

The Mind in the Machine: Anthropomorphism Increases Trust
in an Autonomous Vehicle

Adam Waytz
Northwestern University

Joy Heafner
University of Connecticut

Nicholas Epley
University of Chicago

(uncorrected proof, in press at *Journal of Experimental Social Psychology*)

Corresponding Author:

Adam Waytz
Northwestern University
2001 Sheridan Rd
Evanston, IL 60208
United States
Email: a-waytz@kellogg.northwestern.edu

Word Count: 3,170

Abstract

Sophisticated technology is increasingly replacing human minds to perform complicated tasks in domains ranging from medicine to education to transportation. We investigated an important theoretical determinant of people's willingness to trust such technology to perform competently—the extent to which a nonhuman agent is anthropomorphized with a humanlike mind—in a domain of practical importance, autonomous driving. Participants using a driving simulator drove either a normal car, an autonomous vehicle able to control steering and speed, or a comparable autonomous vehicle augmented with additional anthropomorphic features—name, gender, and voice. Behavioral, physiological, and self-report measures revealed that participants trusted that the vehicle would perform more competently as it acquired more anthropomorphic features. Technology appears better able to perform its intended design when it seems to have a humanlike mind. These results suggest meaningful consequences of humanizing technology, and also offer insights into the inverse process of objectifying humans.

Word Count: 148

Technology is an increasingly common substitute for humanity. Sophisticated machines now perform tasks that once required a thoughtful human mind, from grading essays to diagnosing cancer to driving a car. As engineers overcome design barriers to creating such technology, important psychological barriers that users will face when using this technology emerge. Perhaps most important, will people be willing to trust competent technology to replace a human mind, such as a teacher's mind when grading essays, or a doctor's mind when diagnosing cancer, or their own mind when driving a car?

Our research tests one important theoretical determinant of trust in any nonhuman agent: anthropomorphism (Waytz, Cacioppo, & Epley, 2010). Anthropomorphism is a process of inductive inference whereby people attribute to nonhumans distinctively human characteristics, particularly the capacity for rational thought (agency) and conscious feeling (experience; Gray, Gray, & Wegner, 2007). Philosophical definitions of personhood focus on these mental capacities as essential to being human (Dennett, 1978; Locke, 1841/1997). Furthermore, studies examining people's lay theories of humanness show that people define humanness in terms of emotions that implicate higher order mental process such as self-awareness and memory (e.g., humiliation, nostalgia; Leyens et al., 2000) and traits that involve cognition and emotion (e.g., analytic, insecure; Haslam, 2006). Anthropomorphizing a nonhuman does not simply involve attributing superficial human characteristics (e.g., a humanlike face or body) to it, but rather attributing *essential* human characteristics to the agent (namely a humanlike mind, capable of thinking and feeling).

Trust is a multifaceted concept that can refer to belief that another will behave with benevolence, integrity, predictability, or competence (McKnight & Chervany, 2001). Our prediction that anthropomorphism will increase trust centers on this last component of trust in another's competence (akin to confidence) (Siegrist, Earle, & Gutscher, 2003; Twyman, Harvey, & Harries,

2008). Just as a patient would trust a thoughtful doctor to diagnose cancer more than a thoughtless one, or would rely on mindful cab driver to navigate through rush hour traffic more than a mindless cab driver, this conceptualization of anthropomorphism predicts that people would trust easily anthropomorphized technology to perform its intended function more than seemingly mindless technology. An autonomous vehicle (one that that drives itself) for instance, should seem better able to navigate through traffic when it seems able to think and sense its surroundings than when it seems to be simply mindless machinery. Or a “warbot” intended to kill should seem more lethal and sinister when it appears capable of thinking and planning than when it seems to be simply a computer mindlessly following an operator’s instructions. The more technology seems to have humanlike mental capacities, the more people should trust it to perform its intended function competently, regardless of the valence of its intended function (Epley, Caruso, & Bazerman, 2006; Pierce, Kilduff, Galinsky, & Sivanathan, 2013).

This prediction builds on the common association between people’s perceptions of others’ mental states and of competent action. Because mindful agents appear capable of controlling their own actions, people judge others to be more responsible for successful actions they perform with conscious awareness, foresight, and planning (Cushman, 2008; Malle & Knobe, 1997) than for actions they perform mindlessly (see Alicke, 2000; Shaver, 1985; Weiner, 1995). Attributing a humanlike mind to a nonhuman agent should therefore more make the agent seem better able to control its own actions, and therefore better able to perform its intended functions competently. Our prediction also advances existing research on the consequences of anthropomorphism by articulating the psychological processes by which anthropomorphism could affect trust in technology (Nass & Moon, 2000), and by both experimentally manipulating anthropomorphism as well as measuring it as a critical mediator. Some experiments have manipulated the humanlike appearance of robots and

assessed measures indirectly related to trust. However, such studies have not measured whether such superficial manipulations actually increases the attribution of essential humanlike qualities to that agent (the attribution we predict is critical for trust in technology; Hancock, Billings, Schaeffer, Chen, De Visser, 2011), and therefore cannot explain factors found ad-hoc to moderate the apparent effect of anthropomorphism on trust (Pak, Fink, Price, Bass, & Sturre, 2012). Another study found that individual differences in anthropomorphism predicted differences in willingness to trust technology in hypothetical scenarios (Waytz et al., 2010), but did not manipulate anthropomorphism experimentally. Our experiment is therefore the first to test our theoretical model of how anthropomorphism affects trust in technology.

We conducted our experiment in a domain of practical relevance: people's willingness to trust an autonomous vehicle. Autonomous vehicles—cars that control their own steering and speed—are expected to account for 75% of vehicles on the road by 2040 (Newcomb, 2012). Employing these autonomous features means surrendering personal control of the vehicle and trusting technology to drive safely. We manipulated the ease with which a vehicle, approximated by a driving simulator, could be anthropomorphized by merely giving it independent agency, or by also giving it a name, gender, and a human voice. We predicted that independent agency alone would make the car seem more mindful than a normal car, and that adding further anthropomorphic qualities would make the vehicle seem even more mindful. More important, we predicted that these relative increases in anthropomorphism would increase physiological, behavioral, and psychological measures of trust in the vehicle's ability to drive effectively.

Because anthropomorphism increases trust in the agent's ability to perform its job, we also predicted that increased anthropomorphism of an autonomous agent would mitigate blame for an agent's involvement in an undesirable outcome. To test this, we implemented a virtually unavoidable

accident during the driving simulation in which participants were struck by an oncoming car, an accident clearly caused by the other driver. We implemented this to maintain experimental control over participants' experience because everyone in the autonomous vehicle conditions would get into the same accident, one clearly caused by the other driver. Indeed, when two people are potentially responsible for an outcome, the agent seen to be more competent tends to be credited for a success whereas the agent seen to be less competent tends to be blamed for a failure (Beckman, 1970; Wetzel, 1972). Because we predicted that anthropomorphism would increase trust in the vehicle's competence, we also predicted that it would reduce blame for an accident clear caused by another vehicle.

Experiment

Method

One hundred participants (52 female, $M_{\text{age}}=26.39$) completed this experiment using a National Advanced Driving Simulator. Once in the simulator, the experimenter attached physiological equipment to participants and randomly assigned them to condition: Normal, Agentic, or Anthropomorphic. Participants in the Normal condition drove the vehicle themselves, without autonomous features. Participants in the Agentic condition drove a vehicle capable of controlling its steering and speed (an "autonomous vehicle"). The experimenter followed a script describing the vehicle's features, suggesting when to use the autonomous features, and describing what was about to happen. Participants in the Anthropomorphic condition drove the same autonomous vehicle, but with additional anthropomorphic features beyond mere agency—the vehicle was referred to by name (Iris), was given a gender (female), and was given a voice through human audio files played at predetermined times throughout the course. The voice files followed the same script used by the experimenter in the Agentic condition, modified where necessary (See Supplemental Online Material [SOM]).

All participants first completed a driving history questionnaire and a measure of dispositional anthropomorphism (Waytz et al., 2010). Scores on this measure did not vary significantly by condition, so we do not discuss them further.

Participants in the Agentic and Anthropomorphic conditions then drove a short practice course to familiarize themselves with the car's autonomous features. Participants could engage these features by pressing buttons on the steering wheel. All participants then drove two courses each lasting approximately six minutes. After the first course, participants completed a questionnaire (all on 0-10 scales, see SOM for all items) that assessed anthropomorphism, liking, and trust.

Perceived Anthropomorphism. Four items measured anthropomorphism, defined as attributing humanlike mental capacities of agency and experience to it (Epley et al., 2007; Gray et al., 2007; Waytz et al., 2010). These asked how *smart* the car was, how well it could *feel* what was happening around it, how well it could *anticipate* what was about to happen, and how well it could *plan* a route. These items were averaged into a composite ($\alpha=.89$).

Liking. Four items measured liking: how enjoyable their driving was, how comfortable they felt driving the car, how much participants would like to own a car like this one, and what percentage of cars in 2020 they would like to be [autonomous] like this one. These items were standardized and averaged to form a single composite ($\alpha=.90$).

Self-reported trust. Eight items measured trust in the vehicle: how safe participants felt they and others would be if they actually owned a car like this one, how much they trust the vehicle to drive in heavy and light traffic conditions, how confident they are about the car driving the next course safely, and their willingness to give up control to the car. These items were standardized and averaged to form a single composite ($\alpha=.91$)

After approximately six minutes of driving a second course along a rural highway, a vehicle pulled quickly in front of the car and struck their right side. We designed this accident to be unavoidable so that all participants would experience the same outcome (indeed, only one participant, in the Normal condition, avoided it). Ensuring that everyone got into this accident, however, meant that the accident was clearly the *other* vehicle's fault rather than participants' own vehicle. Throughout the experiment, we measured participants' heart rate using electrocardiography (ECG) and videotaped their behavior unobtrusively to assess responses to this accident.

Heart Rate Change. We reasoned that if participants trusted the vehicle, they should be more relaxed in an arousing situation (namely, the accident), showing an attenuated heart rate increase and startle response. We measured heart rate change to the accident as a percentage change of beats per minute for twenty seconds immediately following the collision (or until they concluded their simulation), in comparison to a forty-five second baseline period immediately following the earlier practice course.

Startle. To assess startle response, we first divided our participants into two random samples. We then recruited 42 independent raters from an undergraduate population to watch all videos from one or the other sample and rate how startled each participant appeared during the video (0=*not at all startled* to 10=*extremely startled*). We then averaged startle ratings for each participant across all of these raters to obtain a startle response measure. Percentage heart rate change and startle were standardized and reverse-scored (multiplied by -1) and then averaged to form a behavioral measure of trust ($r(90)=.28, p<.01$). To assess overall trust, we averaged all standardized measures of trust (the eight self-report measures and the two behavioral measures) into a single composite ($\alpha=.87$).

Blame for Vehicle. After the accident, all participants also assessed how responsible they, the car, the people who designed the car, and the company that developed the car were for the accident (all

0-10 scales, see SOM for exact questions). To assess punishment for the accident, participants were asked to imagine that this accident occurred in the real world, with a different driver behind the wheel of their car. Participants reported how strongly they felt that the driver should be sent to jail, how strongly they felt that the car should be destroyed, how strongly they felt that the car's engineer should be punished, and how strongly they felt that the company that designed the car should be punished. The six items measuring the vehicle's responsibility and resulting punishment for a similar accident were standardized and averaged to form a single composite ($\alpha=.90$).

Distraction. Finally, we used the videotape mentioned above to measure participants' distraction while driving during the second course, measured as the time spent looking away from the simulator rather than paying attention while driving. Results showed a floor effect with very little distraction across conditions (less than 3% of the overall time in the two autonomous vehicle conditions). See Table 1 for these means as well as means from all analyses below.¹

Results

All primary analyses involved planned orthogonal contrasts examining differences between the Normal, Agentic, and Anthropomorphic conditions.

Perceived Anthropomorphism. As predicted, participants in the Anthropomorphic condition anthropomorphized the vehicle more than those in the Agentic condition, $t(97)=3.21, p=.002, d=.65$, who in turn anthropomorphized the vehicle more than in the Normal condition, $t(97)=7.11, p<.0001, d=1.44$.

Liking. Participants in the Anthropomorphic and Agentic conditions liked the vehicle more than did participants in the Normal condition, $t(97)=3.92, p<.0001, d=.80$ and $t(97)=3.29, p=.001, d=.67$, but the autonomous vehicle conditions did not differ significantly from each other ($p=.55$).

Trust. As predicted, on the measure of overall trust, those in the Anthropomorphic condition trusted their vehicle more than did those in the Agentic condition, $t(97)=2.34, p=.02, d=.48$, who in turn trusted their vehicle more than those in the Normal condition, $t(97)=4.56, p<.0001, d=.93$. For behavioral trust, participants in the Anthropomorphic condition trusted their vehicle more than did those in the Agentic condition, $t(97)=3.36, p=.001, d=.68$ and Normal condition, $t(97)=2.78, p<.01, d=.56$, although the Agentic and Normal conditions did not differ significantly ($p=.56$). For self-reported trust, participants in the Anthropomorphic condition and the Agentic condition did not differ significantly ($p=.14$), but both participants in the Agentic condition and Anthropomorphic conditions reported greater trust than participants in Normal condition, $ts(97)=4.83$ and 6.35 , respectively, $ps<.01, ds=.98$ and 1.29 . Table 1 reports the self-report measures and the behavioral measures of trust separately.

To assess whether the vehicle's effect on overall trust was statistically mediated by perceived anthropomorphism, we used Preacher and Hayes' (2008) bootstrapping method and coded condition as Normal=0, Agentic=1, and Anthropomorphic=2 (see Hahn-Holbrook, Holt-Lunstad, Holbrook, Coyne, and Lawson, 2011; Legault, Gutsell, & Inzlicht, 2011 for similar analyses). This analysis confirmed that anthropomorphism statistically mediated the relationship between vehicle condition and overall trust in the vehicle (95% CI=.31 to .55; see Figure 1; 20,000 resamples).

Blame for Vehicle. As noted, we programmed the driving simulation so that all participants would experience the same virtually unavoidable accident clearly caused by the other driver, but it is important to keep the nature of the accident in mind. If a trusted, competent driver were hit by another vehicle, one would hold the competent driver less responsible for the accident because it would clearly appear to be the other driver's fault. Thus, we predicted that anthropomorphism would mitigate blame for an accident clearly caused by the other vehicle. It is important to note, however, that our

prediction would be different if the vehicle were able to *avoid* this accident, in which case we would predict that anthropomorphism would increase the tendency to credit the vehicle for this success.

Participants in the Agentic and Anthropomorphic conditions blamed their car more for the accident than did those in the Normal condition, $t_s(96)=6.30$ and 4.18 , respectively, $ps<.01$, $ds=1.29$ and $.85$. This is consistent with the relationship between agency and perceived responsibility. An object with no agency cannot be held responsible for any actions, and so this comparison is not particularly interesting. More interesting is that participants blamed the vehicle significantly *less* in the Anthropomorphic condition than in the Agentic condition, $t(96)=2.18$, $p=.03$, $d=.44$, in which the perceived thoughtfulness of the fully anthropomorphic vehicle mitigated the responsibility that comes from independent agency (given that the accident was clearly caused by the other vehicle). This shows a clear relationship between anthropomorphism and perceptions of responsibility, but the exact nature of that relationship cannot be tested in this particular paradigm because we are unable to create a uniform accident across conditions clearly caused by participants themselves.

General Discussion

Technological advances blur the line between human and nonhuman, and this experiment suggests blurring this line even further could increase users' willingness to trust technology in place of humans. Amongst those who drove an autonomous vehicle, those who drove a vehicle that was named, gendered, and voiced rated their vehicle as having more humanlike mental capacities than those who drove a vehicle with the same autonomous features but without anthropomorphic cues. In turn, those who drove the anthropomorphized vehicle with enhanced humanlike features (name, gender, voice) reported trusting their vehicle even more, were more relaxed in an accident, and blamed their vehicle and related entities less for an accident caused by another driver. These findings provide further support for the theoretical connection between perceptions of mental capacities in others and

assessments of competence, trust, and responsibility. Attributing a mind to a machine matters because it could create a machine to which users might entrust their lives.

This finding is also of clear practical relevance given the rapidly changing interface between the technological world and the social world. No longer merely mindless tools, modern technology now taps human social skills directly. People ask their phones for driving directions, restaurant recommendations, and baseball scores. Automated customer service agents help people purchase flights, pay credit card bills, and obtain prescription medicine. Robotic pets even provide social support and companionship, sometimes in the place of actual human companionship (Melson, Kahn, Beck, Friedman, Roberts, Garrett, & Gill, 2009). Our research identifies one important consequence of considering the psychological dimensions of technological design. Even the greatest technology, such as vehicles that drive themselves, is of little benefit if consumers are unwilling to use it.

Finally, our research at this human-technology frontier also informs the inverse effect in which people are treated more like technology—as objects or relatively mindless machines (Cikara, Eberhardt, & Fiske, 2011; Loughnan & Haslam, 2007). Adding a human voice to technology, for instance, makes people treat it as more humanlike agent (Takayama & Nass, 2008), which suggests that removing a human voice like one's own from interpersonal communication may make another person seem relatively mindless. Indeed, in one series of recent experiments, participants rated another person as being less mindful (e.g., less thoughtful, less rational) when they read a transcript of an interview than when they heard the audio of the same interview (Schroeder & Epley, 2014). Similarly, verbal accents that differ from one's own trigger prejudice and distrust compared to accents similar to one's own (Anisfeld, Bogo, & Lambert, 1962; Dixon, Mahoney, & Cocks, 2002; Giles & Powesland, 1975; Kinzler, Dupoux, & Spelke, 2007; Kinzler, Corriveau, & Harris, 2011; Lev-Ari &

Keysar, 2010), an effect that may be partially mediated by differences in the attribution of humanlike mental states.

Few divides in social life are more important than the one between us and them, between human and nonhuman. Perceptions of this divide are not fixed but flexible. Understanding when technology crosses that divide to become more humanlike matters not just for how people treat increasingly humanlike technology, but also for understanding why people treat other humans as mindless objects.

References

- Anisfeld, M., Bogo, N., & Lambert, W.E. (1962). Evaluational reactions to accented English speech. *Journal of Abnormal and Social Psychology, 65*, 223-231.
- Alicke, M. D. (2000). Culpable control and the psychology of blame. *Psychological Bulletin, 126*, 556.
- Beckman, L. (1970). Effects of students' performance on teachers' and observers' attributions of causality. *Journal of Educational Psychology, 61*, 76-82.
- Cikara, M., Eberhardt, J. L., & Fiske, S. T. (2011). From agents to objects: Sexist attitudes and neural responses to sexualized targets. *Journal of Cognitive Neuroscience, 23*, 540–551.
- Cushman, F. (2008). Crime and punishment: Distinguishing the roles of causal and intentional analyses in moral judgment. *Cognition, 108*, 353-380.
- Dennett, D. C. (1978). *Brainstorms: Philosophical Essays on Mind and Psychology*. Cambridge: Bradford Books/MIT Press.
- Dixon, J. A., Mahoney, B., & Cocks, R. (2002). Accents of guilt? Effects of regional accent, race, and crime type on attributions of guilt. *Journal of Language and Social Psychology, 21*, 162-168.
- Epley, N., Caruso, E., & Bazerman, M. H. (2006). When perspective taking increases taking: reactive egoism in social interaction. *Journal of personality and social psychology, 91*, 872.
- Epley, N., Waytz, A., & Cacioppo, J. T. (2007). On seeing human: a three-factor theory of anthropomorphism. *Psychological Review, 114*, 864.
- Giles, H., & Powesland, P. F. (1975). *Speech style and social evaluation* (Vol. 7). London: Academic Press.
- Gray, H. M., Gray, K., & Wegner, D. M. (2007). Dimensions of mind perception. *Science, 315*, 619-619.

- Hahn-Holbrook, J., Holt-Lunstad, J., Holbrook, C., Coyne, S. M., & Lawson, E. T. (2011). Maternal defense: Breast feeding increases aggression by reducing stress. *Psychological Science, 22*, 1288-1295.
- Hancock, P. A., Billings, D. R., Schaefer, K. E., Chen, J. Y., De Visser, E. J., & Parasuraman, R. (2011). A meta-analysis of factors affecting trust in human-robot interaction. *Human Factors: The Journal of the Human Factors and Ergonomics Society, 53*, 517-527.
- Haslam, N. (2006). Dehumanization: An integrative review. *Personality and Social Psychology Review, 10*, 252-264.
- Kinzler, K.D., Corriveau, K.H., & Harris, P.L. (2011). Children's selective trust in native accented speakers. *Developmental Science, 14*, 106-111.
- Kinzler, K.D., Dupoux, E., & Spelke, E.S. (2007). The native language of social cognition. *Proceedings of the National Academy of Sciences, 104*, 12577-12580.
- Legault, L., Gutsell, J. N., & Inzlicht, M. (2011). Ironic effects of antiprejudice messages: How motivational interventions can reduce (but also increase) prejudice. *Psychological Science, 22*, 1472-1477.
- Lev-Ari, S., & Keysar, B. (2010). Why don't we believe non-native speakers? The influence of accent on credibility. *Journal of Experimental Social Psychology, 46*, 1093-1096.
- Leyens, J. P., Paladino, P. M., Rodriguez-Torres, R., Vaes, J., Demoulin, S., Rodriguez-Perez, A., & Gaunt, R. (2000). The emotional side of prejudice: The attribution of secondary emotions to ingroups and outgroups. *Personality and Social Psychology Review, 4*, 186-197.
- Locke, J. (1997). *An Essay Concerning Human Understanding*. Harmondsworth, England: Penguin Books. (Original work published 1841).
- Loughnan, S., & Haslam, N. (2007). Animals and androids: Implicit associations between social

- categories and nonhumans. *Psychological Science*, *18*, 116-121.
- McKnight, D. H., & Chervany, N. L. (2001). Trust and distrust definitions: One bite at a time. In *Trust in Cyber-societies* (pp. 27-54). Springer Berlin Heidelberg.
- Malle, B. F., & Knobe, J. (1997). The folk concept of intentionality. *Journal of Experimental Social Psychology*, *33*, 101-121.
- Melson, G. F., Kahn, P. H., Beck, A., Friedman, B., Roberts, T., Garrett, E., & Gill, B. T. (2009). Children's behavior toward and understanding of robotic and living dogs. *Journal of Applied Developmental Psychology*, *30*, 92-102.
- Nass, C., & Moon, Y. (2000). Machines and mindlessness: Social responses to computers. *Journal of Social Issues*, *56*, 81-103.
- Newcomb, D. (2012). Retrieved from <http://www.cnn.com/2012/09/18/tech/innovation/ieee-2040-cars>
- Pak, R., Fink, N., Price, M., Bass, B., & Sturre, L. (2012). Decision support aids with anthropomorphic characteristics influence trust and performance in younger and older adults. *Ergonomics*, *55*, 1059-1072.
- Pierce, J. R., Kilduff, G. J., Galinsky, A. D., & Sivanathan, N. (2013). From glue to gasoline: How competition turns perspective takers unethical. *Psychological Science*, *24*, 1986-1994.
- Preacher, K. J., & Hayes, A. F. (2008). Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models. *Behavior Research Methods*, *40*, 879–891.
- Schroeder, J., & Epley, N. (2013). Speaking louder than words: Voice reveals the presence of a humanlike mind. Unpublished manuscript, *University of Chicago*.
- Shaver, K. G. (1985). *The attribution of blame: Causality, responsibility, and blameworthiness*. New York: Springer-Verlag.

- Siegrist, M., Earle, T. and Gutscher, H. (2003). Test of a trust and confidence model in the applied context of electromagnetic field (EMF) risks. *Risk Analysis*, 23, 705–716.
- Takayama, L., & Nass, C. (2008). Driver safety and information from afar: An experimental driving simulator study of wireless vs. in-car information services. *International Journal of Human-Computer Studies*, 66, 173-184.
- Twyman, M., Harvey, N., & Harries, C. (2008). Trust in motives, trust in competence: Separate factors determining the effectiveness of risk communication. *Judgment and Decision Making*, 3, 111-120.
- Waytz, A., Cacioppo, J., & Epley, N. (2010). Who sees human? The stability and importance of individual differences in anthropomorphism. *Perspectives on Psychological Science*, 5, 219-232.
- Weiner, B. (1995). *Judgments of responsibility: A foundation for a theory of social conduct*. Guilford Press. New York, NY.
- Wetzel, C.G. (1982). Self-serving biases in attribution: A Bayesian analysis. *Journal of Personality and Social Psychology*, 43, 197-209.

Notes

1. Degrees of freedom across analyses vary slightly because of missing responses. We measured skin conductance throughout the experiment through electrodes on participants' wrists, but unanticipated artifacts from movement while driving rendered these results impossible to interpret and so we do not report them.

Acknowledgements

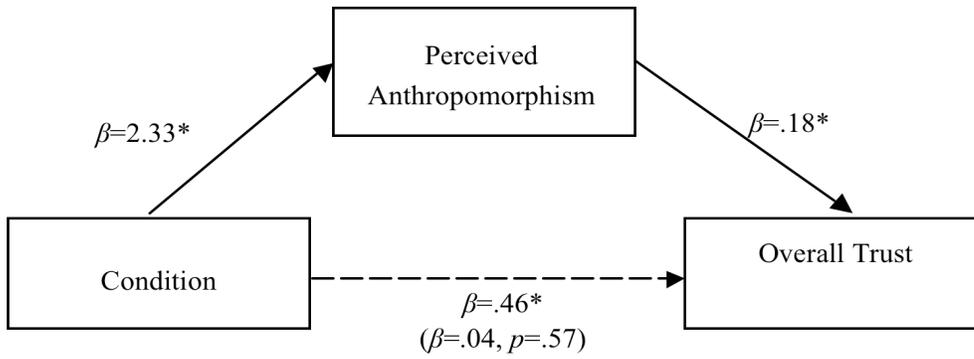
This research was funded by the University of Chicago's Booth School of Business and a grant from the General Motors Company. We thank Julia Hur for assistance with data coding.

Table 1. Means by Condition for Main Dependent Measures.

<u>Measure</u>	<u>Condition</u>		
	<u>Normal</u>	<u>Agentic</u>	<u>Anthropomorphic</u>
Anthropomorphism	2.63 _a	5.86 _b	7.30 _c
Trust:			
Overall	-0.52 _a	0.09 _b	0.41 _c
Self-Reported	-0.60 _a	0.18 _b	0.41 _b
Behavioral (ECG & Startle)	-0.23 _a	-0.35 _a	0.39 _b
Liking	-0.49 _a	0.18 _b	0.30 _b
Blame for the vehicle	-0.60 _a	0.47 _b	0.11 _c
Distraction (instances)	0.16 _a	2.07 _b	2.13 _b
Distraction (percent of total time)	0.1 _a	2.8 _b	2.7 _b

Means that do not share a subscript differ significantly at $p < .05$.

Figure 1. The results of a mediation analysis of condition on overall trust (Betas standardized).



Note 1: Anthropomorphic condition coded as 2; Agentic condition coded as 1; Normal condition coded as 0. $*p \leq 0.0001$