A little anthropomorphism goes a long way: Effects of oxytocin on trust, compliance and team performance with automated agents

Ewart J. de Visser[1,2], **Samuel S. Monfort**[1], **Kimberly Goodyear**[3,6], **Li Lu**[1], **Martin O'Hara**[4], **Mary R. Lee**[5], **Raja Parasuraman**[1], and **Frank Krueger**[1,6]

[1]Department of Psychology, George Mason University, Fairfax, VA, USA

[2]Perceptronics Solutions, Inc., Falls Church, VA, USA

[3]Center for Alcohol and Addiction Studies, Department of Behavioral and Social Sciences, Brown University, Providence, RI, USA

[4]Virginia Hospital Center, Fairfax Hospital, Arlington, VA, USA

[5]Section on Clinical Psychoneuroendocrinology and Neuropschopharmacology, National Institute on Alcohol Abuse and Alcoholism, Bethesda, MD, USA

[6]Molecular Neuroscience Department, George Mason University, Fairfax, VA, USA

Abstract

Objective—We investigated the effects of exogenous oxytocin on trust, compliance, and team decision making with agents varying in anthropomorphism (computer, avatar, human) and reliability (100%, 50%).

Background—Recent work has explored psychological similarities in how we trust human-like automation compared to how we trust other humans. Exogenous administration of oxytocin, a neuropeptide associated with trust among humans, offers a unique opportunity to probe the anthropomorphism continuum of automation to infer when agents are trusted like another human or merely a machine.

Method—Eighty-four healthy male participants collaborated with automated agents varying in anthropomorphism that provided recommendations in a pattern recognition task.

Results—Under placebo, participants exhibited less trust and compliance with automated aids as the anthropomorphism of those aids increased. Under oxytocin, participants interacted with aids on the extremes of the anthropomorphism continuum similarly to placebos, but increased their trust, compliance, and performance with the avatar, an agent on the midpoint of the anthropomorphism continuum.

Correspondence concerning this article should be addressed to Ewart de Visser, Director of Human Factors and UX Research, Perceptronics Solutions, Inc, 3141 Fairview Park Drive, Suite #415, Falls Church, VA 22042, 910-200-8596 (cell), 703-342-4657 (office), 703-346-4661 (fax), edevisser@percsolutions.com; edevisse@gmu.edu.
Ewart J. de Visser, George Mason University, Perceptronics Solutions, Inc.

Conclusion—This study provided the first evidence that administration of exogenous oxytocin affected trust, compliance, and team decision making with automated agents. These effects provide support for the premise that oxytocin increases affinity for social stimuli in automated aids.

Application—Designing automation to mimic basic human characteristics is sufficient to elicit behavioral trust outcomes that are driven by neurological processes typically observed in human-human interactions. Designers of automated systems should consider the task, the individual, and the level of anthropomorphism to achieve the desired outcome.

Keywords

Trust in Automation; Oxytocin; Autonomous Agents; Compliance and Reliance; Human-Automation Interaction; Neuroergonomics; Virtual Humans

Once devoted solely to sensing and mechanical actions (Sheridan & Verplank, 1978), automated systems routinely carry out complex information processing and decision-making functions (Parasuraman, Sheridan, & Wickens, 2000, 2008). Many robotic agents are now specifically designed to imitate human behavior and to take advantage of the human tendency to anthropomorphize and ascribe intentionality to inanimate agents (Epley, Waytz, & Cacioppo, 2007). These systems listen, talk, express emotion, and otherwise interact with human users in apparently intelligent ways. Capitalizing on the human tendency to anthropomorphize may be an efficient and effective way to enable long-term bonds between humans and human-like automated virtual agents (Kurzweil, 2005; Nof, 2009). In some cases, a human-like interaction partner might be more effective than a purely human one: virtual therapists, for example, are sometimes more successful in soliciting disclosure from patients than their human counterparts (Lucas, Gratch, King, & Morency, 2014). In other cases, developments in machine learning may give rise to more advanced personal assistants included in mobile firmware or our homes (e.g., Apple Siri, Android Cortana, or Amazon Echo). Virtual interaction partners may also become necessary in scenarios with limited human habitability, such as in Antarctic winter-over missions or in long-duration space flight to Mars (Sandal, Leon, & Palinkas, 2006; Wu et al., 2015). Greater dependency on these agents may stimulate deeper forms of trust that resemble those experienced in long-term intimate relationships between humans (Bickmore & Cassell, 2001; Cassell & Bickmore, 2003; Gratch, Wang, Gerten, Fast, & Duffy, 2007; Qiu & Benbasat, 2010; Traum, Rickel, Gratch, & Marsella, 2003). As the nature of human-machine social interaction becomes more complex, a greater understanding of the role of machine anthropomorphism is needed (de Visser et al., 2012, 2016; Pak, Fink, Price, Bass, & Sturre, 2012; Pak, McLaughlin, & Bass, 2014).

Early work on human-computer interaction demonstrated that people readily apply social norms to non-human interaction partners if those agents exhibit human characteristics (Nass, Fogg, & Moon, 1996; Nass, Steuer, & Tauber, 1994). Recent research has expanded on this paradigm, finding that more salient human characteristics in automation (e.g., appearance, background story, observable behavior, etiquette) are associated with a tendency to treat that automation like another human (Hayes & Miller, 2011; Parasuraman & Miller, 2004). For example, human-like automation seems to resist breakdowns in trust, having greater *trust*

resilience than computer-like automation (de Visser et al., 2012, 2016; Madhavan, Wiegmann, & Lacson, 2006). A theoretical explanation for this result is that people activate different schemas for humans compared to automation (Madhavan & Wiegmann, 2007). For example, humans are typically expected to be fallible and, as a result, are forgiven more easily than automated systems when mistakes are made (Madhavan & Wiegmann, 2007). Thus, automation designed to mimic human characteristics seems to inherit certain effects typically observed among real humans. Despite these differences found for agents varying in anthropomorphism, it is not clear where the level becomes sufficient for people to perceive automated agents as if they are human (Nass & Moon, 2000; Nass et al., 1994). It is possible that there is a threshold at which people categorically switch their attitudes, intentions, and behaviors towards non-human agents (Blascovich et al., 2002; Dennett, 1989; Wiese, Wykowska, Zwickel, & Müller, 2012). To better support longer-term interactions with human-like automated agents, it will be important to understand how perceptions of anthropomorphism affect trust in those agents.

Forming social bonds with other agents involves processes of motivation, interpreting social information, and creation of social memories. A fundamental driver of this social bonding process is oxytocin. Oxytocin is a hormone and neurotransmitter that mediates social cognition and pro-social behaviors in humans (Kosfeld, Heinrichs, Zak, Fischbacher, & Fehr, 2005; Meyer-Lindenberg, Domes, Kirsch, & Heinrichs, 2011; Zak, Kurzban, & Matzner, 2004, 2005). Spawning a decade of cognitive social neuroscience research with oxytocin, several studies reported the trust-enhancing effect of oxytocin (Kosfeld et al., 2005; Zak et al., 2004, 2005), apparent in a two-person economic investment game paradigm (Berg, Dickhaut, & McCabe, 1995), was absent when a human participant interacted with a computer, that simulated a lottery. Kosfeld et al's comparison between social and non-social agents—between a human and a computer—is a binary representation of context that has been adopted by the majority of oxytocin researchers to date (Bartz, Zaki, Bolger, & Ochsner, 2011; Bethlehem, Baron-Cohen, van Honk, Auyeung, & Bos, 2014; Carter, 2014; Kanat, Heinrichs, & Domes, 2014). Rather than depict context dichotomously, a more nuanced approach is adopted in the current study consistent with recent research on automation anthropomorphism that manipulates social context with greater granularity (de Visser et al., 2016; Pak et al., 2012, 2014; Wiese et al., 2012). Under this design, oxytocin can be used as a drug probe to help determine the number and type of anthropomorphic features that are required to elicit the known biological effect of the peptide. For automated agents of sufficient anthropomorphism, exogenous administration of oxytocin should trigger an increase in outcomes related to social bonding. Various mechanisms have been proposed to explain how oxytocin affects social bonding including increased affiliative motivation, increased salience towards social stimuli, or a decrease in anxiety (Bartz et al., 2011; Bethlehem et al., 2014; Carter, 2014; Grillon et al., 2013; Kanat et al., 2014). Our research question was therefore to examine if oxytocin affects a person's perception of anthropomorphism and the subsequent trust, compliance, and performance during interaction with automated cognitive agents.

The current study investigated, for the first time, the effects of oxytocin on social interactions between humans and automation varying in anthropomorphism (computer, avatar, human). Trust, compliance, and team performance were examined in a pattern

recognition task in which participants were assisted by an automated aid. Synthetic oxytocin (or a placebo) was administered intranasally to healthy male participants in a double-blind between-subjects design. We hypothesized that under placebo, trust, compliance, and team performance would be highest when interacting with a computer, lowest when interacting with a human and with the assessment of the avatar agent falling between those two agents (de Visser et al., 2016; Madhavan et al., 2006; Madhavan & Wiegmann, 2007). In contrast, we predicted that oxytocin would increase trust, compliance, and team performance, but only for automation with a sufficient degree of anthropomorphism (i.e., the avatar or human). Consistent with previous studies (Kosfeld et al., 2005; Van Ijzendoorn & Bakermans-Kranenburg, 2012; Veening & Olivier, 2013; Zak et al., 2004, 2005), we expected no differences between oxytocin and placebo conditions for computer agents. In addition to the anthropomorphism of the automation, we manipulated its reliability, decreasing it from 100% to 50% halfway through the experiment. This drop-off was incorporated to provide greater contextual variance with regard to the appropriateness of compliance with automated advice, a key research issue in the collaboration between human and automation (Hoff & Bashir, 2015; J. Lee & See, 2004; Parasuraman & Riley, 1997). We chose 100% reliability to be compatible with a previous oxytocin study that distinguished between the dichotomous categories of reliable and unreliable human partners (Mikolajczak et al., 2010). We chose 50% reliability as a control condition to match the lottery condition typically used in oxytocin research with economic games and because automation below 70% reliability is generally considered unreliable (Wickens & Dixon, 2007). We did not expect any effect of oxytocin in unreliable conditions (Mikolajczak et al., 2010).

Method

Participants

Eighty-four healthy male college students (M age = 24.2 years, SD = 2.7) from George Mason University (GMU) gave informed written consent and participated for financial compensation. The study was approved by Chesapeake and GMU Institutional Review Board (IRB) committees. Participants were given a starting fund of $35 at the beginning of the experimental task and were penalized $0.35 for every incorrect answer. This study used a method for oxytocin administration identical to a previous study (Goodyear et al., 2016). In particular, participants were excluded if they had evidence of medical or psychiatric disorder, used any drugs that may interact with oxytocin, had a history of hypersensitivity to oxytocin or vehicle, showed the presence of or history of clinically significant allergic rhinitis as assessed by the study physician, smoked more than 10 cigarettes per day, had a current or past history of drug or alcohol abuse or dependence, or had any experiences of trauma involving either injury or threat of injury to themselves or a close family or friend.

Forty-two participants were randomly assigned each to the placebo and oxytocin conditions. Forty-five minutes before the start of the experiment, participants were instructed to self-administer a single 40-IU dose of oxytocin or placebo (Syntocinon spray, Pharmaworld, Zurich, Switzerland). Previous studies have used 40-IU doses in both humans and rodents (Born et al., 2002; Ditzen et al., 2009; Neumann, Maloumby, Beiderbeck, Lukas, &

Landgraf, 2013; Zak, Stanton, & Ahmadi, 2007). Upon completing the study, a physician checked on the well-being of each participant before they departed the experimental suite.

Stimuli and Procedure

Story stimuli—Background stories were used to establish an agent's intent and character; these have been used successfully in previous research (de Visser et al., 2016; Delgado, Frank, & Phelps, 2005). The agent stories contained a brief life history, a paragraph about each agent's personality, and a fictional excerpt of a news report about a characteristic behavior.

Video stimuli—Brief video clips were used to provide additional cues for an agent's thinking process and to promote perceptions of agency identical to a previous study (de Visser et al., 2016). For the experimental task, 24 videos were created for each agent to suggest the process of thinking prior to providing participants with the recommendation. For the human condition, a male actor from the University's theatre department was filmed. The actor was instructed to make various sounds and say phrases such as "let me think…" and "hmmm…" to simulate thinking before giving advice. The avatar agent was created using the online website XtraNormal (http://xtranormal.com), although the software is no longer publically available (some alternatives for creating avatars are nawmal.com, sitepal.com, getinsta3d.com, and/or livingactor.com). A standard avatar agent template was used and animated to say "thinking" phrases similar to those used by the human agent. The computer agent consisted of blinking lights compiled from separate GIF images and put in various orders to simulate thinking accompanied by a series of beeps.

Task—A version of the TNO Trust Task (T^3) was adapted to develop a trust paradigm with high task uncertainty, dependency on an automated agent, and moderate financial stakes (de Visser et al., 2012, 2016; van Dongen & van Maanen, 2013). In the T^3, participants were asked to discover a number pattern (1-2-3-1-2) in a sequence of numbers with the help of an automated aid. Each task trial was completed in a series of five steps (see Figure 1). Participants were first asked to select one of three numbers: 1, 2, or 3, which constituted their first choice. Second, a video of the agent was presented showing the aid was thinking. In the third step, the agent recommended a number in the sequence. After this recommendation was shown, participants made a final choice on selecting the number, either keeping their original answer or changing it. Then in the last step, the correct answer was shown for two seconds. After this step, the next task trial would begin, repeating this procedure for the next agent. Each trial lasted about 10 seconds. Participants were not told that these videos were pre-recorded. However, they were told that they could be receiving recommendations from any of the three agents.

Design—The experimental design for this study was a $2 \times 3 \times 2$ mixed measures design with drug condition (oxytocin, placebo) as a between-subjects variable and agent (computer, avatar, human) and reliability (100%, 50%) as within-subjects factors. Each of the two reliability levels was presented twice for a total of four reliability blocks, always starting with two blocks of 100% first and then followed by two blocks of 50% (see Figure 2). Each reliability block contained 18 trials for a total of 72 experimental trials for the four reliability

blocks. Each reliability block contained six agent blocks and each agent block contained three trials for a total of 18 trials per reliability block. Each agent block contained three trials, one trial for each of the three agents. The agent block was randomized so that the presentation order would vary and to ensure that each agent did not appear more than twice in a row (i.e. avatar, human, computer, computer, avatar, human, etc.). Errors in the agents' advice were randomized within each reliability block for the 50% reliability conditions.

Measurement—*Team performance* (human + agent) was calculated as the overall percentage of trials where the participants provided the correct answer for their final decision, after the aid had made the recommendation. *Compliance* was calculated as the percentage of trials in which the participants changed their initial response to match the one provided by the automation; an indicator of an aid's persuasiveness to the participant. Only trials where the participant's first choice disagreed with the aid's recommendation were used in this calculation (van Dongen & van Maanen, 2013). Lastly, subjective trust was measured using a 10-item Likert scale (1–10) after each reliability block (de Vries, Midden, & Bouwhuis, 2003; J. Lee & Moray, 1992; Lewandowsky, Mundy, & Tan, 2000). Because we were interested in relative metrics—the way participants responded to each agent relative to the way that they responded to the others—we mean centered our outcomes for each regression (Paccagnella, 2006). That is, our research was less concerned with between-persons comparisons than it was with within-person comparisons. This procedure is based on previous work (Cronbach, 1976; Morey, 2008) advocating for the procedure to separate individual effects from their context (i.e., agent effects from person effects, in our research), and has been used in many domains since then (see Kreft, de Leeuw, & Aiken, 1995 for a summary). Practically speaking, mean centering involves subtracting each participants' relevant metrics from their own individual means, and is ideal when the primary goal of the research is to make relative, rather than absolute, judgments.

Control measures—A number of control measures were included to assess inherent differences between the experimental groups and to rule out competing influences on any potential effects. These measures were given one to two weeks before the experimental protocol, and included the NEO-PI-R to assess personality (Costa & McCrae, 1992), the Interpersonal Reactivity Index (IRI) for empathy (Davis, 1983), and the Relationship Scales Questionnaire (RSQ) for affiliative motivation (Griffin & Bartholomew, 1994). Immediately before and after the procedure, we also assessed participants' perceptions of the three agents' humanness, intelligence, and likeability using the Godspeed measure (Bartneck, Kuli , Croft, & Zoghbi, 2009), a measure used to assess responses to a wide variety of agents on dimensions of anthropomorphism, animacy, likability, perceived intelligence, and perceived safety. The survey is available online in multiple languages (Bartneck, 2008). These measures were selected because of their potential relevance to responses to automation and to ensure that both groups did not differ on personality measures. Research suggests that personality and other individual differences can have strong effects on a number of automation-related outcomes (Merritt, Heimbaugh, LaChapell, & Lee, 2013; Merritt, Lee, Unnerstall, & Huber, 2015; Merritt & Ilgen, 2008; Szalma & Taylor, 2011; Szalma, 2009).

To exclude the possibility that anxiety reduction is a possible explanation of oxytocin effects, participants also completed a number of state mood measures before, during, and after the experiment, including the STAI for state anxiety (Spielberger, Gorsuch, Lushene, Vagg, & Jacobs, 1983) and the PANAS for positive and negative affect (Watson, Clark, & Tellegen, 1988). Research suggests that state mood can also affect human-automation interaction outcomes, including initial trust in an automated system (Merritt, 2011; Stokes et al., 2010).

Statistical analyses—We used R and *lme4* (Bates, Maechler, & Bolker, 2012; R Core Team, 2013) to produce linear mixed effects models with restricted maximum likelihood (REML) for all analyses (Bryk & Raudenbush, 1987; Laird & Ware, 1982). This procedure better accounts for the potential non-independence of repeated measurements within individuals. Unlike traditional analyses (e.g., ANOVA, linear regression), mixed effects models do not aggregate data across participants. They can, therefore, account for the random variation between individuals that would otherwise contribute to the error term (Baayen, Davidson, & Bates, 2008). Random slope terms were selected via nested model comparison; only those terms that significantly improved model fit were retained. Sattherwaite (1946) approximations were used to determine denominator degrees of freedom for *t*- and *p*-values. Separate analyses were conducted for trust, compliance, and team performance with drug condition (placebo, oxytocin), agent (computer, avatar, human), and automation reliability (100%, 50%) as predictors. Treatment contrasts were applied to agent to better characterize the effects of automation anthropomorphism on the various outcomes; the human aid was used as the reference group because research to date has only examined effects of oxytocin with human-human interactions (Veening & Olivier, 2013). Mean centering was used to show the relative response of an individual to each agent (Paccagnella, 2006). Effect size estimates were provided for all significant effects (Cohen, 1992).

Results

Trust

There was a significant effect of agent, whereby including agent as a predictor significantly improved overall model fit, $\chi^2(2)=33.2$, p<.001. To examine the contribution of agent in greater depth, treatment contrasts were used, which compared oxytocin's effects on each agent relative to its effect on a comparison group. We used the human agent as the comparison group because oxytocin is typically studied in a human-human interaction context. These tests revealed significantly higher trust ratings for both computer, B=.95, SE=.27, t(101.6)=3.56, p<.001, d=.71, and avatar agents, B=.68, SE=.024, t(106.5)=2.81, p<.001, d=.56, compared to the human agent. Although there was no main effect of drug condition, B=−.22, SE=.28, t(93.7)=−.78, p=.44, there was an interaction between the trust effect for avatar and drug condition, B=.90, SE=.35, t(106.5)=2.61, p=.010, d=.51, and a three-way interaction between avatar, drug condition, and reliability, B=−.53, SE=.25, t(742.7)=−2.10, p=.036, d=.15. These interactions represent a differential effect of oxytocin on the avatar compared to the human agent, and signify that these differences depend upon automation reliability, occurring only in the high reliability blocks. Pairwise contrasts reveal that although participants reported equally high trust for the reliable computer agent by drug

condition, B=−15, SE=.32, t(81.0)=−.47, p=.99, and equally low trust in the reliable human agent by drug condition, B=.30, SE=.39, t(81.0)=.76, p=.97, trust in the reliable avatar depended heavily on whether participants had been administered oxytocin, B=−.96, SE=.27, t(81.0)=−3.50, p=.009, d=−.78. Specifically, participants administered oxytocin reported trust in the avatar that was more comparable to their trust in the computer, whereas participants administered the placebo reported trust in the avatar that was more comparable to their trust in the human. This effect seems to be limited to the higher reliability blocks, as oxytocin administration had no effect on trust in the avatar during low-reliability trials, B=−.042, SE=.26, t(248.5)=−.16, p=1.00 (see Figure 3).

There was also a main effect of automation reliability, whereby 100% reliable automation was trusted substantially more than 50% reliable automation, B=−1.41, SE=.21, t(140.1)= −6.80, p<.001, d=1.15. This effect was qualified by a marginally significant interaction with the avatar contrast term, B=−.33, SE=.18, t(742.7)=−1.84, p=.066, d=−.14, and a significant interaction with the computer contrast term, B=−.58, SE=.18, t(742.5)=−3.25, p=.001, d=−.24. Compared to the trust decrement seen for the human agent, participants reported a steeper loss in trust for the computer and avatar agents as a result of the loss in reliability between blocks.

Lastly, to explore potential differences between the computer and the avatar, a new set of contrasts was conducted using the computer as the reference group. This analysis yielded a marginally significant overall trust difference between computer and avatar agents, B=−.28, SE=.16, t(902.6)=−1.69, p=.091, d=.11, which was qualified by an interaction with drug condition, B=.57, SE=.23, t(902.6)=2.47, p=.014, d=.16. These effects represent an overall lower level of trust for avatar compared to computer agents, an effect driven entirely by differences in the placebo condition, B=−.68, SE=.29, t(40)=−2.34, p=.024, d=.74. Thus consistent with the effect described earlier, whether the level of trust for the avatar agent was similar to that for the human or that for the computer depended on whether or not oxytocin was administered (see Figure 3).

Compliance

There was again a main effect of agent; its inclusion as a predictor significantly improved model fit, $\chi^2(2)$=57.4, p<.001. Agent effects were again investigated using treatment contrasts with humans as the reference group. Consistent with the results for trust, participants were significantly more likely to comply with the advice of the computer, B=.15, SE=.034, t(165.1)=4.60, p<.001, d=.72, and the avatar, B=.21, SE=.037, t(144.3)=5.66, p<.001, d=.94, compared to the human. Also consistent with trust, there was an interaction between the avatar contrast term and drug condition, B=.10, SE=.047, t(165.1)=2.11, p=.036, d=.33. Although this effect was not qualified by a three-way interaction with automation reliability, B=−.073, SE=.052, t(819)=−1.41, p=.16, post-hoc contrasts reveal that compliance differences by drug condition only occurred in the high-reliability blocks. Oxytocin again only affected compliance with the avatar, B=−.12, SE=.040, t(400.0)=−3.02, p=.032, d=−.30, and did not vary for the computer agent, B=.037, SE=.040, t(393.3)=.91, p=.94, nor for the human agent, B=.020, SE=.040, t(397.2)=.51, p=.996. Thus again, participants administered oxytocin complied with avatars at a rate more similar to that for

computers, and participants administered the placebo complied with avatars at a rate more similar to that for humans (see Figure 4).

There was also a main effect of automation reliability, B=.083, SE=.026, t(817.5)=3.23, p=. 001, d=.23, as well as an interaction between reliability and the computer contrast term, B= −.11, SE=.037, t(820.6)=−2.92, p=.004, d=.20, and between reliability and the avatar contrast term, B=−.091, SE=.037, t(819.0)=−2.50, p=.013, d=.17. As with trust, participants exhibited a more dramatic compliance decrement with computer and avatar agents following losses in automation reliability compared to the human agent.

We repeated these contrasts using the computer as the reference group to explore potential differences between the avatar and computer agents. Identical to trust, this analysis yielded a marginally significant main effect, B=−.05, SE=.03, t(167.8)=−1.79, p=.075, d=.28, which interacted significantly with drug condition, B=.11, SE=.04, t(167.8)=2.54, p=.012, d=.39. These moderately-sized effects mirror those found for trust. Compliance rates for avatars tended to be somewhat lower than those for computers overall, but these differences were driven by the placebo condition, which saw the largest difference between the two agents. In contrast, compliance rates for avatars and computers in the oxytocin condition were roughly the same (see Figure 4).

Team Performance

Effects for team performance mirrored those for trust and compliance. The main effect of Agent was again significant, with model fit significantly improved by the addition of agent as a predictor, $\chi^2(2)$=6.57, p=.037. Treatment contrasts revealed that participants performed significantly better with computer and avatar agents than they did with the human agent, B=. 12, SE=.017, t(996)=7.07, p<.001, d=.45, and B=.093, SE=.017, t(996)=5.59, p<.001, d=.35, respectively. There was also an interaction between the contrast term for avatar and drug condition, B=.066, SE=.024, t(996)=2.79, p=.005, d=.18. This effect again seems to have been primarily driven by participant behavior in the 100% reliability blocks, an assertion supported by a marginally significant three-way interaction with automation reliability, B=. 017, SE=.010, t(996)=1.80, p=.072, d=.11. Consistent with the effects for trust and compliance, oxytocin administration only affected performance with the avatar, B=−.067, SE=.025, t(408.1)=−2.57, p=.011, d=−.25. Compared to their average performance rate, participants administered oxytocin performed with the avatar at a level similar to their performance with the computer, and participants administered the placebo performed with the avatar at a level more similar to their performance with the human (see Figure 5).

There were no differences by drug condition for computer nor for human agents, B=−.002, SE=.026, t(408.1)=−.083, p=.934, and B=.026, SE=.026, t(408.1)=1.00, p=.319, respectively. In sum, oxytocin improved team performance relative to placebo, but for the avatar agent only.

There was a large main effect for automation reliability: participants performed substantially worse when the automation was unreliable than when it was reliable, B=−.30, SE=.017, t(996)=−17.9, p<.001, d=−1.13. Again like trust and compliance, the loss in performance following reliability loss was more severe when participants were paired with the computer

or avatar agents compared to the human agent, B=−.14, SE=.024, t(996)=−6.07, p<.001, d= −.38, and B=−.12, SE=.024, t(996)=−4.89, p<.001, d=−.31.

Lastly, repeating these contrasts with computer as the reference group yielded no overall differences between the avatar and the computer, B=−.02, SE=.02, t(996)=−1.53, p=.13, but similar to trust and compliance, the interaction with drug condition was significant, B=.05, SE=.02, t(996)=2.01, p=.045, d=.13. This effect was similar to the previous two, as performance with the avatar was closer to that with the computer for oxytocin compared to placebo (see Figure 5).

Godspeed Scales

Humanness—Overall, participants in both conditions rated the computer and the avatar agents as significantly less human than the human agent, B=−2.94, SE=.14, t(246)=−20.84, p<.001, d=−2.65, and B=−2.56, SE=.14, t(246)=−18.14, p<.001, d=−2.31, respectively. Planned comparisons by drug condition within each agent subsequently revealed that participants administered oxytocin rated the avatar agent as somewhat more human than the participants administered the placebo did, B=−.37, SE=.20, t(245.9)=−1.87, p=.06, d=.24. Similar to the effects for trust, compliance, and performance, there were no differences in perceived humanness between conditions for computer or human agents, B=.20, SE=.20, t(245.9)=.99, p=.32, and B=−.21, SE=.20, t(245.9)=−1.03, p=.30, respectively.

Liking—There were also differences in liking by agent. Participants reported more positive feelings for the human compared to the computer and the avatar, B=−.058, SE=.01, t(246)= −3.34, p<.001, d=−.43, and B=−.033, SE=.01, t(246)=−5.85, p<.001, d=−.75, respectively. Planned comparisons show a divergence for avatars by drug condition: participants administered oxytocin rated avatars as significantly more likeable than placebo participants did, B=−.28, SE=.14, t(245.9)=−1.97, p=.049, d=−.25, and there were no differences by drug condition for computer or human agents, B=−.063, SE=.14, t(245.9)=−.45, p=.653, and B=−.064, SE=.14, t(245.9)=−.45, p=.654, respectively.

Intelligence—Participants rated the computer agent as less intelligent than the human agent, B=−.50, SE=.13, t(246)=−3.93, p<.001, d=−.50, and the intelligence ratings for human and avatar agents were statistically equivalent, B=−.17, SE=.13, t(246)=−1.31, p=.19. Planned comparisons within each agent found that only the avatar differed in intelligence ratings by drug condition. Specifically, oxytocin participants rated the avatar as significantly more intelligent than placebo participants did, B=−.38, SE=.18, t(245.9)=−2.14, p=.034, d= −.27. Consistent with the null effects for humanness and liking, there were no differences in intelligence ratings by drug condition for computer or human agents, B=.091, SE=.18, t(245.9)=.51, p=.613, and B=.14, SE=.18, t(245.9)=.77, p=.441, respectively.

In sum, the administration of oxytocin only affected participants' perceptions of the avatar agent (see Figure 6). Participants given oxytocin rated the avatar as more human-like, likeable, and of greater intelligence than participants given a placebo. Oxytocin did not affect these ratings for humans nor for computers.

Control measures

A number of analyses were conducted on the control measures to rule out competing influences on our observed effects. There were no significant group differences in any of these measures (see Appendix, Table 1). Raw means and standard deviations were also calculated to support future meta-analytic efforts (see Appendix, Table 2).

Discussion

The current study provides the first evidence that administration of exogenous oxytocin affects trust, compliance, and team performance with non-human, automated agents. Our data replicate and extend previous oxytocin research, suggesting that similar processes are involved with trusting human-like machines compared to humans. These effects are heavily dependent on a number of contextual factors, however. First, we found that oxytocin only augmented trust, compliance, and performance among non-human agents that appeared somewhat human. Second, oxytocin did not influence interactions with unreliable automation (i.e., whose performance did not exceed chance levels). These results represent an effect of oxytocin that is consistent with past research: that oxytocin does not increase trust in entirely non-human interaction partners (Kosfeld et al., 2005), and that oxytocin does not override important cues regarding an interaction partner's untrustworthiness (Mikolajczak et al., 2010).

Mechanisms of Oxytocin

The mechanisms of action for oxytocin are diverse and not well-understood (Bartz et al., 2011). Indeed, a great number of proposed mechanisms has arisen as a result of the elusive and complex oxytocin effects depicted in recent reviews including the affiliative motivation, social salience, and anxiety hypotheses (Bartz et al., 2011; Bethlehem et al., 2014; Carter, 2014; Grillon et al., 2013; Kanat et al., 2014).

Our results are mostly consistent with the affiliative motivation hypothesis of oxytocin, which proposes that oxytocin augments motivation to engage in pro-social behavior such as trust (Carter, 2014; Ebstein, Israel, Chew, Zhong, & Knafo, 2010; Gamer, Zurowski, & Büchel, 2010; Heinrichs, von Dawans, & Domes, 2009; Kosfeld et al., 2005; H. J. Lee, Macbeth, Pagani, & Young, 2009; Taylor, 2006; Theodoridou, Rowe, Penton-Voak, & Rogers, 2009). However, our effects were specific to the avatar; affiliative motivation was increased towards the reliable avatar alone. It is possible that the affiliative motivation increased because only avatars were trusted to be competent at the pattern-matching task as compared to the human, consistent with the findings of the control measures that showed an increase in liking and perceived intelligence for the avatar under oxytocin compared to the other agents. The increased anthropomorphic features of the avatar may have been enough to interact with oxytocin to increase motivation to work with the agent. Oxytocin may not have increased motivation or trust towards the human agent because it was perceived as being generally less reliable in this context (i.e. the computational task) compared to the other agents as shown with both the placebo in this study and in previous studies (de Visser et al., 2012, 2016). This conception is consistent with past research that suggests oxytocin makes people more social, but not more gullible towards untrustworthy agents (Mikolajczak et al.,

2010) and that the situational context is important to understand the effect of oxytocin (Bartz et al., 2011). Motivation to engage with technology in general and automation specifically represents an important, but under-studied domain (Szalma, 2014; Weiner, 2012): whether or not users interact with a system may depend on their desire just as much as their ability to do so. Just as system designers endeavor to match their designs to the limitations of the human user, they should also aim to encourage motivation and willful engagement, such as with human-automation etiquette approaches (Hayes & Miller, 2011). Oxytocin's effect on motivation—and the consequent benefits to performance—provide further support to the relevance of motivation in these interactions.

Our results are not consistent with the predictions of the social salience hypothesis, which would predict that oxytocin heightens sensitivity to social cues, either positive or negative, and affects downstream social behavior towards others (Bartz et al., 2011). Consistent with our reasoning above, the hypothesis would predict a decrease in trust for the human because negative cues would receive more attention with oxytocin. The hypothesis would also predict additional decreases for untrustworthy agents because of a heightened sensitivity of untrustworthiness as a negative cue. The results did not show any of these additional decreases above and beyond the placebo. The control measures also did not show a change in perceptions of humanness of any of the agents under oxytocin, which is inconsistent with the predictions of the social salience hypothesis.

Finally, our results also do not support the anxiety reduction hypothesis of oxytocin (Campbell, 2010), which postulates that greater affiliative behaviors are driven by reduced anxiety. The mood of participants did not differ by drug condition and did not significantly change throughout the experiment.

Towards Determining a Continuum of Anthropomorphism

The present study supports the idea that people categorically switch their attitudes for agents varying in anthropomorphism. This result is consistent with models assuming that perceptions of anthropomorphism are divided by a set threshold where an agent *is* or *is not* human (Blascovich et al., 2002; Dennett, 1989; Wiese et al., 2012). As with the avatar in our experiment, oxytocin may lower the anthropomorphic requirements needed to observe and treat agents as social entities. Although this is strong support that oxytocin stimulates a categorical switch in attitudes, it is not enough evidence to support that oxytocin has a dosed effect such that people gradually increase their social stance towards an agent with the amount of oxytocin given. This is a suitable topic for further exploration for anthropomorphic dimensions such as appearance (Martini, Gonzalez, & Wiese, 2016) and biological motion (Thompson, Trafton, & McKnight, 2011).

The fact that the computer was not affected by oxytocin is further support that automated agents without obvious physical human-like characteristics do not activate attributions of anthropomorphism. The lack of an oxytocin effect for computer agents is consistent with past research that found endogenous oxytocin release to be reduced in young children interacting with their mothers through a text interface compared to through a phone or in person (Seltzer, Prososki, Ziegler, & Pollak, 2012). Inconsistent with past research is the null effect for the human agent. Although we expected oxytocin to improve trust-related

outcomes for participants interacting with the human agent as well, there were no differences here by drug condition. This null effect might be interpreted as a result of social context. That is, effects of oxytocin are typically qualified by both the properties of the interaction partner and the context of the interaction (Bartz et al., 2011). The lack of oxytocin differences for the human agent may follow in this tradition, as overall trust levels were lower for human agents. Indeed, a previous study using the same pattern-matching task found that participants held a strong bias against collaborating with the human (de Visser et al., 2016), an effect that seems to be driven by initially lower expectations regarding the human agent's ability to complete the pattern-matching task accurately (Smith, Allaham, & Wiese, 2016). Lower levels of trust and compliance observed in the current study suggest that the human agent was again seen as untrustworthy: administering oxytocin may not be able to override such a bias consistent with previous research (Mikolajczak et al., 2010). Finally, the subjective social threshold may vary by individual, the identification of which may be incorporated into training paradigms. Incorporating the examination of individual differences systematically into oxytocin research with automation may lead to insights for both oxytocin researchers and human factors researchers alike. Both oxytocin and human factors research is heavily moderated by person and context (Bartz et al., 2011; Merritt & Ilgen, 2008; Szalma, 2009). Investigating specific types of populations will be helpful in establishing specific treatments and the design of adaptive automation systems tailored to the needs of an individual (de Visser & Parasuraman, 2011; Parasuraman, Cosenzo, & de Visser, 2009).

Limitations and Future Directions

Our study had several limitations that future researchers of this topic area should consider. The agents used in this study were pre-recorded agents. The observed effects in this study will likely be stronger if participants interacted with real human participants and higher fidelity of the appearance, voice, behavior, agency, and intentionality of virtual agents (Brainbridge, Hart, Kim, & Scassellati, 2011; de Melo, Gratch, & Carnevale, 2014; Fox et al., 2014; Gray, Gray, & Wegner, 2007; Hancock et al., 2011; Lucas et al., 2014; Muralidharan, de Visser, & Parasuraman, 2014; van den Brule, Dotsch, Bijlstra, Wigboldus, & Haselager, 2014). While some have argued that the mere situational presentation of an agent is sufficient to elicit social responses towards a computer (Nass & Moon, 2000; Nass et al., 1994; Reeves & Nass, 1996), others have shown strong physiological responses with more high-fidelity embodied agents such as robots (Li, Ju, & Reeves, 2016). Future research should examine whether oxytocin interacts with the fidelity of the agent.

The environmental context of our interaction with automated or robotic agents may be another important factor that affected our results. For instance, oxytocin produces strong responses to out-group agents when the bond between in-group members is very strong (De Dreu & Kret, 2016). Research on trust with automated and robotic agents is typically still limited to assessing performance aspects and less on more human dimensions such as integrity and benevolence (Mayer, Davis, & Schoorman, 1995). Intimate bonds with machines and robots may become more common when such agents are an essential part of our daily lives (Garreau, 2007; Kahn et al., 2012; Leite et al., 2013; Meadows, 2010; Sung, Guo, Grinter, & Christensen, 2007). In these cases, relationships with machines may be just

as strong as those with other humans and oxytocin would be the driving process for those social bonds. Future research should examine effects of oxytocin in highly social tasks. Finally, our study was limited to examining behavioral effects of oxytocin administration (Van Ijzendoorn & Bakermans-Kranenburg, 2012; Veening & Olivier, 2013). Neuroimaging studies examining oxytocin on specific brain regions are helpful in differentiating and specifying underlying mechanisms of trust (Baumgartner, Heinrichs, Vonlanthen, Fischbacher, & Fehr, 2008; Bethlehem et al., 2014; Bethlehem, van Honk, Auyeung, & Baron-Cohen, 2013; Carter, 2014; De Dreu & Kret, 2016; Heinrichs et al., 2009; Kanat et al., 2014; Meyer-Lindenberg et al., 2011; Zink & Meyer-Lindenberg, 2012). For the purposes of this research, the brain can be used effectively as a social influence detector to differentiate the degree of social response towards an agent based on the pattern of brain activation in relevant neural systems. One study has already shown that activation in the medial frontal cortex and the right-temporo junction increased linearly with anthropomorphism (Krach et al., 2008), both regions implicated with theory-of-mind abilities (Baron-Cohen, 1995; Fletcher, 1995). Furthermore, endogenous oxytocin measurement may be a promising method to assess the neural correlates of the social response towards human-like automated agents (Crockford, Deschner, Ziegler, & Wittig, 2014; De Dreu & Kret, 2016; Striepens et al., 2013). In sum, assessing both performance and brain function can therefore provide more information than either alone, consistent with a neuroergonomic approach to understanding human performance with automation (Parasuraman, 2003, 2011).

Conclusion

This study showed that oxytocin increases social-affiliative behavior between humans and reliable human-like automated agents. The present study provides a greater understanding of the interaction between people and collaborative automation technology. Understanding these dynamics is critical in a society that is increasingly dependent on such technologies to be productive, social, efficient, and safe. Discovering the role of human characteristics in human-machine relationships will allow for more efficient and effective design.

Acknowledgments

Appendix

Table 1

Means, standard deviations, and regression results for various control variables predicting drug condition

	Oxytocin	Placebo	B	SE	t	df	p
Age (yea	22.50 (2.90)	21.83 (2.30)	.468	.424	1.10	77	.273
Education (years)	16.08 (2.49)	15.88 (2.24)	.125	.383	.325	77	.746
Interpersonal Reactivity Index (IRI)							
Perspective taking	3.49 (.70)	3.51 (.73)	−.004	.115	−.039	77	.969
Fantasy	3.41 (.97)	3.61 (.79)	−.142	.143	−.989	77	.326
Emphatic concern	3.70 (.61)	3.64 (.71)	.048	.107	.446	77	.657
Personal distress	2.60 (.76)	2.62 (.77)	−.012	.124	−.098	77	.922
Relationship Scale Questionnaire (RSQ)							
Secure	3.19 (.59)	3.11 (.65)	.056	.100	.561	77	.577
Fearful	2.71 (.90)	2.81 (.76)	−.090	.133	−.672	77	.503
Preoccupied	2.81 (.84)	2.93 (.66)	−.101	.121	−.827	77	.411
Dismissing	3.24 (.63)	3.55 (.56)	−.095	.094	−1.02	77	.312
NEO Five-Factor Inventory (NEO-FFI)							
Openness	3.33 (.63)	3.30 (.49)	.008	.090	.093	77	.926
Conscientiousness	3.72 (.72)	3.64 (.60)	.032	.106	.309	77	.758
Extraversion	3.38 (.58)	3.49 (.61)	−.083	.096	−.868	77	.388
Agreeableness	3.65 (.58)	3.59 (.52)	.039	.090	.440	77	.661
Neuroticism	2.68 (.80)	2.79 (.71)	−.066	.123	−.539	77	.591
State-Trait Anxiety Inventory (STAI)							
Trait Total	2.26 (.231)	2.20 (.223)	.040	.037	1.09	77	.28
State Time 1	1.11 (.23)	1.14 (.29)	−.013	.042	−.312	77	.756
State Time 2	1.08 (.25)	1.05 (.13)	.021	.033	.656	77	.514
State Time 3	1.05 (.13)	1.11 (.45)	−.040	.053	−.747	77	.457
Positive and Negative Affect Schedule (PANAS)							
Time 1	2.46 (.45)	2.53 (.47)	−.038	.073	−.514	77	.609

	Oxytocin	Placebo	B	SE	t	df	p
Time 2	2.34 (.53)	2.39 (.45)	−.025	.079	−.312	77	.756
Time 3	2.29 (.54)	2.31 (.53)	−.011	.087	−.124	77	.902

Table 2

Means and standard deviations for raw trust, team performance, and compliance values by drug condition and agent. Note that all values in this table are positive values and reflect conventional reporting of means and standard deviations. This reporting style contrasts with the mean centered values reported in Figures 3, 4, and 5 that include both negative and positive values.

100% Reliable

Drug Condition	Agent	Trust M	SD	Compliance (%) M	SD	Team Performance (%) M	SD
Placebo	Computer	8.51	1.75	83.71	25.93	88.41	18.65
	Avatar	7.83	2.23	70.85	34.35	82.72	21.67
	Human	7.78	2.28	62.08	38.53	78.05	24.97
Oxytocin	Computer	8.53	1.99	82.68	30.38	89.73	17.74
	Avatar	8.66	1.60	83.43	26.89	90.50	15.66
	Human	7.37	2.87	61.75	36.01	76.55	24.44

50% Reliable

Drug Condition	Agent	Trust M	SD	Compliance (%) M	SD	Team Performance (%) M	SD
Placebo	Computer	6.65	2.25	81.48	30.30	44.51	12.57
	Avatar	6.45	2.38	74.27	31.65	45.33	13.68
	Human	6.36	2.36	69.94	33.89	48.58	14.63
Oxytocin	Computer	6.29	2.23	77.64	32.90	45.54	12.37
	Avatar	6.44	2.33	77.05	33.29	45.35	11.90
	Human	5.85	2.68	69.11	39.17	46.51	15.36

Biographies

Ewart J. de Visser is a director of human factors and user experience research at Perceptronics Solutions, Inc. He is affiliate/adjunct faculty and the director of the Truman Lab at George Mason University. He received his Ph.D. in human factors and applied cognition in 2012 from George Mason University. He received his B.A. in film studies from the University of North Carolina Wilmington in 2005.

Samuel S. Monfort is a doctoral student at George Mason University in the Human Factors and Applied Cognition program. He received his Master's degree in Psychology from Wake Forest University in 2012.

Kimberly Goodyear is a postdoctoral research fellow at Brown University. She received her Ph.D. in neuroscience at George Mason University in 2016. She received her B.S. in psychology at San Diego State University in 2005.

Li Lu is a user experience designer of Ecompex, Inc. She received her Master's degree in human factors and applied cognition in 2014 from George Mason University.

Martin J. O'Hara is a cardiologist in practice at Northern Virginia Hospital Center. He specializes in nuclear cardiology.

Mary R. Lee is a clinical researcher at National Institute on Alcohol Abuse and Alcoholism, Section on Clinical Psychoneuroendocrinology and Neuropsychopharmacology. She is external affiliate at the Krasnow Institute for Advanced Study at George Mason University.

Raja Parasuraman was a University Professor of Psychology. He was the Director of the Graduate Program in Human Factors and Applied Cognition and Director of the Center of Excellence in Neuroergonomics, Technology, and Cognition (CENTEC). Previously he held appointments as Professor and Associate Professor of Psychology at The Catholic University of America, Washington DC from 1982 to 2004. He received a B.Sc. (1st Class Honors) in Electrical Engineering from Imperial College, University of London, U.K. (1972) and a Ph.D. in Psychology from Aston University, Birmingham, U.K. (1976).

Frank Krueger is Associate Professor of Social Cognitive Neuroscience at the Department of Psychology at George Mason University. He is Chief of the Social Cognition and Interaction: Functional Imaging Lab and Co-Director of the Center for the Study of Neuroeconomics at the Krasnow Institute for Advanced Study at George Mason University. He received his Ph.D. in cognitive psychology in 2000 from Humboldt University Berlin, Germany.

References

Baayen RH, Davidson DJ, Bates DM. Mixed-effects modeling with crossed random effects for subjects and items. Journal of Memory and Language. 2008; 59:390–412.

Baron-Cohen, S. Mindblindness: An essay on autism and theory of mind. MIT press; 1995.

Bartneck, C. The Godspeed Questionnaire Series. 2008. Retrieved from http://www.bartneck.de/2008/03/11/the-godspeed-questionnaire-series/

Bartneck C, Kuli D, Croft E, Zoghbi S. Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots. International Journal of Social Robotics. 2009; 1:71–81.

Bartz J, Zaki J, Bolger N, Ochsner K. Social effects of oxytocin in humans: context and person matter. Trends in Cognitive Sciences. 2011; 15:301–309. [PubMed: 21696997]

Bates, D., Maechler, M., Bolker, B. R package version 0.999999.0. Vienna, Austria: R Foundation; 2012. lme4: Linear mixed-effects models using S4 classes.

Baumgartner T, Heinrichs M, Vonlanthen A, Fischbacher U, Fehr E. Oxytocin shapes the neural circuitry of trust and trust adaptation in humans. Neuron. 2008; 58:639–650. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/18498743. [PubMed: 18498743]

Berg J, Dickhaut J, McCabe K. Trust, Reciprocity and Social History. Games and Economic Behavior. 1995; 10:122–142. http://doi.org/10.1006/game.1995.1027.

Bethlehem RAI, Baron-Cohen S, van Honk J, Auyeung B, Bos PA. The oxytocin paradox. Frontiers in Behavioral Neuroscience. 2014; 8:1–5. http://doi.org/10.3389/fnbeh.2014.00048. [PubMed: 24478648]

Bethlehem RAI, van Honk J, Auyeung B, Baron-Cohen S. Oxytocin, brain physiology, and functional connectivity: A review of intranasal oxytocin fMRI studies. Psychoneuroendocrinology. 2013; 38:962–974. [PubMed: 23159011]

Bickmore T, Cassell J. Relational agents: a model and implementation of building user trust. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. 2001:396–403.

Blascovich J, Loomis J, Beall A, Swinth K, Hoyt C, Bailenson J. Immersive virtual environment technology as a methodological tool for social psychology. Psychological Inquiry. 2002; 13:103–124.

Born J, Lange T, Kern W, McGregor GP, Bickel U, Fehm HL. Sniffing neuropeptides: a transnasal approach to the human brain. Nature Neuroscience. 2002; 5:514–516. http://doi.org/10.1038/nn0602-849. [PubMed: 11992114]

Brainbridge W, Hart J, Kim E, Scassellati B. The benefits of interactions with physically present robots over video-displayed agents. International Journal of Social Robotics. 2011; 3:41–52. http://doi.org/DOI10.1007/s12369-010-0082-7.

Bryk AS, Raudenbush SW. Application of hierarchical linear models to assessing change. Psychological Bulletin. 1987; 101:147–158.

Campbell A. Oxytocin and human social behavior. Personality and Social Psychology Review. 2010; 14:281–295. [PubMed: 20435805]

Carter CS. Oxytocin pathways and the evolution of human behavior. Annual Review of Psychology. 2014; 65:17–39. http://doi.org/10.1146/annurev-psych-010213-115110.

Cassell J, Bickmore T. Negotiated collusion: Modeling social language and its relationship effects in intelligent agents. User Modeling and User-Adapted Interaction. 2003; 13(2):89–132.

Cohen J. A power primer. Psychological Bulletin. 1992; 112:155–159. http://doi.org/10.1038/141613a0. [PubMed: 19565683]

Costa, P., McCrae, R. Revised NEO Personality Inventory (NEO PI-R) and NEO Five-Factor Inventory (NEO-FFI) professional manual. Odessa, FL: Psychological Assessment Resources, Inc; 1992.

Crockford C, Deschner T, Ziegler T, Wittig R. Endogenous peripheral oxytocin measures can give insight into the dynamics of social relationships: a review. Frontiers in Behavioral Neuroscience. 2014; 8:1–14. [PubMed: 24478648]

Cronbach, L. Occasional paper. Stanford Evaluation Consortium; Stanford, CA: 1976. Research in classrooms and schools: Formulation of questions, designs and analysis.

Davis MH. Measuring individual differences in empathy: Evidence for a multidimensional approach. Journal of Personality and Social Psychology. 1983; 44:113–126. http://doi.org/10.1037/0022-3514.44.1.113.

De Dreu CKW, Kret M. Oxytocin Conditions Intergroup Relations Through Upregulated In-Group Empathy, Cooperation, Conformity, and Defense. Biological Psychiatry. 2016; 79:165–173. http://doi.org/10.1016/j.biopsych.2015.03.020. [PubMed: 25908497]

de Melo C, Gratch J, Carnevale P. Humans vs. Computers: impact of emotion expressions on people's decision making. IEEE Transactions on Affective Computing. 2014; 6(2):127–136. Retrieved from http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=6853335.

de Visser EJ, Krueger F, McKnight P, Scheid S, Smith M, Chalk S, Parasuraman R. The World is not Enough: Trust in Cognitive Agents. Proceedings of the Human Factors and Ergonomics Society Annual Meeting. 2012; 56:263–267. http://doi.org/10.1177/1071181312561062.

de Visser EJ, Monfort SS, Mckendrick R, Smith MAB, McKnight P, Krueger F, Parasuraman R. Almost Human: Anthropomorphism Increases Trust Resilience in Cognitive Agents. Journal of Experimental Psychology: Applied. 2016; 22:331–349. http://doi.org/http://dx.doi.org/10.1037/xap0000092. [PubMed: 27505048]

de Visser EJ, Parasuraman R. Adaptive Aiding of Human-Robot Teaming: Effects of Imperfect Automation on Performance, Trust, and Workload. Journal of Cognitive Engineering and Decision Making. 2011; 5:209–231. http://doi.org/10.1177/1555343411410160.

de Vries P, Midden C, Bouwhuis D. The effects of errors on system trust, self-confidence, and the allocation of control in route planning. International Journal of Human-Computer Studies. 2003; 58:719–735. http://doi.org/10.1016/S1071-5819(03)00039-9.

Delgado M, Frank R, Phelps E. Perceptions of moral character modulate the neural systems of reward during the trust game. Nature Neuroscience. 2005; 8:1611–1618. [PubMed: 16222226]

Dennett, DC. The intentional stance. Chicago, IL: MIT Press; 1989.

Ditzen B, Schaer M, Gabriel B, Bodenmann G, Ehlert U, Heinrichs M. Intranasal oxytocin increases positive communication and reduces cortisol levels during couple conflict. Biological Psychiatry. 2009; 65:728–731. http://doi.org/10.1016/j.biopsych.2008.10.011. [PubMed: 19027101]

Ebstein RP, Israel S, Chew SH, Zhong S, Knafo A. Genetics of human social behavior. Neuron. 2010; 65:831–844. [PubMed: 20346758]

Epley N, Waytz A, Cacioppo J. On seeing human: A three-factor theory of anthropomorphism. Psychological Review. 2007; 114:864–886. [PubMed: 17907867]

Fletcher P. Other minds in the brain: a functional imaging study of "theory of mind" in story comprehension. Cognition. 1995; 57:109–128. http://doi.org/10.1016/0010-0277(95)00692-R. [PubMed: 8556839]

Fox J, Ahn SJ, Janssen JH, Yeykelis L, Segovia KY, Bailenson JN. Avatars Versus Agents: A Meta-Analysis Quantifying the Effect of Agency on Social Influence. Human-Computer Interaction. 2014; 30(5):37–41. http://doi.org/10.1080/07370024.2014.921494.

Gamer M, Zurowski B, Büchel C. Different amygdala subregions mediate valence-related and attentional effects of oxytocin in humans. Proceedings of the National Academy of Sciences of the United States of America. 2010; 107:9400–9405. http://doi.org/10.1073/pnas.1100561108. [PubMed: 20421469]

Garreau, J. Bots on The Ground. 2007. Retrieved September 23, 2010, from http://www.washingtonpost.com/wp-dyn/content/article/2007/05/05/AR2007050501009_pf.html

Goodyear K, Lee MR, O'Hara M, Chernyak S, Walter H, Parasuraman R, Krueger F. Oxytocin influences intuitions about the relationship between free will and moral responsibility. Social Neuroscience. 2016; 11:88–96. [PubMed: 25916658]

Gratch, J., Wang, N., Gerten, J., Fast, E., Duffy, R. Intelligent Virtual Agents. Springer; Berlin Heidelberg: 2007. Creating rapport with virtual agents; p. 125-138.

Gray HM, Gray K, Wegner DM. Dimensions of Mind Perception. Science. 2007; 315:619–619. http://doi.org/10.1126/science.1134475. [PubMed: 17272713]

Griffin DW, Bartholomew K. Models of the self and other: Fundamental dimensions underlying measures of adult attachment. Journal of Personality and Social Psychology. 1994; 67:430–445. http://doi.org/10.1037/0022-3514.67.3.430.

Grillon C, Krimsky M, Charney DR, Vytal K, Ernst M, Cornwell B. Oxytocin increases anxiety to unpredictable threat. Molecular Psychiatry. 2013; 18:958–960. http://doi.org/10.1038/mp.2012.156. [PubMed: 23147382]

Hancock P, Billings DR, Schaefer KE, Chen JY, de Visser EJ, Parasuraman R. A meta-analysis of factors affecting trust in human-robot interaction. Human Factors. 2011; 53:517–527. [PubMed: 22046724]

Hayes, C., Miller, C. Human-Computer Etiquette: Cultural Expectations and the Design Implications They Place on Computers and Technology. Boca Raton, Florida: Auerbach Publications; 2011.

Heinrichs M, von Dawans B, Domes G. Oxytocin, vasopressin, and human social behavior. Frontiers in Neuroendocrinology. 2009; 30:548–557. http://doi.org/10.1016/j.yfrne.2009.05.005. [PubMed: 19505497]

Hoff K, Bashir M. Trust in Automation: Integrating Empirical Evidence on Factors That Influence Trust. Human Factors. 2015; 57:407–434. http://doi.org/10.1177/0018720814547570. [PubMed: 25875432]

Kahn, PH., Severson, RL., Kanda, T., Ishiguro, H., Gill, BT., Ruckert, JH., Freier, NG. Do people hold a humanoid robot morally accountable for the harm it causes?. Proceedings of the Seventh Annual ACM/IEEE International Conference on Human-Robot Interaction -HRI '12.33. 2012. http://doi.org/10.1145/2157689.2157696

Kanat M, Heinrichs M, Domes G. Oxytocin and the social brain: neural mechanisms and perspectives in human research. Brain Research. 2014; 1580:160–171. http://doi.org/10.1016/j.brainres.2013.11.003. [PubMed: 24216134]

Kosfeld M, Heinrichs M, Zak PJ, Fischbacher U, Fehr E. Oxytocin increases trust in humans. Nature. 2005; 435:673–676. [PubMed: 15931222]

Krach S, Hegel F, Wrede B, Sagerer G, Binkofski F, Kircher T. Can machines think? Interaction and perspective taking with robots investigated via fMRI. PloS One. 2008; 3(7):e2597. http://doi.org/10.1371/journal.pone.0002597. [PubMed: 18612463]

Kreft IG, de Leeuw J, Aiken LS. The Effect of Different Forms of Centering in Hierarchical Linear Models. Multivariate Behavioral Research. 1995; 30:1–21. http://doi.org/10.1207/s15327906mbr3001_1. [PubMed: 26828342]

Kurzweil, R. The singularity is near: When humans transcend biology. New York, NY: The Viking Press; 2005.

Laird NM, Ware JH. Random-effects models for longitudinal data. Biometrics. 1982; 38:963–974. [PubMed: 7168798]

Lee HJ, Macbeth AH, Pagani JH, Young WS. Oxytocin: the great facilitator of life. Progress in Neurobiology. 2009; 88:127–151. [PubMed: 19482229]

Lee J, Moray N. Trust, control strategies and allocation of function in human-machine systems. Ergonomics. 1992; 35:1243–1270. [PubMed: 1516577]

Lee J, See K. Trust in automation : Designing for appropriate reliance. Human Factors. 2004; 46:50–80. [PubMed: 15151155]

Leite I, Pereira A, Mascarenhas S, Martinho C, Prada R, Paiva A. The influence of empathy in human-robot relations. International Journal of Human-Computer Studies. 2013; 71(3):250–260. http://doi.org/10.1016/j.ijhcs.2012.09.005.

Lewandowsky S, Mundy M, Tan GPA. The dynamics of trust: Comparing humans to automation. Journal of Experimental Psychology: Applied. 2000; 6:104–123. [PubMed: 10937315]

Li, J., Ju, W., Reeves, B. Touching a Mechanical Body: Tactile Contact With Intimate Parts of a Human-Shaped Robot is Physiologically Arousing. 66th Annual International Communication Association Conference; Fukuoka, Japan. 2016.

Lucas GM, Gratch J, King A, Morency LP. It's only a computer: Virtual humans increase willingness to disclose. Computers in Human Behavior. 2014; 37:94–100. http://doi.org/10.1016/j.chb.2014.04.043.

Madhavan P, Wiegmann D. Similarities and differences between human-human and human-automation trust: an integrative review. Theoretical Issues in Ergonomics Science. 2007; 8:277–301. http://doi.org/10.1080/14639220500337708.

Madhavan P, Wiegmann D, Lacson F. Automation Failures on Tasks Easily Performed by Operators Undermine Trust in Automated Aids. Human Factors. 2006; 48:241–256. http://doi.org/10.1518/001872006777724408. [PubMed: 16884046]

Martini MC, Gonzalez CA, Wiese E. Seeing minds in others – Can agents with robotic appearance have human-like preferences? PLoS ONE. 2016; 11(1):1–23. http://doi.org/10.1371/journal.pone.0146310.

Mayer R, Davis J, Schoorman F. An integrative model of organizational trust. Academy of Management Review. 1995; 20:709–734.

Meadows, M. We, Robot: Skywalker's Hand, Blade Runners, Iron Man, Slutbots, and How Fiction Became Fact. Lyon, France: Lyons Press; 2010.

Merritt S. Affective Processes in Human-Automation Interactions. Human Factors. 2011; 53:356–370. http://doi.org/10.1177/0018720811411912. [PubMed: 21901933]

Merritt S, Heimbaugh H, LaChapell J, Lee D. I Trust It, But I Don't Know Why: Effects of Implicit Attitudes Toward Automation on Trust in an Automated System. Human Factors. 2013; 55:520–534. http://doi.org/10.1177/0018720812465081. [PubMed: 23829027]

Merritt S, Ilgen DR. Not All Trust Is Created Equal: Dispositional and History-Based Trust in Human-Automation Interactions. Human Factors. 2008; 50:194–210. http://doi.org/10.1518/001872008X288574. [PubMed: 18516832]

Merritt S, Lee D, Unnerstall JL, Huber K. Are Well-Calibrated Users Effective Users? Associations Between Calibration of Trust and Performance on an Automation-Aided Task. Human Factors. 2015; 57:34–47. http://doi.org/10.1177/0018720814561675. [PubMed: 25790569]

Meyer-Lindenberg A, Domes G, Kirsch P, Heinrichs M. Oxytocin and vasopressin in the human brain: social neuropeptides for translational medicine. Nature Reviews Neuroscience. 2011; 12:524–538. [PubMed: 21852800]

Mikolajczak M, Gross JJ, Lane A, Corneille O, de Timary P, Luminet O. Oxytocin Makes People Trusting, Not Gullible. Psychological Science. 2010; 21:1072–1074. [PubMed: 20631321]

Morey RD. Confidence Intervals from Normalized Data: A correction to Cousineau (2005). Tutorials in Quantitative Methods for Psychology. 2008; 4(2):61–64. http://doi.org/10.3758/s13414-012-0291-2.

Muralidharan, L., de Visser, EJ., Parasuraman, R. Proceedings of the extended abstracts of the 32nd annual ACM conference on Human factors in computing systems – CHI EA '14. New York, New York, USA: ACM Press; 2014. The effects of pitch contour and flanging on trust in speaking cognitive agents; p. 2167-2172. http://doi.org/10.1145/2559206.2581231

Nass C, Fogg B, Moon Y. Can computers be teammates? International Journal of Human-Computer Studies. 1996; 45:669–678.

Nass C, Moon Y. Machines and mindlessness: Social responses to computers. Journal of Social Issues. 2000; 56:81–103.

Nass, C., Steuer, J., Tauber, E. CHI '94 Proceedings of the SIGCHI conference on Human factors in computing systems. Boston, MA: ACM; 1994. Computers are social actors; p. 73-78.

Neumann ID, Maloumby R, Beiderbeck DI, Lukas M, Landgraf R. Increased brain and plasma oxytocin after nasal and peripheral administration in rats and mice. Psychoneuroendocrinology. 2013; 38:1985–1993. http://doi.org/10.1016/j.psyneuen.2013.03.003. [PubMed: 23579082]

Nof, SY. Springer handbook of automation. Springer Science & Business Media; 2009.

Paccagnella O. Centering or not centering in multilevel models? The role of the group mean and the assessment of group effects. Evaluation Review. 2006; 30:66–85. http://doi.org/10.1177/0193841X05275649. [PubMed: 16394187]

Pak R, Fink N, Price M, Bass B, Sturre L. Decision support aids with anthropomorphic characteristics influence trust and performance in younger and older adults. Ergonomics. 2012; 55:1059–72. http://doi.org/10.1080/00140139.2012.691554. [PubMed: 22799560]

Pak R, McLaughlin AC, Bass B. A multi-level analysis of the effects of age and gender stereotypes on trust in anthropomorphic technology by younger and older adults. Ergonomics. 2014; 57:1277–89. http://doi.org/10.1080/00140139.2014.928750. [PubMed: 24935771]

Parasuraman R. Neuroergonomics: Research and practice. Theoretical Issues in Ergonomics Science. 2003; 4:5–20. http://doi.org/10.1080/14639220210199753.

Parasuraman R. Neuroergonomics: Brain, Cognition, and Performance at Work. Current Directions in Psychological Science. 2011; 20:181–186. http://doi.org/10.1177/0963721411409176.

Parasuraman R, Cosenzo KA, de Visser EJ. Adaptive automation for human supervision of multiple uninhabited vehicles: Effects on change detection, situation awareness, and mental workload. Military Psychology. 2009; 21:270–297.

Parasuraman R, Miller C. Trust and etiquette in high-criticality automated systems. Communications of the ACM. 2004; 47:51–55.

Parasuraman R, Riley V. Humans and automation: Use, misuse, disuse, abuse. Human Factors. 1997; 39:230–253.

Parasuraman R, Sheridan T, Wickens C. A model for types and levels of human interaction with automation. IEEE Transactions on Systems, Man and Cybernetics, Part A: Systems and Humans. 2000; 30:286–297.

Parasuraman R, Sheridan T, Wickens C. Situation awareness, mental workload, and trust in automation: Viable, empirically supported cognitive engineering constructs. Journal of Cognitive Engineering and Decision Making. 2008; 2:140–160.

Qiu L, Benbasat I. A study of demographic embodiments of product recommendation agents in electronic commerce. International Journal of Human-Computer Studies. 2010; 68:669–688. http://doi.org/10.1016/j.ijhcs.2010.05.005.

R Core Team, R. R: A language and environment for statistical computing. R Foundation for Statistical Computing; Vienna, Austria: 2013.

Reeves, B., Nass, C. The media equation: how people treat computers, television, and new media like real people and places. Cambridge, MA: Cambridge University Press; 1996.

Sandal GM, Leon GR, Palinkas L. Human challenges in polar and space environments. Reviews in Environmental Science and Bio/Technology. 2006; 5:281–296.

Satterthwaite FE. An approximate distribution of estimates of variance components. Biometrics Bulletin. 1946:110–114. [PubMed: 20287815]

Seltzer LJ, Prososki AR, Ziegler TE, Pollak SD. Instant messages vs. speech: hormones and why we still need to hear each other. Evolution and Human Behavior. 2012; 33:42–45. [PubMed: 22337755]

Sheridan, TB., Verplank, WL. Human and computer control of undersea teleoperators. Cambridge, MA: 1978.

Smith MA, Allaham MM, Wiese E. Trust in Automated Agents is Modulated by the Combined Influence of Agent and Task Type. Proceedings of the Human Factors and Ergonomics Society Annual Meeting. 2016; 60:206–210. http://doi.org/10.1177/1541931213601046.

Spielberger, CD., Gorsuch, RL., Lushene, R., Vagg, PR., Jacobs, GA. Manual for the State-Trait Anxiety Inventory. Palo Alto, CA: Consulting Psychologists Press; 1983.

Stokes, CK., Lyons, JB., Littlejohn, K., Natarian, J., Case, E., Speranza, N. 2010 International Symposium on Collaborative Technologies and Systems. IEEE; 2010. Accounting for the human in cyberspace: Effects of mood on trust in automation; p. 180-187.http://doi.org/10.1109/CTS.2010.5478512

Striepens N, Kendrick KM, Hanking V, Landgraf R, Wüllner U, Maier W, Hurlemann R. Elevated cerebrospinal fluid and blood concentrations of oxytocin following its intranasal administration in humans. Scientific Reports. 2013; 3(3440):1–5. http://doi.org/10.1038/srep03440.

Sung, JY., Guo, L., Grinter, RE., Christensen, HI. "My Roomba Is Rambo": Intimate Home Appliances. In: Krumm, J.Abowd, GD.Seneviratne, A., Strang, T., editors. UbiComp 2007: Ubiquitous Computing: Lecture Notes in Computer Science. Vol. 4717. Berlin, Heidelberg: Springer Berlin Heidelberg; 2007. http://doi.org/10.1007/978-3-540-74853-3

Szalma J. Individual differences in human-technology interaction: incorporating variation in human characteristics into human factors and ergonomics research and design. Theoretical Issues in Ergonomics Science. 2009; 10:381–397. http://doi.org/10.1080/14639220902893613.

Szalma J. On the Application of Motivation Theory to Human Factors/Ergonomics: Motivational Design Principles for Human-Technology Interaction. Human Factors. 2014; 56:1453–1471. http://doi.org/10.1177/0018720814553471. [PubMed: 25509824]

Szalma J, Taylor G. Individual differences in response to automation- The big five factors of personality. Journal of Experimental Psychology: Applied. 2011; 17:71–96. [PubMed: 21688932]

Taylor SE. Tend and Befriend: Biobehavioral Bases of Affiliation Under Stress. Current Directions in Psychological Science. 2006; 15:273–277.

Theodoridou A, Rowe A, Penton-Voak I, Rogers P. Oxytocin and social perception: Oxytocin increases perceived facial trustworthiness and attractiveness. Hormones and Behavior. 2009; 56:128–132. [PubMed: 19344725]

Thompson JC, Trafton JG, McKnight P. The perception of humanness from the movements of synthetic agents. Perception. 2011; 40:695–704. http://doi.org/10.1068/p6900. [PubMed: 21936298]

Traum, D., Rickel, J., Gratch, J., Marsella, S. Proceedings of the second international joint conference on Autonomous agents and multiagent systems. Melbourne, Australia: ACM; 2003. Negotiation over tasks in hybrid human-agent teams for simulation-based training; p. 441-448.

van den Brule R, Dotsch R, Bijlstra G, Wigboldus DHJ, Haselager P. Do Robot Performance and Behavioral Style affect Human Trust? International Journal of Social Robotics. 2014; 6:519–531. http://doi.org/10.1007/s12369-014-0231-5.

van Dongen K, van Maanen PP. A framework for explaining reliance on decision aids. International Journal of Human-Computer Studies. 2013; 71:410–424. http://doi.org/10.1016/j.ijhcs.2012.10.018.

Van Ijzendoorn MH, Bakermans-Kranenburg MJ. A sniff of trust: Meta-analysis of the effects of intranasal oxytocin administration on face recognition, trust to in-group, and trust to out-group. Psychoneuroendocrinology. 2012; 37:438–443. [PubMed: 21802859]

Veening JG, Olivier B. Intranasal administration of oxytocin: Behavioral and clinical effects, a review. Neuroscience and Biobehavioral Reviews. 2013; 37:1445–1465. [PubMed: 23648680]

Watson D, Clark LA, Tellegen A. Development and validation of brief measures of positive and negative affect: the PANAS scales. Journal of Personality and Social Psychology. 1988; 54:1063–1070. [PubMed: 3397865]

Weiner, B. An Attribution Theory of Motivation. In: Van Lange, PAM.Kruglanski, AW., Higgins, ET., editors. Handbook of theories of social psychology. Los Angeles, CA: SAGE Publications; 2012. p. 135-155.

Wickens CD, Dixon SR. The benefits of imperfect diagnostic automation: a synthesis of the literature. Theoretical Issues in Ergonomics Science. 2007; 8:201–212. http://doi.org/10.1080/14639220500370105.

Wiese E, Wykowska A, Zwickel J, Müller HJ. I see what you mean: how attentional selection is shaped by ascribing intentions to others. PloS One. 2012; 7(9):e45391. http://doi.org/10.1371/journal.pone.0045391. [PubMed: 23049794]

Wu, P., Suite, NA., Morie, J., Wall, P., Chance, E., Ladwig, J., Ott, T. Maintaining Psycho-Social Health on the Way to Mars and Back. Virtual Reality International Conference (VRIC) Laval; France. 2015.

Zak PJ, Kurzban R, Matzner WT. The neurobiology of trust. Annals of the New York Academy of Sciences. 2004; 1032:224–227. http://doi.org/10.1196/annals.1314.025. [PubMed: 15677415]

Zak PJ, Kurzban R, Matzner WT. Oxytocin is associated with human trustworthiness. Hormones and Behavior. 2005; 48:522–527. http://doi.org/10.1016/j.yhbeh.2005.07.009. [PubMed: 16109416]

Zak PJ, Stanton AA, Ahmadi S. Oxytocin increases generosity in humans. PloS One. 2007; 2(11):e1128. http://doi.org/10.1371/journal.pone.0001128. [PubMed: 17987115]

Zink C, Meyer-Lindenberg A. Human neuroimaging of oxytocin and vasopressin in social cognition. Hormones & Behavior. 2012; 61:400–409. [PubMed: 22326707]

Key points

- As the nature of social interactions between humans and machines becomes more complex, a greater understanding of the role of machine anthropomorphism is needed

- The peptide oxytocin can be used as a probe to determine the *active ingredients*, that is the human-like characteristics of an agent, that have the most significant impact to promoting pro-sociality such as trust

- This study provided the first evidence that administration of exogenous oxytocin affects trust, compliance, and team decision making with automated agents, in particular one with intermediate levels of anthropomorphism.

- Designing automation to mimic basic human characteristics is sufficient to elicit behavioral trust outcomes that are driven by neurological processes typically observed in human-human interactions.

- Designers of automated systems should consider the task, the individual, and the level of anthropomorphism to achieve the desired outcome.
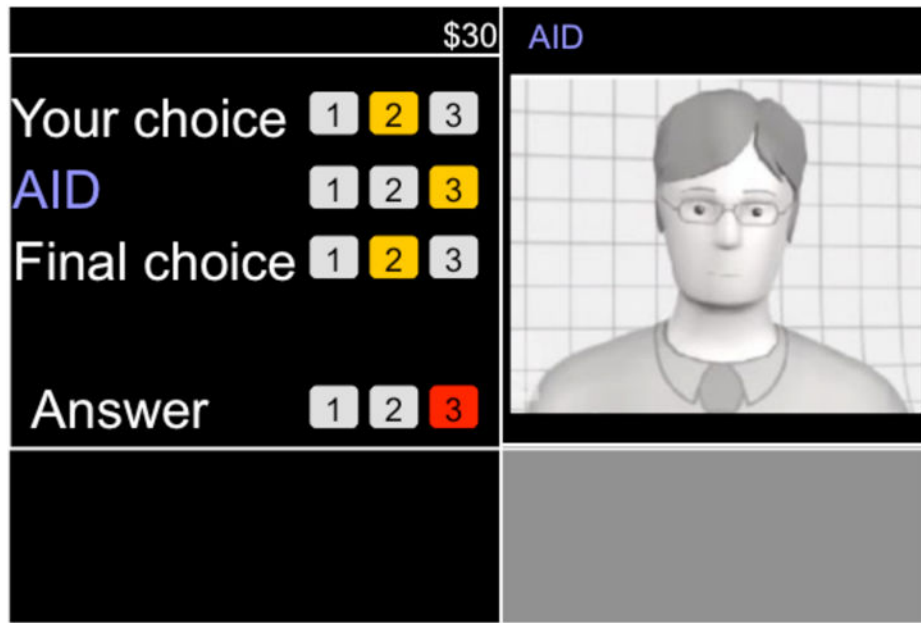
Figure 1.
The TNO Trust Task (T^3) with a five-step task sequence including 1) selecting a number in a sequence, 2) watching an agent video, 3) observing the agent recommended number, 4) make a final number choice, and 5) observing the correct answer.
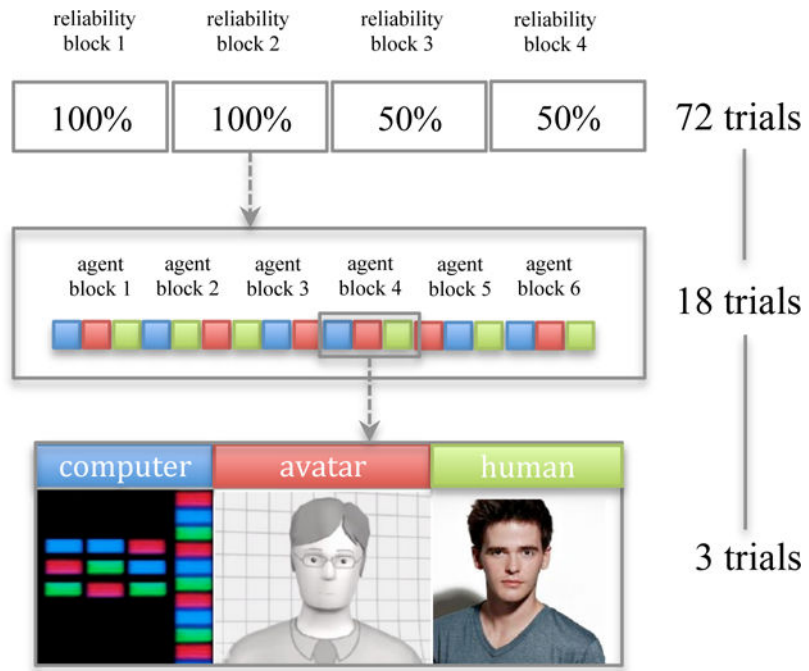
Figure 2.
Block design for the experiment with 3 trials per agent block, 18 trials per reliability block, and 72 trials total.
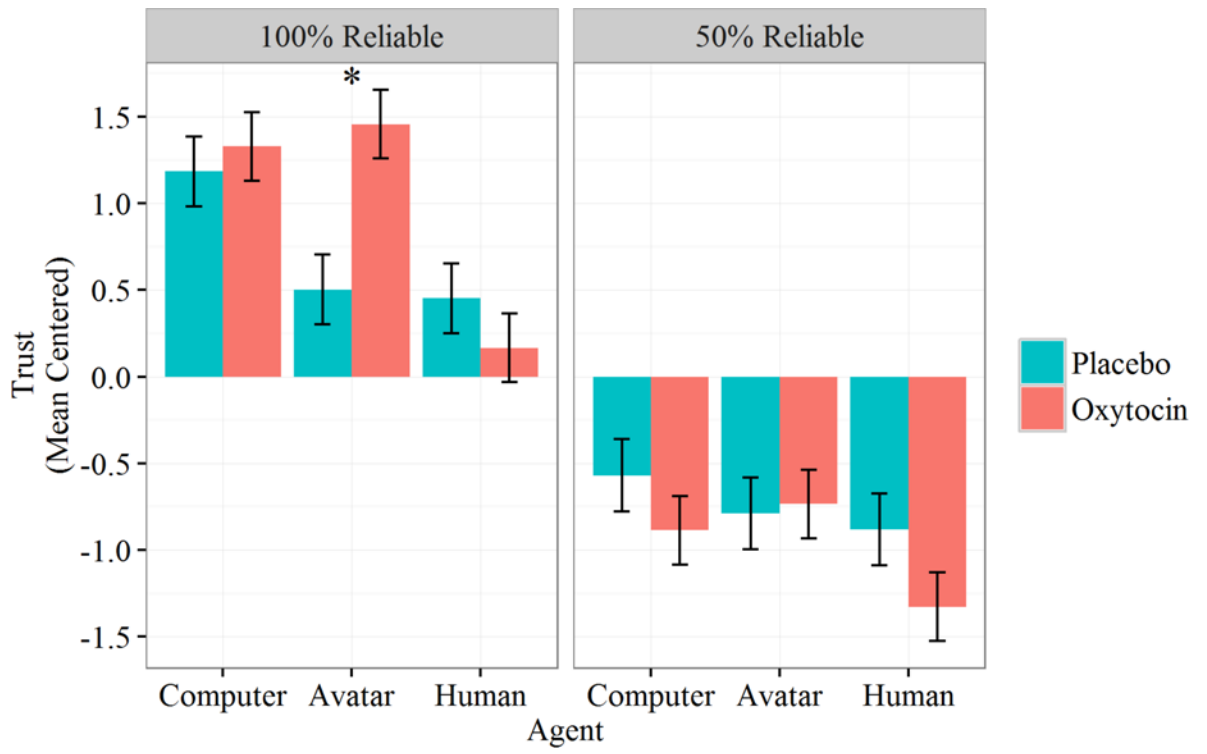
Figure 3.
Mean centered trust by Agent, Drug Condition, and Reliability. Error bars represent standard
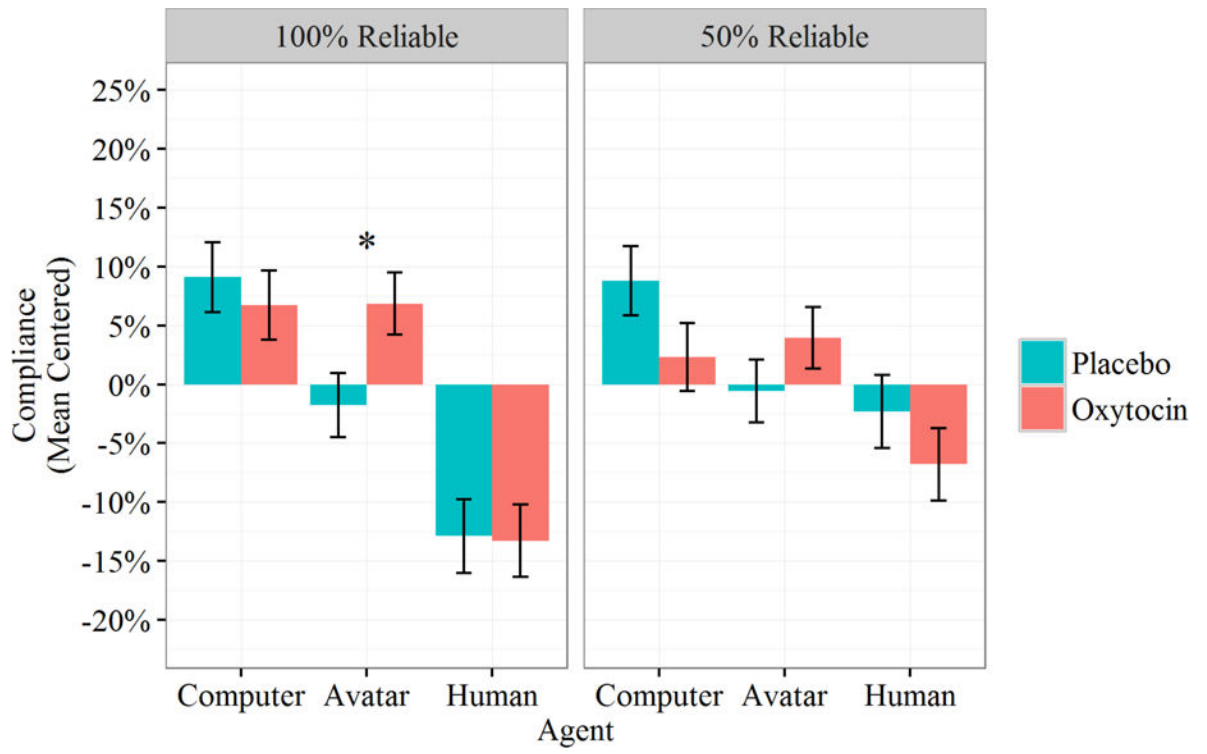errors; asterisk denotes significant difference between conditions.

Figure 4.
Mean centered compliance by Agent, Drug Condition, and Reliability. Error bars represent standard errors; asterisk denotes significant difference between conditions.
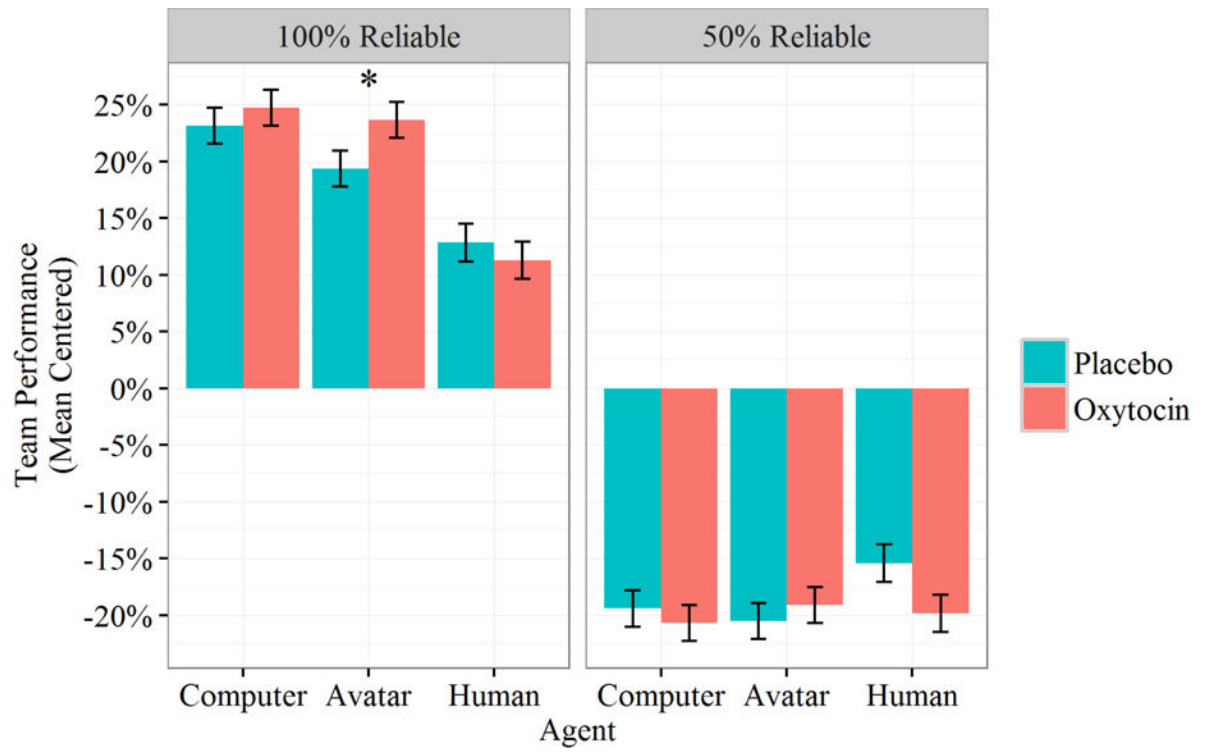
Figure 5.
Mean centered team performance by Agent, Drug Condition, and Reliability. Error bars represent standard errors; asterisk denotes significant difference between conditions.
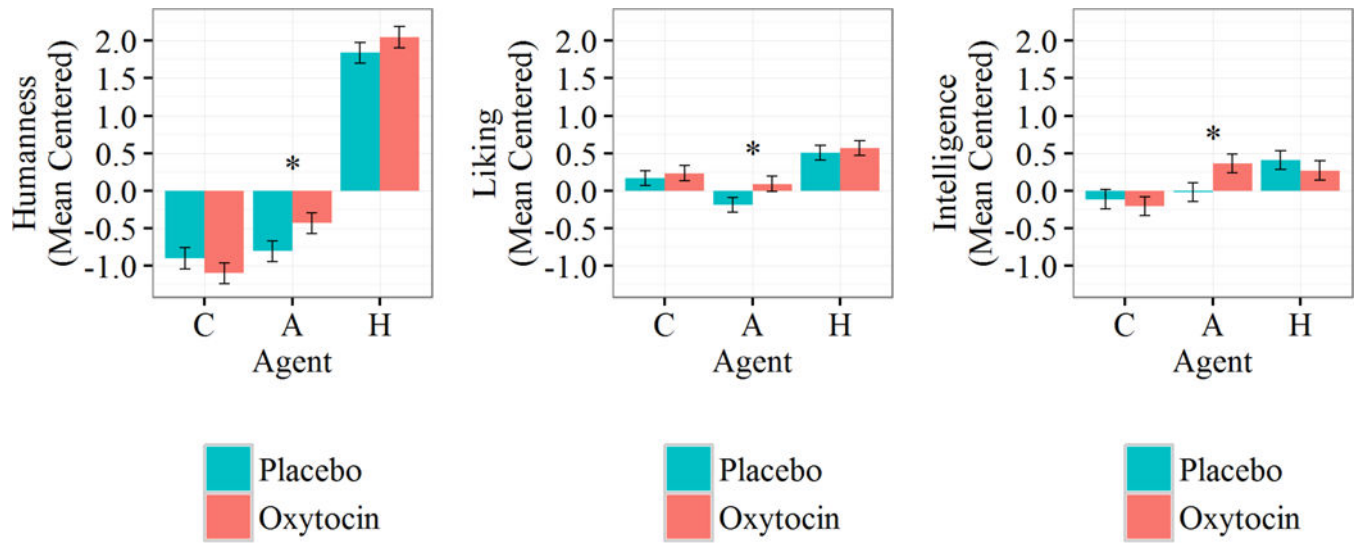
Figure 6.
Mean centered Godspeed ratings for humanness (left), liking (middle), and intelligence (right) by Agent and Drug Condition. Error bars represent standard errors; asterisks denote significant difference between conditions; C = computer, A = avatar; H = human.