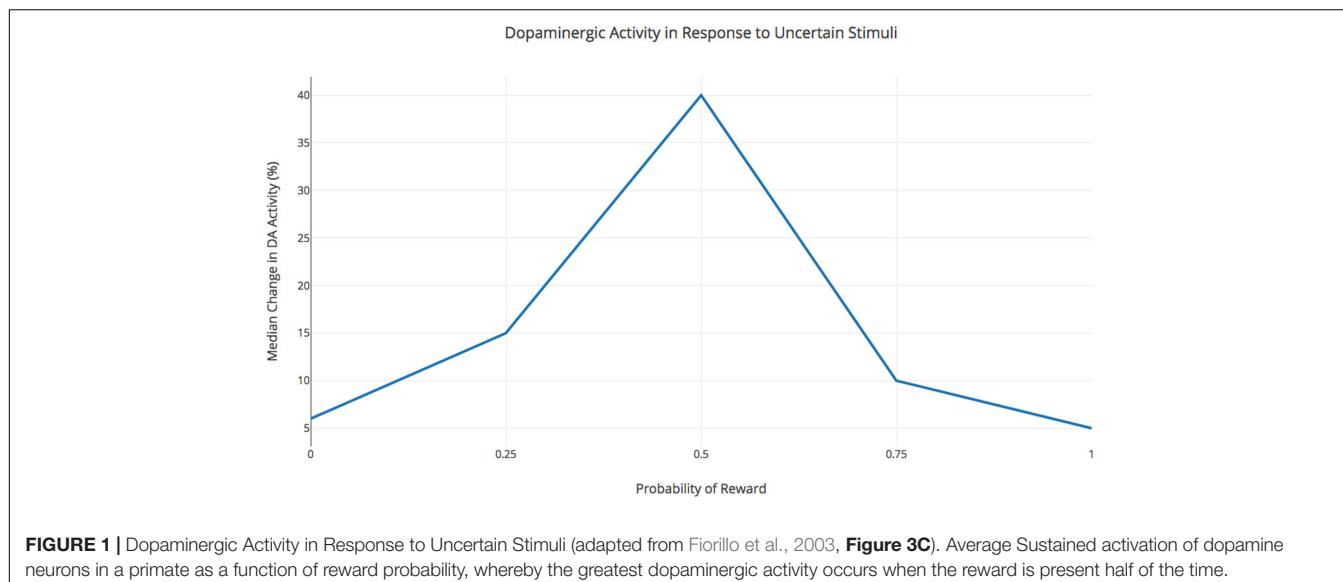
The dopaminergic system regulates two functions that govern addiction: the *anticipation of reward* and *outcome evaluation* (Linnet, 2014). An important finding about dopamine and addiction, however, is that dopaminergic surges typically occur *before the reward*, or more precisely when a cue (e.g., a beep indicating that one can press a lever) signals the reliable delivery of a reward (e.g., from pulling a lever). Because arousal decreases with frequent and predictable exposure, reward anticipation is a much more powerful mediator of strong addictions than outcome evaluations of the stimulus itself (Fiorillo et al., 2003; van Holst et al., 2012). According to this finding, addictions become strongest when we cannot figure out the pattern of when to reliably expect them (van Holst et al., 2012). Behavioral scientists call these addiction-inducing patterns *intermittent reinforcement* or *variable ratio schedules* (Zuriff, 1970). Neuroscientists have identified that a cue triggering a behavior that yields a reward 50% of the time is by far the most anxiety-inducing of delivery schedules. A reward delivered 75% of the time, for example can be reliably expected to deliver *most of the time*. A cue signaling a reward that delivers 25% of the time can similarly be expected *not* to deliver most of the time. Such high-predictability schedules (when the brain can reliably predict what is going to happen) typically trigger low arousal. At a 50% delivery rate, a reward schedule is still predictable enough to be enticing, but unpredictable enough to be anxiety-inducing (Fiorillo et al., 2003).

The point to take home here is that arousal is more highly correlated with reward anticipation than with the reward itself. When rewards become most unpredictable, in turn, arousal typically becomes negative, giving rise to anxiety (**Figure 1**).

Indeed, the beeps and buzzes of smartphone notifications provide just such an intermittent, variable, unpredictable, but uniquely desirable schedule of rarely met anticipation rewards, thus providing chaotic patterns of reward anticipation that trigger very strong modes of arousal. Because of the deeply social nature of the rewards our phones make us crave, we often become entrenched in vicious cycle of addiction (**Figure 1**).

Cravings as Prediction Errors

According to predictive-processing and free-energy theories of cognition, we do not immediately perceive the world as it is. Rather than directly respond to environmental stimulus, we first process information through our *expectations*. Immediate perception, in other words, first occurs through behavioral self-predictions modulated by prior experience (Friston and Kiebel, 2009; Ramstead et al., 2016). On this view, our brains generate statistical models of the world based on prior learning to provide us with predictions of what will arise in experience and how to act accordingly. In doing so our brains predict upcoming sensory states and compare them with actual sensory states, minimizing the differences between these distributions through constant updates of priors and actions (i.e., learning) (Ramstead et al., 2016, 2017). As our perceptual system constantly attempts to reduce uncertainty by computing abysmal amounts of disordered information to make it predictable, discrepancies between prediction and perception – *prediction errors* in the lingo – become commonplace. Cravings, on this view, could

FIGURE 1 | Dopaminergic Activity in Response to Uncertain Stimuli (adapted from Fiorillo et al., 2003, **Figure 3C**). Average Sustained activation of dopamine neurons in a primate as a function of reward probability, whereby the greatest dopaminergic activity occurs when the reward is present half of the time.

FIGURE 2 | Cue-activated reward anticipation and prediction errors and subsequent dopaminergic activity (adapted from Keiflin and Janak, 2015). **(A)** Before the cue is conditioned, the unexpected reward results in phasic activation of dopamine neurons and a positive reward prediction error. **(B)** Once a reward is conditioned, the cue (and not the reward) results in a positive reward anticipation and increased dopamine activity. **(C)** When the cue occurs but is met without the expected award, the result is a negative prediction error and a reduction of dopamine activity below baseline.

be conceptualized as prediction errors (Tobler et al., 2006) (**Figures 2**, **3**).
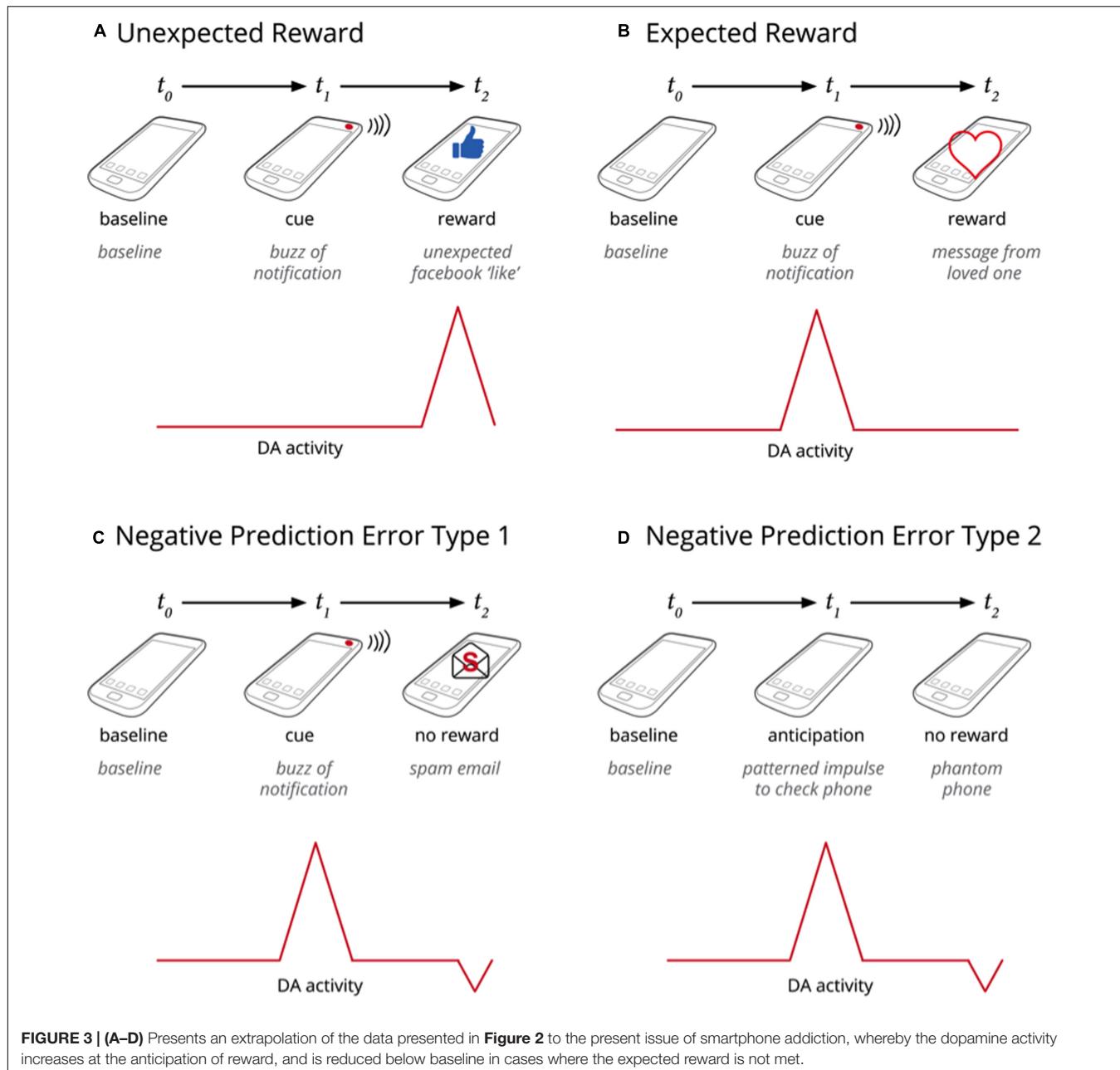
As we mentioned above, associative learning and free-energy models can explain the pervasive expectation that the *anticipation* of smartphone notifications predicts an upcoming social reward. In turn, the intermittent schedule of smartphone notifications promotes stronger anticipations and more compulsive expectations, subsequently inducing prediction errors and affective disappointment.

Notifications are cues for checking behavior that eventually becomes habitual, even without the initial alert (Oulasvirta et al., 2012; Elhai et al., 2017). Recent studies reveal the magnitude of this habitual checking behavior, with the average individual spending over 3 h a day on their smartphone (Alter, 2017), tapping, typing, or swiping an average of 2617 times every day (dscout, 2016). The majority of users go on to experience prediction errors in the form of hallucinations that their phone is vibrating, a phenomenon entitled *phantom phone* (Sauer et al., 2015). These prediction errors reinforce habitual phone checking

behaviors, which are a common gateway to smartphone addiction (Oulasvirta et al., 2012). Prediction errors can also occur in more subtle, but equally frequent and distressing way when precise patterned expectations are not met: a beep that we hope may be a message from a loved one or a Instagram 'like', for example, may turn out to be an incoming spam email or a message from one's boss about an overdue task.

THE DARK SIDE OF SOCIAL MONITORING?

Key models of ordinary cognition, like predictive processing, free-energy, associative learning, and social rehearsal, all offer clues to elucidate the newfangled phenomenon of smartphone addiction. We have seen that smartphone addiction harnesses basic human proclivities for social monitoring and associative learning. While we largely intend this paper to add a hopeful note about potentially healthy social causes of smartphone addiction

FIGURE 3 | (A–D) Presents an extrapolation of the data presented in **Figure 2** to the present issue of smartphone addiction, whereby the dopamine activity increases at the anticipation of reward, and is reduced below baseline in cases where the expected reward is not met.

amidst current panics, we cannot dismiss the growing consensus described above on such negative outcomes as depression, anxiety, and loneliness.

Smartphone use and depression are strongly correlated, and one causal theory suggests that smartphones, which are frequently used to access social networks, provide a platform for which to frequently (often negatively) compare oneself to others (Steers et al., 2014). We have argued, however, that social monitoring is a fundamentally normal – indeed necessary – part of ordinary human cognition. Classical evolutionary accounts of this propensity have emphasized the human fondness for gossip (Dunbar, 2004) and social comparison (Festinger, 1954) as conferring adaptive advantages to assess threats, track trends

and shifts in others' social status, and locate credible sources of cultural information and behavioral guides (Henrich, 2016). We add that comparing ourselves to others and against cultural norms also enables us to derive meaning, motivation, purpose, and a sense of identity. With socially connected smartphones, this evolutionary process simply runs on overdrive. We can now constantly and relentlessly engage in hyper-speed comparisons with social media content that is biased toward positivity. As media researchers have suggested, this continual stream of positive information about others allows users to repeatedly perform upward social comparisons and negative self-evaluations against a so-called "highlight reel" (Steers et al., 2014). Despite the obvious antigenic nature of cyber-mediated social comparisons,

these accounts fail to acknowledge that the desire to socially connect is an even stronger motivator of smartphone use than the desire to do better than others.

To further address the non-benign concerns of smartphone overuse, the following section will once again employ theories of ordinary cognition to propose actions individuals can take to build happy, healthy relationships with mobile technology.

FEEDING OUR HUNGRY GHOSTS

If smartphone addiction rests on the fundamentally human proclivity toward prosociality, we can also learn to harness our social nature to pacify our cravings – or as Buddhic philosophies would put it, we can learn to sate our hungry ghosts.

In classical Buddhism, all creatures are said to undergo six life cycles, or go through six realms of existence (Levitt, 2003; Maté, 2008). They begin in Hell, where their life is described as constant torture, before moving on to the realm of Hungry Ghosts, where they are plagued by insatiable thirst, hunger, and cravings. Next comes the realm of Animals: a world of servitude and stupidity. This realm is followed by Asura, a world of anger, jealousy, and never-ending conflict. The Human realm comes next: a world of contradictions and indecisiveness; sweet and sour, hot and cold, happy and sad, good and evil. The human realm is a world of almost-thereness – wisdom and enlightenment are within reach, but never quite attained. Whether the next world of Deva-gati, or Heavenly Beings, offers final relief is open for debate (Levitt, 2003). It is world of intense pleasures, with intense miseries to match. Freedom from suffering, in the end, seems nowhere to be found. On a contemporary psychological reading, the Six Realms metaphor can also describe the quality and intentionality (aboutness) of the various states of consciousness and affect one will routinely encounter throughout the course of a day.

The Hungry Ghosts in this story can be understood as the state that regulates our cravings. This idea likely predates Buddhic philosophies, and is found in earlier Indian religions under the Sanskrit name *Preta* (Levitt, 2003). Pretas are supernatural creatures plagued by insatiable hunger and thirst. They have enormous stomachs, but very thin necks that can only support eating tiny things. In many Buddhist and Zen rituals, such as the *Oryoki* approach to eating and living, a single grain of rice is offered to Hungry Ghosts to acknowledge their existence and appease them a little (Levitt, 2003). The key here is to feed our Hungry Ghosts, and to find *just the right amount*. As we discuss further in our conclusion, this is consistent with harm-reduction approaches to addiction treatment that advocate responsible use over abstinence (Marlatt, 1996; Marlatt et al., 2011).

Recognizing smartphone cravings as Hungry Ghosts presents the opportunity to turn phone addiction into a intentional, just-enough ritual.

Set Intentional Protocols

Many smartphone users feel trapped by their phones (Harmon and Mazmanian, 2013). The first step toward freedom from phone Hungry Ghosts, as we have seen, is to regain control of the pattern and make it predictable again. Switching off

all sounds and notifications can help to 'un-ring' Pavlov's proverbial bell and cull habitual checking behaviors. As we described above, smartphone addiction is mediated by the grasp of intermittent reinforcement schedules of social rewards. With this in mind, setting regular intervals to check one's phone can reduce the strong cravings that arise from chaotic patterns of reward anticipation. When it comes to instant phone-mediated communication, we can also make our intentions and expectations transparent, and agree on protocols with others. Clear workplace communication policies, for example, those that prohibit evening and weekend emails, or setting clear expectation for time-windows in replying have been shown to be effective in reducing stress and increasing productivity (Mark et al., 2012). Similar 'policies' and clear expectations for when to text or not to text – what we call 'intentional protocols' – can be devised among friends, families and lovers.

CONCLUSION

Like all natural proclivities, social monitoring and rehearsal can turn into Hungry Ghosts. The parallel with natural hunger and eating bear relevance to our argument about mobile technology. Blaming the rice, utensils, or kitchenware for one's insatiable gluttony does not so much deflate the problem as miss the mark entirely. The root of addictions, as we have seen, is not in substances or rewards themselves, and much less in the technologies that deliver such rewards, but in the *anticipation* of rewards and in delivery schedules and rituals. The hard truth about cravings is that they are ultimately self-referential: cravings are about cravings first and foremost.

Smartphones and mobile technologies are not the root cause of modern distress. In post-industrial environments where foods are abundant and readily available, our cravings for fat and sugar sculpted by distant evolutionary pressures can easily go into insatiable overdrive and lead to obesity, diabetes, and rampant heart disease (Henrich, 2016; Harari, 2017). As we argued in this paper, the prosocial needs and rewards of a physically weak species that relied on collective parenting (Hrdy, 2009) and distributed knowledge (Tomasello, 2014; Henrich, 2016) to survive and carve a moral niche in a harsh world can similarly be hijacked to produce a manic theater of hyper-social monitoring. Smartphones may be equated to hyper-efficient kitchenware. Both technologies help optimize the processing and delivery of specific kinds of basic needs: food on the one hand, and social information on the other. The key to eating well and being good social beings lies in finding the quality and intensity of *consumption rituals*. As in the *oriyoki* 'just the right amount' hungry ghost feeding ritual, the recipe lies in setting appropriate intentions, quality of awareness, and pacing for the time, place, and amount of information, connection, and comparison one will consume. Turning off notifications, as we have seen, has been shown to help users regain control of when and why to check their devices intentionally (Alter, 2017). When used to judicious social ends, smartphone and social media use can yield many positive outcomes, from increased subjective well-being (Kim and Lee, 2011) to better romantic relationships (Steers et al., 2014).

To conclude, we recognize that there is a controversy in addiction research between abstinence-based and harm-reduction approaches (Marlatt, 1996; Marlatt et al., 2011). The latter approach, which we advocate in this article, supports safe and responsible use, and consideration of the complexities of the social context in which people are drawn to substance use. While recent studies have shown that temporarily giving up certain social media activities could increase subjective well-being (see Alter, 2017, for a review), the professional and social consequences of giving up smartphone use altogether are currently not known, and are likely to be costly in a age that requires instant connection in so many domains of social life.

Individuals, rather, can mobilize their intrinsic drive toward sociality to mitigate the negative and increase the positive effects of smartphone use. Pursuing healthy social connection is the antidote. Rather than use smartphones to compare our lives to the distorted slice of reality others present, we can use them as communication tools to foster genuine emotional relationships. When competitive comparison seems inevitable, we can subvert into a motivator or reminder of our own unique skills – or better yet, we can cultivate genuine joy for the achievements of others (Chandra, 2017).

AUTHOR CONTRIBUTIONS

SV provided the theoretical framework based on his previous work on cultural affordances and internet sociality. MS helped refine the theoretical framework and further ground it in neuroscience. SV and MS contributed equally to the writing.

REFERENCES

Alter, A. (2017). *Irresistible: The Rise of Addictive Technology and the Business of Keeping us Hooked*. London: Penguin Books.

Andreassen, C. S., Billieux, J., Griffiths, M. D., Kuss, D. J., Demetrovics, Z., Mazzoni, E., et al. (2016). The relationship between addictive use of social media and video games and symptoms of psychiatric disorders: a large-scale cross-sectional study. *Psychol. Addict. Behav.* 30, 252–262. doi: 10.1037/adb0000160 doi: 10.1037/adb0000160

Andreoni, J., and Vesterlund, L. (2001). Which is the fair sex? Gender differences in altruism. *Q. J. Econ.* 116, 293–312. doi: 10.1162/003355301556419

Andrews, P. W., Bharwani, A., Lee, K. R., Fox, M., and Thomson, J. A. Jr. (2015). Is serotonin an upper or a downer? The evolution of the serotonergic system and its role in depression and the antidepressant response. *Neurosci. Biobehav. Rev.* 51, 164–188. doi: 10.1016/j.neubiorev.2015.01.018

Andrews, P. W., and Thomson, J. A. Jr. (2009). The bright side of being blue: depression as an adaptation for analyzing complex problems. *Psychol. Rev.* 116, 620–654. doi: 10.1037/a0016242

Andrews, P. W., Thomson, J. A. Jr., Amstadter, A., and Neale, M. C. (2012). *Primum non nocere*: an evolutionary analysis of whether antidepressants do more harm than good. *Front. Psychol.* 3:117. doi: 10.3389/fpsyg.2012.00117

Atran, S., and Henrich, J. (2010). The evolution of religion: how cognitive by-products, adaptive learning heuristics, ritual displays, and group competition generate deep commitments to prosocial religions. *Biol. Theory* 5, 18–30. doi: 10.1162/BIOT_a_00018

Belin, D., Jonkman, S., Dickinson, A., Robbins, T. W., and Everitt, B. J. (2009). Parallel and interactive learning processes within the basal ganglia: relevance for the understanding of addiction. *Behav. Brain Res.* 199, 89–102. doi: 10.1016/j.bbr.2008.09.027

Bell, A. V., Richerson, P. J., and McElreath, R. (2009). Culture rather than genes provides greater scope for the evolution of large-scale human prosociality. *Proc. Natl. Acad. Sci. U.S.A.* 106, 17671–17674. doi: 10.1073/pnas.0903232106

Billieux, J., Maurage, P., Lopez-Fernandez, O., Kuss, D. J., and Griffiths, M. D. (2015). Can disordered mobile phone use be considered a behavioral addiction? An update on current evidence and a comprehensive model for future research. *Curr. Addict. Rep.* 2, 156–162. doi: 10.1007/s40429-015-0054-y

Boyer, P. (2008). *Religion Explained*. New York, NY: Random House.

Chandra, R. (2017). *Facebuddha: Transcendence in the Age of Social Networks*. San Francisco, CA: Pacific Heart Books.

dscout (2016). "*Mobile Touches: dscout's Inaugural Study on Humans and Their Tech,*" *Research Report*. Available at: https://blog.dscout.com/hubfs/downloads/dscout_mobile_touches_study_2016.pdf

Dunbar, R. I. (2004). Gossip in evolutionary perspective. *Rev. Gen. Psychol.* 8, 100–110. doi: 10.1037/1089-2680.8.2.100

Eckel, C. C., and Grossman, P. J. (1998). Are women less selfish than men?: Evidence from dictator experiments. *Econ. J.* 108, 726–735. doi: 10.1111/1468-0297.00311

Elhai, J. D., Dvorak, R. D., Levine, J. C., and Hall, B. J. (2017). Problematic smartphone use: a conceptual overview and systematic review of relations with anxiety and depression psychopathology. *J. Affect. Disord.* 207, 251–259. doi: 10.1016/j.jad.2016.08.030

Emanuel, R., Bell, R., Cotton, C., Craig, J., Drummond, D., Gibson, S., et al. (2015). The truth about smartphone addiction. *Coll. Stud. J.* 49, 291–299.

Espinosa, M. P., and Kovářík, J. (2015). Prosocial behavior and gender. *Front. Behav. Neurosci.* 9:88. doi: 10.3389/fnbeh.2015.00088

Festinger, L. (1954). A theory of social comparison processes. *Hum. Relat.* 7, 117–140. doi: 10.1177/001872675400700202

Fiorillo, C. D., Tobler, P. N., and Schultz, W. (2003). Discrete coding of reward probability and uncertainty by dopamine neurons. *Science* 299, 1898–1902. doi: 10.1126/science.1077349

Friston, K., and Kiebel, S. (2009). Predictive coding under the free-energy principle. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 364, 1211–1221. doi: 10.1098/rstb.2008.0300

Frith, C. (2002). Attention to action and awareness of other minds. *Conscious. Cogn.* 11, 481–487. doi: 10.1016/S1053-8100(02) 00022-3

Ginott, H. G. (1965/2009). Between Parent and Child: Revised and Updated: The Bestselling Classic That Revolutionized Parent-Child Communication. New York, NY: Harmony.

Harari, Y. N. (2017). *Homo deus*. Paris: Albin Michel.

Harmon, E., and Mazmanian, M. (2013). "Stories of the smartphone in everyday discourse: conflict, tension & instability," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, (New York, NY: ACM), 1051–1060.

Hebb, D. O. (1976). Physiological learning theory. *J. Abnorm. Child Psychol.* 4, 309–314. doi: 10.1007/BF00922529

Henrich, J. (2016). *The Secret of Our Success: How Culture is Driving Human Evolution, Domesticating our Species, and Making us Smarter*. Princeton, NJ: Princeton University Press.

Hrdy, S. B. (2007). "Evolutionary context of human development: the cooperative breeding model," in *Family Relationships: An Evolutionary Perspective*, eds C. A. Salmon and T. K. Shackelford (New York, NY: Oxford University Press), 39–68.

Hrdy, S. B. (2009). *Mothers and Others*. Cambridge, MA: Harvard UP.

Hussain, Z., Griffiths, M. D., and Sheffield, D. (2017). An investigation into problematic smartphone use: the role of narcissism, anxiety, and personality factors. *J. Behav. Addict.* 6, 378–386. doi: 10.1556/2006.6.201 7.052

Kawa, L. (2018). *Two Major Apple Shareholders Push for Study of iPhone Addiction in Children*. Available at: https://www.bloomberg.com/news/articles/2018-01-08/jana-calpers-push-apple-to-study-iphone-addiction-in-children

Keiflin, R., and Janak, P. H. (2015). Dopamine prediction errors in reward learning and addiction: from theory to neural circuitry. *Neuron* 88, 247–263. doi: 10.1016/j.neuron.2015.08.037

Killingsworth, M. A., and Gilbert, D. T. (2010). A wandering mind is an unhappy mind. *Science* 330:932. doi: 10.1126/science.1192439

Kim, J., and Lee, J. R. (2011). The facebook paths to happiness: effects of the number of facebook friends and self-presentation. *Cyberpsychol. Behav. Soc. Netw.* 14, 359–364. doi: 10.1089/cyber.2010.0374

Krach, S., Paulus, F. M., Bodden, M., and Kircher, T. (2010). The rewarding nature of social interactions. *Front. Behav. Neurosci.* 4:22. doi: 10.3389/fnbeh.2010. 00022

Laasch, O., and Conaway, R. (2009). Gender differences in preferences. *J. Econ. Lit.* 47, 448–474. doi: 10.1257/jel.47.2.448

Lee, Y. K., Chang, C. T., Lin, Y., and Cheng, Z. H. (2014). The dark side of smartphone usage: psychological traits, compulsive behavior and technostress. *Comput. Hum. Behav.* 31, 373–383. doi: 10.1016/j.chb.2013.1 0.047

Levitt, P. (2003). *Fingerpainting on the Moon: Writing and Creativity as a Path to Freedom*. New York, NY: Harmony.

Li, S., and Chung, T. (2006). Internet function and internet addictive behavior. *Comput. Hum. Behav.* 22, 1067–1071. doi: 10.1016/j.chb.2004.0 3.030

Linnet, J. (2014). Neurobiological underpinnings of reward anticipation and outcome evaluation in gambling disorder. *Front. Behav. Neurosci.* 8:100. doi: 10.3389/fnbeh.2014.00100

Lopez-Fernandez, O., Honrubia-Serrano, L., Freixa-Blanxart, M., and Gibson, W. (2014). Prevalence of problematic mobile phone use in British adolescents. *CyberPsychol. Behav. Soc. Netw.* 17, 91–98. doi: 10.1089/cyber.2012. 0260

Lu, J. M., and Lo, Y. C. (2017). "Investigation of smartphone use while walking and its influences on one's behavior among pedestrians in Taiwan," in *Proceedings of the International Conference on Human-Computer Interaction*, (Cham: Springer), 469–475. doi: 10.1007/978-3-319-58753-0_67

Mar, R. A., Mason, M. F., and Litvack, A. (2012). How daydreaming relates to life satisfaction, loneliness, and social support: the importance of gender and daydream content. *Conscious. Cogn.* 21, 401–407. doi: 10.1016/j.concog.2011. 08.001

Mark, G., Voida, S., and Cardello, A. (2012). "A pace not dictated by electrons: an empirical study of work without email," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, (Austin, TX: ACM), 555–564. doi: 10.1145/2207676.2207754

Marlatt, G. A. (1996). Harm reduction: come as you are. *Addict. Behav.* 21, 779–788. doi: 10.1016/0306-4603(96)00042-1

Marlatt, G. A., Larimer, M. E., and Witkiewitz, K. (eds). (2011). *Harm Reduction: Pragmatic Strategies for Managing High-Risk Behaviors*. New York, NY: Guilford Press.

Maté, G. (2008). *In the Realm of Hungry Ghosts*. Berkeley, CA: North Atlantic Books.

Meier, S. (2007). Do women behave less or more prosocially than men? evidence from two field experiments. *Public Finance Rev.* 35, 215–232. doi: 10.1177/1091142106291488

Mercier, H., and Sperber, D. (2017). *The Enigma of Reason*. Cambridge: Harvard University Press.

Moll, H., and Tomasello, M. (2007). Cooperation and human cognition: the Vygotskian intelligence hypothesis. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 362, 639–648. doi: 10.1098/rstb.2006.2000

Mrazek, M. D., Phillips, D. T., Franklin, M. S., Broadway, J. M., and Schooler, J. W. (2013). Young and restless: validation of the Mind-Wandering Questionnaire (MWQ) reveals disruptive impact of mind-wandering for youth. *Front. Psychol.* 4:560. doi: 10.3389/fpsyg.2013.00560

Norenzayan, A., Henrich, J., and Slingerland, E. (2013). "Religious prosociality: a synthesis," in *Cultural Evolution: Society, Technology, Language and Religion*, eds P. J. Richerson and M. H. Christiansen (Cambridge, MA: MIT Press), 365–378. doi: 10.7551/mitpress/9780262019750.003. 0019

Norenzayan, A., and Shariff, A. F. (2008). The origin and evolution of religious prosociality. *Science* 322, 58–62. doi: 10.1126/science. 1158757

Oulasvirta, A., Rattenbury, T., Ma, L., and Raita, E. (2012). Habits make smartphone use more pervasive. *Pers. Ubiquitous Comput.* 16, 105–114. doi: 10.1007/s00779-011-0412-2

Pearson, C., and Hussain, Z. (2015). Smartphone use, addiction, narcissism, and personality: a mixed methods investigation. *Int. J. Cyber Behav. Psychol. Learn.* 5:17 doi: 10.4018/ijcbpl.2015010102

Poerio, G. L., and Smallwood, J. (2016). Daydreaming to navigate the social world: What we know, what we don't know, and why it matters. *Soc. Pers. Psychol. Compass* 10, 605–618. doi: 10.1111/spc3.1 2288

Ramstead, M. J., Veissière, S. P., and Kirmayer, L. J. (2016). Cultural affordances: scaffolding local worlds through shared intentionality and regimes of attention. *Front. Psychol.* 7:1090. doi: 10.3389/fpsyg.2016.01090

Ramstead, M. J. D., Badcock, P. B., and Friston, K. J. (2017). Answering Schrödinger's question: a free-energy formulation. *Phys. Life Rev.* doi: 10.1016/j.plrev.2017.09.001 [Epub ahead of print].

Rand, D. G., Brescoll, V. L., Everett, J. A., Capraro, V., and Barcelo, H. (2016). Social heuristics and social roles: intuition favors altruism for women but not for men. *J. Exp. Psychol. Gen.* 145, 389–396. doi: 10.1037/xge000 0154

Richerson, P., Baldini, R., Bell, A. V., Demps, K., Frost, K., Hillis, V., et al. (2016). Cultural group selection follows Darwin's classic syllogism for the operation of selection. *Behav. Brain Sci.* 39:e58. doi: 10.1017/S0140525X1500 0606

Roberts, J. A., Pullig, C., and Manolis, C. (2015). I need my smartphone: a hierarchical model of personality and cell-phone addiction. *Pers. Individ. Dif.* 79, 13–19. doi: 10.1016/j.paid.2015.01.049

Sauer, V. J., Eimler, S. C., Maafi, S., Pietrek, M., and Krämer, N. C. (2015). The phantom in my pocket: determinants of phantom phone sensations. *Mob. Media Commun.* 3, 293–316. doi: 10.1177/2050157914562656

Seger, C. A. (2006). The basal ganglia in human learning. *Neuroscientist* 12, 285–290. doi: 10.1177/1073858405285632

Snodgrass, J. G., Lacy, M. G., Dengah, H. J., Batchelder, G., Eisenhower, S., and Thompson, R. S. (2016). Culture and the jitters: guild affiliation and online gaming eustress/distress. *Ethos* 44, 50–78. doi: 10.1111/etho.1 2108

Song, X., and Wang, X. (2012). Mind wandering in Chinese daily lives–an experience sampling study. *PLoS One* 7:e44423. doi: 10.1371/journal.pone. 0044423

Soutschek, A., Burke, C. J., Beharelle, A. R., Schreiber, R., Weber, S. C., Karipidis, I. I., et al. (2017). The dopaminergic reward system underpins gender

differences in social preferences. *Nat. Hum. Behav.* 1, 819–827. doi: 10.1038/s41562-017-0226-y

Steers, M. L. N., Wickham, R. E., and Acitelli, L. K. (2014). Seeing everyone else's highlight reels: how Facebook usage is linked to depressive symptoms. *J. Soc. Clin. Psychol.* 33, 701–731. doi: 10.1521/jscp.2014.33.8.701

The Economist (2017). *Parents Now Spend Twice as Much Time with Their Children as 50 years ago.* Available at: https://www.economist.com/blogs/graphicdetail/2017/11/daily-chart-20 [accessed January 22, 2018].

Thompson, S. H., and Lougheed, E. (2012). Frazzled by Facebook? an exploratory study of gender differences in social network communication among undergraduate men and women. *Coll. Stud. J.* 46, 88–98.

Tobler, P. N., O'Doherty, J. P., Dolan, R. J., and Schultz, W. (2006). Human neural learning depends on reward prediction errors in the blocking paradigm. *J. Neurophysiol.* 95, 301–310. doi: 10.1152/jn.00762.2005

Tomasello, M. (2009). *Why We Cooperate.* Cambridge, MA: MIT press.

Tomasello, M. (2014). *A Natural History of Human Thinking.* Cambridge, MA: Harvard University Press. doi: 10.4159/9780674726369

Tomasello, M., Melis, A. P., Tennie, C., Wyman, E., Herrmann, E., Gilby, I. C., et al. (2012). Two key steps in the evolution of human cooperation: the interdependence hypothesis. *Curr. Anthropol.* 53, 687–688. doi: 10.1086/668207

Tufekci, Z. (2008). Grooming, gossip, Facebook, and MySpace. *Inf. Commun. Soc.* 11, 544–564. doi: 10.1080/13691180801999050

Twenge, J. M. (2017). *IGen: Why Today's Super-Connected Kids Are Growing Up Less Rebellious, More Tolerant, Less Happy–and Completely Unprepared for Adulthood–and What That Means for the Rest of Us.* New York, NY: Simon and Schuster.

Van Deursen, A. J., Bolle, C. L., Hegner, S. M., and Kommers, P. A. (2015). Modeling habitual and addictive smartphone behavior: The role of smartphone usage types, emotional intelligence, social stress, self-regulation, age, and gender. *Comput. Hum. Behav.* 45, 411–420. doi: 10.1016/j.chb.2014.12.039

van Holst, R. J., Veltman, D. J., Büchel, C., van den Brink, W., and Goudriaan, A. E. (2012). Distorted expectancy coding in problem gambling: is the addictive in the anticipation? *Biol. Psychiatry* 71, 741–748. doi: 10.1016/j.biopsych.2011.12.030

Veissière, S. (2016a). "Varieties of tulpa experiences: the hypnotic nature of human sociality, personhood, and interphenomenality," in *Hypnosis and Meditation: Towards an Integrative Science of Conscious Planes*, eds A. Raz and M. Lifshitz (Oxford: Oxford University Press).

Veissière, S. (2016b). "The internet is not a river: space, movement and personhood in a wired world," in *Click and Kin: Transnational Identity and Quick Media*, eds M. Friedman and S. Schultermandl (Toronto: University of Toronto Press).

Veissière, S. P. L. (2017). Cultural Markov Blankets? Mind the Other Minds Gap!: comment on "Answering Schrödinger's question: a free-energy formulation" by Maxwell James Désormeau Ramstead et al. *Phys. Life Rev.* doi: 10.1016/j.plrev.2017.11.001 [Epub ahead of print].

Weiser, E. B. (2015). # Me: Narcissism and its facets as predictors of selfie-posting frequency. *Pers. Individ. Dif.* 86, 477–481. doi: 10.1016/j.paid.2015.07.007

West, G. L., Drisdelle, B. L., Konishi, K., Jackson, J., Jolicoeur, P., and Bohbot, V. D. (2015). Habitual action video game playing is associated with caudate nucleus-dependent navigational strategies. *Proc. Biol. Sci.* 282:20142952. doi: 10.1098/rspb.2014.2952

Whitehouse, H. (2004). *Modes of Religiosity: A Cognitive Theory of Religious Transmission.* Walnut Creek, CA: Rowman Altamira.

Yin, H. H., and Knowlton, B. J. (2006). The role of the basal ganglia in habit formation. *Nat. Rev. Neurosci.* 7, 464–476. doi: 10.1038/nrn1919

Zuriff, G. E. (1970). A comparison of variable-ratio and variable-interval schedules of reinforcement. *J. Exp. Anal. Behav.* 13, 369–374. doi: 10.1901/jeab.1970.13-369

Check for
updates

Developing Emotional Design: Emotions as Cognitive Processes and their Role in the Design of Interactive Technologies

Stefano Triberti[1]*, Alice Chirico[1], Gemma La Rocca[2] and Giuseppe Riva[1,3]

[1] Department of Psychology, Università Cattolica del Sacro Cuore, Milan, Italy, [2] Independent Researcher, Milan, Italy,
[3] Applied Technology for Neuro-Psychology Laboratory, Istituto Auxologico Italiano, Milan, Italy

In the last 20 years, the debate on the role of *emotions* in the field of industrial design has grown exponentially. Emotional Design emerged as the effort to promote positive emotions (Norman, 2007) or pleasure in users (Jordan, 2002; Green and Jordan, 2003) by means of design properties of products and services. According to Van Gorp and Adams (2012), design based on emotions can affect overall user experience deeply, since emotions influence decision making, affect attention, memory, and generate meaning. It is possible to identify two main approaches to applied emotional design. The first is based on the modification of object's aesthetic appearance or interface, the latter focuses on promoting fluent and engaging interactions.

Both these approaches pertain to technology design, which includes especially common-use technological products. Regarding the first approach, several studies showed the importance of emotional aspects as drivers of market success, enjoinment, and active usage of technologies. For instance, Desmet et al. (2007) demonstrated that users attributed a "wow effect" (i.e., the combination of fascination, pleasant surprise, and desire) to those cellphones having some pleasant features in their exteriors. Studies in multimedia learning (Um et al., 2012; Plass et al., 2013) showed that embedding emotional stimuli (e.g., face-like shapes, vibrant colors) into interfaces elicited positive emotions in learners and improved learning outcomes.

The second perspective considers fluid interactions as a fundamental factor for an overall positive experience of use (Hancock et al., 2005; Hassenzahl and Tractinsky, 2006). This approach includes design based on the concept of psychological flow, namely an optimal experience of total absorption in a task when agent's skills and environmental challenges are both at a high level and balanced (Csikszentmihalyi, 1988; Csikszentmihalyi, 2002). Research demonstrated that flow experience is quite common in technology usage (Pilke, 2004; Triberti et al., 2016), such as in video games (Cowley et al., 2008; Jin, 2012; Argenton et al., 2014) and personal computer-mediated activities (Voiskounsky and Smyslova, 2003; Skadberg and Kimmel, 2004). For this reason, flow-inspired design models have been created and applied to the design of interactive digital technologies such as educational games and augmented reality (Alexiou et al., 2012; Neal, 2012). Other approaches for promoting emotions by engagement are gamification or the inclusion of game mechanics in interfaces (such as, prizes, achievements…) and interactive storytelling, which frames interaction within emotional scenarios with compelling characters, events, and motives (Morford et al., 2014).

The objective of the present contribution is to extend the discourse on emotional design, highlighting that technology designers can rely on other components beyond the above-mentioned aesthetic and engagement ones, in order to create innovative and effective devices. Indeed, emotions have further aspects that could be exploited by emotional designers. For instance, emotions are also cognitive processes—based on appraisal component—with a notable influence on the overall quality of interaction. According to this perspective, new technologies can be considered

and treated as opportunities to manipulate, enhance and trigger different discrete, and even complex emotional states. Finally, emotions can "participate" to interactions (instead of being a mere byproduct of it), by providing inputs to digital technologies to modify or influence final outputs.

This contribution explores opportunities provided by conceiving emotions as cognitive processes and active agents of interactions, in the field of emotional design.

Since *Affective computing* studies (Picard, 2003; Tao and Tan, 2005), designer have developed computers able to sense, recognize, and express emotions. New technologies combined with ubiquitous and wearable sensing become able to adapt to users' actual emotional states. For example, video games content changes (e.g., becoming more or less challenging) according to gamers' emotional state (e.g., bored or frustrated; Gilleade et al., 2005). Also mobile apps have been integrated with biofeedback sensors to promote positive emotions and relaxation (Serino et al., 2014). For instance, users can learn to monitor and control their emotional states by looking at virtual environments features (e.g., a burning fire) changing according to their psychophysiological activation. *Affective Design* (Reynolds and Picard, 2001) has shown that "emotional design" could be conceived not only as the inclusion of pleasant and/or engaging aspects in interfaces to augment pleasure, but also as the recognition and measurement of emotions to provide inputs to the technology and modify its functioning.

However, we argue that this approach, which is mainly based on general affect and moods, can be extended to discrete emotions, each characterized by a specific pattern of appraisal (i.e., emotion's cognitive profile). Studies on appraisal showed that an emotional episode emerges when one evaluates his/her own relationship with the surroundings (Roseman, 1991; Smith and Lazarus, 1993; Aue and Scherer, 2008; So et al., 2015). This automatic and subjective evaluation is based on specific properties of the stimulus such as relevance and congruence to personal goals or agency (oneself, others, or impersonal causes of the event), coping potential and control (Moors et al., 2013). The results of such evaluations bring about specific discrete emotions. Discrete emotional events are separable, distinguishable, and identifiable emotional state inducing changes into psychophysiology, behavior, motivation, judgment, and experience (Lench et al., 2011). Specifically, a discrete emotional event such as surprise, disgust, fear, would emerge after this first evaluation of the stimulus. After the appraisal component has been activated, a motivation to approach or avoid the stimulus follows (Moors et al., 2013). Furthermore, also changes in physiological parameters are involved, ranging from perspiration to muscle contraction. Finally, emotions are subjectively felt, since they can be described by the subject or can be quantified through numerical scales (Harmon-Jones et al., 2016), usually based on arousal (high/low intensity) and valence (positive/negative) aspects of the emotion at least (Mattek et al., 2017).

In our opinion, the scientific knowledge of discrete emotions based on their cognitive components—*appraisal*—can be easily translated into initial guidelines to develop a cognitive science-informed emotional design.

For instance, a field in which a partial discrete emotional approach was combined with affective is automotive technologies design (Ho and Spence, 2013). Nasoz et al. (2010) successfully tested a multi-modal intelligent car interface based on psychophysiological signals, able to classify driver's discrete emotional state as fear, boredom or anger that can be used to tune multisensory features of the car environment accordingly to help prevent accidents. In this case, technologies provide unprecedented opportunities to record even discrete users' emotional states (**monitoring emotions**), in order to tailor final outcomes. Future research in emotional design may explore how the continuous measurement of specific emotions can be exploited to influence ongoing interaction with common-use technology, for example modifying real-time easiness of use of devices or selecting digital content depending on the users' ongoing emotional responses.

A lot have been done, but we argue that still more can be done relying on an appraisal-based discrete emotion design approach. Indeed, appraisal theories of emotion have a lot to offer emotional design (Desmet, 2003; Bordegoni et al., 2014; Oatley and Johnson-Laird, 2014). Drawing on the scientific literature on discrete emotions as cognitive process, it is possible to *expand the kinds of emotions* that designers can reproduce and promote. Insofar emotions are considered as discrete events emerging from a specific pattern of appraisal themes (Smith and Lazarus, 1993), the more these themes are detailed, the higher the number of emotions and *emotional nuances* a designer can detect and control. For instance, sadness' core appraisal concerns an irrevocable loss (Smith and Ellsworth, 1985; Lazarus, 1991). If we detail this core appraisal, we can distinguish different kinds of sadness, such as melancholy, disappointment.

Such approach not only allows distinguishing different emotional nuances but it can also provide suggestions about reaching and promoting specific *complex* emotional states which include several single discrete emotional sub-components. Indeed, intervening on aesthetic appeal of interfaces allows designers to promote a general positive feeling in users, that is what has been done by most current approaches. However, the scientific literature can provide indications to elicit even specific complex emotions simply basing on their pattern of appraisal. For instance, one is the emotion of *awe* or the deep feeling of wonder, astonishment and fear people experience when facing stimuli perceived as incredible and incommensurable (Keltner and Haidt, 2003) (e.g., looking at vast panoramas; witnessing childbirth; etc.). Emotional appraisal leading to the experience of awe includes two distinctive elements, namely the feeling of vastness (perceptual or conceptual) and need for accommodation (i.e., the need for updating one's mental schemas to adapt them to the extraordinary). Recent research demonstrated that immersive technologies (e.g., Virtual Reality and 360° immersive videos) can be used to induce profound awe experiences in controlled environments, such as the lab (Gallagher et al., 2014; Chirico et al., 2016, 2017; Gaggioli et al., 2016). For instance, Chirico et al. (2017) were able to grasp subtle differences in the emergence of awe considering both self-reported and psychophysiological measures of this emotion. Awe resulted in a "freezing" response in front of something perceived vast and whose intensity can
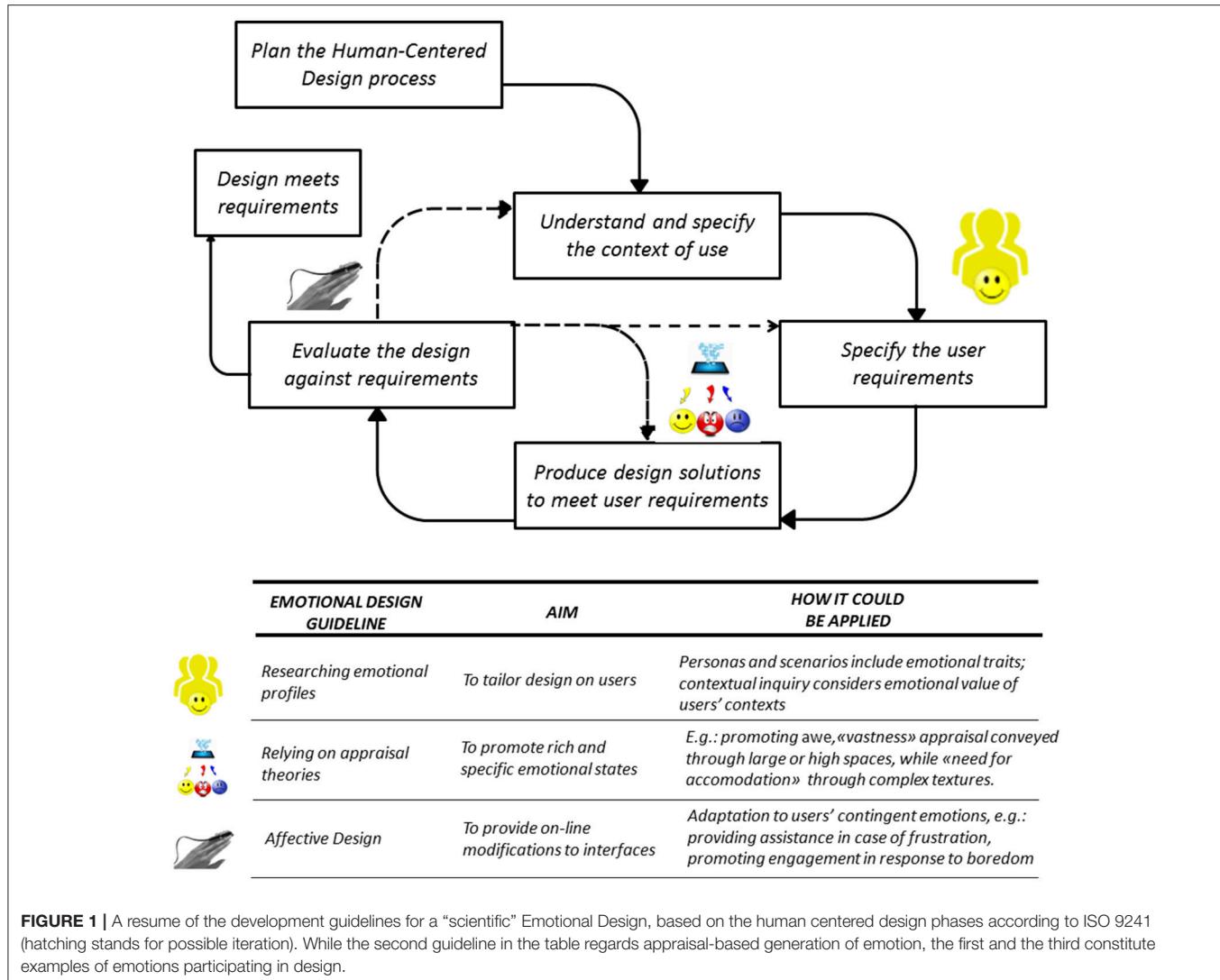
FIGURE 1 | A resume of the development guidelines for a "scientific" Emotional Design, based on the human centered design phases according to ISO 9241 (hatching stands for possible iteration). While the second guideline in the table regards appraisal-based generation of emotion, the first and the third constitute examples of emotions participating in design.

be enhanced by placing a user inside a 360° immersive virtual environment even with a low degree of interactivity. Appraisal dimensions of this emotion were analyzed in relation with the psychophysiological ones, thus providing a clearer picture of the emotional process.

In the emotional design, another important aspect concerns that emotions are closely intertwined over a continuous stream within subjects' experience. The sub-components of emotional episodes influence each other and subsequent emotional responses. For example, sad people are more likely to attribute agency of subsequent stimuli to others and the external world, because sadness is an emotion experienced toward events one cannot control (Han et al., 2007). Angry people are more likely to transfer anger to the next event to be evaluated in the surroundings (Beaudry et al., 2010; Darban and Polites, 2016).

In other words, emotions do not appear "out of nowhere" as the simple byproduct of a given stimulus and its appraisal. Instead, they are influenced by previous emotional states, or pre-existing individual traits, dispositions, and contextual

factors (Verduyn and Brans, 2012; Kim et al., 2016). Therefore, a technology designer working with emotions should be able to identify and measure *emotional profiles* or pre-existing individual/contextual characteristics that can influence the effectiveness of emotion-based technological services. For example, smartphones can be designed to elicit reactions such as surprise (Desmet et al., 2007). Nevertheless, such emotional state is not lasting in time, rather it tends to disappear shortly after the first encounters with the stimulus, since surprise arises from unexpected and novel events (Horstmann, 2006). Emotional designer should be able to create technologies updating according to users' personal information, in order to renovate the emotion of surprise continuously. In other words, they should design technological products able to actively adapt their outcomes to users' everyday life in line with individuals' peculiarities. This would allow designers promoting lasting emotional benefits such as loyalty, satisfaction, and possibly happiness and well-being. Although such ability largely depends on the designer's ability, it is possible to empower one's capacity to analyze emotional

profiles of users by employing User Centered Design research techniques (Abras et al., 2004; Garrett, 2010; Lowdermilk, 2013; Triberti and Liberati, 2014; Triberti and Barello, 2016), especially those involving the observation of users in the context of use (Viitanen, 2011) and those resuming typical users' needs and emotional benefits (Osterwalder and Pigneur, 2010; Miaskiewicz and Kozar, 2011). Collecting data on users' habits, intentions and context could help the designer to tailor technologies on their pre-existing emotional stream, within a user-centered design framework.

Finally, the advancement of common-use technology, combined with the knowledge available in cognitive science literature, could provide designers with extraordinary possibilities to fully exploit emotions' potential for user experience (see **Figure 1** for resume). In our opinion, this new approach could be based on: (1) the assessment of discrete emotions in an ongoing interaction to provide on-line modifications of interfaces (affective computing/affective design); (2) relying on scientific literature on emotions as discrete cognitive processes, to promote even complex emotions, and (3) analyzing users' "emotional profiles" to tailor technologies on their pre-existing emotional traits, within a user-centered design framework.

AUTHOR CONTRIBUTIONS

ST conceived the ideas presented in the article and wrote the first draft. AC assisted in drafting the manuscript and contributed with important intellectual content. GLR edited the manuscript from a design perspective and created the image. GR supervised the whole process and contributed with important intellectual content.

ACKNOWLEDGMENTS

REFERENCES

Abras, C., Maloney-Krichmar, D., and Preece, J. (2004). "User-centered design," in *Encyclopedia of Human-Computer Interaction*, ed W. Bainbridge (Thousand Oaks, CA: Sage), 1–14.

Alexiou, A., Schippers, M., and Oshri, I. (2012). Positive psychology and digital games: the role of emotions and psychological flow in serious games development. *Psychology* 3, 1243–1247. doi: 10.4236/psych.2012.312A184

Argenton, L., Triberti, S., Serino, S., Muzio, M., and Riva, G. (2014). "Serious games as positive technologies for individual and group flourishing," in *Technologies of Inclusive Well-Being*, eds A. L. Brooks, S. Brahnam, and L. C. Jain (New York, NY: Springer), 221–244.

Aue, T., and Scherer, K. R. (2008). Appraisal-driven somatovisceral response patterning: effects of intrinsic pleasantness and goal conduciveness. *Biol. Psychol.* 79, 158–164. doi: 10.1016/j.biopsycho.2008.04.004

Beaudry, A., Pinsonneault, A., and Molson, J. (2010). The other side of acceptance: studying the direct and indirect effects of emotions on information technology use. *MIS Q.* 34, 689–710. doi: 10.1287/mnsc.1100.1195

Bordegoni, M., Cugini, U., Ferrise, F., and Graziosi, S. (2014). A method for bringing user experience upstream to design. *Virtual Phys. Prototyp.* 9, 181–194. doi: 10.1080/17452759.2014.934574

Chirico, A., Cipresso, P., Yaden, D. B., Biassoni, F., Riva, G., and Gaggioli, A. (2017). Effectiveness of immersive videos in inducing awe: an experimental study. *Sci. Rep.* 7:1218. doi: 10.1038/s41598-017-01242-0

Chirico, A., Yaden, D. B., Riva, G., and Gaggioli, A. (2016). The potential of virtual reality for the investigation of awe. *Front. Psychol.* 7:1766. doi: 10.3389/fpsyg.2016.01766

Cowley, B., Charles, D., Black, M., and Hickey, R. (2008). Toward an understanding of flow in video games. *Comput. Entertain.* 6:1. doi: 10.1145/1371216.1371223

Csikzentmihalyi, M. (1988). "The flow experience and its significance for human psychology," in *Optimal Experience*, eds M. Csikzentmihalyi and I. S. Csikzentmihalyi (Cambridge, UK: Cambridge University Press), 15–35.

Csikszentmihalyi, M. (2002). *Flow: The Classic Work on How to Achieve Happiness*. London: Rider Books.

Darban, M., and Polites, G. L. (2016). Do emotions matter in technology training? Exploring their effects on individual perceptions and willingness to learn. *Comput. Human Behav.* 62, 644–657. doi: 10.1016/j.chb.2016.04.028

Desmet, P. (2003). A multilayered model of product emotions. *Des. J.* 6, 4–13. doi: 10.2752/146069203789355480

Desmet, P. M. A., and Porcelijn, R., and van Dijk, M. B. (2007). Emotional design; application of a research-based design approach. *Knowl. Technol. Policy* 20, 141–155. doi: 10.1007/s12130-007-9018-4

Gaggioli, A., Chirico, A., Triberti, S., and Riva, G. (2016). Transformative interactions: designing positive technologies to foster self-transcendence and meaning. *Annu. Rev. Cyber Ther. Telemed.* 14, 169–175.

Gallagher, S., Reinerman-Jones, L., Sollins, B., and Janz, B. (2014). Using a simulated environment to investigate experiences reported during space travel. *Theor. Issues Ergon. Sci.* 15, 376–394. doi: 10.1080/1463922X.2013.869370

Garrett, J. J. (2010). *Elements of User Experience The: User-Centered Design for the Web and Beyond*. London: Pearson Education.

Gilleade, K. M., Dix, A., and Allanson, J. (2005). "Affective videogames and modes of affective gaming : assist me, challenge me, emote me," in *DIGRA* (Vancouver, BC), 16–20.

Green, W., and Jordan, P. (2003). *Pleasure with Products: Beyond Usability*. London: Taylor and Francis.

Han, S., Lerner, J. S., and Keltner, D. (2007). Feelings and consumer decision making: the appraisal-tendency framework. *J. Consum. Psychol.* 17, 158–168. doi: 10.1016/S1057-7408(07)70023-2

Hancock, P. A., Pepe, A., and Murphy, L. L. (2005). Hedonomics: the power of positive and pleasurable ergonomics. *Ergon. Design* 13, 8–14. doi: 10.1177/106480460501300104

Harmon-Jones, C., Bastian, B., and Harmon-Jones, E. (2016). The discrete emotions questionnaire: a new tool for measuring state self-reported emotions. *PLoS ONE* 11:e0159915. doi: 10.1371/journal.pone.0159915

Hassenzahl, M., and Tractinsky, N. (2006). User experience - a research agenda. *Behav. Inform. Technol.* 25, 91–97. doi: 10.1080/01449290500330331

Ho, C., and Spence, C. (2013). Affective multisensory driver interface design. *Int. J. Vehicle Noise Vibrat.* 9, 61–74. doi: 10.1504/IJVNV.2013.053817

Horstmann, G. (2006). Latency and duration of the action interruption in surprise. *Cogn. Emot.* 20, 242–273. doi: 10.1080/02699930500262878

Jin, S.-A. A. (2012). Toward integrative models of flow: effects of performance, skill, challenge, playfulness, and presence on flow in video games. *J. Broadcast. Electron. Media* 56, 169–186. doi: 10.1080/08838151.2012.678516

Jordan, P. W. (2002). *Designing Pleasurable Products—An Induction to the New Human Factors*. London: Taylor and Francis.

Keltner, D., and Haidt, J. (2003). Approaching awe, a moral, spiritual, and aesthetic emotion. *Cogn. Emot.* 17, 297–314. doi: 10.1080/02699930302297

Kim, S., Park, G., Lee, Y., and Choi, S. (2016). Customer emotions and their triggers in luxury retail: understanding the effects of customer emotions before and after entering a luxury shop. *J. Bus. Res.* 69, 5809–5818. doi: 10.1016/j.jbusres.2016.04.178

Lazarus, R. S. (1991). *Emotion and Adaptation*. New York, NY: Oxford University Press.

Lench, H. C., Flores, S. A., and Bench, S. W. (2011). Discrete emotions predict changes in cognition, judgment, experience, behavior, and physiology:

a meta-analysis of experimental emotion elicitations. *Psychol. Bull.* 137, 834–855. doi: 10.1037/a0024244

Lowdermilk, T. (2013). *User-Centered Design: A Developer's Guide to Building User-Friendly Applications*. Sebastopol, CA: O'Reilly.

Mattek, A. M., Wolford, G., and Whalen, P. J. (2017). A mathematical model captures the structure of subjective affect. *Perspect. Psychol. Sci.* 12, 508–526. doi: 10.1177/1745691616685863

Miaskiewicz, T., and Kozar, K. A. (2011). Personas and user-centered design: how can personas benefit product design processes? *Des. Stud.* 32, 417–430. doi: 10.1016/j.destud.2011.03.003

Moors, A., Ellsworth, P. C., Scherer, K. R., and Frijda, N. H. (2013). Appraisal theories of emotion: state of the art and future development. *Emot. Rev.* 5, 119–124. doi: 10.1177/1754073912468165

Morford, Z. H., Witts, B. N., Killingsworth, K. J., and Alavosius, M. P. (2014). Gamification: the intersection between behavior analysis and game design technologies. *Behav. Anal.* 37, 25–40. doi: 10.1007/s40614-014-0006-1

Nasoz, F., Lisetti, C. L., and Vasilakos, A. V. (2010). Affectively intelligent and adaptive car interfaces. *Inf. Sci.* 180, 3817–3836. doi: 10.1016/j.ins.2010.06.034

Neal, M. (2012). "Creating and maintaining psychological flow state in augmented reality applications," in *EEE International Conference e-Learning, e-Business, Enterp. Inf. Syst.* (Las Vegas: e-Government).

Norman, D. A. (2007). *Emotional Design: Why We Love (or Hate) Everyday Things*. Available online at: https://books.google.com/books?id=QtWSu3zBtPoC&pgis=1 (Accessed December 22, 2015).

Oatley, K., and Johnson-Laird, P. N. (2014). Cognitive approaches to emotions. *Trends Cogn. Sci.* 18, 134–140. doi: 10.1016/j.tics.2013.12.004

Osterwalder, A., and Pigneur, Y. (2010). *Business Model Generation*. New York: John Wiley & Sons.

Picard, R. W. (2003). Affective computing: challenges. *Int. J. Hum. Comput. Stud.* 59, 55–64. doi: 10.1016/S1071-5819(03)00052-1

Pilke, E. M. (2004). Flow experiences in information technology use. *Int. J. Hum. Comput. Stud.* 61, 347–357. doi: 10.1016/j.ijhcs.2004.01.004

Plass, J. L., Heidig, S., Hayward, E. O., Homer, B. D., and Um, E. (2013). Emotional design in multimedia learning: effects of shape and color on affect and learning. *Learn. Instrum.* 29, 128–140. doi: 10.1016/j.learninstruc.2013.02.006

Reynolds, C., and Picard, R. W. (2001). "Designing for affective interactions," in *Proceedings from the 9th International Conference on Human-Computer Interaction* (New Orleans, LA), 6.

Roseman, I. J. (1991). Appraisal determinants of discrete emotions. *Cogn. Emot.* 5, 161–200. doi: 10.1080/02699939108411034

Serino, S., Cipresso, P., Gaggioli, A., Pallavicini, F., Cipresso, S., Campanaro, D., et al. (2014). Smartphone for self-management of psychological stress: a preliminary evaluation of positive technology app. *Rev. Psicopatol. Psicol. Clin.* 19, 253–260. doi: 10.5944/rppc.vol.19.num.3.2014.13906

Skadberg, Y. X., and Kimmel, J. R. (2004). Visitors' flow experience while browsing a Web site: its measurement, contributing factors and consequences. *Comput. Hum. Behav.* 20, 403–422. doi: 10.1016/S0747-5632(03)00050-5

Smith, C. A., and Ellsworth, P. C. (1985). Patterns of cognitive appraisal in emotion. *J. Personal. Soc. Psychol.* 48, 813–838. doi: 10.1037/0022-3514.48.4.813

Smith, C. A., and Lazarus, R. S. (1993). Appraisal components, core relational themes, and the emotions. *Cogn. Emot.* 7, 233–269. doi: 10.1080/02699939308409189

So, J., Achar, C., Han, D., Agrawal, N., Duhachek, A., and Maheswaran, D. (2015). The psychology of appraisal: specific emotions and decision-making. *J. Consum. Psychol.* 25, 359–371. doi: 10.1016/j.jcps.2015.04.003

Tao, J., and Tan, T. (2005). "Affective computing: a review," in *Affective Computing and Intelligent Interaction*, eds J. Tao, T. Tan, and R. W. Picard (Berlin; Heidelberg: Springer), 981–995.

Triberti, S., and Barello, S. (2016). The quest for engaging AmI: patient engagement and experience design tools to promote effective assisted living. *J. Biomed. Inform.* 63, 150–156. doi: 10.1016/j.jbi.2016.08.010

Triberti, S., and Liberati, E. G. (2014). "Patient centered virtual reality: an opportunity to improve the quality of patient's experience," in *Virtual Reality: Technologies, Medical Applications and Challenges*, eds P. Cipresso and S. Serino (New York, NY: Nova Science), 3–30.

Triberti, S., Chirico, A., and Riva, G. (2016). "New technologies as opportunities for flow experience: a framework for the analysis," in *Flow Experience: Empirical Research and Applications* eds L. Harmat, F. Ø. Andersen, F. Ullén, J. Wright, and G. Sadlo (New York, NY: Springer), 249–263.

Um, E., Plass, J. L., Hayward, E. O., and Homer, B. D. (2012). Emotional design in multimedia learning. *J. Educ. Psychol.* 104, 485–498. doi: 10.1037/a0026609

Van Gorp, T., and Adams, E. (2012). *Design for Emotions*. Waltham, MA: Elsevier.

Verduyn, P., and Brans, K. (2012). The relationship between extraversion, neuroticism and aspects of trait affect. *Pers. Individ. Dif.* 52, 664–669. doi: 10.1016/j.paid.2011.12.017

Viitanen, J. (2011). Contextual inquiry method for user-centred clinical IT system design. *Stud. Health Technol. Inform.* 169, 965–969. doi: 10.3233/978-1-60750-806-9-965

Voiskounsky, A. E., and Smyslova, O. V (2003). Flow in computer hacking : a model. *Web Commun. Technol. Internet-related Soc.* 2713, 176–186. doi: 10.1007/3-540-45036-X_18

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.

A Usability Study of a Serious Game in Cognitive Rehabilitation: A Compensatory Navigation Training in Acquired Brain Injury Patients

Milan N. A. van der Kuil[1]*, Johanna M. A. Visser-Meily[2,3], Andrea W. M. Evers[1] and Ineke J. M. van der Ham[1]

[1] Department of Health, Medical and Neuropsychology, Leiden University, Leiden, Netherlands, [2] Center of Excellence in Rehabilitation Medicine, Brain Center Rudolf Magnus, University Medical Center Utrecht and De Hoogstraat Rehabilitation, Utrecht, Netherlands, [3] Department of Rehabilitation, Physical Therapy Science & Sports, Brain Center Rudolf Magnus, University Medical Center Utrecht, Utrecht, Netherlands

Acquired brain injury patients often report navigation impairments. A cognitive rehabilitation therapy has been designed in the form of a serious game. The aim of the serious game is to aid patients in the development of compensatory navigation strategies by providing exercises in 3D virtual environments on their home computers. The objective of this study was to assess the usability of three critical gaming attributes: movement control in 3D virtual environments, instruction modality and feedback timing. Thirty acquired brain injury patients performed three tasks in which objective measures of usability were obtained. Mouse controlled movement was compared to keyboard controlled movement in a navigation task. Text-based instructions were compared to video-based instructions in a knowledge acquisition task. The effect of feedback timing on performance and motivation was examined in a navigation training game. Subjective usability ratings of all design options were assessed using questionnaires. Results showed that mouse controlled interaction in 3D environments is more effective than keyboard controlled interaction. Patients clearly preferred video-based instructions over text-based instructions, even though video-based instructions were not more effective in context of knowledge acquisition and comprehension. No effect of feedback timing was found on performance and motivation in games designed to train navigation abilities. Overall appreciation of the serious game was positive. The results provide valuable insights in the design choices that facilitate the transfer of skills from serious games to real-life situations.

Keywords: spatial navigation, acquired brain injury, usability, serious game, rehabilitation, cognitive training

INTRODUCTION

Serious games are games that are designed for a primary purpose other than entertainment (Michael and Chen, 2005). The key concept of serious gaming is the implementation of game attributes and game mechanisms to engage users toward achieving real-life goals. While many of these game attributes and mechanics are adapted from the entertainment video games, their underlying concepts correspond well to ideas originating in fields such as behaviorism,

constructivism, and neuroscience (Yusoff et al., 2009). As such, effective implementation of goals, feedback, rules, challenges and fantasy elements enhances the motivation and engagement of users toward achieving learning outcomes (Garris et al., 2002; Yusoff et al., 2009; Charsky, 2010).

Over the past decade, serious gaming has proliferated into different areas such as healthcare, military, corporate, education and government (Susi et al., 2007). A notable application of serious gaming is its introduction into the field of neuropsychological rehabilitation. Acquired brain injuries (e.g., stroke, traumatic brain injury and brain tumors) are highly prevalent in modern society (Ma et al., 2014; Peeters et al., 2015). Cognitive and behavioral deficits resulting from acquired brain injury have a profound effect on many daily life activities of these patients (Fann et al., 1995). The aim of neuropsychological rehabilitation is to aid brain injured patients in overcoming impairments and disabilities and to facilitate a return to usual self-care and daily activities (Dobkin and Dorsch, 2013). Rehabilitation programs often span over several months and require patients to engage in repeated exercises or mental rehearsals. Furthermore, patients are often required to continue with home-based therapies after they are discharged from hospital care (Trialists, 2004). The combination of home-training, repetition of exercises, and high treatment costs provide interesting opportunities for innovative approaches such as serious gaming in rehabilitation.

A distinction can be made between physical and cognitive rehabilitation. Physical rehabilitation focusses on motor abilities and sensorimotor functioning. Serious games have been developed to aid in the rehabilitation of balance impairments (Betker et al., 2006), motor functions of the hand (Afyouni et al., 2017) and the upper limbs (Broeren et al., 2008; Yoo et al., 2014), for instance. Motor rehabilitation games take a restitution-based rehabilitation approach, in which the aim is to restore impaired functions through intense and repeated stimulation of that function (Wolf et al., 2002). Consequently, the application of serious games in physical rehabilitation benefits from the motivational and engaging components of video games. Furthermore, adaptive difficulty systems implemented through game mechanics, allow for the presentation of adequate challenges, further tailoring to the need of patients in the program.

Serious gaming in cognitive rehabilitation is less common. As of now, several serious games in cognitive rehabilitation have been developed with the intention of directly training cognitive functions by incorporating mental exercises in games ('brain training'). Brain training games such as "Lumosity" aim to strengthen attention, working memory and executive functions (Sternberg et al., 2013). The approach taken in these programs is similar to the restitution-based rehabilitation approach taken in serious games for motor rehabilitation, as patients repeatedly perform short task with increasing difficulty. Most brain training games have been developed for healthy elderly and persons with mild cognitive impairments. Randomized controlled trial studies have been performed to assess the effectiveness of brain training games in patients with cognitive impairments as a result of brain injuries. Evidence for the effectiveness of these brain training

games in this population is inconclusive, as the effects of the training generally do not generalize beyond the training itself (Zickefoose et al., 2013; van de Ven et al., 2017).

Contrary to restitution-based rehabilitation, compensation-based rehabilitation has not been thoroughly explored with serious games. Compensation training is based on the concept that cognitive deficits can be overcome by substituting different latent skills or by acquiring new skills (Dixon and Bäckman, 1999). Compensatory training is one of the most important techniques in neurologic rehabilitation of acquired brain injury (Cicerone et al., 2000, 2005, 2011). Accordingly, the Cognitive Rehabilitation Task Force of the American Congress of Rehabilitation Medicine Brain Injury Interdisciplinary Special Interest Group has recommended compensation training as standard practice for memory impairments after traumatic brain injury and stroke (Cicerone et al., 2011).

Serious games designed to train compensation strategies will have additional design considerations compared to games designed to stimulate engagement. Aside from the affective components, emphasis is placed on the cognitive and educational components of the applications. Compensation strategies trained in serious games need to be transferred to daily activities. This requires patients to have a general understanding of the cognitive function that will be compensated and their own impairments regarding this function. Novel strategies will need to be introduced and trained. Finally, patients need to learn how and when a novel strategy can be applied in real-life situations (Geusgens et al., 2007).

In the current project, we have developed a serious game for the rehabilitation of spatial navigation impairments after acquired brain injury. Navigation impairments are common among stroke patients and have profound effects on the quality of life, as patients experience reduced mobility, autonomy and spatial anxiety (van der Ham et al., 2013). Even though navigational impairments in stroke patients are prevalent, no standardized rehabilitation training is currently available. A recent article advocates a compensatory approach to the rehabilitation of navigation impaired patients (Claessen et al., 2016). Instead of focusing on the rehabilitation of impaired cognitive function (such as memory or attention), The authors propose that the rehabilitation training should focus on training patients to use an alternative navigation strategy. Claessen et al. (2016) identified patients' impaired components of the navigational ability through an extensive diagnosis procedure in a simulated virtual environment. Based on a profile resulting from this diagnosis, patients were trained to adopt a more advantageous navigation strategy in a series of virtual reality therapy sessions provided by a neuropsychologist. The results of the navigation compensation training were promising, as patients reported that they successfully adopted novel navigation strategies in real-life situations and improved on the trained navigation abilities.

As an extension to this therapy, we have developed a serious game that trains compensatory strategy use by providing multiple navigation exercises in combination with psycho-education. The goal of this serious game is to change patients' navigation strategy in order to improve their navigation ability in daily life.

The key concepts of the virtual reality therapy are adapted into a serious game that can be used at home, without supervision of a therapist. In order to ensure the usability of the application by the target patient population, an extensive user interaction test was conducted. In this usability study, three core principles of the application were examined: interaction in 3D environments, instruction modality and feedback timing.

The game's training components take place in open, 3D environments, which patients view and interact with from a first-person perspective. In order to promote presence and stimulate the transfer of skills trained in the game, unrestricted, realistic movement in 3D environments is required. Effective movement within the 3D environments requires intuitive and accessible human–computer interaction. The manner in which users use buttons and sensors of input devices to control software events is referred to as a control scheme. Effective control schemes are believed to have a positive effect on game performance and the affective components of a game such as enjoyment, frustration and feelings of competence (Limperos et al., 2011; McEwan et al., 2012; Rogers et al., 2015; Shin and Chung, 2017). Furthermore, input modality can affect working memory, presence and experienced realism during gameplay (Kent et al., 2012; Shafer et al., 2014; Shin and Chung, 2017). In terms of compensatory strategy training, suboptimal movement control might frustrate patients, reduce engagement, and shift attention away from the educative goals of the exercises. The first aim of current experiment was to assess the subjective experience and objective performance of movement in 3D environments using two simple movement control schemes.

The navigation training application consists of different training games. In each of the games a specific spatial skill is trained. In order for patients to integrate these skills into a compensatory strategy, patients require knowledge about the concepts that underlie the training. The concepts used in spatial cognition (e.g., egocentric navigation, mental mapping, landmark knowledge, etc.) can be particularly hard to grasp for the average user. Therefore, it was important that instructions and background information about the training concepts were presented in a format that was easy to understand for patients. As the games were presented on a multimedia computer, we had the option of presenting information using text-based or video-based instructions. Video-based instructions have the advantage of conveying graphical information supporting a narrative verbal instruction, which can be particularly useful for illustrating concepts in spatial cognition. However, the stream of information from video's might exceed the processing capacity of viewers and have a adverse effect on comprehension and knowledge organization (Mayer and Moreno, 2003; Chiu et al., 2016). This might be of importance as working memory is particularly vulnerable for impairment after acquired brain injury (McDowell et al., 1997; Christodoulou et al., 2001). Consequently, we expected that the self-pacing nature of text-based information would allow for a more optimal transfer of knowledge in acquired brain injury patients. The second aim of the study was to determine whether text-based instructions are more effective than video-based instructions by assessing objective performance

and subjective preferences in an instruction comprehension task.

Feedback presentation is an important component of effective serious game design (Garris et al., 2002; Yusoff et al., 2009; Charsky, 2010). The type, amount and timing of feedback has been shown to be of influence on learning efficacy and motivation in computer-based learning (Erhel and Jamet, 2013). The effect of feedback timing is often studied in the context of knowledge and skill assessments, where feedback is given directly after an answer is given or after a delayed period of time. Advantages and disadvantages of feedback timing on learning efficiency have been identified. Direct feedback allows learners to instantly correct erroneous responses, contributing to knowledge acquisition (Kulik and Kulik, 1988). However, processing direct feedback competes with cognitive resources required for learning process and can disrupt the learning process (Schooler and Anderson, 2008). Inversely, delayed feedback has been shown to facilitate knowledge retention over longer periods of time, but performance during knowledge acquisition is reduced (Shute, 2008). Feedback timing effects have predominantly been studied in educational scenario's such as classroom settings, quizzes and programming courses. In these scenarios responses can be directly evaluated and responses are often clearly correct or false. Less is known about the effects of feedback timing in games where skills are taught through interaction with a virtual game world. Responses are seldom binary in games, but rather expressed in a variable such as a score. Therefore, scoreboards are often implemented to allow users to monitor their performance during the gameplay. The timing and prevalence of this scoreboard can be controlled.

The current study focused on two methods of feedback timing: cumulative feedback and delayed feedback. Cumulative feedback refers to the explicit presentation of a patient's overall performance during gameplay. Cumulative feedback is shown directly after completing each challenge on an interval basis. Delayed feedback refers to explicit presentation of a patient's overall performance after gameplay. The third aim of the study was to determine whether feedback timing affects objective performance and motivation (engagement and self-efficacy) during a navigation strategy training game. Cumulative feedback has been shown to positively affect performance in a working memory task compared to a no feedback condition (Adam and Vogel, 2016). Furthermore, cumulative feedback is similar to direct feedback described in more traditional feedback timing studies in the sense that patients can adjust their behavior during tasks. We hypothesized that cumulative feedback leads to increased performance during gameplay compared to delayed feedback.

The serious game will serve as a home-based rehabilitation treatment which patients will use over an extended period of time without supervision. In this usability study, three core principles of the application were examined: interaction in 3D environments, instruction modality, and feedback timing. As the game required patients to interact with 3D virtual environments, we have determined what type of movement control was most intuitive: mouse controlled movement or

keyboard controlled movement. In order for the training to be effective, an understanding of complex spatial concepts was required. We therefore determined what instruction modality was most effective for the acquisition of knowledge in acquired brain injury patients: video-based instructions or text-based instructions. Furthermore, we have determined how performance and perceived competence were affected by cumulative and delayed feedback. Finally, as the serious game was designed to be effective for all patients with brain injuries, regardless of the nature of the brain injury, we assessed whether differences between brain injury types exist in the appreciation of the application.

MATERIALS AND METHODS

Patients

A total of 30 acquired brain injury patients participated in the study (**Table 1**). All patients were included by occupational therapists at the Department of Rehabilitation of the University Medical Center Utrecht. Inclusion criteria were: (a) clinically diagnosed with acquired brain injury (e.g., cerebrovascular accident, traumatic brain injury, hypoxic-anoxic brain injury), (b) in the non-acute phase of brain injury, (c) between 18 and 80 years of age, (d) capable of operating a computer system using their left or right hand, (e) sufficient communication, comprehension and taxability (judged by an occupational therapist), (f) no visual impairments interfering with the tasks (e.g., blindness, neglect). All participants gave written informed consent before participating in the study. Patients did not receive monetary compensation for study participation.

This study was exempted from ethical approval by the Medical Ethics Committee of the University Medical Center Utrecht in accordance with the Dutch WMO law. This study was performed in accordance with the Declaration of Helsinki and the ICH guidelines for good clinical practice.

TABLE 1 | Characteristics of patients in study (n = 30).

Variable	
Gender, male N	15 (50%)
Age in years, mean (range)	47.2 (23 − 68)
Education*, mean (SD)	5.4 (1.07)
Brain Injury Type	
- Cerebrovascular accident	16 (53.3%)
- Traumatic brain injury	9 (30%)
- Brain tumor	4 (13%)
- Brain hypoxia	1 (3.33%)
Brain injury location	
- Left	9 (30%)
- Right	11 (36.67%)
- Bilateral	3 (10%)
- Unspecified/Unknown	7 (23.33%)
Months after brain injury, mean (SD)	26.43 (52.71)

*Education scores used the Verhage scale. This is a Dutch education classification system including 7 categories (Verhage, 1964): 1, lowest; 7, highest.

Tasks and Material

Three tasks were employed to assess different aspects of the software's usability: movement control, instruction modality and feedback timing. Each task was comprised of an objective component, performance on the task, and a subjective component, a questionnaire with questions regarding a patient's user experience (**Tables 2**, **3**). Furthermore, a questionnaire was used to assess the menu-interaction experience (**Table 4**). Additional questionnaires were presented at the start and end of the experimental session to measure computer experience and general appreciation, respectively (**Table 5** and Supplementary Table 1).

Movement Control

The movement control task was designed to assess usability differences between mouse controlled and keyboard controlled movement in 3D environments. A virtual environment was created resembling a sandy desert (**Figure 1**). A bordered plateau was placed in the middle of this environment. The plateau consisted of three distinct components: A broad meandering road, a large circular environment and a building consisting of narrow corridors and 8 90-degree turns (**Figure 2**). Three colored cubes (red, green, blue) were placed in the circular environment. The starting-location was placed at the beginning of the meandering road and the end-location was placed at the end of the corridor inside the building. Following the one-way road lead to the end-location as no junction points or crossroads were present. A geometrically mirrored version of environment was created to facilitate comparable environments for the two movement conditions.

Keyboard controlled movement was performed by pressing the four arrow keys on the keyboard. "Up" corresponded to forward movement, "down" corresponded to backward movement and the "left" and "right" buttons corresponded to left and right rotation. Mouse controlled movement was performed by using the left and right mouse button and by utilizing the optical sensor. Left mouse button corresponded to forward movement, right mouse button corresponded to backward movement, moving the mouse left or right corresponded with rotation in the respective direction. Similar to the keyboard input condition, participants were unable to look up or down using the mouse. Movement speed was set to 5 in both conditions. This corresponded to a walking velocity of approximately 5 km/hour.

Patients were placed at the start of the meandering road and were asked to travel to the end-location which was placed at the end of the corridors in the building. Before entering the building, all colored cubes had to be picked up. Cubes were picked up by bumping into them. Patients were instructed to travel to the end-location as fast as possible, without touching the walls. Time required to finish the task (seconds) and number of collisions with the walls were recorded. Patients performed a single trial in each condition. A usability questionnaire was filled in following each movement tasks. This questionnaire measured the following concepts: *ease of use*, *experienced improvement*, *similarity with other software*, *enjoyment* and *presence* on a 5 point Likert scale (**Table 2**). After both the mouse controlled and keyboard control tasks were completed, patients were presented with an open

TABLE 2 | Movement control questionnaire (*n* = 30).

Variable	Question	Mouse** mean (*SD*)	Keyboard** mean (*SD*)	*p**
Ease of use	I thought walking around in the environment was easy	4.2 (1.35)	3.33 (1.49)	**<0.01**
Improvement	Over time I felt I improved at walking around in the environment	4.3 (1.09)	3.9 (1.37)	0.14
Other software	The controls of this application were similar to other software I have used	3.33 (1.77)	2.86 (1.59)	0.11
Enjoyment	I enjoyed walking in the environment	4.24 (1.06)	3.72 (1.22)	**<0.01**
Presence	I could imagine myself walking in the environment	3.7 (1.26)	3.3 (1.44)	**<0.05**

*Significant differences are printed in bold letters. **Ratings on a Likert scale with 1 corresponding to "completely disagree" and 5 corresponding to "completely agree." Standard deviations appear in parentheses next to means.

TABLE 3 | Feedback timing questionnaire (*n* = 21).

Type	Question	Cumulative** mean (*SD*)	Delayed** mean (*SD*)	*p**
Interest	I thought the task was interesting	4.33 (0.80)	4.57 (0.93)	0.17
Enjoyment	When I performed the task. I enjoyed myself.	4.38 (1.12)	4.57 (0.98)	0.26
Perceived difficulty	I thought the task was easy	2.90 (1.34)	2.90 (1.41)	0.96
Effort	I put a lot of effort into completing the task	3.86 (1.35)	3.48 (1.25)	0.32
Strive	I did the best I could during this task	4.62 (0.59)	4.48 (0.87)	0.41
Competence	I had the feeling I was good at the task	3.71 (1.27)	3.62 (1.36)	0.69
Accept results	I am content with my performance	3.57 (1.33)	3.52 (1.44)	0.79
Competition	I think my performance was above-average	2.67 (1.15)	3.10 (1.37)	0.11
Desire to improve	I wish I was better at the task	3.81 (1.36)	3.62 (1.32)	0.47

*Differences between responses in the delayed and cumulative feedback timing condition were compared per item using the Wilcoxon Signed Rank test. **Ratings on a Likert scale with 1 corresponding to "completely disagree" and 5 corresponding to "completely agree." Standard deviations appear in parentheses next to means.

questionnaire consisting of four questions: (1) What method of movement did you like best? (2) Why did you prefer this method over the other? (3) Do you have suggestions on how we could further improve the movement in the game? (4) What method of movement control would like to see in the training?

TABLE 4 | Menu-interaction experience (*n* = 29).

Statement	Response* mean (*SD*)
The text was easy to read	4.41 (1.09)
The information was placed where I expected it to be	4.14 (0.88)
The color and layout used in the application was distracting**	4.62 (0.78)
The terms used in the application were comprehensible	3.93 (1.36)
I understood what was meant with the term "levels"	4.38 (1.12)
I knew what the training was about by reading the names of the games	3.89 (1.26)
It was easy to navigate between different menus	4.03 (1.27)
It was easy to view the progression that was made on different challenges	3.97 (1.35)
I thought logging in was difficult**	4.48 (1.24)
Controlling the application was easy to learn	4.69 (0.81)
Learning what the terms meant was easy	4.14 (1.30)

*Ratings on a Likert scale with 1 corresponding to "completely disagree" and 5 corresponding to "completely agree." Standard deviations appear in parentheses next to means. **Data shown on a reversed scale, higher score indicate higher ratings of usability.

Instruction Modality

As the serious game was designed for desktop computers, instructions could be provided using narrated video (tutorial video) as well as more traditional texts. The instruction modality task was designed to assess differences in knowledge acquisition between text-based instructions and video-based instructions. The instructions of 2 existing navigation training games were used ("sense of direction game" and the "map use game"). Text-based and video-based instructions were constructed for both games. In the video version, the text was read aloud by a narrator and supported by a video montage of a person playing the

TABLE 5 | Overall appreciation questionnaire (*n* = 24).

Variable	Statement	Response* mean (*SD*)
Ease of use	The software was use to use	3.63 (0.25)
Enjoyment	I enjoyed the experience	4.17 (0.23)
Clear goals	The goals were clearly defined	4.00 (0.24)
Rewarding	The experience was rewarding	3.92 (0.22)
Control	I had a feeling of total control	3.29 (0.26)
Attention	My attention was completely directed on the task at hand	4.79 (0.10)
Concentration	I was concentrated	4.54 (0.19)
Willingness to play again	I would like to play the game again	4.13 (0.23)
Challenge	The game was challenging	4.08 (0.21)

*Ratings on a Likert scale with 1 corresponding to "completely disagree" and 5 corresponding to "completely agree." Standard deviations appear in parentheses next to means.

FIGURE 1 | Design of the environment used in the movement control task. The environment can be subdivided in a meandering part, a circular area and a building featuring sharp turns. A mirrored version was created to accommodate for the two conditions.

FIGURE 2 | Design of the corridor with sharp turns used in the movement control task. The corridors inside the building are made up of 8, 90°; turns. The blue icon with arrows indicates the entrance of the building. The blue icon with the square indicates the end location of the task.
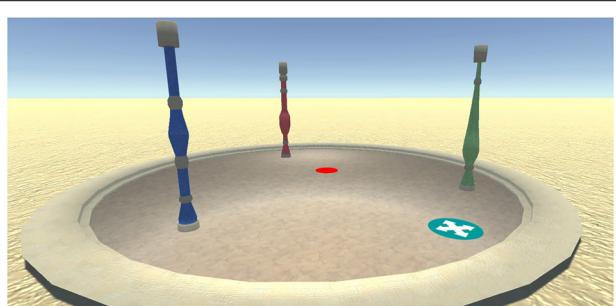
FIGURE 3 | Design of the environment used in the feedback timing task. In this version of the task, participants study a map to remember the location of the goal (red dot) in relation to the landmarks (pillars). Patients were then placed on the starting location (blue dot). The goal and start locations are not visible during a round.

game. In the text version, text was printed on the screen and patients could scroll through the text at their own pace. When presented with the video version, patients were asked to watch and memorize the video. When presented with the text only version of the instructions, patients were instructed to read and memorize the text. No time limit was set. The order in which

patients received the video-based or text-based instruction, as well as the combination of instruction modality and version of the game was counterbalanced across patients.

After observing the instructions, patients were shown 12 statements about the objectives of the game and the implications of using the navigation strategy that was trained in the game (Supplementary Tables 2, 3). Patients determined whether these statements were true or false. Following the true or false statements for both instruction modalities, participants answered three open questions: (1) What instruction type did you find most effective? (2) Why did you prefer this type of instructions? (3) Do you have suggestions on how we could further improve the instructions?

Feedback Timing

The feedback timing task was designed to assess the effect of cumulative vs. delayed feedback on performance and motivation during a play-through of a training game. A virtual environment was created resembling a sandy desert. In this middle of the environment, a bordered circular plateau was placed. Two versions of the game were used. In the first version, 4 distinct landmarks were placed in the north, south east and west of the plateau. These landmarks resembled the Horse of Troy, a Greek galley, a Greek temple and the Colossus. In the second version, 3 local landmarks were placed inside of the plateau. The landmarks resembled different colored pillars (red, green, blue). A hidden goal location was placed on the plateau (**Figure 3**).

At the start of a trial, a 2D map of the environment was shown on which the hidden plateau and the landmarks were highlighted. Patients were then placed in the 3D environment and were tasked to walk toward the hidden plateau, by orienting on the landmarks. The movement control was similar to the keyboard controlled movement described in the movement task above. A pedometer bar was shown at the top of the screen to indicate the amount of distance a patient had traveled. The amount of coins in possession corresponded to the size of the pedometer bar. As such, patients were instructed to take as few steps as necessary to reach the end-location. Between 0 and 2 coins could be earned in each round. The goal of the game was to earn as many coins as possible over the course of 3 rounds. In the cumulative feedback condition, a large scoreboard was presented between rounds. This scoreboard showed the percentage of coins collected over the whole trial (so if patients collected 3 coins at the end of round 2, the score would show 75%). The scoreboard allowed patients to monitor their performance of the span of 3 rounds. In the delayed feedback condition, no overall score feedback was given between rounds.

At the end of the three rounds, an overall score was shown in both conditions. The total amount of coins earned was used the measure of performance. After completing a task, patients filled in a questionnaire that measured motivational components related to engagement: *interest in task, enjoyment, effort invested while playing, strive* (I did the best I could during this task), *desire to improve*, and components related to self-efficacy: *perceived difficulty, competence, result acceptance, comparative score* (**Table 3**). The items were rated on a scale from 1 to 5,

1 corresponding to "completely disagree" to 5 corresponding to "completely agree."

Additional Measures

The menu interaction task was designed to assess the comprehensibility of the menu structure and phrasing of terms used in the game. Patients were required complete seven tasks by navigating through the menu tabs. In each task, specific information needed to be found or specific actions were required. Patients were asked to conduct the following activities: (1) log in, (2) start a specific game, (3) locate background information about the application, (4) determine the current level on a specific game, (5) start another game, (6) determine the amount of coins (score) currently in possession, (7) quit the application. Patients were instructed to think out loud while navigating the menu screens. When patients navigated to a wrong menu or when they indicated they were unable to find the requested information, the experiment would show the correct method of finding the information. Following the menu interaction task, a usability questionnaire was filled in (**Table 4**). The questionnaire was specifically designed to address layout, comprehensibility and interaction with important items of the menu interface.

The computer experience questionnaire consisted of nine items and was rated on a 5-point Likert scale (Supplemental **Table 2**). The items in this questionnaire were inspired by the Computer Attitude Scale and the Computer User Self Efficacy scale (Nickell and Pinto, 1986; Cassidy and Eachus, 2002). The first four items of this question addressed a patient's exposure to computers. Items 5–8 concerned a patient's self-reported knowledge of operating software and hardware. The ninth item addressed feelings of anxiety when using a computer.

The overall appreciation questionnaire consisted of nine items and was rated on a 5-point Likert scale (**Table 5**). Six items in this questionnaire were adapted from the Flow State Scale and three items constructed in context of the usability test (Jackson and Marsh, 1996). The items addressed the overall appreciation of the application and the experience of flow during the tasks. The items were rated on a scale from 1 to 5, 1 corresponding to "completely disagree" to 5 corresponding to "completely agree."

The tasks were constructed in the Unity 3D game engine, version 5.3.4.4.f1, and run as standalone applications. The application was run on a HP EliteBook 8760w laptop with a NVIDEA Quadro 3000M graphic processing unit. The laptop's screen size was 17.3-inch wide screen (15.5* 8.98) inch. The laptop's keyboard and a standard desktop mouse model (Dell Optical Mouse – MS116) were used as input devices. All questionnaires were constructed in Qualtrics and presented using an internet browser.

Procedure

The data was collected in a therapy room of the Department of Rehabilitation of the University Medical Center Utrecht. All patients read the study's information letter in advance and gave written informed consent prior to the session. All experimental sessions were planned prior to or after a patient's scheduled appointment with a doctor or occupational therapist. In order to comply to a patient's schedule during the visit to the medical center, each experimental session was brought to an end after approximately 60 min of testing.

At the start of the experimental session, patients were informed about the nature of the study. Patients were explicitly informed about the study's objective of tailoring the software to patients' capability and needs. As such, patients were encouraged to ask questions about the software, discuss design choices and propose suggestions for changes in the software's design. To stimulate communication with the patients, an informal and relaxed atmosphere was pursued.

The experiment started with the computer experience questionnaire. This was followed by the movement control task, the instruction modality task, the menu navigation task and finally the feedback task. Patients then filled in the overall appreciation questionnaire.

Statistical Analyses

Analysis of Objective Performance

Objective performance in the movement control, instruction modality and feedback timing tasks was analyzed using within-subject tests. Data were tested for normality using Kolmogorov–Smirnov tests. Normally distributed data were analyzed using a three-way mixed model ANOVAs with (condition) as within subject factor and (brain injury type) and (brain injury location) as between subject factors. Non-normal data were analyzed using Wilcoxon signed-rank tests, in which conditions were contrasted. Separate Kruskal–Wallis H Tests were used to assess the effects of brain injury type and brain injury location on performance in non-normal datasets.

Analysis of Subjective Measures

Internal reliability analyses were performed on all questionnaires. Non-parametric tests were used to analyze the effect of condition on subjective measures. Additionally, the proportion of responses for the preference (*what condition did you prefer*?) items in the open questionnaires were analyzed using Chi-square tests of independence. The effects of brain injury type and brain injury location on subjective responses were assessed using Kruskal–Wallis tests.

Exploratory Analysis

Exploratory analyses were performed to inspect the relation between objective performance and subjective measures for the movement control task and the feedback timing task. Pearson correlations analyses were conducted to investigate the relation between objective performance and items of the subjective measure questionnaires.

Attrition

Six patients were unable to complete all tasks of the experiment within 60 min. Additionally, 2 patients were unable to complete the instruction modality task due to reading impairments. One patient was unable to complete the feedback timing task due to severe navigation impairments. Technical difficulties lead to missing data of 1 patient in the movement control task and 2

patients in the feedback timing task. As such, the sample size for the objective performance analysis for the movement task was 29 (30 for the subjective measures), the sample size of the instruction task was 27(29 for the preference response) and the sample size of the feedback timing task was 21.

RESULTS

Movement Control

In order to compare objective movement performance in the mouse and keyboard controlled conditions, time required to finishing the task (time) and the number of collisions with the walls (wall bumps) were analyzed as main measures. A Kolmogorov–Smirnov test indicated that the data for time (mouse), $D(29) = 0.21$, $p < 0.01$ and wall bumps (keyboard), $D(29) = 0.17$, $p < 0.05$, were both significantly non-normal.

A Wilcoxon signed-rank test revealed that time in the mouse control condition ($M = 85.29$, $SD = 44.19$) was significantly shorter than time in the keyboard control condition ($M = 132.42$, $SD = 58.63$), $z = -4.68$, $p < 0.01$, $r = -0.61$ (**Figure 4**). No significant effects of condition were found on the number of wall bumps $z = -0.92$, $p = 0.36$, $r = -0.12$. Additional Wilcoxon signed-rank tests were performed to compare the effects of movement control type within in the three sections of the environment. Mouse controlled movement was faster than keyboard-controlled movement in the meandering area ($p < 0.01$) the circular area ($p < 0.01$) and the area with the sharp turns ($p < 0.01$) (**Figure 4**).

A Kruskal–Wallis H Test revealed that there was no effect of brain injury type on performance in the keyboard [$\chi^2(3) = 3.71$, $p = 0.29$] and mouse controlled [$\chi^2(3) = 5.49$, $p = 0.14$] movement tasks. Similarly, no effect of brain injury location was found on performance in the keyboard [$\chi^2(3) = 1.99$, $p = 0.57$] and mouse controlled [$\chi^2(3) = 2.94$, $p = 0.40$] movement task.

After completing the movement task, patients filled in a subjective preference questionnaire. A reliability analysis was performed and revealed an internal reliability of $\alpha = 0.85$ for the keyboard condition and $\alpha = 0.69$ for the mouse condition. Each of the 5 items of the questionnaire were compared for the mouse control and keyboard control condition using a Wilcoxon signed-rank test. A significant effect of condition was found for ease of use, as the mouse controls ($M = 4.2$, $SD = 1.35$) were rated as easier to use than keyboard controls ($M = 3.33$, $SD = 1.49$), $z = -2.67$ $p < 0.01$, $r = -0.34$. Mouse control ($M = 4.24$, $SD = 1.06$) was also rated as significantly more *enjoyable* than keyboard control ($M = 3.72$, $SD = 1.22$), $z = -2.67$, $p < 0.01$, $r = -0.34$. Furthermore, a higher level of *presence* was experienced during mouse controlled movement ($M = 3.7$, $SD = 1.26$) compared to the keyboard control ($M = 3.3$, $SD = 1.44$), $z = -2.36$, $p < 0.05$, $r = -0.30$ (**Table 2**).

Analysis of the open questionnaire revealed that 90.0% of the patients reported a preference for mouse controls, 10% of the patients reported a preference of keyboard control and 0% of the patients did not have a clear preference. A Chi-square test of independence revealed a significant difference in proportions, $\chi^2(1) = 19.20$, $p < 0.01$.
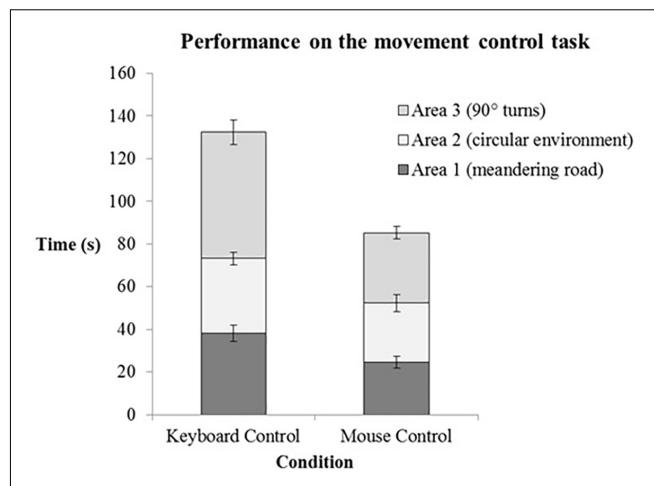

FIGURE 4 | Performance on the movement task for keyboard and mouse controlled movement ($n = 29$). The average time spend (seconds) in each area is indicated by the different colored stacks in the graph. The error bars represent the standard error of the mean.

Using Spearman correlation analyses, the relation between objective performance (time) in the movement tasks and the ratings on the 5 items of the questionnaire was explored. A correlation between objective performance and *enjoyment* was found for both the mouse control, $r = 0.43$, $p < 0.05$, and keyboard control $r = 0.39$, $p < 0.05$, conditions. Additionally, a correlation between objective performance and *presence* was found for both the mouse control, $r = 0.41$, $p < 0.05$, and keyboard control $r = 0.40$, $p < 0.05$, condition.

Instruction Modality

In order to determine the effect of instruction modality on learning, patients answered 12 true of false questions about the content of the instructions. Percentage correct was compared for the video-based and text-based condition. A Kolmogorov–Smirnov test indicated that the video-based instruction data was significantly non-normal $D(27) = 0.19$, $p < 0.05$. A Wilcoxon signed rank test was used to compare percentage correct for the video-based and text-based condition. No significant effect of instruction modality was found, $z = -0.82$, $p = 0.41$, $r = -1.12$. Percentage correct did not differ between the video-based ($M = 70.20$, $SD = 15.64$) and text-based ($M = 66.13$, $SD = 17.25$) condition.

A Kruskal–Wallis H Test revealed that there was no effect of brain injury type on percentage correct in the video-based [$\chi^2(2) = 1.78$, $p = 0.41$] and text-based [$\chi^2(2) = 1.01$, $p = 0.60$] conditions. Furthermore, no effect of brain injury location was found on the percentage correct in the video-based [$\chi^2(3) = 0.9$, $p = 0.83$] and text-based [$\chi^2(3) = 1.09$, $p = 0.78$] conditions.

The proportion of self-reported instruction preference was investigated using a chi-square test of independence. 65.51% of participants indicated a preference for the video-based instructions compared while 20.69% of the patients preferred the text-based instructions. 13.79% of the participants did not have a

clear preference. The chi-square test revealed that this difference in proportions was significant, $\chi^2(2) = 13.72$, $p < 0.01$.

Feedback Timing

The effect of feedback timing on objective performance was investigated by comparing the total amount of coins between the cumulative and delayed feedback condition. The total score was calculated by summing the amount of coins over three rounds for the cumulative feedback ($M = 3.48$, $SD = 1.63$) and delayed feedback ($M = 3.95$, $SD = 1.75$) tasks (Supplementary Table 4). A Kolmogorov–Smirnov test indicated that the total score (cumulative), $D(21) = 0.15$, $p = 0.2$ and total (delayed), $D(21) = 0.17$, $p = 0.14$ were normally distributed.

A three-way repeated measures ANOVA was performed to compare the effect of feedback timing on total score in the delayed and cumulative feedback condition with brain injury type and brain injury location as between subject factors. No significant main effect of condition was found $F(1,12) = 0.13$, $p = 0.27$, $\eta^2_P = 0.10$. No significant interaction effect was found for brain injury type and condition ($p = 0.41$) and brain injury location ($p = 0.73$).

After completing the feedback timing task, patients filled in the motivation questionnaire. Each of the 9 items of the questionnaire were compared between the cumulative and delayed feedback conditions using a Wilcoxon signed-rank test. No significant effect of condition was found in any of the items (**Table 3**).

In an explorative analysis, the relation between objective scores on the feedback tasks and ratings on the questionnaire were analyzed using Spearman correlations. In delayed feedback condition, a significant relation was found between objective score and ratings in *perceived difficulty*, $r = 0.59$, $p < 0.01$, *competence*, $r = 0.55$, $p < 0.01$, *result acceptance*, $r = 0.74$, $p < 0.01$ and *competition*, $r = 0.73$, $p < 0.01$. The subjective raring on the items correlated in a positive linear fashion with the objective score.

Similar relations were found between objective score and self-reported ratings on the cumulative feedback condition. Objective score significantly related to *perceived difficulty*, $r = 0.61$, $p < 0.01$, *competence*, $r = 0.64$, $p < 0.01$, *result acceptance*, $r = 0.72$, $p < 0.01$ and *competition*, $r = 0.57$, $p < 0.01$. The subjective raring on the items correlated in a positive linear fashion with the objective score. Additionally, a strong negative relation was found between *desire to improve*, $r = -0.65$, $p < 0.01$, and objective performance. The rating on the desire to improve item correlated negatively with objective score in linear fashion.

Additional Measures

After performing the menu interaction tasks, patients rated the usability of the menu navigation (**Table 4**). The 11 item questionnaire showed a high internal reliability of $\alpha = 0.81$. An overall score of the menu-navigation was computed by averaging the ratings of each item. A Kruskal–Wallis test was conducted to compare appreciation ratings between brain injury type and between brain injury location. No effect of brain injury type or location was found on the ratings on the overall menu interaction questionnaire.

The overall appreciation questionnaire was filled in at the end of the session to obtain ratings of overall appreciation and the experience of flow (**Table 5**). The 9 items of this questionnaire yielded a reliability rating of $\alpha = 0.76$. An overall rating of appreciation questionnaire was computed by averaging the ratings of each item. A Kruskal–Wallis test was conducted to compare usability rating between brain injury types and brain injury locations. No effect of brain injury type or location was found on the ratings on the overall appreciation of the game.

DISCUSSION

The usability of a serious game designed to train compensatory navigation strategies in acquired brain injury patients was investigated. The usability of three core principles of the application was examined using objective and subjective measures: movement control, instruction modality and feedback timing.

Intuitive control schemes in games contribute to motivation, engagement and reduction of cognitive load (Limperos et al., 2011; McEwan et al., 2012). The importance of responsive controls in serious games has been identified by several guidelines and frameworks concerned with usability (Pinelle et al., 2008). In order to optimize interactivity with the virtual environments used in the game, two control types were assessed: mouse and keyboard. The acquired brain injury patients clearly preferred mouse controlled movement over keyboard controlled movement. Mouse controlled movement was rated easier to use, more enjoyable and a stronger feeling of presence in the environment was experienced. While there is no consensus about the positive effects of presence in training programs, several studies have suggested that high levels of presence might aid in the transfer of skills acquired during the training (Youngblut and Huie, 2003; Alexander et al., 2005; Stevens and Kincaid, 2015). The advantages of mouse controlled movement over keyboard controlled movements were reflected in the objective performance measurements. Time required to finish the tasks was lower is using the mouse, while the number of wall collisions between control type did not differ. This indicates that patients did not lower accuracy in favor of speed when using mouse controlled input. Additionally, mouse controlled movement was faster in all three areas of the environment, revealing that the advantages of mouse movement were not specific to a single maneuver, such as taking sharp turns. An exploratory analysis revealed a positive relation between objective performance and ratings of enjoyment and presence in the environment in both movement control conditions. This finding further supports the notion that effective interaction results in a more enjoyable and natural gameplay experience. In sum, the implementation of simple, mouse controlled movement in 3D environments is recommended over keyboard-controlled movement based on objective and subjective evidence in this study.

Unrestricted movement in virtual environments allows patients to develop and experiment with novel navigation strategies. However, patients can only progress through the game when specific strategies are successfully adapted. It is therefore

important that the underlying concepts of the compensatory strategies are clearly communicated. Computers are multimedia systems that allow for different instruction modalities. In the current experiment, we examined the effects of video-based and text-based instruction on knowledge acquisition. No clear learning advantages of video-based instructions over text-based instruction were found. Similar results are found in studies that assess knowledge acquisition of complex topics (the news) through printed text and video (Furnham and Gunter, 1985; Walma van der Molen and Van Der Voort, 2000). While the results do not indicate an advantage for either modality, a clear preference for the video-based instructions was found in the questionnaire responses. During conversations with the patients about their preferred instruction modality, patients mentioned the advantage of visual information in explaining spatial concepts. This discrepancy between performance and preference can be explained in terms of cognitive capacity. Patients recognized that more information was presented to them in the video condition compared to the text condition. However, this additional information was not effectively maintained. We suspect that the continuous stream of information in the instruction video might have disrupted the information encoding process. Capacity constrains were not limited to the video-based instructions. Two patients were unable to complete the text-based instruction task due to their impairments. While these patients were able to read short texts, they were incapable of maintaining their attention when reading extensive bodies of text. The overload of cognitive capacity can be managed by providing patients with additional control over the pacing of the video (Mayer and Moreno, 2003). The aim for the instructions in the current game is to provide short and effective information before starting a gaming module. In this context, requiring patients to systematically analyze a video might not be an optimal solution. Subsequently, the addition of visual static images to text-based instructions might be more effective than both video-based and solely text-based instruction. This suggestion is supported by studies with healthy subjects (Mayer et al., 2005). More research is required to determine if this combination will indeed enhance knowledge acquisition in acquired brain injury patients. Overall, in this study we have established that patients prefer video-based instructions over text-based instructions. Video-based instructions are not more effective in context of knowledge acquisition and comprehension.

Feedback presentation is an important component in education and serious gaming (Garris et al., 2002; Yusoff et al., 2009; Charsky, 2010). Contrary to our expectation, we did not find a beneficial effect of cumulative feedback on objective performance. Updating patients on their overall score between rounds did not enhance performance in the task. Furthermore, the motivational components of the game were not affected by the timing of feedback as cumulative feedback did not affect engagement and self-efficacy. An earlier study showed beneficial effects of cumulative feedback on performance in a working memory tasks when compared to a no-feedback condition (Adam and Vogel, 2016). There might be several reasons why this effect was not observed in the current study. First, the current task included only 3 trials per condition,

whereas Adam and Vogel (2016) employed 150 short trials. It is possible that the beneficial effects of cumulative feedback only arise after participants are familiar with the task and start performing at a stable level. In the current task, it is possible that participants were still experimenting with strategies to complete the task. Second, the current task was considerably more complex than the working memory task employed by Adam and Vogel (2016). This might have lead to a greater variation in performance in both feedback timing conditions. Another explanation for this finding is that patients were not heavily invested in their performance within the game, as patients were explicitly informed that the goal of the study was to test the usability of the application. However, further analysis revealed positive linear relations between objective score and result acceptance ("*I am happy with my performance*"), indicating that patients were indeed concerned with their score. The exploratory analysis also revealed a negative linear relation between willingness to improve ("*I wish I was better at the task*") and the objective score in the cumulative feedback condition. This finding hints at a subtle effect of cumulative feedback on motivation. It is, however, unclear whether this effect is beneficial or disadvantageous, as this statement can be interpreted as a lack in confidence induced by the feedback or an increase in motivation to perform better. Overall, the current experiment did not provide evidence for the advantageous learning or motivational effects of cumulative feedback over delayed feedback.

Interaction with the menu screens and the overall appreciation of the game were evaluated positively. Importantly, neither the type of brain injury nor the location of the brain injury affected ratings on the appreciation and menu interaction questionnaires. Similarly, no effect of brain injury location and type were found on any of the objective tasks. The results suggest that the overall design and interaction with the serious game was suitable for all types of brain injury patients in the sample.

Summarizing, in this study we have established what design choices should be made in order to enhance the usability of a serious game designed to train navigation strategies. From this first study, we can conclude that mouse controlled movement in 3D environments is more accessible than keyboard controlled movement. Video-based instructions are strongly preferred over text-based instructions, but not more effective in transferring knowledge. Feedback timing did not affect performance and motivation in the current training games. Based on the scores and usability questionnaires, the results suggest that usability of the serious game is adequate for the target patient population after the implementation of the appropriate features as determined in this study.

AUTHOR CONTRIBUTIONS

MvdK, JV-M, and IvdH developed the theoretical framework and conceived of the presented idea. JV-M contributed to organizing the experiments. MvdK carried out the experiment. MvdK wrote the manuscript with support from AE, JV-M, and IvdH. AE, JV-M, and IvdH provided critical feedback on the drafted manuscript. AE, JV-M, and IvdH supervised the project.

FUNDING

ACKNOWLEDGMENTS

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2018.00846/full#supplementary-material

REFERENCES

Adam, K. C. S., and Vogel, E. K. (2016). Reducing failures of working memory with performance feedback. *Psychon. Bull. Rev.* 23, 1520–1527. doi: 10.3758/s13423-016-1019-4

Afyouni, I., Rehman, F. U., Qamar, A. M., Ghani, S., Hussain, S. O., Sadiq, B., et al. (2017). A therapy-driven gamification framework for hand rehabilitation. *User Model. User Adapt. Interact.* 27, 215–265. doi: 10.1007/s11257-017-9191-4

Alexander, A. L., Brunyé, T., Sidman, J., and Weil, S. A. (2005). From gaming to training: a review of studies on fidelity, immersion, presence, and buy-in and their effects on transfer in pc-based simulations and games. *DARWARS Train. Impact Group* 5, 1–14.

Betker, A. L., Szturm, T., Moussavi, Z. K., and Nett, C. (2006). Video game–based exercises for balance rehabilitation: a single-subject design. *Arch. Phys. Med. Rehabil.* 87, 1141–1149. doi: 10.1016/j.apmr.2006.04.010

Broeren, J., Bjorkdahl, A., Claesson, L., Goude, D., Lundgren-Nilsson, A., Samuelsson, H., et al. (2008). Virtual rehabilitation after stroke. *Stud. Health Technol. Inform.* 136, 77–82.

Cassidy, S., and Eachus, P. (2002). Developing the computer user self-efficacy (CUSE) scale: investigating the relationship between computer self-efficacy, gender and experience with computers. *J. Educ. Comput. Res.* 26, 133–153. doi: 10.2190/JGJR-0KVL-HRF7-GCNV

Charsky, D. (2010). From edutainment to serious games: a change in the use of game characteristics. *Games Cult.* 5, 177–198. doi: 10.1177/1555412009354727

Chiu, P.-S., Chen, H.-C., Huang, Y.-M., Liu, C.-J., Liu, M.-C., and Shen, M.-H. (2016). A video annotation learning approach to improve the effects of video learning. *Innov. Educ. Teach. Int.* 1–11. doi: 10.1080/14703297.2016.1213653

Christodoulou, C., DeLuca, J., Ricker, J., Madigan, N., Bly, B., Lange, G., et al. (2001). Functional magnetic resonance imaging of working memory impairment after traumatic brain injury. *J. Neurol. Neurosurg. Psychiatry* 71, 161–168. doi: 10.1136/jnnp.71.2.161

Cicerone, K. D., Dahlberg, C., Kalmar, K., Langenbahn, D. M., Malec, J. F., Bergquist, T. F., et al. (2000). Evidence-based cognitive rehabilitation: recommendations for clinical practice. *Arch. Phys. Med. Rehabil.* 81, 1596–1615. doi: 10.1053/apmr.2000.19240

Cicerone, K. D., Dahlberg, C., Malec, J. F., Langenbahn, D. M., Felicetti, T., Kneipp, S., et al. (2005). Evidence-based cognitive rehabilitation: updated review of the literature from 1998 through 2002. *Arch. Phys. Med. Rehabil.* 86, 1681–1692. doi: 10.1016/j.apmr.2005.03.024

Cicerone, K. D., Langenbahn, D. M., Braden, C., Malec, J. F., Kalmar, K., Fraas, M., et al. (2011). Evidence-based cognitive rehabilitation: updated review of the literature from 2003 through 2008. *Arch. Phys. Med. Rehabil.* 92, 519–530. doi: 10.1016/j.apmr.2010.11.015

Claessen, M. H., van der Ham, I. J., Jagersma, E., and Visser-Meily, J. M. (2016). Navigation strategy training using virtual reality in six chronic stroke patients: a novel and explorative approach to the rehabilitation of navigation impairment. *Neuropsychol. Rehabil.* 26, 822–846. doi: 10.1080/09602011.2015.1045910

Dixon, R. A., and Bäckman, L. (1999). "Principles of compensation in cognitive neurorehabilitation," in *Cognitive Neurorehabilitation*, eds D. T. Stuss, G. Winocur, and I. H. Robertson (New York, NY: Cambridge University Press), 59–72.

Dobkin, B. H., and Dorsch, A. (2013). New evidence for therapies in stroke rehabilitation. *Curr. Atheroscler. Rep.* 15:331. doi: 10.1007/s11883-013-0331-y

Erhel, S., and Jamet, E. (2013). Digital game-based learning: impact of instructions and feedback on motivation and learning effectiveness. *Comput. Educ.* 67, 156–167. doi: 10.1016/j.compedu.2013.02.019

Fann, J. R., Katon, W. J., Uomoto, J. M., and Esselman, P. C. (1995). Psychiatric disorders and functional disability in outpatients with traumatic brain injuries. *Am. J. Psychiatry* 152, 1493–1499. doi: 10.1176/ajp.152.10.1493

Furnham, A. F., and Gunter, B. (1985). Sex, presentation mode and memory for violent and non-violent news. *J. Educ. Telev.* 11, 99–105. doi: 10.1080/0260741850110203

Garris, R., Ahlers, R., and Driskell, J. E. (2002). Games, motivation, and learning: a research and practice model. *Simul. Gaming* 33, 441–467. doi: 10.1177/1046878102238607

Geusgens, C. A., Winkens, I., van Heugten, C. M., Jolles, J., and van den Heuvel, W. J. (2007). Occurrence and measurement of transfer in cognitive rehabilitation: a critical review. *J. Rehabil. Med.* 39, 425–439. doi: 10.2340/16501977-0092

Jackson, S. A., and Marsh, H. W. (1996). Development and validation of a scale to measure optimal experience: the Flow State Scale. *J. Sport Exerc. Psychol.* 18, 17–35. doi: 10.1123/jsep.18.1.17

Kent, T. M., Marraffino, M. D., Najle, M. B., Sinatra, A. M., and Sims, V. K. (2012). "Effects of input modality and expertise on workload and video game performance," in *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, (Thousand Oaks, CA: SAGE Publications). doi: 10.1177/1071181312561233

Kulik, J. A., and Kulik, C.-L. C. (1988). Timing of feedback and verbal learning. *Rev. Educ. Res.* 58, 79–97. doi: 10.3102/00346543058001079

Limperos, A. M., Schmierbach, M. G., Kegerise, A. D., and Dardis, F. E. (2011). Gaming across different consoles: exploring the influence of control scheme on game-player enjoyment. *Cyberpsychol. Behav. Soc. Netw.* 14, 345–350. doi: 10.1089/cyber.2010.0146

Ma, V. Y., Chan, L., and Carruthers, K. J. (2014). Incidence, prevalence, costs, and impact on disability of common conditions requiring rehabilitation in the United States: stroke, spinal cord injury, traumatic brain injury, multiple sclerosis, osteoarthritis, rheumatoid arthritis, limb loss, and back pain. *Arch. Phys. Med. Rehabil.* 95, 986–995.e1. doi: 10.1016/j.apmr.2013.10.032

Mayer, R. E., Hegarty, M., Mayer, S., and Campbell, J. (2005). When static media promote active learning: annotated illustrations versus narrated animations in multimedia instruction. *J. Exp. Psychol. Appl.* 11, 256–265. doi: 10.1037/1076-898X.11.4.256

Mayer, R. E., and Moreno, R. (2003). Nine ways to reduce cognitive load in multimedia learning. *Educ. Psychol.* 38, 43–52. doi: 10.1207/S15326985EP3801_6

McDowell, S., Whyte, J., and D'Esposito, M. (1997). Working memory impairments in traumatic brain injury: evidence from a dual-task paradigm. *Neuropsychologia* 35, 1341–1353. doi: 10.1016/S0028-3932(97)00082-1

McEwan, M., Johnson, D., Wyeth, P., and Blackler, A. (2012). "Videogame control device impact on the play experience," in *Proceedings of the 8th Australasian Conference on Interactive Entertainment: Playing the System*, eds D. Cermak-Sassenrath and C. Walker (Auckland: Aotea Centre). doi: 10.1145/2336727.2336745

Michael, D. R., and Chen, S. L. (2005). *Serious Games: Games that Educate, Train, and Inform*. Cincinnati, OH: Muska & Lipman.

Nickell, G. S., and Pinto, J. N. (1986). The computer attitude scale. *Comput. Hum. Behav.* 2, 301–306. doi: 10.1016/0747-5632(86)90010-5

Peeters, W., van den Brande, R., Polinder, S., Brazinova, A., Steyerberg, E. W., Lingsma, H. F., et al. (2015). Epidemiology of traumatic brain injury in Europe. *Acta Neurochir.* 157, 1683–1696. doi: 10.1007/s00701-015-2512-7

Pinelle, D., Wong, N., and Stach, T. (2008). Heuristic evaluation for games: usability principles for video game design. *Paper Presented at the Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, Florence.

Rogers, R., Bowman, N. D., and Oliver, M. B. (2015). It's not the model that doesn't fit, it's the controller! The role of cognitive skills in understanding the links between natural mapping, performance, and enjoyment of console video games. *Comput. Hum. Behav.* 49, 588–596. doi: 10.1016/j.chb.2015.03.027

Schooler, L. J., and Anderson, J. R. (2008). "The disruptive potential of immediate feedback," in *Proceedings of the Cognitive Science Society*, (Hillsdale, NJ: Erlbaum).

Shafer, D. M., Carbonara, C. P., and Popova, L. (2014). Controller required? The impact of natural mapping on interactivity, realism, presence, and enjoyment in motion-based video games. *Presence* 23, 267–286. doi: 10.1162/PRES_a_00193

Shin, D.-H., and Chung, K.-M. (2017). The effects of input modality and story-based knowledge on users' game experience. *Comput. Hum. Behav.* 68, 180–189. doi: 10.1016/j.chb.2016.11.030

Shute, V. J. (2008). Focus on formative feedback. *Rev. Educ. Res.* 78, 153–189. doi: 10.3102/0034654307313795

Sternberg, D. A., Ballard, K., Hardy, J. L., Katz, B., Doraiswamy, P. M., and Scanlon, M. (2013). The largest human cognitive performance dataset reveals insights into the effects of lifestyle factors and aging. *Front. Hum. Neurosci.* 7:292. doi: 10.3389/fnhum.2013.00292

Stevens, J. A., and Kincaid, J. P. (2015). The relationship between presence and performance in virtual simulation training. *Open J. Model. Simul.* 3, 41–48. doi: 10.1080/00140139.2015.1122234

Susi, T., Johannesson, M., and Backlund, P. (2007). *Serious Games: An Overview.* Skövde: Institutionen för kommunikation och information.

Trialists, O. S. (2004). Rehabilitation therapy services for stroke patients living at home: systematic review of randomised trials. *Lancet* 363, 352–356. doi: 10.1016/S0140-6736(04)15434-2

van de Ven, R. M., Buitenweg, J. I. V., Schmand, B., Veltman, D. J., Aaronson, J. A., Nijboer, T. C. W., et al. (2017). Brain training improves recovery after stroke but waiting list improves equally: a multicenter randomized controlled trial of a computer-based cognitive flexibility training. *PLoS One* 12:e0172993. doi: 10.1371/journal.pone.0172993

van der Ham, I. J., Kant, N., Postma, A., and Visser-Meily, J. (2013). Is navigation ability a problem in mild stroke patients? Insights from self-reported navigation measures. *J. Rehabil. Med.* 45, 429–433. doi: 10.2340/16501977-1139

Verhage, F. (1964). *Intelligentie en Leeftijd: Onderzoek Bij Nederlanders van Twaalf tot Zevenenzeventig Jaar.* Assen: Van Gorcum.

Walma van der Molen, J. H., and Van Der Voort, T. H. (2000). Children's and adults' recall of television and print news in children's and adult news formats. *Commun. Res.* 27, 132–160. doi: 10.1177/009365000027002002

Wolf, S. L., Blanton, S., Baer, H., Breshears, J., and Butler, A. J. (2002). Repetitive task practice: a critical review of constraint-induced movement therapy in stroke. *Neurologist* 8, 325–338.

Yoo, J. W., Lee, D. R., Sim, Y. J., You, J. H., and Kim, C. J. (2014). Effects of innovative virtual reality game and EMG biofeedback on neuromotor control in cerebral palsy. *Biomed. Mater. Eng.* 24, 3613–3618. doi: 10.3233/BME-141188

Youngblut, C., and Huie, O. (2003). The Relationship between presence and performance in virtual environments: results of a VERTS study. *Proceedings of the IEEE Virtual Reality Conference, 2003* Washington, DC: IEEE Computer Society. doi: 10.21236/ADA427495

Yusoff, A., Crowder, R., Gilbert, L., and Wills, G. (2009). A Conceptual Framework For Serious Games. *Proceedings of the ICALT 2009, 9th IEEE International Conference on Advanced Learning Technologies*, eds I. Aedo, N. S. Chen, K. D. Sampson, L. Zaitseva (Southampton: University of Southampton), 21–23. doi: 10.1109/ICALT.2009.19

Zickefoose, S., Hux, K., Brown, J., and Wulf, K. (2013). Let the games begin: a preliminary study using Attention Process Training-3 and Lumosity[TM] brain games to remediate attention deficits following traumatic brain injury. *Brain Inj.* 27, 707–716. doi: 10.3109/02699052.2013.775484

frontiers in Psychology

Young Children and Voice Search: What We Know From Human-Computer Interaction Research

Silvia B. Lovato* and Anne Marie Piper

Communication Studies, Northwestern University, Evanston, IL, United States

Young children are prolific question-askers. The growing ubiquity of voice interfaces (e.g., Apple's Siri, Amazon's Alexa), as well as the availability of voice input in search fields, now make it possible for children to ask questions via Internet search when they are able to speak clearly, but before they have learned to read and write, typically between 3 and 6 years of age. The prevalence of *voice search* makes it important to understand children's changing conceptions of digital devices as a source of information and the role of technology-mediated question-asking in development. While limited research has focused on young children's use of voice interfaces, reviewing two related bodies of literature sheds light on how this use might unfold. This paper brings together studies of how children look for information, and of how they perceive and understand the informational and social roles of technology, drawing on human-computer interaction research. We conclude by highlighting lines of questioning for future work on younger children's interaction through voice search.

Keywords: children, question-asking, voice interfaces, internet search, information-seeking

VOICE SEARCH AND CHILDREN'S QUESTIONS

Young children are curious and prolific question-askers. They are known to ask factual and causal questions about the world around them when they perceive a gap in their understanding (Tizard and Hughes, 1984; Callanan and Oakes, 1992; Chouinard et al., 2007). Voice interfaces powered by natural language processing, such as Apple's Siri, Amazon's Alexa, and the Google Assistant, as well as the availability of microphone input features on the search fields of Google, YouTube, and other services make it possible for children to press a button, or use a "wake word," and simply ask a question or perform an Internet search. The present paper refers to this interaction as *voice search*.

Voice search is now a common part of many interfaces on traditional computers, connected home speakers, and mobile devices. Smartphones and tablet computers in particular have become ubiquitous in the lives of American children. According to a report from Common Sense Media (Rideout, 2017), almost all (98%) American children now live in a home with a tablet or smartphone, and this trend includes low-income families adopting technology at similar rates (Kabali et al., 2015). Unlike keyboards and mice, touch screens are immediately intuitive and operated with gestures such as pointing and swiping, which develop in the first year of life. Indeed, children aged 12 to 17 months are able to navigate simple tablet-based applications with moderate ability (Hourcade et al., 2015).

While text-based Internet search results might not be accessible to pre- and emerging readers, the explosive growth of online video content supports young children's ability to independently find and consume online information. Robert Kyncl, YouTube's Chief Business Officer, predicted at the 2012 Consumer Electronics Show that by 2020, 90% of Internet traffic would be used by video, a prediction later anticipated by Cisco to 2019 (Tribbey, 2016). Google search results come with a video tab, from which children can choose a video based on a representative picture, press the play button and watch a video related to their query. Thus, this dramatic shift in young children's ability to search online is due to both the prevalence of voice-based, natural language search features and the increasing volume of video-based search results.

This shift in children's ability to find information through connected devices makes it important to understand their changing conceptions of digital devices as sources of information and how they might fare as they attempt to use them to find answers to their questions. However, research has yet to understand the behaviors of young children using voice search. By young children, here, we refer to children who are able to speak fully formed sentences but have yet to learn how to read and write with enough fluency to perform internet searches by typing, typically between the ages of 3 and 6 years.

To help bridge this gap and identify promising lines of future work, this review examines the existing literature on children's search behavior as well as studies of children's perceptions of technology. These existing studies largely focus on children ages 7 and older, because until recently, searching had required reading, writing and typing skills, making it out of reach for younger children. However, the findings of these studies can help shed some light on what might happen when younger children attempt to ask questions of technology independently.

We start by reviewing studies that focus on how children as young as age 7 have searched the Internet at various points during its history, including child-specific web directories like Yahooligans (Bilal, 2000, 2001, 2002), keyword searches (Druin et al., 2009, 2010) and the more recent use of natural language (Kammerer and Bohnacker, 2012). To complement this literature, we review studies of how children understand technology, including their ideas about computers in general (e.g., Van Duuren et al., 1998; Rücker and Pinkwart, 2016 for a comprehensive review), how they understand robots (Kahn et al., 2012, 2013) and media technology (Reeves and Nass, 1996; Chiasson and Gutwin, 2005). We include these studies because children's perceptions of technology impact their expectations of whether such technology might serve as sources of information. We end by putting what we know in the context of child development and suggesting new areas for future research and design involving developmentally appropriate interactions through voice search.

CHILDREN AND INTERNET SEARCH

Studies of how children aged 7 and older search for information in digital interfaces began with the CD-ROM encyclopedias and digital libraries of the 1980s and 1990s, where the realm of information available was limited (e.g., Marchionini, 1989). Even then, elementary-aged children showed a tendency to use natural language in search fields (Marchionini, 1989). In a system that was designed to find keywords, this strategy failed, generating no results (Solomon, 1993).

In a series of studies of seventh graders using the web directory Yahooligans, a child-focused resource managed by Yahoo, Inc., from 1996 to 2006, Bilal (2000, 2001, 2002) found that children consistently preferred to browse the directory than to use the search functionality. Only 50% of the students succeeded at finding answers to specific, fact-based queries given by a science teacher, while 69% partially succeeded at researching a topic more generally using their own queries and 73% succeeded at finding answers to an undirected, self-generated query. Bilal (2002) also reported that 13% of children in the third study, who were using their own queries, used natural language instead of keywords, something seen as a liability at the time, leading to the conclusion that students should receive better web search training.

In a more recent study about how children ages 7, 9, and 11 used keyword interfaces to search the Internet (Druin et al., 2009), the researchers found that children had trouble typing, spelling and deciding which words to use as search terms. Specifically, children tended to look at the keyboard while typing, making it difficult to catch typos until the entire word or phrase had been entered and to see the predictive terms offered by the search engine. Parents in their study suggested voice-input as a solution to children's typing and spelling problems. The study also found other difficulties that might not be eliminated by voice input: for example, children had difficulty choosing which words to use and breaking down a complex query into multiple steps when needed (query reformulation). When asked to find what day of the week the vice-president's birthday would fall on the following year, none of the children were able to find the answer; the youngest children, age 7, did not even try.

In a larger study including 83 children, again aged 7, 9, and 11, and their parents (Druin et al., 2010), these findings were confirmed and expanded: the researchers identified seven distinct search "roles," or search behavior patterns, displayed by the children, in isolation or combined with one or more other roles. Each of these roles is associated with specific behaviors, triggers (motivation for using search), obstacles (such as typing, spelling and reading difficulties, lack of motivation and self-imposed limiting rules) and influencer, or parent, roles (demonstrator, fixer, mentor). The most common role was that of a *developing* searcher, displayed by 58 of the children. Developing searchers were found to be willing to search but possess a limited command of search tools and, again, a tendency to use natural language. The developing role was most often displayed at the same time as that of *domain-specific* searcher, in which children are comfortable with a few "tried-and-true" resources, usually related to personal interests, and tend to return to those websites repeatedly, even when searching for unrelated information. For example, children attempted to find information about dolphins and about the vice-president of the United States at a games website and on spongebob.com. Other roles identified were *power searcher*, *distracted*, *non-motivated*, *visual,* and *rule-based searcher*.

While Druin et al. found that children's use of natural language in search engines was problematic, Kammerer and Bohnacker (2012) compared natural language to keyword searches performed by 21 children aged 8 to 10 using Google in German and found that natural language users were more successful than those using keywords. They gave children a two-part task in which the first part was a simple yes/no question (do all kangaroos have pouches?) and the second required a more complex strategy and answer (how do baby kangaroos stay in pouches?). Tasks were given orally and children could choose what to enter in the search field. Of the 13 natural language users, 8 were able to answer both parts of the task correctly, 4 were able to answer only the first and one was unable to answer either. The 8 keyword users fared far worse, with only 3 being able to answer both queries correctly, 3 answering only the first and 2 being unable to answer either.

As we consider younger children using voice interfaces to search, some of the mechanical obstacles identified by prior work (e.g., typing and spelling difficulties) may lose importance while the discrepancy between the intended users of the interface, by and large adults, and younger children, who now have access to search, increases. For example, Druin's *domain-specific* searchers might become *app-specific* in this generation. Young children who become comfortable searching inside an application such as *YouTube Kids* could attempt to use it for queries that would be better served by a different tool. Younger users also have a less developed vocabulary and may be less precise in how they formulate queries. Additionally, while videos and spoken responses may dispense with reading requirements, such audio and video content was likely not produced with young children in mind, creating the potential for comprehension difficulties. These obstacles, however, only matter if younger children indeed perceive these technologies as sources of information and attempt to ask questions of them.

CHILDREN'S PERCEPTIONS OF DIGITAL DEVICES

To predict whether young children might see the devices they use as potential sources of answers to their questions and not just game and video players, we consider how they conceptualize computing devices. Existing work (Van Duuren et al., 1998; Papastergiou, 2005; Yan, 2005; Diethelm et al., 2012; see Rücker and Pinkwart, 2016 for a review) about older children's understanding of computers and the Internet has found that they perceive computers as capable of containing infinite amounts of information (e.g., Van Duuren et al., 1998); however, results of a few recent studies (McKenney and Voogt, 2010; Eisen and Lillard, 2016, 2017) with younger children have been mixed.

Rücker and Pinkwart (2016) present a systematic, interdisciplinary review of studies of children's conceptions of computers. They identify five main ideas children have expressed in studies over the years, between 1968 and 2012: (1) intelligent machines; (2) omniscient databases; (3) mechanical devices; (4) wire networks and (5) programmable machines. As we consider the notion of computer-like devices as information

sources to young children, the most relevant concepts are those of an intelligent machine and an omniscient database. Studies included in the review found that children aged 8 and 11 (but not 5-year-olds) believed computers had the results of all possible mathematical calculations already stored in their memory (Van Duuren et al., 1998) and that 12-to-16 year-olds believed that the entire Internet was stored in one single computer, either the user's own or another accessible through the network (Papastergiou, 2005; Diethelm et al., 2012).

Studies with younger children, however, present a more mixed picture. A study of Dutch children's perceptions of their own computer use including 4- to 7-year-olds, most of whom had daily access to computers both in and out of school, found that the overwhelming majority of young children used computers to play games and that using the computer for a creative or communicative activity or to search the Internet was far less common (McKenney and Voogt, 2010). Eisen and Lillard (2016, 2017) performed two studies to understand which functions preschool children attribute to touch screen devices when compared to other media such as television and books. In the first study, they found that children tend to attribute fewer functions to most objects than adults. When asked to identify the best object for learning about dogs, hearing Spanish or looking at a map, touch screen devices were not their top choice. The computer was the preferred method for seeing a map. In the second study, children were asked to choose between a tablet computer and a book for several learning tasks (e.g., cooking, the weather, Virginia, yesterday's football game). While the younger children in the sample showed no clear preference, 6-year-olds preferred the tablet computer for most tasks. However, children did not take into account whether the information sought was timely (i.e., the weather, yesterday's football game), with even 6-year-olds preferring books to learn about the game.

While voice input is available, for example via the Google mobile application, as well as YouTube and YouTube Kids, these interfaces don't respond verbally, but show the user's query as text input in the search field and then display search results after the query is submitted. Conversational agents (like Siri or Alexa), on the other hand, are programmed to respond as a person would. Research on how children understand and interact with robots and with other media provides insights into how machines that attempt to act like humans are perceived by children.

Kahn et al. (2013) argue that social robots are establishing a new ontological category, distinct from humans, animals or simple artifacts. As children interact with a social robot, they tend to believe that it has rights and feelings (Kahn et al., 2012). At the same time, they are aware of the robot's machine status. Through a number of experiments, Reeves and Nass (1996) found that people tend to respond to computers and other media as they would to humans. They refer to this phenomenon as *the media equation*. The set of interactions that are specific to computers, whose responses, unlike those of television, are contingent on user input, are studied under an area of research called CASA (Computers as Social Actors). But is the tendency the same in children? Some critics of these theories argue that only inexperienced users would respond to machines as if they were people. Children, then, could easily be expected

to act this way. Chiasson and Gutwin (2005) predicted that children would be even more affected by the media equation than adults, since they are more likely to anthropomorphize objects and accept fictional characters as real. They also predicted that providing social cues in interfaces that made interactions closer to those with people would help children stay engaged in educational activities. To test this, they replicated two classic Nass and Reeves CASA experiments comparing groups of adults to children aged 10 to 12. In both experiments, they measured the impact of social language – praise in one case and treating the participant as part of the computer's team in another – on users' assessments of their own experiences playing simple games. Surprisingly, they found that, while social language had a positive impact on adults, it had no impact on the children. They proposed two explanations for this: one is that children are so affected by the media equation that this overwhelms any difference between experimental conditions (i.e., they would have had a positive experience regardless of the social language in the game). The other explanation is that people who have grown up with computers, as was the case of the child participants, are less susceptible to the media equation than those who learned to use computers later in life, as was the case of the adult participants.

VOICE SEARCH IN THE CONTEXT OF DEVELOPMENT: DIRECTIONS FOR FUTURE RESEARCH

While voice input removes mechanical obstacles to Internet searching, such as typing and spelling, there are other developmental factors to take into account as we consider younger children using voice search. First, children who are able to make themselves understood by language processing software are still developing theory of mind skills, broadly defined as the set of skills that allows us to understand the mental states of others. From our own prior work (Lovato and Piper, 2015), we know that one of the obstacles young children face when using voice search is not fully understanding what the system can and cannot answer (i.e., what the system knows) and how much context to provide. For example, systems cannot usually answer questions about the location of specific people or objects – at least not yet – and cannot answer questions about undescribed objects or referents it cannot see (e.g., "where was this made?"). Understanding what someone knows is an aspect of theory of mind that is still in development in young children (Wellman and Liu, 2004).

Preschoolers' trust in technology sources has been found to be largely based on previous experience, as it is with people (Danovitch and Alzahabi, 2013). This behavior evolves with age, with 4- and 5-year-olds being more likely to use past experience as a reference than 3-year-olds (Mills et al., 2011). The imperfect ability of voice agents to understand children's speech, combined with the agents' inability to ask for more information or context, could have an impact on how much children learn to rely on conversational agents as sources: if Siri or Alexa misunderstands a child and responds with an answer that

doesn't make sense, the child might lose trust in it as a source of answers.

While the existing literature on older children's Internet search and perceptions of technology as information sources seems to support the potential for younger children to use voice search, it also points to two central lines of inquiry regarding what happens when younger children ask questions of voice interfaces or conversational agents. The first relates to the distinct obstacles young children might face when using this technology and how voice interfaces can better support children in their developmental needs. The second, equally important question, relates to how the particular use of language required by search engines and conversational agents might shape how children learn to use language to obtain information.

As mentioned, young children ask questions when they perceive an inconsistency, or a gap, in their understanding of the world (Tizard and Hughes, 1984; Callanan and Oakes, 1992; Chouinard et al., 2007). Chouinard et al. (2007) found that children's levels of persistence in question-asking are high when they receive responses that do not contain the information requested and low when they do receive such information, suggesting that children really are looking for information (as opposed to simply adult attention). In understanding young children's goals when asking information-seeking questions (i.e., filling gaps in understanding), it is important to consider what an optimal answer would be: would a piece of information stated in a way the child can understand suffice? Or is a conversation indispensable?

It is possible that when children ask questions, at least some of the time, a simple factual answer is not the best answer. When children direct factual questions at adults, these serve as "more knowledgeable others," who help children advance their state of development (Vygotsky, 1978). Parents and teachers might ask a child why she is asking a question, or what she thinks the answer is, scaffolding the child as she figures out the answer, partly on her own. Through dialog, children not only develop understanding, but also language and reasoning. Can conversational agents serve as more knowledgeable others? Future research should consider how child-friendly conversational agents should respond to children's queries for optimal child development outcomes.

There is no question that voice search and conversational agents will continue to develop. It is not impossible for this technology to be made more child-friendly by, for example, learning to distinguish between child and adult voices and responding to children in ways that are more supportive. A system could explain what it cannot answer or request additional information in order to respond to a query. Such developments could encourage young children to use these systems more frequently, in turn increasing our need to understand how such use could impact language development and cognitive development more broadly.

AUTHOR CONTRIBUTIONS

REFERENCES

Bilal, D. (2000). Children's use of the Yahooligans! Web search engine: I. Cognitive, physical, and affective behaviors on fact-based search tasks. *J. Am. Soc. Inf. Sci.* 51, 646–665. doi: 10.1002/(SICI)1097-4571(2000)51:7<646::AID-ASI7>3.0.CO;2-A

Bilal, D. (2001). Children's use of the Yahooligans! Web search engine: II. Cognitive and physical behaviors on research tasks. *J. Am. Soc. Inf. Sci. Technol.* 52, 118–136. doi: 10.1093/nar/gkw1066 doi: 10.1093/nar/gkw1066

Bilal, D. (2002). Children's use of the Yahooligans! Web search engine. III. Cognitive and physical behaviors on fully self-generated search tasks. *J. Am. Soc. Inf. Sci. Technol.* 53, 1170–1183. doi: 10.1002/asi.10145

Callanan, M. A., and Oakes, L. M. (1992). Preschoolers' questions and parents' explanations: causal thinking in everyday activity. *Cogn. Dev.* 7, 213–233. doi: 10.1016/0885-2014(92)90012-G

Chiasson, S., and Gutwin, C. (2005). "Testing the media equation with children," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ed. W. E. Mackay (New York, NY: ACM), 829–838.

Chouinard, M. M., Harris, P. L., and Maratsos, M. P. (2007). Children's questions: a mechanism for cognitive development. *Monogr. Soc. Res. Child Dev.* 72, vii–ix, 1–112; discussion 113–26.

Danovitch, J. H., and Alzahabi, R. (2013). Children show selective trust in technological informants. *J. Cogn. Dev.* 14, 499–513. doi: 10.1080/15248372.2012.689391

Diethelm, I., Wilken, H., and Zumbrügel, S. (2012). "An investigation of secondary school students' conceptions on how the internet works," in *Proceedings of the 12th Koli Calling International Conference on Computing Education Research* (New York, NY: ACM), 67–73. doi: 10.1145/2401796.2401804

Druin, A., Foss, E., Hatley, L., Golub, E., Guha, M. L., Fails, J., et al. (2009). "How children search the internet with keyword interfaces," in *Proceedings of the 8th International Conference on Interaction Design and Children* (New York, NY: ACM), 89–96. doi: 10.1145/1551788.1551804

Druin, A., Foss, E., Hutchinson, H., Golub, E., and Hatley, L. (2010). "Children's roles using keyword search interfaces at home," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (New York, NY: ACM), 413–422. doi: 10.1145/1753326.1753388

Eisen, S., and Lillard, A. S. (2016). Just google it: young children's preferences for touchscreens versus books in hypothetical learning tasks. *Front. Psychol.* 7:1431. doi: 10.3389/fpsyg.2016.01431

Eisen, S., and Lillard, A. S. (2017). Young children's thinking about touchscreens versus other media in the US. *J. Child. Media* 11, 167–179. doi: 10.1080/17482798.2016.1254095

Hourcade, J. P., Mascher, S. L., Wu, D., and Pantoja, L. (2015). "Look, my baby is using an iPad! an analysis of youtube videos of infants and toddlers using tablets," in *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (New York, NY: ACM), 1915–1924. doi: 10.1145/2702123.2702266

Kabali, H. K., Irigoyen, M. M., Nunez-Davis, R., Budacki, J. G., Mohanty, S. H., Leister, K. P., et al. (2015). Exposure and use of mobile media devices by young children. *Pediatrics* 136, 1044–1050. doi: 10.1542/peds.2015-2151

Kahn, P. H. Jr., Kanda, T., Ishiguro, H., Freier, N. G., Severson, R. L., Gill, B. T., et al. (2012). "Robovie, you'll have to go into the closet now": children's social and moral relationships with a humanoid robot. *Dev. Psychol.* 48, 303–314. doi: 10.1037/a0027033

Kahn, P. H., Gary, H. E., and Shen, S. (2013). Children's social relationships with current and near future robots. *Child Dev. Perspect.* 7, 32–37. doi: 10.1111/cdep.12011

Kammerer, Y., and Bohnacker, M. (2012). "Children's web search with Google: the effectiveness of natural language queries," in *Proceedings of the 11th International Conference on Interaction Design and Children* (New York, NY: ACM), 184–187. doi: 10.1145/2307096.2307121

Lovato, S., and Piper, A. M. (2015). "Siri, is this you?": understanding young children's interactions with voice input systems," in *Proceedings of the 14th International Conference on Interaction Design and Children* (New York, NY: ACM), 335–338. doi: 10.1145/2771839.2771910

Marchionini, G. (1989). Information-seeking strategies of novices using a full-text electronic encyclopedia. *J. Am. Soc. Inf. Sci.* 40, 54–66. doi: 10.1002/(SICI)1097-4571(198901)40:1<54::AID-ASI6>3.0.CO;2-R

McKenney, S., and Voogt, J. (2010). Technology and young children: how 4-7 year-olds perceive their own use of computers. *Comput. Hum. Behav.* 26, 656–664. doi: 10.1016/j.chb.2010.01.002

Mills, C. M., Legare, C. H., Grant, M. G., and Landrum, A. R. (2011). Determining who to question, what to ask, and how much information to ask for: the development of inquiry in young children. *J. Exp. Child Psychol.* 110, 539–560. doi: 10.1016/j.jecp.2011.06.003

Papastergiou, M. (2005). Students' mental models of the Internet and their didactical exploitation in informatics education. *Educ. Inf. Technol.* 10, 341–360. doi: 10.1007/s10639-005-3431-7

Reeves, B., and Nass, C. (1996). *How People Treat Computers, Television, and New Media Like Real People and Places*. Cambridge: CSLI Publications and Cambridge university press, 9–36.

Rideout, V. (2017). *The Common Sense census: Media Use by Kids Age Zero to Eight Chap*. San Francisco, CA. Common Sense Media.

Rücker, M. T., and Pinkwart, N. (2016). Review and discussion of children's conceptions of computers. *J. Sci. Educ. Technol.* 25, 274–283. doi: 10.1007/s10956-015-9592-2

Solomon, P. (1993). Children's information retrieval behavior: a case analysis of an OPAC. *J. Am. Soc. Inf. Sci.* 44, 245–264. doi: 10.1002/(SICI)1097-4571(199306)44:5<245::AID-ASI1>3.0.CO;2-#

Tizard, B., and Hughes, M. (1984). *Young Children Learning: Talking and Thinking at Home and at School*. London: SAGE Publishing.

Tribbey, C. (2016). *YouTube Exec Doubles Down on Online Video Growth Projection. Variety*. Available at: http://variety.com/2016/digital/news/youtube-exec-doubles-down-on-online-video-growth-projection-1201674923/ [accessed January 7, 2016].

Van Duuren, M., Dossett, B., and Robinson, D. (1998). Gauging children's understanding of artificially intelligent objects: a presentation of "counterfactuals". *Int. J. Behav. Dev.* 22, 871–889. doi: 10.1080/016502598384207

Vygotsky, L. S. (1978). *Mind in Society: The Development of Higher Mental Process*. Cambridge, MA: Harvard University Press.

Wellman, H. M., and Liu, D. (2004). Scaling of theory-of-mind tasks. *Child Dev.* 75, 523–541. doi: 10.1111/j.1467-8624.2004.00691.x

Yan, Z. (2005). Age differences in children's understanding of the complexity of the Internet. *J. Appl. Dev. Psychol.* 26, 385–396. doi: 10.1016/j.appdev.2005.04.001

Guided Embodiment and Potential Applications of Tutor Systems in Language Instruction and Rehabilitation

Manuela Macedonia [1,2]*, Florian Hammer [3] and Otto Weichselbaum [1,4]

[1] Information Engineering, Johannes Kepler Universität Linz, Linz, Austria, [2] Neural Mechanisms of Human Communication, Max-Planck-Institut für Kognitions- und Neurowissenschaften, Leipzig, Germany, [3] Linz Center of Mechatronics GmbH, Linz, Austria, [4] Sew Systems Gmbh, Linz, Austria

Intelligent tutor systems (ITSs) in mobile devices take us through learning tasks and make learning ubiquitous, autonomous, and at low cost (Nye, 2015). In this paper, we describe guided embodiment as an ITS essential feature for second language learning (L2) and aphasia rehabilitation (ARe) that enhances efficiency in the learning process. In embodiment, cognitive processes, here specifically language (re)learning are grounded in actions and gestures (Pecher and Zwaan, 2005; Fischer and Zwaan, 2008; Dijkstra and Post, 2015). In order to guide users through embodiment, ITSs must track action and gesture, and give corrective feed-back to achieve the users' goals. Therefore, sensor systems are essential to guided embodiment. In the next sections, we describe sensor systems that can be implemented in ITS for guided embodiment.

Keywords: tutor systems, language instruction, aphasia therapy, intelligent tutor system, gesture production, gesture recognition, learning

Today in L2 learning, ITSs transpose classroom activities as reading, listening, and making exercises in electronic environments (Holland et al., 2013). Similarly in ARe, a virtual therapist in a tablet helps patients in the treatment of verbal anomia by presenting pictures (Lavoie et al., 2016). Virtual therapists do basically what a human therapist would do, i.e., they ask patients to name the pictures presented (Brandenburg et al., 2013; Kurland et al., 2014; Szabo and Dittelman, 2014).

Both domains, L2 and ARe, still treat language a purely mentalistic process, a manipulation of symbols in our minds (Fodor, 1976, 1983). Consequently, symbols such as written words or pictures representing the word's semantics are the base of main stream language educational and rehabilitation methods. Despite this, in the last three decades, a growing number of studies have converged to suggest that language as a cognitive capacity is grounded in our bodily experiences in the environment, in perception and action (Lakoff, 2012; Dijkstra and Post, 2015; Borghi and Zarcone, 2016). Words are not symbols any more. Instead, they have been described as "experience related brain networks" (Pulvermüller, 2002). Interestingly, not only concrete but also abstract vocabulary is rooted in the body. In a comprehensive review of neuroscientific studies, Meteyard and colleagues show that simple recognition of abstract words elicits activity in sensorimotor brain regions (Meteyard et al., 2012). This is explained by the fact that abstract concepts are also internalized by real experiences that in their turn are related to the body. Take for example the word *love*: it is embodied because acquired from concrete and experienced concepts, i.e., perceiving the partner physically, doing things with the partner, and so on. All these experiences converge to a metaphorical extension which is labeled as *love*.

In fact, first language acquisition is tightly connected to sensorimotor experiences (Inkster et al., 2016; Thill and Twomey, 2016). In infancy, the body is the main vehicle that collects experiences related to language units as nouns and verbs (Tomasello et al., 2017). Furthermore, very early in development, gestures make their appearance. They are precursors of spoken language (Mattos and Hinzen, 2015) and tightly bound to it. Language and gestures represent the two sides of the human communicative system (Kelly et al., 2010).

In adult age, the body can be used as a tool to enhance memory for verbal information (Zimmer, 2001). This is achieved by performing gestures to words or phrases that are to be memorized. The effect of gestures on memory for verbal information has been named "enactment effect" (EE) Engelkamp and Zimmer (1985) and "self-performed task effect" (Cohen, 1981). The EE is robust and has been extensively investigated with different materials, tests, and populations (Von Essen and Nilsson, 2003). In memory research, the EE effect has been reconducted to a motor trace that the gesture leaves in words' representations (Engelkamp, 1998).

Also, in second language learning, self-performed gestures accompanying words enhance memory performance compared to just reading the words and/or listening to them (Macedonia, 2014), in the short and in the long term (Macedonia and Klimesch, 2014). In a study with functional Magnet Resonance Imaging (fMRI), Macedonia and Mueller (2016) have shown that passive recognition of second language words trained with gestures activates extended sensorimotor networks. These networks involve motor cortices and subcortical structures as the basal ganglia, and the cerebellum. They all participate to a large motor network. It is thus conceivable that retention is superior because words learned with gestures might engage procedural memory in addition to declarative memory (Nilsson and Bäckman, 1989). Interestingly, recent studies on patients with impaired procedural memory have demonstrated that the patients could not take advantage of learning through gestures (Klooster et al., 2014).

In aphasia, gestures produced by patients trying to communicate can easily be observed. These gestures fulfill compensatory functions (Göksun et al., 2015; Rose et al., 2016) if the patients' language is impoverished or omitted (Pritchard et al., 2015). However, because of the high variance in lesion patterns, age of the patients, patho-linguistic profile, intensity of intervention, etc., studies employing gestures and studies employing other therapeutic instruments are difficult to compare. Hence, effects of gestures on rehabilitation can be diverging (Kroenke et al., 2013). Main stream aphasia therapy is still constrained to the verbal modality and bans gestures as tool that might help to restore language networks (Pulvermüller, 2002). Nevertheless, a growing number of studies show that action and gesture can help support the missing side of the communicative coin (Rose, 2013). Whereas simple observation of action has a positive impact on word recovery (Bonifazi et al., 2013), observation followed by execution of action leads to better recovery results (Marangolo and Caltagirone, 2014). These studies pave the way for a novel understanding of aphasia therapy in which the body helps the mind to regain language

functions, as long as brain structures serving procedural memory are not compromised (Klooster et al., 2014).

This is to say that humans need the body to acquire first language, to support memory for verbal information, to learn a second language, and to reacquire language functions disrupted by brain lesions. At this point, a core issue is to stress that embodiment of language needs active experience. In enactment research, it has long been known that it is not enough to observe gestures and actions, one must perform them (Cohen, 1981; Engelkamp et al., 1994). When interacting with an ITS, the user is first presented with the language to be trained and the gestures to be performed. Thereafter, the user must perform the actions and the gestures. Monitoring can make action performance accurate in execution. Thus, one component of the ITS must detect motion and gesture, compare it with a template and give feedback on execution accuracy. Execution monitoring needs sensor systems.

TECHNOLOGIES FOR GESTURE PERFORMANCE MONITORING

Guided embodiment requires an interaction between ITS and user: A gesture representing a concept is performed by an ITS avatar. The user observes the gesture and imitates it. The user's gesture must be sensed during performance. Performance is evaluated by the system on the base of a template. Visual, auditory and or tactile feedback is given by the ITS (please see **Figure 1**).

Audio-Visual Gesture Presentation (AVGP)
First, a written word is presented to the user on a display simultaneously with a video in which an actor performs a representational gesture. The gesture can be presented by a human through a video or by an avatar, or an agent (Bergmann,
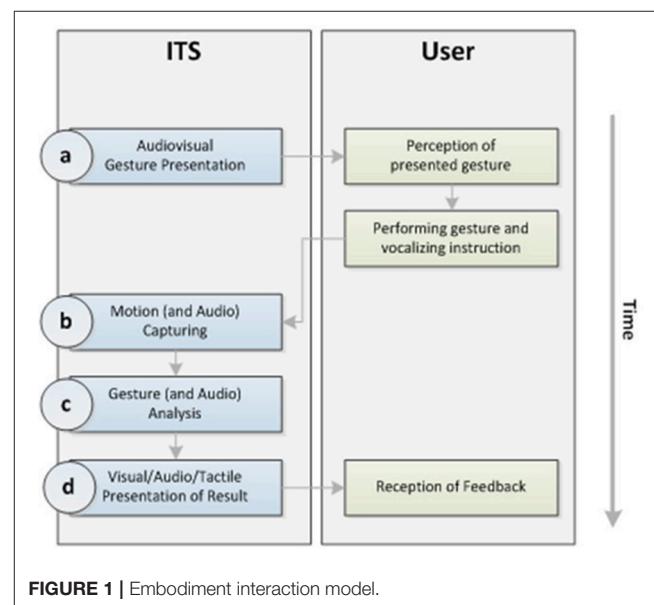
FIGURE 1 | Embodiment interaction model.

2015). Synchronously, an audio file of the word is played via loudspeaker.

Motion Capturing

Motion is the change of body position in time. Motion capturing occurs as a two-phases process. First, a single motion is sensed generating data (*motion sensing*) (Moeslund et al., 2006). Secondly, the data are sampled (*motion sampling*) and sequenced in time into a movement path, a so-called motion trajectory model. Depending on the location of the sensors used to detect the motion, Motion capturing can be subdivided into two categories: *infrastructure based* or through *wearables*. Infrastructure-based systems rely on hardware that is rigidly mounted inside a room as high-speed infrared cameras in a gait analysis laboratory, or sensors in a blue screen environment. Infrastructure based systems use sensors with high power consumption.

Systems based on **microwave, ultrasonic or radar sensors** operate by emitting electromagnetic or sonic waves and sensing the echo received. Depending on the purpose of motion capturing, sensor technologies can vary. For example, ultrasonic motion detection is quite common in prenatal diagnostics (Birnholz et al., 1978). For remote vital sign detection radar-based motion detection is frequently used (Lubecke et al., 2002).

Vision-based systems (VBS), including single camera, multiple cameras, and depth camera systems, play the most important role in human motion capture. Sensors detect light which can be visible or invisible to the human eye which is emitted or reflected by the body or an object (Moeslund et al., 2006).

Single camera-based motion detection systems are present in notebooks, tablets, and mobiles. Although these systems often have a high-quality resolution, they operate with a single camera. A single camera cannot capture the motion of body parts that are occluded by other body parts. This results in an inaccurate or incomplete analysis of the motion.

Multiple camera systems with two or more cameras allow 3D capturing. Algorithms combining 2D images from the cameras calculate a 3D-resolution (Aggarwal and Cai, 1997; Cai and Aggarwal, 1999). In the 3D-resolution, the synchronized recordings are combined. The combination includes the positions of the cameras relative to each other and their angles of view. Multiple camera systems are used in rigid mounted setups, in laboratories or dedicated rooms for example in rehabilitation (gait analysis), and sports (motion analysis).

Depth cameras sense 3D-information by means of infrared light. They calculate the distance between the camera and a body in two ways. They project an invisible grid onto the scenery and sense the grid's deformations. Alternatively, they measure the distance to the scenery and they calculate the transfer time of the infrared light from the camera to the object. This second kind of depth camera is also called "Time-of-flight"-camera (ToF) (Barnachon et al., 2014; Cunha et al., 2016; Garn et al., 2016).

Depth camera systems with a single device do not overcome the problem of occluded parts (Han et al., 2013). However, they have an advantage: they provide information about the distance of each object or body within the camera's view relative to the camera's position. These systems do not rely on heuristics about proportions of the object in order to determine its distance. This information increases accuracy in calculating the position of a human body or object.

Wearables are sensors worn on the body. They are light-weighted and have low power consumption. They are often used in sports (Roetenberg et al., 2013). Among wearables, we find inertial measurement units (IMUs) and sensing textiles.

Inertial Measurement Units (IMUs) are small electronic devices that measure acceleration, angular changes and changes in the magnetic field surrounding the body or object (Roetenberg, 2006; Shkel, 2011). If the starting position is known, an approximate position at time t is can be calculated by implementing the changes in forces, angles and magnetic field from the starting position up to t. IMUs differ from camera-based systems: while the latter measure the absolute position of the body at every time point t, IMUs acquire a starting position and the movement's sequence.

IMUs are integrated into wearable objects and respond on minimal deviations of the sensors by showing a drift. This drift can sum up to false positions over time. Fusion algorithms combining filtering and validation of sensor are used to compensate, respectively minimize drifts values (Luinge and Veltink, 2005; Sabatini, 2011; Roetenberg et al., 2013).

Sensing textiles represent a novel way of capturing motion. They consist of fabrics containing enwoven pressure sensitive fibers. These fibers change their electric resistance depending on the pressure changes that they sense (Mazzoldi et al., 2002; Parzer et al., 2016). Clothes tailored with these fabrics enable to calculate movements of the body in a fine-grained way (Parzer et al., 2016). The choice of the adequate type of motion sensing technology depends on the application domain. In our case, sensing of human body movements for an ITS can be accomplished with four sensor technologies: camera, depth-camera, IMUs, and sensing textiles.

Vision-based systems (VBS) take pictures over time and analyze them in order to detect body parts. Thereafter, VBS transform the detected body parts into digital representations, into human body models. Common models are skeletal, joint-based (Badler and Smoliar, 1979; Han et al., 2017), and mesh-based (de Aguiar et al., 2007). For an overview and classification of the major techniques used for sampling 3D data, please see Aggarwal and Xia (2014).

Additionally, VBS can increase the accuracy of the human body model by markers as light-emitting diodes, passive reflectors or patterns. These markers are fixed on pre-defined body parts and map them to the according representation within the model. Marker-less systems use heuristics about shapes, dimensions, and relations between body parts estimating and calculating the model according to these constraints.

Body data are sampled and thereafter transferred into a digital form in constant periods of time. This is done in order to obtain the motion trajectory model needed. It represents the body parts and their changes in posture over the time of recording (Poppe, 2010). Hence, motion sampling results in a motion trajectory model.

	Single camera	Multiple cameras	Depth camera	Sensing textiles	IMUs
Accuracy	0	++	+	++	0
Setup	++	+	+	++	+
Mobility	+	+	+	++	++
Size	+	+	0	++	+

0, moderately fulfilling the users' requirements; +, fulfilling the requirements; ++, fulfilling the requirements very well.

Gesture (and Audio) Analysis

In the literature, different approaches for matching motion trajectory models are discussed. Kollorz et al. (2008) ground their model on projections of image depth. Mitra and Acharya (2007) describe the use of hidden Markov models (Rabiner and Juang, 1986), finite-state machines (Marvin, 1967) and, neural networks (Lippmann, 1987). Other authors use a support-vector machine-based approach (Cristianini and Shawe-Taylor, 2000; Schuldt et al., 2004; Miranda et al., 2014). A template-based method for matching motion has been developed by Müller and Röder (2006). Stiefmeier et al. (2007) convert the motion trajectory model into strings of symbols. This is done in order to apply string matching algorithms that are faster in running analyses. Detailed reviews on vision-based human motion recognition methods are provided by Poppe (2010) and Weinland et al. (2011).

Embodiment-based ITS employed in language learning and rehabilitation need real-time processing of sensed gestures because of the immediate feedback on gesture accuracy that users need (Ganapathi et al., 2010).

Accuracy in sound reproduction is an important issue in both, second language learning and aphasia rehabilitation. Language output by the user is recorded and analyzed by different methods (Rabiner and Juang, 1993). Recent approaches employ complex models as neural networks for speech recognition (Hinton et al., 2012; Graves et al., 2013).

After a match between the sensed gesture or the voice and the template within the representing motion trajectory model has occurred, feedback can follow. It can be visual via the display, acoustical with sound through a speaker (built-in or external), and tactorial by means of a vibration given by the device. Feedback can be simple (i.e., a sound or synthesized speech).

Evaluation of Sensor Technologies

In order to give an overview of the sensor technologies presented in the preceding sections, we created **Table 1**. It describes the degree of following characteristics: accuracy in motion sensing, ease of set up for an expert, mobility and size. Note that the description is done for the use of a professional (lab technician) and for an institution (language school or hospital). We do not consider ITS software, software processes, and design patterns, or aspects of user-interface design. For further reading, please see (Oppermann, 2002; Dillon, 2003; Carroll, 2006; Smith-Atakan, 2006; Preece et al., 2015).

In this paper, we describe two application domains for ITS following principles of guided embodiment: language (re-)learning and aphasia rehabilitation. So far, we have focused on the possible use of the ITS in an institution (school vs. hospital). However, considering that language learning and rehabilitation need massed practice (Pulvermüller et al., 2001; Kurland et al., 2014), ITS should accompany users during the learning task in their homes. Sensing textiles can represent an emerging field in guided embodiment for language learning and aphasia rehabilitation. A learning t-shirt could combine a few advantages: high accuracy in sensing motion, ease of use and possible vibration feedback. However, to our knowledge no such system is present to date on the market, even as a prototype.

To present, only single camera systems present in tablets and mobile phones are affordable and easy to use. Also, nearly everyone has an own device. Because of their size, single camera systems can be carried where users need them. Despite the fact that presently single cameras are not very accurate in motion capturing as described in the preceding section, they might become the instruments used in a near future.

Altogether, this brief overview highlights the fact that guided embodiment of language could be the way to enhance performance in learning and rehabilitation. However, more research in the field is needed.

AUTHOR CONTRIBUTIONS

MM has laid down the structure of this paper and written the sections on embodiment. FH and OW have written the sections on technologies for gesture performance monitoring.

FUNDING

REFERENCES

Aggarwal, J. K., and Cai, Q. (1997). "Human motion analysis: a review," in *IEEE Nonrigid and Articulated Motion Workshop, IEEE Comput. Soc* (San Juan), 90–102.

Aggarwal, J. K., and Xia, L. (2014). Human activity recognition from 3D data : a review. *Patt. Recogn. Lett.* 48, 70–80. doi: 10.1016/j.patrec.2014.04.011

Badler, N. I., and Smoliar, S. W. (1979). Digital representations of human movement. *ACM Comput. Surveys* 11, 19–38. doi: 10.1145/356757.356760

Barnachon, M., Bouakaz, S., Boufama, B., and Guillou, E. (2014). Ongoing human action recognition with motion capture. *Patt. Recogn.* 47, 238–247. doi: 10.1016/j.patcog.2013.06.020

Bergmann, K. (2015). "Towards gesture-based literacy training with a virtual agent," in *Proc. Symposium on Multimodal Communication* (Linköping), 113–121.

Birnholz, J. C., Stephens, J. C., and Faria, M. (1978). Fetal movement patterns: a possible means of defining neurologic developmental milestones in utero. *Am. J. Roentgenol.* 130, 537–540. doi: 10.2214/ajr.130.3.537

Bonifazi, S., Tomaiuolo, F., Altoè, G., Ceravolo, M. G., Provinciali, L., and Marangolo, P. (2013). Action observation as a useful approach for enhancing recovery of verb production: new evidence from aphasia. *Eur. J. Phys. Rehabil. Med.* 49, 473–481.

Borghi, A. M., and Zarcone, E. (2016). Grounding abstractness: abstract concepts and the activation of the mouth. *Front. Psychol.* 7:1498. doi: 10.3389/fpsyg.2016.01498

Brandenburg, C., Worrall, L., and Rodriguez, A. D. (2013). Mobile computing technology and aphasia: an integrated review of accessibility and potential uses. *Aphasiology* 27, 444–461. doi: 10.1080/02687038.2013.772293

Cai, Q., and Aggarwal, J. K. (1999). Tracking human motion in structured environments using a distributed-camera system. *IEEE Trans. Patt. Anal. Mach. Intell.* 21, 1241–1247. doi: 10.1109/34.809119

Carroll, J. M. (2006). "Human-computer interaction," in *Encyclopedia of Cognitive Science,* ed L. Nadel (Chichester: John Wiley & Sons, Ltd), 1–4.

Cohen, R. L. (1981). On the generality of some memory laws. *Scand. J. Psychol.* 22, 267–281. doi: 10.1111/j.1467-9450.1981.tb00402.x

Cristianini, N., and Shawe-Taylor, J. (2000). *An Introduction to Support Vector Machines and Other Kernel-based Learning Methods.* Cambridge, MA: Cambridge University Press.

Cunha, J. P., Choupina, H. M., Rocha, A. P., Fernandes, J. M., Achilles, F., Loesch, A. M., et al. (2016). NeuroKinect: a novel low-cost 3Dvideo-EEG System for epileptic seizure motion quantification. *PLoS ONE* 11:e0145669. doi: 10.1371/journal.pone.0145669

de Aguiar, E., Theobalt, C., Stoll, C., and Seidel, H.-P. (2007). "Marker-less deformable mesh tracking for human shape and motion capture," in *2007 IEEE Conference on Computer Vision and Pattern Recognition* (IEEE), 1–8.

Dijkstra, K., and Post, L. (2015). Mechanisms of embodiment. *Front. Psychol.* 6:1525. doi: 10.3389/fpsyg.2015.01525

Dillon, A. (2003). "User interface design," in *MacMillan Encyclopedia of Cognitive Science,* ed E. Nadel (Chichester: John Wiley & Sons, Ltd), 453–458.

Engelkamp, J. (1998). *Memory for Actions.* Hove: Psychology Press/Taylor & Francis.

Engelkamp, J., and Zimmer, H. D. (1985). Motor programs and their relation to semantic memory. *German J. Psychol.* 9, 239–254.

Engelkamp, J., Zimmer, H. D., Mohr, G., and Sellen, O. (1994). Memory of self-performed tasks: self-performing during recognition. *Mem. Cogn.* 22, 34–39. doi: 10.3758/BF03202759

Fischer, M. H., and Zwaan, R. A. (2008). Embodied language: a review of the role of the motor system in language comprehension. *Q. J. Exp. Psychol.* 61, 825–850. doi: 10.1080/17470210701623605

Fodor, J. A. (1976). *The language of Thought.* Hassocks: Harvester Press.

Fodor, J. A. (1983). *The modularity of Mind : An Essay on Faculty Psychology.* Cambridge, MA; London: MIT Press.

Ganapathi, V., Plagemann, C., Koller, D., and Thrun, S. (2010). "Real time motion capture using a single time-of-flight camera," in *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (San Francisco, CA: IEEE), 755–762.

Garn, H., Member, S., Kohn, B., Dittrich, K., Wiesmeyr, C., Kloesch, G., et al. (2016). "3D Detection of Periodic Limb Movements in Sleep," in *2016 IEEE 38th Annual International Conference of the Engineering in Medicine and Biology Society (EMBC),* (Orlando, FL), 427–430.

Göksun, T., Lehet, M., Malykhina, K., and Chatterjee, A. (2015). Spontaneous gesture and spatial language: Evidence from focal brain injury. *Brain Lang.* 150, 1–13. doi: 10.1016/j.bandl.2015.07.012

Graves, A., Mohamed, A., and Hinton, G. (2013). Speech "Recognition with deep recurrent neural networks," in *2013 IEEE International Conference on Acoustics, Speech and Signal Processing* (Vancouver, BC), 150. 6645–6649.

Han, F., Reily, B., Hoff, W., and Zhang, H. (2017). Space-time representation of people based on 3D skeletal data : a review. *Comp. Vis. Image Understanding,* 158, 85–105. doi: 10.1016/j.cviu.2017.01.011

Han, J., Shao, L., Xu, D., and transactions on Shotton, - J. (2013). "Enhanced computer vision with microsoft kinect sensor: a review," in *IEEE Transactions on Cybernetics.* Available online at: http://ieeexplore.ieee.org/abstract/document/6547194/

Hinton, G., Deng, L., Yu, D., Dahl, G., Mohamed, A. R., Jaitly, N., et al. (2012). Deep neural networks for acoustic modeling in speech recognition: the shared views of four research groups. *IEEE Signal Process. Mag.* 29, 82–97. doi: 10.1109/MSP.2012.2205597

Holland, M. V., Sams, M. R., and Kaplan, J. D. (2013). *Intelligent Language Tutors: Theory Shaping Technology.* Available online at: http://books.google.com/books?hl=enandlr=andid=Db9dAgAAQBAJandoi=fndandpg=PP1andots=ZVkvAFWNfzandsig=mYinl4JTfiNalluHWsiTv0BhDBs

Inkster, M., Wellsby, M., Lloyd, E., and Pexman, P. M. (2016). Development of embodied word meanings: sensorimotor effects in children's lexical processing. *Front. Psychol.* 7:317. doi: 10.3389/fpsyg.2016.00317

Kelly, S. D., Ozyürek, A., and Maris, E. (2010). Two sides of the same coin: speech and gesture mutually interact to enhance comprehension. *Psychol. Sci.* 21, 260–267. doi: 10.1177/0956797609357327

Klooster, N. B., Cook, S. W., Uc, E. Y., and Duff, M. C. (2014). Gestures make memories, but what kind? Patients with impaired procedural memory display disruptions in gesture production and comprehension. *Front. Hum. Neurosci.* 8:1054. doi: 10.3389/fnhum.2014.01054

Kollorz, E., Penne, J., Hornegger, J., and Barke, A. (2008). Gesture recognition with a Time-Of-Flight camera. *Int. J. Intell. Syst. Technol. Appl.* 5: 334. doi: 10.1504/IJISTA.2008.021296

Kroenke, K. M., Kraft, I., Regenbrecht, F., and Obrig, H. (2013). Lexical learning in mild aphasia: gesture benefit depends on patholinguistic profile and lesion pattern. *Cortex* 49, 2637–2649. doi: 10.1016/j.cortex.2013.07.012

Kurland, J., Wilkins, A. R., and Stokes, P. (2014). iPractice: Piloting the effectiveness of a tablet-based home practice program in aphasia treatment. *Semin. Speech Lang.* 35, 51–63. doi: 10.1055/s-0033-1362991

Lakoff, G. (2012). Explaining embodied cognition results. *Top. Cogn. Sci.* 4, 773–785. doi: 10.1111/j.1756-8765.2012.01222.x

Lavoie, M., Routhier, S., Légaré, A., and Macoir, J. (2016). Treatment of verb anomia in aphasia: efficacy of self-administered therapy using a smart tablet. *Neurocase* 22, 109–118. doi: 10.1080/13554794.2015.1051055

Lippmann, R. (1987). An introduction to computing with neural nets. *IEEE ASSP Mag.* 4, 4–22. doi: 10.1109/MASSP.1987.1165576

Lubecke, O. B., Ong, P. W., and Lubecke, V. M. (2002). "10 GHz Doppler radar sensing of respiration and heart movement," in *Proceedings of the IEEE Annual Northeast Bioengineering Conference, NEBEC, Vol. 2002* (Janua, IEEE), 55–56.

Luinge, H. J., and Veltink, P. H. (2005). Measuring orientation of human body segments using miniature gyroscopes and accelerometers. *Med. Biol. Eng. Comput.* 43, 273–282. doi: 10.1007/BF02345966

Macedonia, M. (2014). Bringing back the body into the mind: gestures enhance word learning in foreign language. *Front. Psychol.* 5:1467. doi: 10.3389/fpsyg.2014.01467

Macedonia, M., and Klimesch, W. (2014). Long-term effects of gestures on memory for foreign language words trained in the classroom. *Mind Brain Educ.* 8, 74–88. doi: 10.1111/mbe.12047

Macedonia, M., and Mueller, K. (2016). Exploring the neural representation of novel words learned through enactment in a word recognition task. *Front. Psychol.* 7:953. doi: 10.3389/fpsyg.2016.00953

Marangolo, P., and Caltagirone, C. (2014). Options to enhance recovery from aphasia by means of non-invasive brain stimulation and action observation therapy. *Expert Rev. Neurother.* 14, 75–91. doi: 10.1586/14737175.2014.864555

Marvin, L. M. (1967). *Computation : Finite and Infinite Machines.* Prentice-Hall. Available online at: https://dl.acm.org/citation.cfm?id=1095587

Mattos, O., and Hinzen, W. (2015). The linguistic roots of natural pedagogy. *Front. Psychol.* 6:1424. doi: 10.3389/fpsyg.2015.01424

Mazzoldi, A., De Rossi, D., Lorussi, F., Scilingo, E. P., and Paradiso, R. (2002). Smart textiles for wearable motion capture systems. *Autex Res. J.* 2, 199–203.

Meteyard, L., Cuadrado, S. R., Bahrami, B., and Vigliocco, G. (2012). Coming of age: a review of embodiment and the neuroscience of semantics. *Cortex* 48, 788–804. doi: 10.1016/j.cortex.2010.11.002

Miranda, L., Vieira, T., Martínez, D., Lewiner, T., Vieira, A. W., and Mario, M. F. (2014). Online gesture recognition from pose kernel learning and decision forests. *Patt. Recogn. Lett.* 39, 65–73. doi: 10.1016/j.patrec.2013.10.005

Mitra, S., and Acharya, T. (2007). Gesture recognition: a survey. *IEEE Trans. Syst. Man Cybernet.* 37, 311–324. doi: 10.1109/TSMCC.2007.893280

Moeslund, T. B., Hilton, A., and Krüger, V. (2006). A survey of advances in vision-based human motion capture and analysis. *Comput. Vis. Image Understand.* 104, 90–126. doi: 10.1016/j.cviu.2006.08.002

Müller, M., and Röder, T. (2006). "Motion templates for automatic classification and retrieval of motion capture data," in *Proceedings of the ACM SIGGRAPH/Eurographics Symposium on Computer Animation* (Eurographics Association), 137–146.

Nilsson, L. G., and Bäckman, L. (1989). *Implicit Memory and the Enactment of Verbal Instructions. Implicit Memory: Theoretical Issues.* Available online at: http://books.google.com/books?hl=enandlr=andid=xmG8XVnxlV8Candoi=fndandpg=PA173anddq=free$+$recall$+$and$+$enactmentandots=g7tKpUmHEnandsig=vKkzXWgReaJxwjvSc2o3j2AWPis

Nye, B. D. (2015). Intelligent tutoring systems by and for the developing world: a review of trends and approaches for educational technology in a global context. *Int. J. Artif. Intell. Educ.* 25:177. doi: 10.1007/s40593-014-0028-6

Oppermann, R. (2002). "User-interface design," in *Handbook on Information Technologies for Education and Training* (Berlin; Heidelberg: Springer), 233–248.

Parzer, P., Probst, K., Babic, T., Rendl, C., Vogl, A., Olwal, A., et al. (2016). "FlexTiles," in *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems - CHI EA'16* (New York, NY: ACM Press), 3754–3757.

Pecher, D., and Zwaan, R. A. (2005). *Grounding Cognition: The Role of Perception and Action in Memory, Language, and Thinking.* Available online at: http://books.google.com/books?hl=enandlr=andid=RaxTkckBnh4Candoi=fndandpg=PP1anddq=embodiment$+$cognitionandots=EHGQcDOX-zandsig=xD68EMdN-rrSqWyMROqqF9mTzxU

Poppe, R. (2010). A survey on vision-based human action recognition. *Image Vis. Comput.* 28, 976–990. doi: 10.1016/j.imavis.2009.11.014

Preece, J., Sharp, H., and Rogers, Y. (2015). *Interaction Design: Beyond Human-Computer Interaction, 4th Edn.* New York, NY: Wiley.

Pritchard, M., Dipper, L., Morgan, G., and Cocks, N. (2015). Language and iconic gesture use in procedural discourse by speakers with aphasia. *Aphasiology* 29, 826–844. doi: 10.1080/02687038.2014.993912

Pulvermüller, F. (2002). *The Neuroscience of Language : On Brain Circuits of Words and Serial Order.* Cambridge; New York, NY: Cambridge University Press.

Pulvermüller, F., Neininger, B., Elbert, T., Mohr, B., Rockstroh, B., Koebbel, P., et al. (2001). Constraint-induced therapy of chronic aphasia after stroke. *Stroke* 32, 1621–1626. doi: 10.1161/01.STR.32.7.1621

Rabiner, L. R., and Juang, B. H. (1986). An introduction to hidden markov models. *IEEE ASSP Mag.* 3, 4–16. doi: 10.1109/MASSP.1986.1165342

Rabiner, L. R., and Juang, B. H. (1993). *Fundamentals of Speech Recognition.* PTR Prentice Hall. Available online at: http://www.citeulike.org/group/10577/article/308923

Roetenberg, D. (2006). *Inertial and Magnetic Sensing of Human Motion.* These de Doctorat, Ph.D., University of Twente.

Roetenberg, D., Luinge, H., and Slycke, P. (2013). *Xsens MVN : Full 6DOF Human Motion Tracking Using Miniature Inertial Sensors.* Technical report, Vol. 3.

Rose, M. L. (2013). Releasing the constraints on aphasia therapy: the positive impact of gesture and multimodality treatments. *Am. J. Speech Lang. Pathol.* 22, S227–S239. doi: 10.1044/1058-0360(2012/12-0091)

Rose, M. L., Mok, Z., and Sekine, K. (2016). Communicative effectiveness of pantomime gesture in people with aphasia. *Int. J. Lang. Commun. Disord.* 52,227–237.doi: 10.1111/1460-6984.12268

Sabatini, A. M. (2011). Estimating three-dimensional orientation of human body parts by inertial/magnetic sensing. *Sensors* 11, 1489–1525. doi: 10.3390/s110201489

Schuldt, C., Barbara, L., and Stockholm, S.-. (2004). "Recognizing Human Actions: A Local SVM Approach * Dept. of Numerical Analysis and omputer Science," in *ICPR 2004. Proceedings of the 17th International Conference on Pattern Recognition, 2004* (Cambridge, UK), 32–36.

Shkel, A. M. (2011). "Precision navigation and timing enabled by microtechnology: are we there yet?" in *IEEE Sensors 2010 Conference* (Kona, HI).

Smith-Atakan, S. (2006). *Human-Computer Interaction.* Thomson. Available online at: https://books.google.at/books?hl=deandlr=andid=tjPHVhncBzYCandoi=fndandpg=PR9anddq=human$+$computer$+$interactionandots=mr7DY7LhEnandsig=Xf5Q_hbo_xyx8CCXOBw5Fq2c_G8#v=onepageandq=human computer interactionandf=false

Stiefmeier, T., Roggen, D., and Tröster, G. (2007). "Gestures are strings: efficient online gesture spotting and classification using string matching," in *Proceedings of 2nd International Conference on Body Area Networks* (Florence), 1–8.

Szabo, G., and Dittelman, J. (2014). Using mobile technology with individuals with aphasia: native iPad features and everyday apps. *Semin. Speech Lang.* 35, 5–16. doi: 10.1055/s-0033-1362993

Thill, S., and Twomey, K. E. (2016). What's on the inside counts: a grounded account of concept acquisition and development. *Front. Psychol.* 7:402. doi: 10.3389/fpsyg.2016.00402

Tomasello, R., Garagnani, M., Wennekers, T., and Pulvermüller, F. (2017). Brain connections of words, perceptions and actions: a neurobiological model of spatio-temporal semantic activation in the human cortex. *Neuropsychologia* 98, 111–129. doi: 10.1016/j.neuropsychologia.2016.07.004

Von Essen, J. D., and Nilsson, L. G. (2003). Memory effects of motor activation in subject-performed tasks and sign language. *Psychon. Bull. Rev.* 10, 445–449. doi: 10.3758/BF03196504

Weinland, D., Ronfard, R., and Boyer, E. (2011). A survey of vision-based methods for action representation, segmentation and recognition. *Comp. Vis. Image Understand.* 115, 224–241. doi: 10.1016/j.cviu.2010.10.002

Zimmer, H. (2001). "Why do Actions Speak Louder Than Words. Action Memory as a Variant of Encoding Manipulations or the Result of a Specific Memory System?" in *Memory for Action: A Distinct Form of Episodic Memory?*, ed H. Zimmer (New York, NY: Oxford University Press), 151–198.

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

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