Trusting Humans and Avatars: A Brain Imaging Study Based on Evolution Theory

RENÉ RIEDL, PETER N.C. MOHR, PETER H. KENNING, FRED D. DAVIS, AND HAUKE R. HEEKEREN

René Riedl is a professor of digital business and innovation at the University of Applied Sciences Upper Austria and an associate professor of business informatics at the University of Linz. He was a guest researcher and guest professor at universities in the United States, Germany, and Canada. He earned his Ph.D. from the University of Linz. His research interests include NeuroIS, human–computer interaction, strategic IT management, and technostress. His work has been published in various journals, including *Advances in Human–Computer Interaction, Behavior Research Methods, Business & Information Systems Engineering, DATA BASE for Advances in Information Systems, Journal of Computer Information Systems, Journal of Information Technology, Journal of Neuroscience, Psychology, and Economics,* and *MIS Quarterly.* Moreover, he has editorial roles in several scientific journals, and he also serves as an ad hoc reviewer for the National Science Foundation, the German Research Foundation (DFG), and for several leading journals.

Peter N.C. Mohr is an assistant professor of information processing and economic decision making at the University of Constance, Germany. He studied business administration in Münster, Germany, and received his Ph.D. in psychology from Freie Universität Berlin. His research interests include situational and contextual influences on economic decision making as well as individual differences in this domain. He focuses on behavioral modeling of decision processes as well as on their neural underpinnings.

Peter H. Kenning is a professor of marketing at Zeppelin University, Germany. He received his Ph.D. in business administration from the University of Münster, Germany. His research interests include consumer behavior, consumer neuroscience, neuroeconomics, and marketing management. His work has been published widely, including in *International Journal of Research in Marketing, MIS Quarterly, Journal of Economic Psychology, Journal of Behavioral Decision Making, Journal of Neuroimaging, NeuroReport,* and the *Brain Research Bulletin.* He received several best-paper awards and grants. Since 2011 he has served as a member of the Scientific Advisory Board on Consumer and Food Policies at the Federal Ministry of Consumer Protection, Food and Agriculture (BMELV) in Germany.

Fred D. Davis is Distinguished Professor and David D. Glass Chair in Information Systems at the Sam M. Walton College of Business, University of Arkansas. He earned his Ph.D. from MIT's Sloan School of Management. He served on the Business School faculties of the University of Michigan, University of Minnesota, and University of Maryland. His research interests include user acceptance of information technology, technology-supported decision making, system development practices, and NeuroIS.

Journal of Management Information Systems / Spring 2014, Vol. 30, No. 4, pp. 83–113.

ISSN 0742–1222 (print) / ISSN 1557–928X (online)
DOI: 10.2753/MIS0742-1222300404

His research has been published in *MIS Quarterly, Information Systems Research, Management Science, Journal of Applied Psychology,* and other leading journals. He is currently a senior editor for *Information Systems Research.*

Hauke R. Heekeren is a professor of affective neuroscience and psychology of emotion at the Cluster of Excellence "Languages of Emotion" at Freie Universität Berlin. He received his Ph.D. in medicine from the Charité University of Medicine Berlin. His primary research interests concern the roles of motivation and affect in decision making, cognitive and affective components in normal and disturbed social cognition, and multimodal neuroimaging methodology. Current themes of his research include the effect of neuromodulators (such as drugs or common genetic polymorphisms) on cognitive and affective functions. His work has been published in various outlets, including *Nature, Nature Reviews Neuroscience, Journal of Neuroscience, PLOS Biology, Current Biology,* and *Proceedings of the National Academy of Sciences.*

Abstract: Avatars, as virtual humans possessing realistic faces, are increasingly used for social and economic interaction on the Internet. Research has already determined that avatar-based communication may increase perceived interpersonal trust in anonymous online environments. Despite this trust-inducing potential of avatars, however, we hypothesize that in trust situations, people will perceive human faces differently than they will perceive avatar faces. This prediction is based on evolution theory, because throughout human history the majority of interaction among people has taken place in face-to-face settings. Therefore, unlike perception of an avatar face, perception of a human face and the related trustworthiness discrimination abilities must be part of the genetic makeup of humans. Against this background, we conducted a functional magnetic resonance imaging experiment based on a multiround trust game to gain insight into the differences and similarities of interactions between humans versus human interaction with avatars. Our results indicate that (1) people are better able to predict the trustworthiness of humans than the trustworthiness of avatars; (2) decision making about whether or not to trust another actor activates the medial frontal cortex significantly more during interaction with humans, if compared to interaction with avatars; this brain area is of paramount importance for the prediction of other individuals' thoughts and intentions (mentalizing), a notably important ability in trust situations; and (3) the trustworthiness learning rate is similar, whether interacting with humans or avatars. Thus, the major implication of this study is that although interaction on the Internet may have benefits, the lack of real human faces in communication may serve to reduce these benefits, in turn leading to reduced levels of collaboration effectiveness.

Key words and phrases: agent, avatar, brain, cognitive neuroscience, evolutionary psychology, evolution theory, functional magnetic resonance imaging (fMRI), medial frontal cortex (MFC), mentalizing, NeuroIS, theory-of-mind (TOM).

Determining who can be trusted and who cannot has been crucial since the days of ancient civilizations. From the time of the emergence of early hominids such as *Australopithecus afarensis,* some 3.5 million years ago, until the recent past, trust in individuals who turned out not to be trustworthy could result in loss, or even death [21, 68]. Thus, it is critical for humans to possess distinctive trustworthiness prediction abilities. Attention to another person's face in a trust situation is of particular importance because the human face has been shown to allow one individual to predict another's emotional state, intention, and behavior (e.g., [33, 57]). This predictive function has demonstrated application to trustworthiness [116, 121, 122]. Thus, all other things being equal, the inability to see another person's face creates more difficulty for predicting the other individual's trustworthiness than there would be in a situation in which facial information is available.

Biological anthropologists (e.g., [14, 20]) report that face-to-face interaction has been the main communication medium for more than 99 percent of human history (i.e., since the emergence of the first hominids). It follows, then, that application of Darwin's theory of evolution [23] would suggest that the human trait of processing facial information, and the related prediction abilities, must be part of the genetic makeup of humans—a notion that is widely supported in the scientific literature [21, 23, 121]. For example, Kock and Hantula, who pioneered the introduction of evolution theory into social science disciplines such as organization science, information systems (IS) research, and economics, wrote: "The most likely scenario is that our brain has been primarily designed to excel in co-located communication, especially where face-to-face interaction takes place" [65, p. iii]. Similarly, Kock argued that "our brain has likely been to a large extent hardwired [i.e., genetically predetermined] for co-located and synchronous communication employing facial expressions" [60, p. 120]. As direct support for these statements, scientific evidence shows that newborns react to human faces in the first hours of life [48]. Clearly, such behavior signifies that the ability to perceive individual faces is innate, and not learned.

The intrinsic human preference toward face-to-face interaction suggests that individuals would not be predisposed to communicate through electronic methods (e.g., e-mail or text-based Internet chat). Evidence based on empirical research supports this view (see, e.g., the works reviewed in [59, 61]). The theoretical arguments underlying this idea, as well as the implications, are summarized under the media naturalness proposition [59], which Kock later refers to as the *media naturalness hypothesis* [60] and the *media naturalness theory* [61]. In essence, this theoretical framework defines the mismatch that exists between the characteristics of face-to-face interaction (which is the benchmark, due to its 100 percent naturalness) and the characteristics of other forms of communication (e.g., e-mail) as independent variables, whereas cognitive effort, communication ambiguity, and physiological arousal are defined as dependent variables.

The degree of naturalness of a communication medium can be evaluated based on the degree to which it incorporates the characteristics of face-to-face interaction [59, 61]. The primary conditions for assessing the level of naturalness include (1) two communicating people share the same context, and they are able to see and hear each other, (2) individuals can quickly (i.e., in real time) exchange communicative stimuli, (3) the situation provides the ability to both convey and observe facial expressions, (4) to convey and observe body language, and (5) to convey and listen to speech. The theory predicts that a decrease in the degree of naturalness (i.e., a higher degree of mismatch) results in three major effects regarding a communications process: (1) increased cognitive effort, (2) increased ambiguity, and (3) decreased physiological arousal (in terms of excitement). Importantly, a low degree of naturalness can negatively affect satisfaction, performance, and productivity in a number of collaborative tasks (e.g., customer support of an online service firm), although in computer-mediated environments humans also have the ability to compensate for lower degrees in naturalness—an ability known as *compensatory adaptation* (e.g., [59]).

To date, both conceptual and empirical articles have investigated the relationship between face-to-face interaction and a number of communications media (see, e.g., research reviewed in [59, 61]). These investigations include e-mail, text-based Internet chat, as well as audio- and videoconferencing. Despite the insights generated by the extensive number of articles investigating these more traditional communications media, our study adds significantly to the extant literature through our focus on the medium of *avatars,* which has gained considerable momentum in use over the past few years, but has not been explored in this regard in the literature.

In human–computer interaction, the term *avatar* is used "as a label for digital representations of humans in online or virtual environments" [6, p. 64]. Typically, avatars are virtual humans that possess realistic faces [29], a fact that is confirmed in several studies published in the IS literature (e.g., [25, 86, 112]). Among the many applications for avatars, both in practical function and as entertainment (e.g., virtual conferences [7], e-learning [34], shopping in virtual malls [73], instant messaging [43], and product innovation [66]), such characters are increasingly used for interactions among humans on the Internet [7, 82]. Against this background, avatars are a ubiquitous phenomenon in today's computerized world, and are important in both private and organizational contexts (e.g., Second Life, World of Warcraft). Because attention to a realistic portrayal of a human face is of particular importance for trust perceptions [117], it has been theorized that the use of avatars may positively affect trust perceptions in computer-mediated interactions, thereby mitigating uncertainty perceptions that are afforded by the anonymity of the Internet. In fact, previous research [12] found that avatar-based communication increases perceived interpersonal trust, relative to the more traditional communications media such as text chat, although an avatar's face does not need to resemble a computer user's real face because it can be chosen freely [87]. Importantly, technologies exist that capture a user's facial expressions via a camera in order to transmit them in real time to the avatar's face (e.g., [120]). Thus, actual facial expressions are transmitted into the virtual world, blurring the boundaries between the real world and cyberspace.

Despite the potential for avatars to induce trust through facial expression, however, evolution theory would suggest that a human face is perceived differently when compared to an avatar face [21, 23, 90, 121], and this is expected to hold true even in a static context (i.e., presentation of pictures of human or avatar faces rather than animations) (e.g., [77, 85, 116, 122]). For human faces, studies indicate that a judgment of reliable trustworthiness can be formed after an exposure time of 100 milliseconds (ms) [121], and because such times are not sufficient for saccadic eye movements, trustworthiness judgments are usually “single glance” impressions [116]. The automatic quality of processing trust judgments suggests that information-processing traits are hardwired (e.g., [59]), which allows the concept of the genetic nature of processing facial information to be extended to trust perceptions.

Against this background, we specifically address three research questions:

RQ1: Do differences in trustworthiness discrimination abilities exist, dependent on whether the interaction partner is human or is an avatar?

RQ2: If so, are these behavioral differences associated with neural differences?

RQ3: Are there differences in learning of trustworthiness in interaction with humans versus avatars?

The remainder of this paper is structured as follows. In the next section, we present a discussion of the theoretical concepts that are related to our research questions. Based on this discussion, we derive three hypotheses. Specifically, we discuss concepts from evolution theory (H1), brain imaging results from social and cognitive neuroscience (H2), and from brain plasticity theory (H3). Subsequently, we describe the methods that we used to test the hypotheses. An outline of the research results precedes a general discussion of the results. Moreover, we detail our specific contribution to the IS literature. Afterward, we outline limitations of our study as well as possible avenues for future research. Finally, we provide concluding comments.

Theories and Hypotheses

Evolution Theory

Darwin’s theory of evolution explains that the human species evolved through natural selection, a process spanning from at least thousands to as much as millions of years, and in which random mutations are introduced in the genetic makeup of offspring, leading to traits that may increase, or may decrease, chances for survival [23, 60]. Those genetic mutations supporting mating and survival are then passed on to offspring, until the mutations become established as species-wide traits. The extremely long neck and legs of the giraffe, for example, are the outcome of such an evolutionary process, because these traits increased foraging abilities in savannah environments (e.g., reaching leaves at the top of trees), thereby increasing the fitness of this species (i.e., the number of surviving offspring). Similarly, perceiving human facial information in order to predict another person’s trustworthiness is a trait that has evolved over time, increasing the fitness in environments in which social interaction takes place face to face. Because this trait is generally found in all healthy individuals independent of cultural background, it can be considered as a universal human trait (for details, see Appendix E in [61]).

The universality of the ability to use facial information to predict trustworthiness cannot be automatically extended to avatars. Because these virtual humans have existed for only a few decades, the time period is much too short to lead to behaviorally relevant changes in the human genotype [14, 20, 65]. Drawing on knowledge from evolution theory, the field of *evolutionary psychology* has been developing since the 1980s (e.g., [16, 20, 72]), and has been shown to provide significant "fresh new insights" for IS theorizing [61, p. 395]. In essence, evolutionary psychology posits that a large number of modern brain functions evolved after the emergence of the first hominids, and that these functions, often in an unconscious fashion, significantly affect behavior in the modern world. Relevant contributions to the literature on the potential of evolutionary psychology for IS research include Abraham et al. [1], Kock [62], and Kock and Chatelain-Jardón [64].

The following section briefly explains the functioning of evolutionary psychology theorizing in relation to the focus of this work—the information-processing trait that has evolved to become what we refer to as "face perception." The explanation is conceptually summarized in Figure 1.

As illustrated at the bottom of Figure 1, genotype influences brain activity. This influence is mediated by the impact of genes on brain structure (e.g., gray matter density), neurotransmitter systems, and other biological mechanisms, because physiology such as brain structure affects both the probability and magnitude of activity in a specific brain area (e.g., [78, 114]). A specific brain activity pattern is, in turn, related to a specific information-processing trait. The present study focuses on the important trait of "face perception."

From the foundation of evolution theory, it is logical to conclude that an information-processing trait is related to task performance. For example, specific brain modules that direct attention to human faces, and that process facial information, have evolved over time [51, 53], because this has turned out to be advantageous for predicting another individual's trustworthiness in face-to-face settings. Several parts of the human brain play a crucial role in face perception, such as the pulvinar, inferior occipital gyrus (occipital face area), middle fusiform gyrus (fusiform face area), posterior superior temporal sulcus, amygdala, and anterior infero-temporal cortex (e.g., [18, 51, 53, 104, 119]). Importantly, research on localized brain damage (lesion studies) indicates that damage in the above-mentioned areas is often related to prosopagnosia [15], a disorder in which the ability to recognize faces is impaired, while general cognitive functioning (e.g., decision making) and other aspects of visual processing (e.g., object discrimination) remain intact [104]. This evidence from two streams of research (brain imaging and lesion studies) supports the conclusion that cognitive neuroscience research has developed a substantial knowledge base on the neural foundations of face processing. Moreover, an experiment with both humans and monkeys as subjects shows that primates, in general, "possess the remarkable ability to differentiate faces of group members and to extract relevant information about the individual directly from the face" [22, p. 2973]. This result strongly supports the idea that face perception can be discussed appropriately from an evolutionary perspective, and substantiates the theoretical foundation of the present paper.

The human species, intriguingly, developed a complex system of facial muscles (22 on each side of the face) that allows humans to generate more than 6,000 expressions for communicating thoughts and feelings [8, 58, p. 10, 60, p. 120, 81]. Trustworthi-

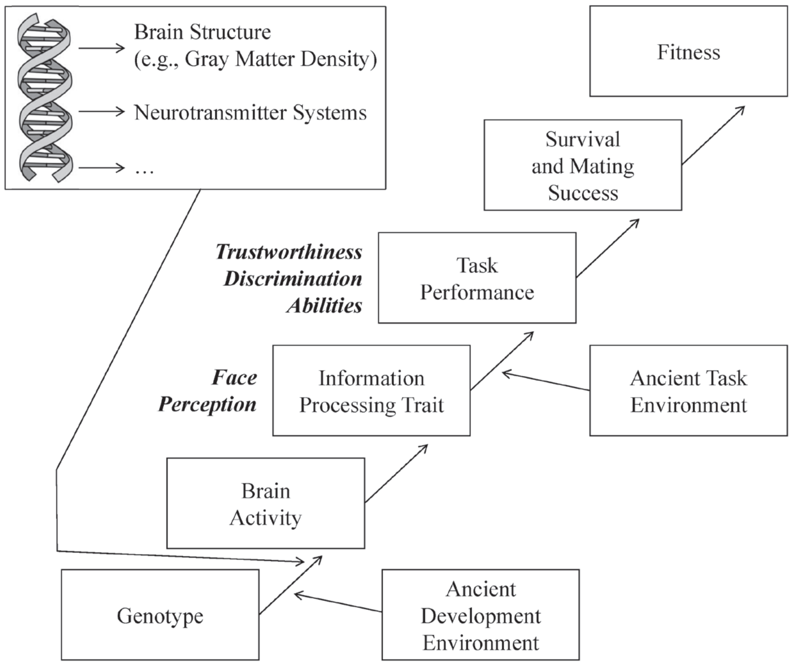

Figure 1. Evolution of Information Processing Trait by Natural Selection

Source: Based on Kock [61, p. 398].

Notes: Complementary to the original theorizing, we added the element "brain activity," thereby signifying that the influence of genotype on a specific information processing trait is mediated by activity in specific brain areas (e.g., [50]). Moreover, we added another element illustrating that the influence of genotype on brain activity is mediated by brain structure (e.g., gray matter density), neurotransmitter systems (i.e., chemicals that transmit signals from a neuron to another one across a synapse), and other neurobiological mechanisms (e.g., [78, 114]). Also, we have added both the information processing trait "face perception," and the corresponding ancient task performance, namely trustworthiness discrimination abilities. Altogether, this figure conceptually integrates the brain, as well as the specific topic of this article—face trustworthiness—into evolutionary psychology theorizing.

ness discrimination abilities, based on encoded facial information, are thought to have increased survival success and mating success, increasing the likelihood that the behavior trait would be replicated in the succeeding generation, with a positive effect on fitness as a result.

Moreover, Figure 1 demonstrates that the influence of genotype on the brain has been affected by the ancient development environment in its broadest sense, ranging from nutrition to social conditions [61]. In general, there is agreement among scientists that most traits that are significantly affected by genetic makeup, including those related to face-to-face communication, need interaction with the environment to become completely developed (e.g., [17, 59]), a phenomenon referred to as *epigenetic.* With respect to the ability to perceive individual human faces, for example, early visual input (e.g., during the first few months of life) is a necessary precondition for the normal development of neural mechanisms that will later specialize to support the processing of faces [74].

Furthermore, as is illustrated in Figure 1, the ancient task environment moderates the influence of an information-processing trait on task performance [61]. Because the ancient task environment was characterized by the nonexistence of technology for communication during almost 100 percent of human history (written communication emerged as recently as approximately 5,000 years ago, and electronic forms of communication were not invented until the very recent past), attention to human faces and understanding facial features has been indispensable for effective communication and for survival until the recent past. Understandably, millions of years of exercise in face-to-face interaction would have resulted in elevated abilities to use facial information to discriminate between trustworthy and untrustworthy actors. In contrast, if humans had interacted only by way of technologies such as telephone or text chat that suppress facial information (task environment), perceptive reading of the human face (information processing trait) would not likely have developed significant trustworthiness discrimination abilities (task performance).

Against the background discussion herein, and through an application of evolution theory and evolutionary psychology, we derive the following hypothesis:

> *Hypothesis 1: Based on facial information, people are better able to predict the trustworthiness of humans than the trustworthiness of avatars.*

H1 pertains to the behavioral level of analysis. However, because all human behavior is—at least partly—determined by biological factors, in particular those related to genes and the brain (e.g., [17]), there is reason to assume that behavioral differences are accompanied by neurobiological differences.

Trust, Mentalizing, and the Medial Frontal Cortex

The concept of *NeuroIS,* recently introduced in the IS literature, describes the idea of applying neuroscience theories, methods, and tools in IS research [27]. Among other contributions (e.g., [27, 28, 97, 100, 102, 118]), this new research field is expected to add to the development of new theories that make possible accurate predictions of behavioral responses to information technology (IT) [102]. Such new theories seek to integrate the behavioral and biological levels of analysis, thereby leading to a deeper understanding of IS phenomena. Specifically, functional magnetic resonance imaging (fMRI) studies have demonstrated that brain activation patterns have the power to explain differences in behavior toward IT artifacts (e.g., [11, 26, 99]).

In this context, we would expect that the behavioral difference specified in H1 is accompanied by activation differences in specific brain areas. Parts of the *medial frontal cortex* (MFC), in particular, have been identified as crucial brain regions for the neural implementation of trustworthiness predictions (e.g., [71, 79]). Specific brain areas within the MFC, which correspond approximately to Brodmann areas (BAs) 9, 10, 24, and 32, facilitate neural implementation of mentalizing, a human ability to infer the internal states of other actors (e.g., intentions) in order to predict their personality traits and behavior (see, e.g., the following reviews on the importance of the MFC for mentalizing: [2, 4, 40, 42, 44]).

In the scientific literature, the concept of mentalizing is also referred to as the *theory-of-mind* [41, 91, 109]. In general, mentalizing is a fundamental cognitive process in trust situations because the decision to trust involves thinking about an interaction partner's intentions to infer his or her trustworthiness. Fehr confirms this view: "Since trust decisions are also likely to involve perspective-taking, they should also activate areas implicated in theory-of-mind tasks" [37, p. 228]. A recent review of the literature on the biology of trust [98] identifies the concept of mentalizing and the corresponding MFC activations as significant determinants of human trust behavior. Importantly, this review integrates research from various scientific fields such as neuroeconomics, social neuroscience, and NeuroIS, which supports the theory that mentalizing and MFC activations are fundamental in trust situations across many different contexts.

Based on results of a brain imaging experiment in which participants had to judge *human* faces regarding their trustworthiness, Winston et al. conclude that "social judgments about faces reflect a combination of brain responses that are stimulus driven . . . and driven by processes relating to inferences concerning the intentionality of others" [122, p. 281]. Similarly, Frith and Frith argue that "faces, in particular, are an important source of information about their inner states. For example, there is agreement about what a trustworthy person looks like" [40, p. 531]. Finally, a review on the neurobiological foundations of mentalizing indicates that "[i]n the case of knowledge of other minds, we appear to begin with . . . perception of a face" [3, p. 696]. These studies suggest that the human face provides rich information for predicting another individual's intentions and behavior. Considering that face-to-face interaction has been the major communications mode for the majority of human history, a specific trait for processing human facial information and corresponding trustworthiness prediction abilities must be part of the human genetic makeup. A number of other scholars share this view; they also regard trust and mentalizing as "evolved capacities" (e.g., [38, p. 277, 115, p. 201]). With respect to possible explanations of *why* such perception of human faces constitutes a basic instinct—and as a complement to the fact that trust in untrustworthy humans could have directly resulted in death—scholars have offered explanations related to the effectiveness of social functioning (for details, see [32, 95, p. 65]).

Building on the brain imaging research presented in this section, and extended by the knowledge that brain activation mediates the influence of genes on behavior (e.g., [17, 50]), we derive the following hypothesis:

> *Hypothesis 2: When determining whether or not to trust another actor, decision making activates the MFC more during interaction with humans, if compared to interaction with avatars.*

Brain Plasticity Theory

So far, we have argued that because the ability to accurately assess another person's trustworthiness on the basis of facial information has increased survival probability over the course of millennia, the human brain has been shaped by that evolution process. However, as explained by scholars in various disciplines such as biology, psychology, and neuroscience (for a review of sources, see [59, p. 336]), evolution not only has given the human species a genetic makeup with a predetermined set of instincts (such as trustworthiness discrimination abilities based on face perception), but has also equipped humans with a "plastic" brain. Plasticity of the brain refers to its ability to change structure and function [67]. Obviously, this general characteristic of the brain is advantageous for survival, thereby increasing fitness. Learning and memory, as well as experience, are concepts that are inseparably associated with brain plasticity, because they imply changes in dendritic length, synapse formation, and metabolic activity, among other modifications [69]. Such neurobiological changes often lead to changes in behavior [67].

The ability of the human species to develop new schemas in the brain pertains to most types of schemas, including those related to computer-mediated communication [59]. Against this background, it is no surprise that predictions of behavioral theories about electronic forms of communication are consistent with predictions based on brain plasticity theory. Kock, for example, explains that channel expansion theory [19] "is compatible with brain plasticity" [59, p. 336]. This theory is in line with the prediction that continued use of an unnatural communication medium (e.g., avatar-based interaction) may, over time, result in learning processes. Such processes may, importantly, lead to lower levels of communication ambiguity, a characteristic that is usually associated with face-to-face communication but is not typically associated with interaction without real facial information.

In this context, Kock coined the phrase "cognitive adaptation proposition" [59, p. 336]. To illustrate the concept (in a later study), Kock cited Carlson and Zmud [19, p. 157] on the example of e-mail: "As individuals develop experience communicating with others using a specific channel, such as e-mail, they may develop a knowledge base for more adroitly applying this communication channel. . . . [U]sers may become aware of how to craft messages to convey differing levels of formality or of how to use channel-specific metalanguage to communicate subtleties. Similarly, these individuals are also likely to interpret messages received on this channel more richly because they can interpret an increasing variety of cues" [61, p. 410]. This statement supports a clear understanding that schemas that are not in the human genetic makeup, but can be learned through direct experience, may affect behavior to a considerable degree. Kock and Hantula even suggest that, due to brain plasticity, electronic forms of communication may become "second nature" [65, p. iii]. Similarly, Schlicht et al., investigating trustworthiness perceptions of opponents' faces while playing poker, argue for the importance of learning mechanisms: "Once rapid impressions have been formed, beliefs can later be updated by direct experience with the individual, to develop a new estimate that will be used going forward" [106, p. 1].

In general, scholars (e.g., [68, pp. 718–746]) indicate that learning is always accompanied by changes in the nervous system. Consequently, learning can be studied at various analytical levels, ranging from observable behavioral changes to alterations in molecular structure. Applying the context of brain plasticity theory, we derive the following hypothesis, which pertains to the behavioral level of analysis:

Hypothesis 3: In a trust situation, people's learning rate of trustworthiness is similar, whether interacting with humans or avatars.

Methods

Tasks and Measurement

In order to test our hypotheses, we used a modified version of the original trust game [13], a multiround trust game in which an investor has an initial endowment. The investor first decides whether to keep her or his endowment, or to send it to the trustee. Then the trustee observes the investor's action (i.e., sending or not sending) and decides whether to keep the amount received or to share it with the investor. The experimenter multiplies the investor's transfer by some factor, so that both players are advantaged, collectively, if the investor transfers money and the trustee sends back a part of it. As a behavioral measure for trust we used the decision of the investor to send money, and as a behavioral measure for trustworthiness we used the trustee's decision of whether or not to return money [36].

The trust game has been used as an experimental paradigm in a number of investigations, particularly in studies on the neurobiological foundations of trust (for a review, see [98]). Moreover, this game has recently been described as an appropriate task for the study of mentalizing in trust situations within brain imaging environments. Sripada et al. wrote: "The medial prefrontal cortex has been implicated as a key brain region that implements mentalizing during the social interactions [and] the 'trust game' serves as a potent probe of mentalizing abilities because it sets up the need to make inferences about the mental state of others. . . . Therefore, functional neuroimaging of the trust game . . . may shed light on a novel behavioral and neural mechanism" [111, p. 984]. Thus, the trust game, if used in an fMRI environment, is a suitable task for investigating both H1 and H2.

In a research agenda addressing trust in online environments, Gefen et al. wrote that "[t]rust develops gradually as people interact with each other. . . . [T]herefore it is also important to study the longitudinal effects of trust on transaction decisions and other behavioral outcomes" [46, p. 277]. The trust game can be played either on a one-shot basis or as a multiround version. In the latter case, two actors play against each other multiple times (e.g., ten consecutive rounds; [55]). As a result of this extended interaction, the longitudinal nature of trust is accentuated and players can learn about the behavior of the other actor. We used such a multiround version of the game, as this is a precondition for the investigation of H3.

Stimulus Material

In our trust game experiment, the participants played the role of investor against both humans and avatars in the role of trustee. The objective of the stimulus selection was therefore to identify a number of human and avatar faces for the fMRI study. Ultimately, both the human and avatar groups comprised four faces with a high degree of trustworthiness and four faces with a low degree of trustworthiness (examples are provided in

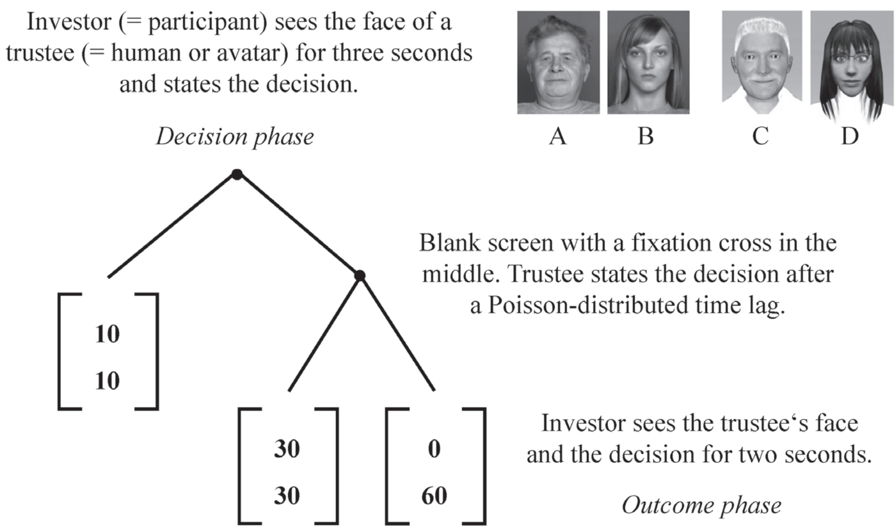

Figure 2. Structure and Payoff Matrix of the Trust Game

Notes: The upper value in the square brackets indicates the investor's payoff, and the lower value shows the trustee's payoff. Participants played against eight humans and eight avatars in the fMRI study. The human and avatar groups were matched for pre-rated trustworthiness based on facial appearance, and for behavioral trustworthiness (i.e., the number of rounds in the trust game in which money was returned). We used four trustworthy humans (see, e.g., Face A) and four untrustworthy humans (e.g., Face B), as well as four trustworthy avatars (e.g., Face C) and four untrustworthy avatars (e.g., Face D). All humans and avatars had a direct gaze, as illustrated in the four example faces. Both the human and avatar groups consisted of four males (two trustworthy and untrustworthy ones) and four females (two trustworthy and untrustworthy ones). The complete set of 16 faces used in our fMRI study is available elsewhere [101].

Figure 2). The complete set of 16 faces used in our fMRI study is available in Riedl et al. [101], as is detailed information on the stimulus material selection process.

Subjects

For the main study, we selected 11 male and 8 female subjects (none of whom had participated in the stimulus selection pretest). All of the subjects were healthy and reported no history of neurological or psychiatric diseases. The mean age of the subjects was 31.83 years. One subject (female) had to be excluded from further analyses, as she indicated after the fMRI session that she was afraid while in the fMRI machine, and as a result was not really thinking of the assigned task. Therefore, the sample size underlying our data analysis is $N = 18$. All of the participants were familiar with the Internet and had been using it for many years (mean = 10.77, SD [standard deviation] = 3.40, minimum = 4, maximum = 18). By design, our investigation is focused on experienced computer and Internet users rather than novices. All of the participants were paid for their participation, and gave written informed consent. The study was approved by the Freiburg Ethics Commission International (FECI), Germany. Further details on the subjects' characteristics are described elsewhere [101].

Experimental Procedure and Stimulus Presentation

In our experiment, the participants played the trust game in the role of the investor against (1) humans and (2) avatars (both playing in the role of trustee, see Figure 2). Our game, therefore, mimics a typical interaction in both social and economic exchange relationships, both in bricks-and-mortar environments (human condition) and in computer and online environments (avatar condition). The participants were told in advance that their playing partners (i.e., the trustees) would not be responsive to their playing strategies. We did stress, however, that each trustee has a specific character that determines his or her trustworthiness. Half the trustees, both humans and avatars, were predetermined by the experimenters to be relatively trustworthy, whereas the other half were predetermined as relatively untrustworthy. Trustworthy actors returned €30 in seven out of ten rounds, whereas untrustworthy actors returned €30 in only three rounds. Apart from their facial appearance, the participants had no information regarding the trustees.

In each round of the game, the participants were asked whether they wanted to keep their initial endowment of €10 (≈ $13 at the time the experiment was conducted), or whether they wanted to give it to the trustee whose face was presented to them. In the case of giving the €10 to the trustee, the amount was multiplied by six (resulting in €60), which the trustee could then either keep, or split (i.e., return €30 to the participant). The participants played ten rounds of the game with each trustee, with three seconds in each round to make the investment decision (= *Decision phase;* see Figure 2); note that decision times of one to three seconds are sufficient to make economic decisions in an fMRI environment (e.g., [122]). After a variable time (varied based on a Poisson distribution) in which a blank screen with a fixation cross in the middle was presented, the trustee's face and the decision was visually presented for two seconds (= *Outcome phase;* see Figure 2). Before the first round, after the fifth round, and after the tenth round, the participants were asked to rate the trustworthiness of the trustee, which was operationalized as the probability that the trustee, in the case of being given money, would behave in a trustworthy manner (i.e., returning €30).

We used the program Presentation (Neurobehavioral Systems, Albany, CA) to present the stimuli in the MRI-scanner (3T Siemens Tim Trio, Erlangen, Germany) and to record the responses on a laptop computer with Windows XP as the operating system. Visual stimuli were presented using video goggles. Details about data collection and analysis are provided in the Appendix.

Results

Before we tested H1, and without making a distinction between trustworthy and untrustworthy actors, we analyzed whether or not the participants differ in their decisions to trust (i.e., to invest their initial endowment of €10) when playing against humans or avatars. In essence, we found that the participants showed considerable trust behavior, independent of playing against humans or avatars (average trust in the human condition: in 52 of 80 games, SD = 11.55; average trust in the avatar condition: in 51.39 of 80 games, SD = 13.57). There was no significant difference between the number of decisions to trust in the human and avatar conditions ($t = 0.445$, df [degrees of freedom] = 17, $p = 0.662$).

Hypothesis 1

In both the human and avatar groups, discrimination between trustworthy and untrustworthy actors revealed a significant difference in trustworthiness prediction ability, as was predicted in H1. We used the trustworthiness evaluations of the pretest (see [101]) as a benchmark (an established procedure in neuroscience face evaluation studies [105, p. 2]). Similar to the results of the pretest, we found a significant ($p < 0.1$) difference between high and low trust faces in the fMRI trustworthiness ratings, indicating a positive manipulation check.

While we found no significant difference in the participants' decisions to trust when playing against trustworthy versus untrustworthy avatars (25.89 versus 25.50, $t = 0.408$, df = 17, $p = 0.689$), we identified a significant difference when playing against trustworthy versus untrustworthy humans (27.61 versus 24.39, $t = 3.502$, df = 17, $p = 0.003$). The difference between the difference values of the number of decisions to trust humans, trustworthy versus untrustworthy (Δ 3.23), and to trust avatars, trustworthy versus untrustworthy (Δ 0.38), is statistically significant ($t = 2.536$, df = 17, $p = 0.021$). Thus, the participants trusted *un*trustworthy avatars to a similar degree as they trusted trustworthy avatars. In the human group, importantly, we could not observe this disadvantageous behavior. Our results therefore indicate that people are better able to predict the trustworthiness of humans than the trustworthiness of avatars. This finding confirms H1.

Hypothesis 2

In the present experiment, we focus on the investigation of brain activation during the *Decision phase* (see Figure 2); that is, we analyzed brain activation during the presentation of the trustees' faces. In this phase, the participants had to decide whether or not to send their initial endowment of €10 to the trustee.

We found that activation in the dorsomedial prefrontal cortex (DMPFC), the rostral part of the anterior cingulate cortex (rACC), and the ventromedial prefrontal cortex (VMPFC) was significantly higher in the human condition compared to the avatar condition ($z > 3.09$; cluster size > 25; see Figure 3 and Table 1). These three brain regions correspond approximately to BAs 9, 10, 24, and 32, and they have been summarized under the label "MFC," a brain structure that plays a significant role in the neural implementation of mentalizing [2, 4, 40, 42, 44]. This finding confirms H2.

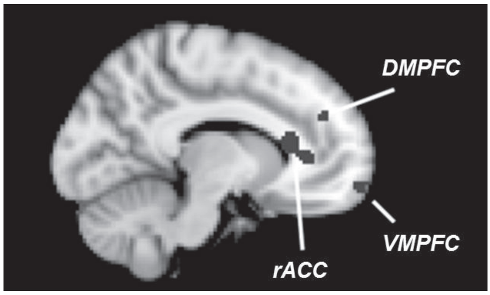

Figure 3. Brain Activation Results (Sagittal Cut)

Notes: Brain activation in the *Decision phase* (fMRI contrast: human–avatar). The reverse contrast (avatar–human) did not result in significant activation differences. The three brain areas play a significant role in the neural implementation of mentalizing. DMPFC = dorsomedial prefrontal cortex, rACC = rostral anterior cingulate cortex, VMPFC = ventromedial prefrontal cortex.

Hypothesis 3

In order to understand the participants' trust learning process, particularly their learning rates, we modeled the perceived trustworthiness of the trustees on a round-by-round basis. The underlying assumption of this approach is that perceived trustworthiness of a trustee is based on the facial appearance before the first round begins, and is adjusted based on its behavioral trustworthiness during the subsequent rounds of the game. Importantly, recent empirical evidence provides support for this assumption [105].

We assume that perceived trustworthiness is reflected in the subjective probability that a trustee will be trustworthy (i.e., will return €30) when a participant invests the initial endowment of €10. Moreover, a participant typically selects the gaming strategy that offers a higher expected value. The expected value of keeping the initial endowment of €10 is €10, because the probability for this gain is 100 percent in our game (see Figure 2, left path). The expected value of investing the endowment of €10 is equal to the product of a trustee's perceived trustworthiness and the possible payoff of €30. Hence, the trust decision depends only on the perceived trustworthiness (see Figure 2, right path).

During the ten rounds of interaction with a specific trustee, trustworthiness is expected to be updated on the basis of that trustee's behavior. In this paper, we model the updating process with a *reinforcement learning model* [9, 10, 96, 113]. Such a model generally assumes that after the decision has been made for one alternative, a received reward $R(t)$ at time t is compared to an expected value $EV(t)$, with the deviation d formalized as prediction error PE: $d(t) = R(t) – EV(t)$. A reinforcement learning model assumes that learning is driven by these deviations; hence, a PE is used to update $EV(t)$, allowing the optimization of reward predictions. The influence of a specific PE on $EV(t)$ regarding the next trust decision is determined by the learning rate.

Table 1. Brain Areas Activated in the Decision Phase (Human–Avatar)

Anatomical region	Cluster size (voxels)	Maximum z-score	MNI coordinates (x, y, z)		
Rostral anterior cingulate cortex	**163**	**4.23**	**–8**	**26**	**14**
Lateral occipital cortex	56	3.70	–38	–66	24
Precuneus/posterior cingulate cortex	49	3.97	12	–54	44
Temporal pole/lateral orbitofrontal cortex	37	3.79	30	20	–28
Precuneus/posterior cingulate cortex	36	3.65	–16	–54	38
Frontal pole/ventromedial prefrontal cortex	**32**	**3.90**	**–8**	**66**	**–12**
Dorsomedial prefrontal cortex	**28**	**3.67**	**–8**	**44**	**28**
Frontal pole	28	3.56	–8	42	–28

Notes: MNI = Montreal Neurological Institute. There was no significant activation in the reverse contrast (avatar–human). The brain regions indicated in bold are summarized under the label medial frontal cortex (MFC); they are discussed in this article because they play a significant role in the neural implementation of mentalizing, a concept of high relevance in trust situations (see, e.g., a review by Riedl and Javor [98]).

Formally, a reinforcement learning model is defined as

$$EV(t) = EV(t-1) + \alpha \cdot d(t-1). \tag{1}$$

From $EV(t)$, it is possible to calculate the subjective probability of trustworthy behavior at time t by dividing $EV(t)$ by €30. In our trust game, a model with a constant learning rate (see formula (1)) would assume that the perceived trustworthiness is updated equally for trustworthy and untrustworthy behavior.

We fitted the reinforcement learning model for the interaction with humans and avatars, and instructed each participant to rate the trustworthiness of each trustee before the first round of the trust game began (based solely on facial appearance). We used these initial ratings as the starting points of the reinforcement learning process. The free parameter α was fitted by minimizing the sum of squared differences between model predictions and a participant's trustworthiness ratings after the fifth and tenth rounds.

We investigated whether learning rates were significantly different between the human and avatar conditions (see Table 2 for a summary of the statistics). As expected, we found no significant differences ($t = 0.1510$, df = 17, $p = 0.8820$), thus confirming H3.

Assuming a deterministic decision strategy specifying that people always choose the alternative that offers the higher expected value when two alternatives offer a higher-than-expected value, the reinforcement learning model, on average, correctly predicts 73.61 percent (71.51 percent) of the trust decisions in the human (avatar)

Table 2. Trustworthiness Learning Rates (Human and Avatar)

Trustworthiness learning rates	Mean (SD)
Sum of squared differences (constant learning rate, human)	0.2343 (0.1322)
Sum of squared differences (constant learning rate, avatar)	0.2773 (0.1950)
Alpha (human)	0.2139 (0.0825)
Alpha (avatar)	0.2122 (0.0916)
Correct model predictions (constant learning rate, human)	73.61% (9.40%)
Correct model predictions (constant learning rate, avatar)	71.51% (13.92%)

condition. Importantly, the predictive power of this model is significantly higher than chance level (human: $t = 10.6540$, df = 17, $p < 0.001$; avatar: $t = 6.870$, df = 17, $p < 0.001$). Further analysis regarding the robustness of our results related to H3 is provided elsewhere [101].

General Discussion

Measuring by means of fMRI in a multiround trust game, we investigated (1) individuals' trustworthiness discrimination abilities in their interaction with humans versus avatars (H1); (2) the underlying neurobiological mechanisms, with a focus on activation in the MFC, a brain area well known for its importance in the neural implementation of mentalizing (H2); and (3) individuals' trustworthiness learning rates (H3). First, based on theories from evolutionary psychology, we hypothesized that people are better able to predict the trustworthiness of humans than the trustworthiness of avatars. Second, based on findings from brain imaging studies in social and cognitive neuroscience, we predicted that making decisions about whether or not to trust another actor will activate parts of the MFC significantly more during interaction with humans, if compared to interaction with avatars (in this context, also see related research from the field of robotics [56, 70]). Third, based on brain plasticity theory, we predicted that, whether interacting with humans or avatars, the trustworthiness learning rate of people is similar. We found support for all three hypotheses.

Trust evaluations that are based on the processing of facial information, as well as corresponding brain activity patterns, are phenomena in social interactions that provided a survival advantage during the past eras of human history. Although only recently have computer-mediated forms of communication and cooperation begun to emerge, they promise to bring revolutionary changes in the organization and interaction of societies. Of particular interest for our research, the exponential increase in the number of individuals who communicate and cooperate via computers and the Internet, and who are beginning to represent themselves in virtual worlds through avatars, signifies a changed understanding of interaction [7]. Genotype, which affects the development of social skills such as mentalizing and trustworthiness predictions, evolved in natural, face-to-face environments. Today's computerized world, however, often lacks natural facial information. Rather, people are afforded the possibility of representing themselves as avatars, thereby hiding their real faces behind a mask, which presents opportunities to actively misrepresent themselves [43]. Therefore, although interaction on the Internet may have benefits (e.g., [54]), the lack of real human faces in communication may serve to reduce these benefits, in turn leading to reduced levels of collaboration effectiveness. A main reason for this resulting condition is that the human brain is preprogrammed for face-to-face interaction, but not for computer-mediated interactions—in the words of evolutionary psychologist David M. Buss, humans "carry around a stone-aged brain in a modern environment" [16, p. 20].

However, despite the fact that some hardwired instincts have made humans less adaptive to life in a computerized world [65], one remarkable feature of the human brain is its flexibility. Interestingly, this powerful feature is itself an outcome of evolution [60], and this flexibility has been described as "mental meta-module . . . favored by natural selection" [63, p. 26]. In situations in which the environment is changing too fast for genetically coded instincts to be patterned in the brain, humans can nevertheless adapt to a situation in short time periods, because experience constantly changes the structure and/or functioning of the brain, leading to behavioral adaptions (e.g., [110]). Importantly, a change in the brain developed through experience is not a phenomenon that exclusively pertains to children; rather, the adult brain is also "plastic" (e.g., [67, 69]).

Despite this cognitive adaptation ability, however, the major differences between hardwired instincts and learned schemas is that the former require less cognitive effort (brain activity), and typically operate on an unconscious level (e.g., [75, 105]). This, in turn, leads to more freely available cognitive (brain) resources that may be used for other purposes, or, if not used, people may experience a lower level of cognitive load. Both factors, obviously, may increase efficiency. However, even though learning may compensate for deficits in hardwired abilities, this compensation does not come without a price (e.g., cognitive effort). Kock, citing works from evolutionary psychology, confirms this view: "As far as human communication is concerned, learned circuits are unlikely to be as efficient as the hardwired circuits endowed on us by evolution. . . . [G]enetically coded circuits are . . . automatic, unconscious, and undistracted by irrelevant aspects of world knowledge" [60, p. 121].

Specific Contributions to the IS Literature

In addition to the contributions discussed in the previous section, which refer to people's interaction with humans and avatars in general, and are therefore related to the scientific discourse in various disciplines (e.g., human–computer interaction, psychology, cognitive neuroscience), the present paper makes four specific contributions to the IS literature. First, we integrate two previously unconnected IS research domains based on the example of avatar research, namely IS evolutionary psychology and NeuroIS. Second, we generate novel insights within the domain of IS avatar research. Third, we create a knowledge base that is relevant for IS trust research. Fourth, in addition to these three theoretical contributions, this paper makes a methodological contribution to the IS literature. We discuss these four contributions in the following subsections.

Contribution 1

In this paper, we connect two previously unconnected IS research domains, namely, IS evolutionary psychology [61] and NeuroIS [27, 28, 97, 100, 102, 118]. Evolutionary psychology, in general, argues that a large number of modern brain functions evolved during the past millions of years, and that these functions, usually without conscious perception, significantly affect human performance and behavior in the modern, highly computerized world. Recently, IS researchers tellingly argued: "[l]et us remember that the person to whom we give today's latest IS has an operating system that has changed little in [the past] 100,000 years [and therefore] humans are the ultimate legacy system" [1, p. 68]. Arguments such as this substantiate the value of evolutionary psychology for IS theorizing.

Importantly, influential theoretical articles in the IS literature (e.g., [61]) as well as empirical contributions (e.g., [64]) have been published on the subject of evolutionary psychology, but to the best of our knowledge no previous IS study has applied brain imaging technology, based on a behavioral research paradigm (the trust game), to directly investigate the neurobiological *and* behavioral consequences of people "carry[ing] around a stone-aged brain in a modern environment" [16, p. 20]. Unlike previous evolutionary psychology research in the IS domain (for a collection of corresponding articles, see [62]; on the value of evolutionary psychology theorizing in the context of technology acceptance research, see [1]) that has *not* investigated brain activity as a mediating factor between genotype and a specific information-processing trait (here perception of human faces in the context of avatar research), we used fMRI to shed light on the nomological network underlying evolutionary psychology theorizing (see Figure 1). This contribution, importantly, is in line with the current call for the application of neuroscience theories and tools to inform IS phenomena (e.g., [27, 28, 102]), and specifically, our study demonstrates how the behavioral *and* biological levels of analysis can be integrated to better understand the antecedents of human behavior toward IT artifacts such as avatars.

Generally, we investigated two constructs that have been identified as major variables in IS research—trust and mentalizing [27, p. 691]. Our comparative study (human versus avatar) therefore contributes to a better understanding of two major IS concepts. Moreover, our contribution is in line with a previous review that identified a "neuropsychological focus" and an "evolutionary psychology lens" as important for IS avatar research to progress [107, pp. 440, 443].

Contribution 2

Based on an extensive analysis of the IS literature, we could identify only a limited number of avatar papers (the analysis may be obtained in electronic form by request from the first author). Thus, avatar research, in general, is an underrepresented topic in the IS literature, and consequently needs more attention within the IS discipline. Against such a research background, our study addresses a topic that has hardly been explored, yet is a highly timely topic that has become increasingly important during the past years, mainly as a consequence of the rising pervasiveness of virtual worlds [108].

Moreover, the present paper generates valuable insights within the domain of contemporary IS avatar research. Our analysis of the IS avatar literature reveals that most empirical investigations manipulated avatar design features (independent variables) in order to study the resulting effects on important outcome variables. For example, Qiu and Benbasat [93] manipulated humanoid embodiment (avatar versus no avatar) and output modality (human voice versus text-to-speech versus text) in order to study the effects on usage intention, a relationship that is mediated by social presence and trusting beliefs. As another example, Hess et al. [52] studied the effects of an avatar's vividness (among other features) on trusting beliefs, and hypothesized that this relationship is mediated by social presence. To state a third example, Nunamaker et al. [89] investigated how an avatar's gender and demeanor affects perceptions of the avatar's power, trustworthiness, likability, and expertise. In addition to these experimental studies, qualitative research has been carried out. Mueller et al. [86], for example, interviewed employees of IBM (who use the virtual world Second Life as a corporate knowledge management platform) in order to identify differences between interaction based on more traditional technologies (e.g., e-mail) and interaction via avatars regarding "knowledge and knowing activities." Thus, this kind of research compares human interaction via avatars to other forms of electronic interaction.

Overall, we conclude from our literature review that previous IS avatar research either studied the effects of design manipulations on outcome variables or compared interaction via avatars with other forms of electronic communication. Despite the valuable insights that these studies have generated, a much more fundamental question has received little analysis in the IS literature, namely, the question about the differences between human interaction in face-to-face settings and human interaction via avatars (we could identify only one study [39], discussed below). We believe that examination of this question is fundamental, as IS research should not only compare the effects of the different forms of electronic communication (e.g., e-mail, text chat, videoconferencing, avatar) but should also compare these various forms of electronic communication (e.g., avatar) with people's natural interaction method, namely, face-to-face communication.

The significance of this topic, importantly, is even reflected in the IS avatar literature itself, as signified by the following statements. Mueller et al., for example, wrote: "[Second Life] allows to share an interactive synthetic environment from a true first-person perspective, which is regarded as the second best alternative to the direct interaction of real people" [86, p. 490]. As another example, Nunamaker et al. indicated that "[w]hile humans are relatively good at identifying expressed emotions from other humans whether static or dynamic, identifying emotions from synthetic faces is more problematic" [89, p. 22]. Similarly, Franceschi et al. argue that "[e]ffective group collaboration requires . . . the existence of an appropriate working environment—an environment that makes all group members feel comfortable and fosters their participation. This is more difficult in distance group collaboration due to the lack of face-to-face contact . . . facial expressions provide rich information about the emotions and intentions of others. For this reason, probably, executives continue to travel thousands of miles to have face-to-face meetings" [39, p. 84]. Finally, as further confirmation of the value of the present study, Davis et al. explain that "[w]hether personal focus or direct contact is the same in a metaverse as in face-to-face environments is yet unanswered" [25, p. 96] (a metaverse is a three-dimensional virtual world in which people interact via avatars). Considering these statements, as well as the fact that human interaction in face-to-face settings has hardly been compared to human interaction via avatars, our study contributes to closing a significant research gap in the IS avatar literature.

As mentioned, we could identify only one article in which face-to-face interaction was directly compared to human interaction via avatars [39]. Specifically, in this field study, three learning environments (a traditional face-to-face classroom setting, a text-based virtual learning environment such as Moodle, and an avatar interaction in a virtual world based on Second Life) were compared regarding their effects on student performance. The major result of the study is that "performance of participants in the virtual world learning environment is significantly better than that of the other two environments" [39, p. 94]. This result is not in line with our finding regarding H1, where we discovered a performance advantage for people's interaction with humans, when compared to interaction with avatars (regarding trustworthiness discrimination abilities). However, the result indicating that students perform better in an avatar environment than in a face-to-face setting might be an artifact of sample selection; Franceschi et al. [39] recognize that a "novelty bias" might have affected their results. Specifically, "we mean that a certain technology, such as virtual worlds, attracts attention simply because it is new. This could potentially affect the students' measured engagement [with potential subsequent effects on performance] . . . admittedly, most of the students in our study were technology novices" [39, p. 96]. However, it might also be possible that their result reflects the brain's flexibility to quickly adapt to new environments, thereby supporting our finding regarding H3. Future studies based on more sophisticated experimental designs are therefore necessary for developing more definitive conclusions. Overall, however, the results of our study suggest that the use of avatars for human interaction and collaboration (e.g., virtual learning) may be less beneficial than previously assumed—a fact that is of particular relevance for managers, software engineers, and policymakers, among others.

Contribution 3

The present paper generated knowledge relevant for IS trust research. Specifically, our study responds to a call in a prominent research agenda paper for more investigation into the "longitudinal nature of trust" [46, p. 277]. In this research agenda for trust in online environments, Gefen et al. explain that "[t]rust, of course, is not only about one-time interactions. Trust develops gradually as people interact with each other. . . . [T]he importance of trust as a key consideration decreases with experience [and] models should have examined how the importance of trust changes over time" [46, p. 277]. Importantly, Qiu and Benbasat [93, p. 166] made a similar recommendation for additional studies employing longitudinal designs, and their call is directly concentrated on avatar trust research. Obviously, such calls confirm the value of the present paper, particularly with respect to H3.

In this context, it is also important to discuss both swift trust and knowledge-based trust, two concepts that were studied in face-to-face settings and virtual teams [103]. Swift trust denotes high levels of initial trust in another individual without any knowledge about that person, whereas knowledge-based trust denotes trust developed through interactions, based on the assessment of actual behavior. In their study, Robert et al. integrated these "two seemingly contradictory views of trust" [103, p. 242]; specifically, they manipulated team member characteristics as well as team member behavior to empirically test a theoretical model of trust formation and the influence of IT on trust formation. One major finding of their study was that perception of team member characteristics (ability, integrity, benevolence) dominated the initial formation of swift trust. However, once individuals gained knowledge, swift trust became less important and knowledge-based trust became dominant.

The results of the present paper are in line with the findings reported in Robert et al.'s [103] study; that is, subjects learned the trustees' trustworthiness based on repeated interaction (knowledge-based trust), and this affects the relevance of initial trustworthiness assessments (swift trust). A major difference between these two examinations is, however, the methodological approach to study swift trust and knowledge-based trust. While Robert et al. used "simulations of real events" (referred to as "vignettes" in the scientific literature [103, p. 251]), we investigated *actual behavior,* based on the trust game. Thus, our study overcomes a major limitation of Robert et al.'s study: "subjects could potentially respond differently to a hypothetical scenario [and] vignettes are sometimes not as powerful a manipulation as traditional experiential experiments" [103, p. 267]. However, despite the fact that vignettes cannot substitute for the investigation of actual behavior, Robert et al.'s study makes a substantial contribution to the IS literature, and we consider our experiment as a valuable complement to this prior investigation. Moreover, the value of our study is further substantiated by Robert et al., who indicate in a section on the implications of their study for future research that "we believe that more research is needed on swift trust and the *factors* that influence initial trust judgments before knowledge of behaviors has been gained" [103, p. 267, emphasis added]. Obviously, an interaction partner's face—the thematic focus of the present paper—is such a factor.

Contribution 4

Neurobiological research on trust is typically based on definitions and measurements that are different from the conceptualizations used in the IS literature. McKnight and

Chervany [80] analyzed trust definitions used in 65 publications, and categorized them by conceptual type. They found that trust is often conceptualized as belief or intention in the IS literature, and that other conceptualizations are used less frequently. Moreover, the study lists corresponding operationalizations, thereby providing a valuable basis for the development of survey instruments. If trust is conceptualized as a belief, for example, it could be operationalized along characteristics of a trustee (e.g., benevolence, integrity, and competence), and possible measurement items could be, for example, "I believe that [the trustee] would act in my best interest," "If I required help, [the trustee] would do its best to help me," and "[the trustee] is interested in my well-being, not just its own" (e.g., [80]). Importantly, in the present paper we base trust measurements on observations of actual behavior (i.e., investment in the trust game). Hence, our measurement approach differs from that in many other IS trust studies, including those in the avatar domain, where trust measurement is either based on people's beliefs or intentions (survey and interview data; see [52, 86, 89, 93]) or on hypothetical scenarios (vignettes; see [103]). By measuring actual behavior, we are making a methodological contribution to the IS literature, thereby complementing the existing set of measurement approaches.

Future Work and Limitations

A possible avenue for future research would be to vary the degree of humanlikeness of avatars. For the present study, we used avatars that have a medium degree of humanlikeness (i.e., they are not completely simplistic cartoon-like characters, nor are they photorealistic portrayals; see Figure 2). It would be reasonable to assume that avatars with increasing humanlikeness should trigger brain activation patterns monotonically more similar to that of real humans. Research suggests, however, that such an assumption would be too simplistic, because as avatars approach photorealistic perfection but do not fully accomplish it, they cause humans to feel negative emotions (e.g., [24]) that neurologically resemble distrust reactions (e.g., [26]). This effect is referred to as the *uncanny valley effect* (e.g., [47, 77, 83]).

The participants in the present study had no information about whether or not the avatars represent actual humans, but were advised that the trustees would be unresponsive to their playing strategy, and that the character of each determines trustworthiness. Thus, another useful extension of our research could be to add an experimental condition in which participants are told that the avatars represent real humans (e.g., [5]). Such a manipulation could affect both trust behavior and activation in mentalizing brain areas, which would have far-reaching practical implications, for example, with respect to reputation mechanisms and the resulting collaboration effectiveness (for details, see [35, 42, 55, 88, 94]).

In our discussion of the theoretical rationale for H3, we indicated that learning is always accompanied by changes in the nervous system. Thus, learning can be studied at various analytical levels, ranging from observable behavioral changes to alterations in molecular structure. In the present study, our research focus was on the behavioral level. Consequently, future studies should adopt the neurobiological level to examine learning of another agent's trustworthiness (human or avatar). Voxel-based morphometry (VBM), for example, has recently been suggested in the IS literature as an appropriate tool for such studies [102], because this tool makes it possible to track gray matter density changes induced by learning processes within an individual (e.g., [30, 31]).

As is common in scientific research, the present fMRI study has limitations that should be taken into account. First, the interpretation of our empirical findings is based on a simple game-playing task in a controlled laboratory environment. Second, during the experiment, the participants were required to lie still and were restrained with pads to prevent motion during measurement sessions. Therefore, the experimental situation was relatively artificial, because in real-life conditions users typically sit in front of their computers. Accordingly, future IS avatar studies could measure users' psychophysiological responses (e.g., skin conductance, heart rate) to reduce experimental artificiality. Third, the present study investigates activation of mentalizing brain mechanisms by using static pictures of humans and avatars. Indeed, we found activation in a network spanning the MFC. However, inferring another actor's thoughts and intentions is not based solely on the act of processing static information. Motion (e.g., gestures) has also been demonstrated to enable inferences regarding another actor's mind (e.g., [42]). Future investigations could replicate our study using nonstatic stimulus material. Considering recent findings regarding the positive effects of both visual and behavioral realism of avatars on outcome variables such as affect-based trustworthiness (e.g., [12, 49, 92]), we hypothesize that, for an avatar, an increased level of humanlikeness that is induced by nonstatic information (e.g., gestures or animated facial expressions) could reduce the neurobiological differences between humans and avatars in mentalizing circuits [84]. In this context, one study [45] has already found that the mirror neuron system—which transforms observed actions into the neural representations of these actions—responds to both human and robotic actions, with no significant differences between these two agents.

Despite the limitations demonstrated by the range of important issues not addressed here, however, we believe that the present study contributes to a better understanding of trust and mentalizing in IS research, and holds significant value as a preliminary study for those that will build on the investigative line that we open here.

Concluding Comments

Trust and mentalizing are of essential value for social interaction [98]. From an evolutionary viewpoint, trust has been critical for survival over millions of years [21]. Most significantly, throughout history humans have faced risks related to trust, as trust in individuals who turned out not to be trustworthy has resulted in loss and even in death. Moreover, as a correlative condition for humans, being part of a social group has supported survival [76]. The division of basic human activities (e.g., one group member takes care of food acquisition, another provides protection from adversarial groups, and another cares for offspring) has resulted in a greater certainty of survival, while social groups without this structure have not achieved the same result. Most notably, the division of activities among group members implies cooperative behavior, which is strongly based on trust and mentalizing. Bearing these explanations in mind, it becomes clear that trust and mentalizing are intriguing—and essential—social phenomena that have evolved in order to secure the survival of humankind. The human face, in particular, has served as a reliable source of information for the prediction of another individual's trustworthiness, and for the consequent increased fitness for survival. Slowly, a trait for processing human facial information has been encoded in human genetic makeup. The recent emergence of computer-mediated forms of communication and cooperation, however, and most particularly the advent of avatars, promises to bring revolutionary changes in the organization and interaction of societies. Whatever new forms of social and economic interaction will confront society in the future, humans hold the ability to cope with the challenges, because evolution has endowed us with a remarkable, flexible brain.

Acknowledgments: The authors thank the guest editors, Ting-Peng Liang and Jan vom Brocke, as well as three anonymous reviewers for their excellent work in providing guidance on ways to improve the paper. Moreover, they are grateful to Cornelia Huber for her support in developing the avatars. They also appreciate the generous and visionary support of Novomind AG, Germany (www.novomind.com). They appreciate the generous support of Schindler Parent Meersburg who, through the Schindler Parent Distinguished Guest Lecturer for Marketing, supported René Riedl's work on NeuroIS and consumer neuroscience projects at Zeppelin University Friedrichshafen. Peter N.C. Mohr's work on this project was supported by the "Collaborative Research Center 649: Economic Risk" funded by the German Research Foundation (DFG). Fred D. Davis's work on this paper was partially supported by Sogang Business School's World Class University Project (R32-20002) funded by the Korean Research Foundation. The authors acknowledge the support of Hock Hai Teo and Hock Chuan Chan, who served as chairs of the track "Human–Computer Interaction" at the 2011 International Conference on Information Systems (ICIS, Shanghai, December 4–7), where an earlier version of this paper was presented. They also thank the conference participants who provided useful comments. Finally, they thank Deborah C. Nester for proofreading the manuscript.

References

1. Abraham, C.; Boudreau, M.-C.; Junglas, I.; and Watson, R. Enriching our theoretical repertoire: The role of evolutionary psychology in technology acceptance. *European Journal of Information Systems, 22,* 1 (2013), 56–75.

2. Abu-Akel, A. A neurobiological mapping of theory of mind. *Brain Research Reviews, 43,* 1 (2003), 29–40.

3. Adolphs, R. The social brain: Neural basis of social knowledge. *Annual Review of Psychology, 60* (2009), 693–716.

4. Amodio, D.M., and Frith, C.D. Meeting of minds: The medial frontal cortex and social cognition. *Nature Reviews Neuroscience, 7,* 4 (2006), 268–277.

5. Appel, J.; von der Pütten, A.M.; Krämer, N.C.; and Gratch, J. Does humanity matter? Analyzing the importance of social cues and perceived agency of a computer system for the emergence of social reactions during human–computer interaction. *Advances in Human–Computer Interaction,* (2012), article ID 324694, 1–10.

6. Bailenson, J.N., and Blascovich, J. Avatars. In W.S. Baindbridge (ed.), *Encyclopedia of Human–Computer Interaction.* Great Barrington, MA: Berkshire, 2004, pp. 64–68.

7. Bainbridge, W.S. The scientific research potential of virtual worlds. *Science, 317* (July 27, 2007), 472–476.

8. Bates, B., and Cleese, J. *The Human Face.* New York: DK, 2001.

9. Behrens, T.E.J.; Hunt, L.T.; and Rushworth, M.F.S. The computation of social behavior. *Science, 324* (May 29, 2009), 1160–1164.

10. Behrens, T.E.J.; Hunt, L.T.; Woolrich, M.W.; and Rushworth, M.F.S. Associative learning of social value. *Nature, 456* (November 13, 2008), 245–249.

11. Benbasat, I.; Dimoka, A.; Pavlou, P.A.; and Qiu, L. Incorporating social presence in the design of the anthropomorphic interface of recommendation agents: Insights from an fMRI study. In *Proceedings of the 31st International Conference on Information Systems.* Atlanta: Association for Information Systems, 2010, pp. 1–22.

12. Bente, G.; Rüggenberg, S.; Krämer, N.C.; and Eschenburg, F. Avatar-mediated networking: Increasing social presence and interpersonal trust in Net-based collaborations. *Human Communication Research, 34,* 2 (2008), 287–318.

13. Berg, J.; Dickhaut, J.; and McCabe, K. Trust, reciprocity, and social history. *Games and Economic Behavior, 10,* 1 (1995), 122–142.

14. Boaz, N.T., and Almquist, A.J. *Biological Anthropology: A Synthetic Approach to Human Evolution.* Upper Saddle River, NJ: Prentice Hall, 2001.

15. Bodamer, J. 1947. Die prosopagnosie. *Archiv für Psychiatrie und Nervenkrankheiten, 179,* 1–2 (1947), 6–53.

16. Buss, D.M. *Evolutionary Psychology: The New Science of the Mind.* Needham Heights, MA: Allyn & Bacon, 1999.

17. Cacioppo, J.T.; Berntson, G.G.; Sheridan, J.F.; and McClintock, M.K. Multilevel integrative analyses of human behavior: Social neuroscience and the complementing nature of social and biological approaches. *Psychological Bulletin, 126,* 6 (2000), 829–843.

18. Caldara, R.; Seghier, M.L.; Rossion, B.; Lazeyras, F.; Michel, C.; and Hauert, C.-A. The fusiform face area is tuned for curvilinear patterns with more high-contrasted elements in the upper part. *NeuroImage, 31,* 1 (2006), 313–319.

19. Carlson, J.R., and Zmud, R.W. Channel expansion theory and the experiential nature of media richness perceptions. *Academy of Management Journal, 42,* 2 (1999), 153–170.

20. Cartwright, J. *Evolution and Human Behavior: Darwinian Perspectives on Human Nature.* Cambridge: MIT Press, 2000.

21. Cosmides, L., and Tooby, J. Cognitive adaptations for social exchange. In J.H. Barkow, L. Cosmides, and J. Tooby (eds.), *The Adapted Mind: Evolutionary Psychology and the Generation of Culture.* London: Oxford University Press, 1992, pp. 163–228.

22. Dahl, C.D.; Logothetis, N.K.; Bülthoff, H.H.; and Wallraven, C. The Thatcher illusion in humans and monkeys. *Proceedings of the Royal Society B: Biological Sciences, 277,* 1696 (2010), 2973–2981.

23. Darwin, C.R. *On the Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life.* London: John Murray, 1859.

24. Davidson, R.J. Affective neuroscience and psychophysiology: Toward a synthesis. *Psychophysiology, 40,* 5 (2003), 655–665.

25. Davis, A.; Murphy, J.; Owens, D.; Khazanchi, D.; and Zigurs, I. Avatars, people, and virtual worlds: Foundations for research in metaverses. *Journal of the Association for Information Systems, 10,* 2 (2009), 90–117.

26. Dimoka, A. What does the brain tell us about trust and distrust? Evidence from a functional neuroimaging study. *MIS Quarterly, 34,* 2 (2010), 373–396.

27. Dimoka, A.; Pavlou, P.A.; and Davis, F.D. NeuroIS: The potential of cognitive neuroscience for information systems research. *Information Systems Research, 22,* 4 (2011), 687–702.

28. Dimoka, A.; Banker, R.D.; Benbasat, I.; Davis, F.D.; Dennis, A.R.; Gefen, D.; Gupta, A.; Ischebeck, A.; Kenning, P.; Müller-Putz, G.; Pavlou, P.A.; Riedl, R.; vom Brocke, J.; and Weber, B. On the use of neurophysiological tools in IS research: Developing a research agenda for NeuroIS. *MIS Quarterly, 36,* 3 (2012), 679–702.

29. Donath, J. Virtually trustworthy. *Science, 317* (July 6, 2007), 53–54.

30. Draganski, B.; Gaser, C.; Busch, V.; Schuierer, G.; Bogdahn, U.; and May, A. Neuroplasticity: Changes in grey matter induced by training. *Nature, 427* (January 22, 2004), 311–312.

31. Driemeyer, J.; Boyke, J.; Gaser, C.; Büchel, C.; and May, A. Changes in gray matter induced by learning—Revisited. *PLoS ONE, 3,* 7 (2008), 1–5.

32. Dunbar, R.I.M. The social brain hypothesis. *Evolutionary Anthropology, 6,* 5 (1998), 178–190.

33. Ekman, P. *Emotion in the Human Face,* 2d ed. New York: Cambridge University Press, 1982.

34. Falloon, G. Using avatars and virtual environments in learning: What do they have to offer? *British Journal of Educational Technology, 41,* 1 (2010), 108–122.

35. Fehr, E. Human behaviour: Don't lose your reputation. *Nature, 432* (November 25, 2004), 449–450.

36. Fehr, E. On the economics and biology of trust. *Journal of the European Economic Association, 7,* 2–3 (2009), 235–266.

37. Fehr, E. Social preferences and the brain. In P.W. Glimcher, C.F. Camerer, E. Fehr, and R.A. Poldrack (eds.), *Neuroeconomics: Decision Making and the Brain.* Amsterdam: Academic Press, 2009, pp. 215–232.

38. Flinn, M.V. Evolution of stress response to social threat. In R.I.M. Dunbar and L. Barrett (eds.), *The Oxford Handbook of Evolutionary Psychology.* Oxford: Oxford University Press, 2007, pp. 273–295.

39. Franceschi, K.; Lee, R.M.; Zanakis, S.H.; and Hinds, D. Engaging group e-learning in virtual worlds. *Journal of Management Information Systems, 26,* 1 (Summer 2009), 73–100.

40. Frith, C.D., and Frith, U. The neural basis of mentalizing. *Neuron, 50,* 4 (2006), 531–534.

41. Frith, U., and Frith, C.D. Development and neurophysiology of mentalizing. *Philosophical Transactions of the Royal Society B: Biological Sciences, 358,* 1431 (2003), 459–473.

42. Frith, U., and Frith, C.D. The social brain: Allowing humans to boldly go where no other species has been. *Philosophical Transactions of the Royal Society B: Biological Sciences, 365,* 1537 (2010), 165–176.

43. Galanxhi, H., and Nah, F.F.-H. Deception in cyberspace: A comparison of text-only vs. avatar-supported medium. *International Journal of Human–Computer Studies, 65,* 9 (2007), 770–783.

44. Gallagher, H.L., and Frith, C.D. Functional imaging of "theory of mind." *Trends in Cognitive Sciences, 7,* 2 (2003), 77–83.

45. Gazzola, V.; Rizzolatti, G.; Wicker, B.; and Keysers, C. The anthropomorphic brain: The mirror neuron system responds to human and robotic actions. *NeuroImage, 35,* 4 (2007), 1674–1684.

46. Gefen, D.; Benbasat, I.; and Pavlou, P.A. A research agenda for trust in online environments. *Journal of Management Information Systems, 24,* 4 (Spring 2008), 275–286.

47. Geller, T. Overcoming the uncanny valley. *IEEE Computer Graphics and Applications, 28,* 4 (2008), 11–17.

48. Goren, C.C.; Sarty, M.; and Wu, P.Y.K. Visual following and pattern discrimination of face-like stimuli by newborn infants. *Pediatrics, 56,* 4 (1975), 544–549.

49. Groom, V.; Nass, C.; Chen, T.; Nielsen, A.; Scarborough, J.K.; and Robles, E. Evaluating the effects of behavioral realism in embodied agents. *International Journal of Human–Computer Studies, 67,* 10 (2009), 842–849.

50. Hamer, D. Rethinking behavior genetics. *Science, 298* (October 4, 2002), 71–72.

51. Haxby, J.V.; Gobbini, M.I.; Furey, M.L.; Ishai, A.; Schouten, J.L.; and Pietrini, P. Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science, 293* (September 28, 2001), 2425–2430.

52. Hess, T.; Fuller, M.; and Campbell, D. Designing interfaces with social presence: Using vividness and extraversion to create social recommendation agents. *Journal of the Association for Information Systems, 10,* 12 (2009), 889–919.

53. Kanwisher, N.; McDermott, J.; and Chun, M.M. The fusiform face area: A module in human extrastriate cortex specialized for face perception. *Journal of Neuroscience, 17,* 11 (1997), 4302–4311.

54. Keeney, R.L. The value of Internet commerce to the customer. *Management Science, 45,* 4 (1999), 533–542.

55. King-Casas, B.; Tomlin, D.; Anen, C.; Camerer, C.F.; Quartz, S.R.; and Montague, P.R. Getting to know you: Reputation and trust in a two-person economic exchange. *Science, 308* (April 1, 2005), 78–83.

56. Kircher, T.; Blumel, I.; Marjoram, D.; Lataster, T.; Krabbendam, L.; Weber, J.; van Os, J.; and Krach, S. Online mentalising investigated with functional MRI. *Neuroscience Letters, 454,* 3 (2009), 176–181.

57. Knutson, B. Facial expression of emotion influence interpersonal trait inferences. *Journal of Nonverbal Behavior, 20,* 3 (1996), 165–182.
58. Kock, N. The ape that used e-mail: Understanding e-communication behavior through evolution theory. *Communications of the Association for Information Systems, 5,* 3 (2001), 1–29.
59. Kock, N. The psychobiological model: Towards a new theory of computer-mediated communication based on Darwinian evolution. *Organization Science, 15,* 3 (2004), 327–348.
60. Kock, N. Media richness or media naturalness? The evolution of our biological communication apparatus and its influence on our behavior toward e-communication tools. *IEEE Transactions on Professional Communication, 48,* 2 (2005), 117–130.
61. Kock, N. Information systems theorizing based on evolutionary psychology: An interdisciplinary review and theory integration framework. *MIS Quarterly, 33,* 2 (2009), 395–418.
62. Kock, N. *Evolutionary Psychology and Information Systems Research: A New Approach to Studying the Effects of Modern Technologies on Human Behavior.* New York: Springer, 2010.
63. Kock, N. Evolutionary psychology and information systems theorizing. In N. Kock (ed.), *Evolutionary Psychology and Information Systems Research: A New Approach to Studying the Effects of Modern Technologies on Human Behavior.* New York: Springer, 2010, pp. 3–37.
64. Kock, N., and Chatelain-Jardón, R. Four guiding principles for research on evolved information processing traits and technology-mediated task performance. *Journal of the Association for Information Systems, 12,* 10 (2011), 684–713.
65. Kock, N., and Hantula, D.A. Do we have e-collaboration genes? *International Journal of e-Collaboration, 1,* 1 (2005), i–ix.
66. Kohler, T.; Matzler, K.; and Fuller, J. Avatar-based innovation: Using virtual worlds for real-world innovation. *Technovation, 29,* 6–7 (2009), 395–407.
67. Kolb, B., and Whishaw, I.Q. Brain plasticity and behaviour. *Annual Review of Psychology, 49* (1998), 43–64.
68. Kolb, B., and Whishaw, I.Q. *Fundamentals of Human Neuropsychology,* 6th ed. New York: Worth, 2009.
69. Kolb, B.; Gibb, R.; and Robinson, T.E. Brain plasticity and behaviour. *Current Directions in Psychological Science, 12,* 1 (2003), 1–5.
70. Krach, S.; Hegel, F.; Wrede, B.; Sagerer, G.; Binkofski, F.; and Kircher, T. Can machines think? Interaction and perspective taking with robots investigated via fMRI. *PLoS ONE, 3,* 7 (2008), 1–11.
71. Krueger, F.; McCabe, K.; Moll, J.; Kriegeskorte, N.; Zahn, R.; Strenziok, M.; Heinecke, A.; and Grafman, J. Neural correlates of trust. *Proceedings of the National Academy of Sciences, 104,* 50 (2007), 20084–20089.
72. Lawrence, P., and Nohria, N. *Driven: How Human Nature Shapes Our Choices.* San Francisco: Jossey-Bass, 2002.
73. Lee, K.C., and Chung, N. A Web DSS approach to building an intelligent Internet shopping mall by integrating virtual reality and avatar. *Expert Systems with Applications, 28,* 2 (2005), 333–346.
74. LeGrand, R.; Mondloch, C.J.; Maurer, D.; and Brent, H.P. Neuroperception: Early visual experience and face processing. *Nature, 410* (April 19, 2001), 890.
75. Lieberman, M.D. Social cognitive neuroscience: A review of core processes. *Annual Review of Psychology, 58* (2007), 259–289.
76. Lieberman, M.D., and Eisenberger, N.I. Pains and pleasures of social life. *Science, 323* (February 13, 2009), 890–891.
77. MacDorman, K.F.; Green, R.D.; Ho, C.-C.; and Koch, C.T. Too real for comfort? Uncanny responses to computer generated faces. *Computers in Human Behavior, 25,* 3 (2009), 695–710.
78. Mattay, V.S., and Goldberg, T.E. Imaging genetic influences in human brain function. *Current Opinion in Neurobiology, 14,* 2 (2004), 239–247.
79. McCabe, K.; Houser, D.; Ryan, L.; Smith, V.; and Trouard, T. A functional imaging study of cooperation in two-person reciprocal exchange. *Proceedings of the National Academy of Sciences, 98,* 20 (2001), 11832–11835.
80. McKnight, D.H., and Chervany, N.L. Trust and distrust definitions: One bite at a time. In R. Falcone, M. Singh, and Y.-H. Tan (eds.), *Trust in Cyper-Societies.* Berlin: Springer, 2001, pp. 27–54.

81. McMeill, D. *The Face: A Natural History.* Boston: Little, Brown, 1998.

82. Miller, G. The promise of parallel universes. *Science, 317* (September 7, 2007), 1341–1343.

83. Mori, M. Bukimi no tani [The uncanny valley]. *Energy, 7,* 4 (1970), 33–35.

84. Morris, J.P.; Pelphrey, K.A.; and McCarthy, G. Regional brain activation evoked when approaching a virtual human on a virtual walk. *Journal of Cognitive Neuroscience, 17,* 11 (2005), 1744–1752.

85. Moser, E.; Derntl, B.; Robinson, S.; Fink, B.; Gur, R.C.; and Grammer, K. Amygdala activation at 3T in response to human and avatar facial expressions of emotions. *Journal of Neuroscience Methods, 161,* 1 (2007), 126–133.

86. Mueller, J.; Hutter, K.; Fueller, J.; and Matzler, K. Virtual worlds as knowledge management platform: A practice-perspective. *Information Systems Journal, 21,* 6 (2011), 470–501.

87. Nowak, K.L., and Rauh, C. The influence of the avatar on online perceptions of anthropomorphism, androgyny, credibility, homophily, and attraction. *Journal of Computer-Mediated Communication, 11,* 1 (2005), 153–178.

88. Nowak, M.A., and Sigmund, K. Evolution of indirect reciprocity. *Nature, 437* (October 27, 2005), 1291–1298.

89. Nunamaker, J.F., Jr.; Derrick, D.C.; Elkins, A.C.; Burgoon, J.K.; and Patton, M.W. Embodied conversational agent-based kiosk for automated interviewing. *Journal of Management Information Systems, 28,* 1 (Summer 2011), 17–48.

90. Oosterhof, N.N., and Todorov, A. The functional basis of face evaluation. *Proceedings of the National Academy of Sciences, 105,* 32 (2008), 11087–11092.

91. Premack, D., and Woodruff, G. Does the chimpanzee have a theory of mind? *Behavioral and Brain Sciences, 1,* 4 (1978), 515–526.

92. Qiu, L., and Benbasat, I. Online consumer trust and live help interfaces: The effects of text-to-speech voice and three-dimensional avatars. *International Journal of Human–Computer Interaction, 19,* 1 (2005), 75–94.

93. Qiu, L., and Benbasat, I. Evaluating anthropomorphic product recommendation agents: A social relationship perspective to designing information systems. *Journal of Management Information Systems, 25,* 4 (Spring 2009), 145–181.

94. Raub, W., and Weesie, J. Reputation and efficiency in social interactions: An example of network effects. *American Journal of Sociology, 96,* 3 (1990), 626–654.

95. Rendall, D.; Notman, H.; and Vokey, J.R. Homologizing the mind: Frodoian or fodorian psychology? In R.I.M. Dunbar and L. Barrett (eds.), *The Oxford Handbook of Evolutionary Psychology.* Oxford: Oxford University Press, 2007, pp. 59–70.

96. Rescorla, R.A., and Wagner, A.R. A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In A.H. Black and W.F. Prokasy (eds.), *Classical Conditioning II: Current Research and Theory.* New York: Appleton-Century-Crofts, 1972, pp. 64–99.

97. Riedl, R. Zum Erkenntnispotenzial der kognitiven Neurowissenschaften für die Wirtschaftsinformatik: Überlegungen anhand exemplarischer Anwendungen [On the potential of cognitive neuroscience for business informatics: Thoughts based on exemplary applications]. *NeuroPsychoEconomics, 4,* 1 (2009), 32–44.

98. Riedl, R., and Javor, A. The biology of trust: Integrating evidence from genetics, endocrinology and functional brain imaging. *Journal of Neuroscience, Psychology, and Economics, 5,* 2 (2012), 63–91.

99. Riedl, R.; Hubert, M.; and Kenning, P. Are there neural gender differences in online trust? An fMRI study on the perceived trustworthiness of eBay offers. *MIS Quarterly, 34,* 2 (2010), 397–428.

100. Riedl, R.; Kindermann, H.; Auinger, A.; and Javor, A. Technostress from a neurobiological perspective: System breakdown increases the stress hormone cortisol in computer users. *Business & Information Systems Engineering, 4,* 2 (2012), 61–69.

101. Riedl, R.; Mohr, P.N.C.; Kenning, P.H.; Davis, F.D.; and Heekeren, H.R. Trusting humans and avatars: Behavioral and neural evidence. In *Proceedings of the 32nd International Conference on Information Systems.* Atlanta: Association for Information Systems, 2011, pp. 1–23.

102. Riedl, R.; Banker, R.D.; Benbasat, I.; Davis, F.D.; Dennis, A.R.; Dimoka, A.; Gefen, D.; Gupta, A.; Ischebeck, A.; Kenning, P.; Müller-Putz, G.; Pavlou, P.A.; Straub, D.W.;

vom Brocke, J.; and Weber, B. On the Foundations of NeuroIS: Reflections on the Gmunden Retreat 2009. *Communications of the Association for Information Systems, 27* (2010), 243–264.

103. Robert, L.P., Jr.; Dennis, A.R.; and Hung, Y.C. Individual swift trust and knowledge trust in face-to-face and virtual team members. *Journal of Management Information Systems, 26,* 2 (Fall 2009), 241–279.

104. Rossion, B.; Hanseeuw, B.; and Dricot, L. Defining face perception areas in the human brain: A large-scale factorial fMRI face localizer analysis. *Brain and Cognition, 79,* 2 (2012), 138–157.

105. Rudoy, J.D., and Paller, K.A. Who can you trust? Behavioral and neural differences between perceptual and memory-based influences. *Frontiers in Human Neuroscience, 3* (2009), 1–6.

106. Schlicht, E.J.; Shimojo, S.; Camerer, C.F.; Battaglia, P.; and Nakayama, K. Human wagering behavior depends on opponents' faces. *PLoS ONE, 5,* 7 (2010), 1–10.

107. Schultze, U. Embodiment and presence in virtual worlds: A review. *Journal of Information Technology, 25,* 4 (2010), 434–449.

108. Schultze, U., and Orlikowski, W.J. Virtual worlds: A performative perspective on globally distributed, immersive work. *Information Systems Research, 21,* 4 (2010), 810–821.

109. Singer, T. Understanding others: Brain mechanisms of theory of mind and empathy. In P.W. Glimcher, C.F. Camerer, E. Fehr, and R.A. Poldrack (eds.), *Neuroeconomics: Decision Making and the Brain.* Amsterdam: Academic Press, 2009, pp. 251–268.

110. Sparrow, B.; Liu, J.; and Wegner, D.M. Google effects on memory: Cognitive consequences of having information at our fingertips. *Science, 333* (August 5, 2011), 776–778.

111. Sripada, C.S.; Angstadt, M.; Banks, S.; Nathan, P.J.; Liberzon, I.; and Phan, K.L. Functional neuroimaging of mentalizing during the trust game in social anxiety disorder. *NeuroReport, 20,* 11 (2009), 984–989.

112. Suh, K.-S.; Kim, H.; and Suh, E.K. What if your avatar looks like you? Dual-congruity perspectives for avatar use. *MIS Quarterly, 35,* 3 (2011), 711–729.

113. Sutton, R.S., and Barto, A.G. *Reinforcement Learning: An Introduction.* Cambridge: MIT Press, 1998.

114. Thompson, P.; Cannon, T.D.; Narr, K.L.; van Erp, T.; Poutanen, V.-P.; Huttunen, M.; Lönnqvist, J.; Standertskjöld-Nordenstam, C.-G.; Kaprio, J.; Khaledy, M.; Dail, R.; Zoumalan, C.I.; and Toga, A.W. Genetic influences on brain structure. *Nature Neuroscience, 4,* 12 (2001), 1253–1258.

115. Todd, P.M., and Gigerenzer, G. Mechanisms of ecological rationality: Heuristics and environments that make us smart. In R.I.M. Dunbar and L. Barrett (eds.), *The Oxford Handbook of Evolutionary Psychology.* Oxford: Oxford University Press, 2007, pp. 197–210.

116. Todorov, A. Evaluating faces on trustworthiness: An extension of systems for recognition of emotions signaling approach/avoidance behaviors. *Annals of the New York Academy of Sciences, 1124* (2008), 208–224.

117. Todorov, A.; Said, C.P.; Engell, A.D.; and Oosterhof, N.N. Understanding evaluation of faces on social dimensions. *Trends in Cognitive Sciences, 12,* 12 (2008), 455–460.

118. Vom Brocke, J.; Riedl, R.; and Léger, P.-M. Application strategies for neuroscience in information systems design science research. *Journal of Computer Information Systems, 53,* 3 (2013), 1–13.

119. Vuilleumier, P.; Armony, J.L.; Driver, J.; and Dolan, R.J. Distinct spatial frequency sensitivities for processing faces and emotional expressions. *Nature Neuroscience, 6,* 6 (2003), 624–631.

120. Wei, X.; Zhu, Z.; Yin, L.; and Ji, Q. A real time face tracking and animation system. In *Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops.* Los Alamitos, CA: IEEE Compuater Society, 2004, pp. 1–8.

121. Willis, J., and Todorov, A. First impressions: Making up your mind after a 100-ms exposure to a face. *Psychological Science, 17,* 7 (2006), 592–598.

122. Winston, J.S.; Strange, B.A.; O'Doherty, J.; and Dolan, R.J. Automatic and intentional brain responses during evaluation of trustworthiness of faces. *Nature Neuroscience, 5,* 3 (2002), 277–283.

Appendix

We acquired 2 runs of 690 functional T2*-weighted echoplanar images (EPI) [TR [repetition time], 2 s; echo time (TE), 40 ms; flip angle, 90°; field of view, 256 mm; matrix, 64 × 64 mm; 26 axial slices approximately parallel to the bicommissural plane; slice thickness, 4 mm]. In addition, for registration purposes, a high-resolution T1-weighted structural image (MPRAGE) was acquired from each participant [TR, 20 ms; TE, 5 ms; flip angle, 30°; 179 sagittal slices; voxel size, 1 × 1 × 1 mm]. Initial analysis was performed using the FSL toolbox from the Oxford Centre for fMRI of the Brain (www.fmrib.ox.ac.uk/fsl/). The image time-course was first realigned to compensate for small head movements. Data were spatially smoothed using an 8 mm full-width-half-maximum Gaussian kernel and were temporally filtered using a high-pass temporal filter (with sigma = 100 s). Registration was conducted through a two-step procedure, whereby EPI images were first registered to the MPRAGE structural image and then to standard MNI (Montreal Neurological Institute) space (MNI152_T1_2mm_brain), using 7 parameters for the first registration step and 12 parameters for the second. Statistical analyses were performed in native space, with the statistical maps normalized to standard space prior to higher-level analysis. Statistical analysis of functional data was performed using a multilevel approach implementing a mixed-effects model treating participants as a random effect. This was initially performed separately for each participant's concatenated runs. Regressors-of-interest were created by convolving a rectangular function representing stimulus duration times with a canonical (double-gamma) hemodynamic response function. Time-series statistical analysis was carried out using FILM (FMRIB's Improved Linear Model) with local autocorrelation correction. The functional analysis was based on two regressors-of-interest and four regressors-of-no-interest. Two binary regressors modeled the decision phase. The first regressor represents the decision in the human condition, whereas the second regressor models the decision in the avatar condition. The durations of the two regressors correspond to the decision time, which was three seconds in all cases. The final four regressors-of-no-interest modeled the trustworthiness ratings for humans and avatars as well as the outcome presentations for both conditions. On the group level (second level of analysis), we integrated the results from the single-subject level (first level), again applying a general linear model. One binary regressor modeled a constant effect of all the first-level parameter estimates on the group level.