Autonomous Agents and Human Interpersonal Trust: Can We Engineer a Human-Machine Social Interface for Trust?

David J. Atkinson and **Micah H. Clark**
Florida Institute for Human & Machine Cognition (IHMC)
15 SE Osceola Avenue, Ocala, FL 34471 USA
{*datkinson*, *mclark*} @*ihmc.us*

Abstract

There is a recognized need to employ autonomous agents in domains that are not amenable to conventional automation and/or which humans find difficult, dangerous, or undesirable to perform. These include time-critical and mission-critical applications in health, defense, transportation, and industry, where the consequences of failure can be catastrophic. A prerequisite for such applications is the establishment of well-calibrated trust in autonomous agents. Our focus is specifically on human-machine trust in deployment and operations of autonomous agents, whether they are embodied in cyber-physical systems, robots, or exist only in the cyber-realm. The overall aim of our research is to investigate methods for autonomous agents to foster, manage, and maintain an appropriate trust relationship with human partners when engaged in joint, mutually interdependent activities. Our approach is grounded in a systems-level view of humans and autonomous agents as components in (one or more) encompassing meta-cognitive systems. Given human predisposition for social interaction, we look to the multi-disciplinary body of research on human interpersonal trust as a basis from which we specify engineering requirements for the interface between human and autonomous agents. If we make good progress in reverse engineering this "human social interface," it will be a significant step towards devising the algorithms and tests necessary for trustworthy and trustable autonomous agents. This paper introduces our program of research and reports on recent progress.

Background

There is a recognized need to employ autonomous agents in domains that are not amenable to conventional automation and/or which humans find difficult, dangerous, or otherwise undesirable to perform (Takayama, Ju, and Nass 2008). These include time-critical and mission-critical applications in health, defense (USAF 2010), transportation (Wing 2008), and industry (Bekey et al. 2006), where the consequences of failure can be catastrophic. Trust in autonomous agents is indeed a very formidable problem, especially when we are tasking agents with difficult, high impact, time- and mission-critical functions. After all, even the best humans sometimes fail in challenging, dynamic, and adversarial environments despite the best training and testing possible. Awareness is growing of the technical and psychological hurdles for establishing confidence and maintaining trust in autonomous agents across the system life cycle, especially when those agents are capable of self-adaptation, optimization and learning. These issues have been cited as serious obstacles to larger scale use of autonomy technology (USAF 2010). Reliance on autonomous agents necessitates calibrated trust, that is, human trust judgments that reflect the objective capabilities of the system and utility in a given situation (Parasuraman and Riley 1997; Lee and See 2004).

We observe that physical and cultural evolution has provided humans with an efficacious ability to judge the trustworthiness of each other and to make good decisions in dynamic and uncertain situations based on that trust. There is a vast body of knowledge in the social sciences regarding the nature of human interpersonal trust, and from multiple disciplines regarding human-machine interaction and reliance.

This research supports the idea that the innate cognitive, emotional, and social predispositions of humans play a strong role in trust of automation (Lee and See 2004). We are predisposed to anthropomorphize and treat machines as social actors (Nass, Fogg, and Moon 1996). The social context and perceived role of actors affect human-machine trust (Wagner 2009; Groom et al. 2011). Individual personality traits, as well as affective state, can affect delegation to autonomous agents (Cramer et al. 2008; 2010; Stokes et al. 2010). Behavioral research has found that intuitive and affective processes create systematic biases and profoundly affect human trust, behavior, and choice (Weber, Malhotra, and Murnighan 2004; Dunn and Schweitzer 2005; Schoorman, Mayer, and Davis 2007; Stokes et al. 2010; Rogerson et al. 2011).

Turkle (2004; 2010) asserts that today's technology "push[es] our Darwinian buttons, exhibiting the kinds of behavior people associate with sentience, intentions, and emotions." As a result, humans readily attribute mental states to technology (Parlangeli, Chiantini, and Guidi 2012). As increasingly intelligent and capable autonomous agents interact with humans in ever more natural ("human-like") ways, perhaps even embodied as humanoid robots, this will increasingly evoke human social treatment (Schaefer, Billings,

and Hancock 2012; DeSteno et al. 2012).

Given human predisposition for anthropomorphizing and social interaction, it is reasonable to ask whether the concept of human interpersonal trust is an anthropomorphic concept that we should now consider applying to autonomous agents. Our answer is yes.

We are especially interested in autonomous agent application domains where task achievement requires interactivity and co-dependency between human and machine; that is, where humans and machines are partners in larger metacognitive systems (Johnson et al. 2011). A good example of this is the application of autonomous agents in decision support systems. Key processes in the "Data to Decision" domain are knowledge seeking, sharing, and transfer. Previous studies on trust in organizations have shown that interpersonal trust is an important factor in these processes (Mayer, Davis, and Schoorman 1995; Kramer and Tyler 1996; Rousseau et al. 1998). Trust increases the likelihood that newly acquired knowledge is usefully absorbed (Mayer, Davis, and Schoorman 1995; Srinivas 2000; Levin, Cross, and Abrams 2002). Optimal reliance of humans on autonomous agent-based decision support systems will occur only when there is appropriate, well-calibrated trust in the agent as a source of knowledge. Little work has been done on how to achieve this with autonomous agents, although it is beginning (Klein et al. 2004).

From a systems engineering point of view, the purpose of trust in a multi-agent system composed of human and machine elements is to achieve optimal overall performance via appropriate interdependency, mutual reliance, and appropriate exchange of initiative and control between the cognitive components (human and/or machine). Our central hypothesis is that the cognitive and affective nature of human interpersonal trust provides useful guidance for the design and development of autonomous agents that engender appropriate human-machine reliance and interdependency, specifically, via correct understanding and use of what we term the "human social interface."

Approach

Our approach is inspired by a social model of trust (Falcone and Castelfranchi 2001) wherein each agent, human or machine, has a need and intention to be reliant upon the other agent in joint activity, and this intention is a consequence of some structure of beliefs in a given task, role and situational context. Trust becomes manifest when there is action: some delegation of responsibility to the other agent. Conversely, as those beliefs change, intention may change and delegation revoked (Falcone and Castelfranchi 2001). We concur that trust is a dynamic process — a reciprocal relationship over time between two or more agents that requires periodic reassessment and maintenance (Lee and See 2004; Hoffman et al. 2009).

In this context, a good human social interface specification will describe assumptions about each agent, communicative signals and interaction protocols including how and when these are used given certain beliefs in specific (operational) contexts, and how the internal state of each agent is consequentially affected. The trust-relevant internal state of a human agent includes a structure of beliefs (Castelfranchi 2000; Levin and Cross 2004), and specific cognitive and affective reasoning processes involved in trust (McAllister 1995). Interaction and signaling along multiple channels and modes between agents conveys essential information that in turn modulates these belief structures (Semin and Marsman 1994; Pentland 2004; Stoltzman 2006; Pentland and Heibeck 2008). Situational factors strongly affect signaling, interaction, and ultimately, judgments regarding trust (Simpson 2007).

The initial focus of our work is on understanding, with an eye towards computational mechanisms, the structure of beliefs that are important to inter-agent trust, including what evidence is required, how it is acquired (e.g., observation, reasoning, reputation, certification, communication, signals, social norms and stereotypes), how credence in a belief is gained or lost, and how such change in the structure of beliefs affects inter-agent reliance and delegation.

Structure of Trust-Relevant Beliefs

What is the necessary and sufficient structure of beliefs required for trust? Such beliefs may cover a broad territory, but previous research suggest belief structures include causal factors, attitudes, evaluations and expectations centered around other agents (especially the potential "trustee"), the situation, goals, and tasks (Castelfranchi 2000; Levin, Cross, and Abrams 2002).

Models and beliefs that one agent has about the attitudes, motives, intentions, future behavior, et cetera of other agents that may differ from the agent's own constitute what is often called a "theory of mind" (Premack and Woodruff 1978; Carruthers and Smith 1996). For trust, two of the most important kinds of beliefs about another "trustee" agent concern that agent's competence (Mayer, Davis, and Schoorman 1995) and predictability (Marble et al. 2004; Feltovich et al. 2007). Other important beliefs center on integrity, benevolence, risk (aka "safety") and transparency (aka "openness") (Levin, Cross, and Abrams 2002).

To investigate belief structures, and the relative importance of different kinds of beliefs (e.g., those related to competence), we are conducting a two-phase experimental program consisting of survey research with follow-up laboratory experiments, including prototype autonomous agents for experimental testbeds.

The first part in our survey research has consisted of interviews with Subject Matter Experts (SME) in several domains with the purpose of quickly identifying the most salient trust-related beliefs for reliance on autonomous systems. The Robonaut robotic astronaut assistant (Ambrose et al. 2000) is a good example of a robot deployed in a life- and mission-critical domain where the addition of autonomous capabilities could yield significant benefits. While astronauts cited safety and predictability as key for their trust in Robonaut, surprisingly the developers said that "similarity" was what ultimately changed the astronauts' distrust into trust. Similarity in this case consisted simply of donning Robonaut in "team colors," i.e., a spacesuit. As for further SME examples: A doctor and a surgical technician,

both familiar with the Da Vinci system (Spinoglio et al. 2012) and other robotic surgical tools, cited predictability and competence as the most important traits they would rely upon in considering a deployment of an autonomous surgical robot in an operating room where delays or errors due to automation are costly and possibly life-threatening. And an automotive industry specialist who is currently involved in planning deployment of autonomous vehicle technologies said that "small, transparent competencies" are most important, as these traits enable incremental introduction of autonomous systems technologies. While our informal interviews echo what might be expected from review of the literature on trust, we note that there are differences of opinion according to role (e.g., developer, deployment decision-maker, user) and variations across application domains that need to be systematically explored with respect to autonomous agents.

The second part in our survey research involved development and administration of a broader, methodical on-line survey on attitudes towards autonomous agents. The survey is designed to elicit attitudes, opinions, and preferences that should shed further light on the belief structures important for trust of autonomous agents. In this survey, our focus is on factors related to perceived competence, predictability, openness, and judgment of risk. Once again, our target population consists of stakeholders and subject matter experts in autonomous agents — individuals involved with autonomous agents at various points in the system life cycle.

The survey is designed around seven hypothetical scenarios that require participants to choose whether to rely on an autonomous agent. The scenarios vary systematically in terms of the four factors cited above and are dilemmas that force the participant to weigh the relative importance of these factors. The survey also includes brief assessments of the participant's personality using short versions of standard personality instruments: Big Five Inventory (BFI), Innovation Inventory (II), and Domain-Specific Risk-Taking Scale (DOSPERT). We anticipate a systematic variation between relative preferences for competence and predictability correlated with personality measures, e.g., risk tolerance, openness to innovation, and participants' perception of risks in each scenario. At the time of this writing, data collection is beginning; the results will be discussed in a later publication.

Trust-Relevant Computational Mechanisms

Beyond understanding human-machine trust, our aim is to contribute to the development of computational mechanisms that enable autonomous agents to exercise the human social interface. Our desiderata for such agents include: (a) representational system rich enough to support a theory of mind (i.e., distinguishes the mental content of others from itself); (b) accurate declarative and procedural models sufficient to anticipate the effects of action on the trust relationship; (c) reasoning and planning capabilities that integrate trust-relevant knowledge/models into action; and (d) ability to reflect on, learn from, and individuate trust relationships based on ongoing experience. While these requirements are certainly ambitious, we do not think they are by any means impossible. Indeed, in keeping with our initial focus on the structure of trust-relevant beliefs, we have begun prototyping a representational system for codifying trust-relevant belief structures. The product of this effort will be a proof-of-concept platform for development and experimentation with trust-relevant computational cognitive models.

Briefly sketched, our prototype will employ the ViewGen (Ballim and Wilks 1991; Wilks 2011) representation paradigm wherein a (conceptual) tree of "topic environments" (collections of beliefs) and "viewpoints" (belief scoping) is used to represent an individual agent's beliefs about the world and about the beliefs of others. Default reasoning (usually via ascription) is used to minimize the necessity of explicitly storing beliefs and allows the system to approximate a doxastic modal logic while avoiding some of the computational complexity such logics usually entail. Then in similar fashion to (Bridewell and Isaac 2011), we will extend this representational scheme with other modalities (e.g., goals, intentions) as necessary to account for the multimodal structure of trust-relevant beliefs.

Our prototype will depart from the ViewGen paradigm with respect to the uniformity of inference. ViewGen traditionally assumes that an agent reasons about others' attitudes using the same methods with which the agent reasons about its own — that is to say, the methods are independent of the belief holder's identity. Like Clark (2010; 2011), we intend for the artificial agent to use different inference methods for reasoning over and about its own attitudes (using, e.g., normative models) versus reasoning about the attitudes of human others (using, e.g., predictive psychological models). Our justification is that while a trustable artificial agent needs to be informed by, anticipate, plan for, and react proactively to beliefs, intentions, and behaviors arising from natural human cognitive processes and their attendant biases (as revealed by social and cognitive psychology studies), there is little reason (and perhaps even great risk) for the machine to adopt these for itself.

Discussion

How much of our knowledge about human interpersonal trust is applicable to human interaction with autonomous agents? What are the significant differences and the consequent limitations, especially with respect to trust and the healthy interdependency that is necessary for effective human-agent teams?

A recent workshop explored these topics and related questions in detail (Atkinson, Friedland, and Lyons 2012). While there has been a long history of work on trust in the fields of psychology, sociology and others, participants from multiple disciplines agreed that far too little has been done to understand what those results mean for autonomous agents, much less how to extend them in computational terms to foster human-autonomous agent trust. That is a prime motivation for the research project we have described.

The human cognitive aspect of trust arises from our ability, based on various dimensions of commonality, to make reasonable inferences about the internal state of other agents (e.g., beliefs, dispositions, intentions) in order to predict future behavior and judge the risk versus benefit of delegation.

It is therefore crucial that autonomous agents not only correctly use the human social interface, but also provide reliable signals that are indicative of the agent's state. Such "honest" signals (Pentland and Heibeck 2008) are necessary for a human partner to construct a set of beliefs about the agent that accurately reflect the internal state of the agent.

However, we are mindful that trust between humans and autonomous agents is not likely to be equivalent to human interpersonal trust regardless of how "human-like" agents become in intelligence, social interaction, or physical form. Autonomous agents are not human, do not have our senses or reason as we do, and do not live in human society or share common human experience, culture, or biological heritage. These differences are potentially very significant for attribution of human-like internal states to autonomous agents. The innate and learned social predispositions and inferential short cuts that work so well for human interpersonal trust are likely to lead us astray in ascribing trustworthiness to autonomous agents insofar as our fundamental differences lead to misunderstanding and unexpected behavior. The foreseeable results could be miscommunication, errors of delegation, and inappropriate reliance.

Therefore what is needed are not only ways to measure, interpret, and accurately portray the internal state of autonomous agents, but to do so in terms that relate meaningfully (e.g., functionally) to the beliefs that humans find essential for judging trustworthiness. For example, how do we measure diligence (an important component of competence)? What does openness or transparency really mean with respect to an autonomous agent? How does an autonomous agent demonstrate its disposition and intentionality? These are key questions to answer, for without accurately communicating the internal state of an autonomous agent in a way that enables well-calibrated trust, we enter forewarned into an ethical and functional minefield (see, e.g., Bringsjord and Clark 2012) where the human social interface is a means for arbitrary manipulation and agent "trust inducing" behavior is dangerous and deceptive puppetry.

Conclusion & Future Research

Our aim is to enable autonomous agents to use the human social interface appropriately to provide humans with insight into an agent's state and thus enable reasonable and accurate judgments of agent trustworthiness. Ultimately, this means creating compatible algorithms that exercise the human social interface. Algorithmic techniques may include, for example, (a) modeling a human partner, (b) anticipating situations where trust will be a determinate factor, and (c) planning for and exchange of social signals through multi-modal channels of interaction with a human.

In this paper, we have introduced our research program on the applicability of human interpersonal trust to trust between humans and autonomous agents. We presented our exploratory survey designed to elicit attitudes towards autonomous systems in the context of several scenarios that challenge trust along one or more dimensions. The results of this survey will lead in the next stage of our research to experiments that we anticipate will begin to give us insight into how change in trust-related belief structures affects reliance and delegation to an autonomous agent. In particular, our planned experiments are aimed at understanding how manipulation of multimodal social signals (those perceived as evidence supporting trust-related beliefs) can be used to modulate trust and, specifically, contribute to an attribution of benevolence to an autonomous agent. A belief in benevolence in an autonomous agent is likely to be important for certain applications, such as urban search and rescue, where rapid acceptance of help from an autonomous agent may be life critical. One of the key questions we hope to explore in the near future is whether an attribution of benevolence requires the human to believe that the autonomous agent has volition, i.e., "a choice" in the matter (Kahn et al. 2007).

Finally, we discussed the necessity of developing ways to measure, interpret, and accurately portray the internal state of autonomous agents in terms that relate meaningfully to the belief structures that humans rely upon for judging trustworthiness. These methods will be essential for honest social behavior by autonomous agents, that is, not mere mimicry. We envision that such measurements and their methodology might also find good use in the development of design guidelines and requirements for trustworthy and trustable autonomous agents (but further discussion of this point is deferred to elsewhere).

Acknowledgments

This research is supported by the Air Force Office of Scientific Research under grant FA9550-12-1-0097 and by the Office of Naval Research under award number N00014-13-1-0225. The view presented in this paper is solely that of the authors and does not necessarily reflect the views of the Air Force Office of Scientific Research, the Office of Naval Research, or the United States Department of Defense.

References

Ambrose, R.; Aldridge, H.; Askew, R. S.; Burridge, R.; Bluethmann, W.; Diftler, M.; Lovchik, C.; Magruder, D.; and Rehnmark, F. 2000. Robonaut: NASA's Space Humanoid. *IEEE Intell. Syst.* 15(4):57–63.

Atkinson, D.; Friedland, P.; and Lyons, J. 2012. Human-Machine Trust for Robust Autonomous Systems. In *Proc. of the 4th IEEE Workshop on Human-Agent-Robot Teamwork*.

Ballim, A., and Wilks, Y. 1991. *Artificial Believers: The Ascription of Belief*. Lawrence Erlbaum.

Bekey, G.; Ambrose, R.; Kumar, V.; Sanderson, A.; Wilcox, B.; and Zheng, Y. 2006. WTEC Panel Report on International Assessment of Research and Development in Robotics. Technical report, World Technology Evaluation Center, Baltimore, MD.

Bridewell, W., and Isaac, A. 2011. Recognizing Deception: A Model of Dynamic Belief Attribution. In *Advances in Cognitive Systems: Papers from the AAAI Fall Symposium*, 50–57.

Bringsjord, S., and Clark, M. 2012. Red-Pill Robots Only, Please. *IEEE Trans. Affect Comput.* 3(4):394–397.

Carruthers, P., and Smith, P., eds. 1996. *Theories of theories of mind*. Cambridge, UK: Cambridge University Press.

Castelfranchi, C. 2000. Artificial liars: Why computers will (necessarily) deceive us and each other. *Ethics Inf. Technol.* 2(2):113–119.

Clark, M. 2010. *Cognitive Illusions and the Lying Machine: A Blueprint for Sophistic Mendacity*. Ph.D. Dissertation, Rensselaer Polytechnic Institute, Troy, NY.

Clark, M. 2011. Mendacity and Deception: Uses and Abuses of Common Ground. In *Building Representations of Common Ground with Intelligent Agents: Papers from the AAAI Fall Symposium*, 2–9.

Cramer, H.; Evers, V.; Kemper, N.; and Wielinga, B. 2008. Effects of Autonomy, Traffic Conditions and Driver Personality Traits on Attitudes and Trust towards In-Vehicle Agents. In *Proc. of the IEEE/WIC/ACM Int. Conf. on Web Intelligence and Intelligent Agent Technology*, volume 3, 477–482.

Cramer, H.; Goddijn, J.; Wielinga, B.; and Evers, V. 2010. Effects of (in)accurate empathy and situational valence on attitudes towards robots. In *Proc. of the 5th ACM/IEEE Int. Conf. on Human-Robot Interaction*, 141–142.

DeSteno, D.; Breazeal, C.; Frank, R.; Pizarro, D.; Baumann, J.; Dickens, L.; and Lee, J. 2012. Detecting the Trustworthiness of Novel Partners in Economic Exchange. *Psychol. Sci.* 23(12):1549–1556.

Dunn, J., and Schweitzer, M. 2005. Feeling and Believing: The Influence of Emotion on Trust. *J. Pers. Soc. Psychol.* 88(5):736–748.

Falcone, R., and Castelfranchi, C. 2001. Social Trust: A Cognitive Approach. In Castelfranchi, C., and Tan, Y.-H., eds., *Trust and Deception in Virtual Societies*. Dordrecht, The Netherlands: Kluwer Academic Publishers. 55–90.

Feltovich, P.; Bradshaw, J.; Clancey, W.; and Johnson, M. 2007. Toward an Ontology of Regulation: Socially-Based Support for Coordination in Human and Machine Joint Activity. In *Engineering Societies in the Agents World VII*, volume 4457 of *LNCS*. Heidelberg, Germany: Springer-Verlag. 175–192.

Groom, V.; Srinivasan, V.; Bethel, C.; Murphy, R.; Dole, L.; and Nass, C. 2011. Responses to robot social roles and social role framing. In *Proc. of the Int. Conf. on Collaboration Technologies and Systems*, 194–203.

Hoffman, R.; Lee, J.; Woods, D.; Shadbolt, N.; Miller, J.; and Bradshaw, J. 2009. The Dynamics of Trust in Cyberdomains. *IEEE Intell. Syst.* 24(6):5–11.

Johnson, M.; Bradshaw, J.; Feltovich, P.; Hoffman, R.; Jonker, C.; van Riemsdijk, B.; and Sierhuis, M. 2011. Beyond Cooperative Robotics: The Central Role of Interdependence in Coactive Design. *IEEE Intell. Syst.* 26(3):81–88.

Kahn, P. J.; Ishiguro, H.; Friedman, B.; Kanda, T.; Freier, N.; Severson, R.; and Miller, J. 2007. What is a human? Toward psychological benchmarks in the field of human-robot interaction. *Interact. Stud.* 8(3):363–390.

Klein, G.; Woods, D.; Bradshaw, J.; Hoffman, R.; and Feltovich, P. 2004. Ten Challenges for Making Automation a "Team Player" in Joint Human-Agent Activity. *IEEE Intell. Syst.* 19(6):91–95.

Kramer, R. M., and Tyler, T. R. 1996. *Trust in Organizations: Frontiers of Theory and Research*. Thousand Oaks, CA: Sage Publications.

Lee, J., and See, K. 2004. Trust in Automation: Designing for Appropriate Reliance. *Hum. Factors* 46(1):50–80.

Levin, D., and Cross, R. 2004. The Strength of Weak Ties You Can Trust: The Mediating Role of Trust in Effective Knowledge Transfer. *Manage Sci.* 50(11):1477–1490.

Levin, D.; Cross, R.; and Abrams, L. 2002. Why Should I Trust You? Predictors of Interpersonal Trust in a Knowledge Transfer Context. In *Academy of Management Meeting*.

Marble, J.; Bruemmer, D.; Few, D.; and Dudenhoeffer, D. 2004. Evaluation of Supervisory vs. Peer-Peer Interaction with Human-Robot Teams. In *Proc. of the 37th Hawaii Int. Conf. on System Sciences*, volume 5, 50130b.

Mayer, R.; Davis, J.; and Schoorman, F. D. 1995. An integrative model of organizational trust. *Acad. Manage Rev.* 20(3):709–734.

McAllister, D. 1995. Affect- and Cognition-Based Trust as Foundations for Interpersonal Cooperation in Organizations. *Acad. Manage J.* 38(1):24–59.

Nass, C.; Fogg, B. J.; and Moon, Y. 1996. Can computers be teammates? *Int. J. Hum. Comput. Stud.* 45(6):669–678.

Parasuraman, R., and Riley, V. 1997. Humans and Automation: Use, Misuse, Disuse, Abuse. *Hum. Factors* 39(2):230–253.

Parlangeli, O.; Chiantini, T.; and Guidi, S. 2012. A mind in a disk: The attribution of mental states to technological systems. *Work* 41(1):1118–1123.

Pentland, A., and Heibeck, T. 2008. Understanding "Honest Signals" in Business. *MIT Sloan Manage Rev.* 50(1):70–75.

Pentland, A. 2004. Social Dynamics: Signals and Behavior. In *Proc. of the 3rd Int. Conf. on Development and Learning*, 263–267.

Premack, D., and Woodruff, G. 1978. Does the chimpanzee have a theory of mind? *Behav. Brain Sci.* 1(4):515–126.

Rogerson, M.; Gottlieb, M.; Handelsman, M.; Knapp, S.; and Younggren, J. 2011. Nonrational processes in ethical decision making. *Am. Psychol.* 66(7):614–623.

Rousseau, D.; Sitkin, S.; Burt, R.; and Camerer, C. 1998. Not so different after all: A cross-discipline view of trust. *Acad. Manage Rev.* 23(3):393–404.

Schaefer, K.; Billings, D.; and Hancock, P. 2012. Robots vs. Machines: Identifying User Perceptions and Classifications. In *Proc. of the IEEE Int. Multi-Disciplinary Conf. on Cognitive Methods in Situation Awareness and Decision Support*, 168–171.

Schoorman, F. D.; Mayer, R.; and Davis, J. 2007. An integrative model of organizational trust: Past, present, and future. *Acad. Manage Rev.* 32(2):344–354.

Semin, G., and Marsman, J. G. 1994. Multiple inference-inviting properties of interpersonal verbs: Event instigation,

dispositional inference, and implicit causality. *J. Pers. Soc. Psychol.* 67(5):836–849.

Simpson, J. 2007. Psychological Foundations of Trust. *Curr. Dir. Psychol. Sci.* 16(5):264–268.

Spinoglio, G.; Lenti, L.; Maglione, V.; Lucido, F.; Priora, F.; Bianchi, P.; Grosso, F.; and Quarati, R. 2012. Single-site robotic cholecystectomy (SSRC) versus single-incision laparoscopic cholecystectomy (SILC): comparison of learning curves. First European experience. *Surg. Endosc.* 26(6):1648–1655.

Srinivas, V. 2000. *Individual Investors and Financial Advice: A Model of Advice-seeking in the Financial Planning Context*. Ph.D. Dissertation, Rutgers University, New Brunswick, NJ.

Stokes, C.; Lyons, J.; Littlejohn, K.; Natarian, J.; Case, E.; and Speranza, N. 2010. Accounting for the human in cyberspace: Effects of mood on trust in automation. In *Proc. of the 2010 Int. Symp. on Colaborative Technologies and Systems*, 180–187.

Stoltzman, W. 2006. Toward a Social Signaling Framework: Activity and Emphasis in Speech. Master's thesis, Massachusetts Institute of Technology, Cambridge, MA.

Takayama, L.; Ju, W.; and Nass, C. 2008. Beyond Dirty, Dangerous and Dull: What Everyday People Think Robots Should Do. In *Proc. of the 3rd ACM/IEEE Int. Conf. on Human-Robot Interaction*, 25–32.

Turkle, S. 2004. How Computers Change the Way We Think. *The Chronicle of Higher Education* 50(21):B26.

Turkle, S. 2010. In good company? On the threshold of robotic companions. In Wilks, Y., ed., *Close Engagements with Artificial Companions: Key Social, Psychological, Ethical and Design Issues*. Amsterdam, The Netherlands: John Benjamins Publishing Company. 3–10.

USAF. 2010. Technology Horizons: A Vision for Air Force Science & Technology During 2010–2030. Technical Report AF/ST-TR-10-01-PR, United States Air Force, Office of Chief Scientist (AF/ST), Washington, DC.

Wagner, A. 2009. *The Role of Trust and Relationships in Human-Robot Social Interaction*. Ph.D. Dissertation, Georgia Institute of Technology, Atlanta, GA.

Weber, J. M.; Malhotra, D.; and Murnighan, J. K. 2004. Normal acts of irrational trust: Motivated attributions and the trust development process. *Res. Organ Behavr.* 26:75–101.

Wilks, Y. 2011. Protocols for Reference Sharing in a Belief Ascription Model of Communication. In *Advances in Cognitive Systems: Papers from the AAAI Fall Symposium*, 337–344.

Wing, J. 2008. Cyber-Physical Systems Research Charge. In *Cyber-Physical Systems Summit*.