Title
Trust and Algorithmic Decision Making

Author
Nguyen, Elise

Publication Date
2023-04-01

Trust and Algorithmic Decision Making

Elise Nguyen

Communication, University of California, Santa Barbara

Abstract

The acceleration and advancement of today's technology has led to the growing use of machine learning algorithms in everyday life. Therefore, our collective trust in algorithmic decision making becomes increasingly important to consider. Current literature suggests that people may be skeptical of relying on algorithmic judgment rather than human judgment, regardless of performance quality or accuracy (Logg, 2018). However, conflicting results have arisen from previous studies regarding this algorithmic aversion or appreciation. An online experiment was conducted using a 2x2 design with 120 adult participants in order to examine how the control and risk environment of an algorithm's decision making process affects human trust towards algorithmic decision making. Results indicate that humans are less trusting, or more averse, of automated systems in situations with higher perceived risk and lower human control. These findings shed light on the evolving relationship between humans and the automated systems we rely upon and have implications for the development and operation of automated systems we generate.

Introduction

Today, machine learning algorithms pervade every aspect of our daily lives, from music recommendations to advertisements to search suggestions. Artificial intelligence (AI) dramatically changes previous user interface paradigms. However, given the degree of unpredictability and uncertainty that comes with reliance upon automated systems, the human user needs to decide whether or not to trust the AI given different situations. The dynamic between this human user and AI system is a relationship that hinges on user trust, which, in turn, becomes imperative to understand.

In spite of this widespread use of algorithmic advice, current literature suggests that many humans distrust algorithmic judgment and decision making, often referred to as “algorithm aversion” (Logg et al. 2018, Dietvorst, Simmons, & Massey, 2015). The concept of algorithm aversion involves consciously or unconsciously disregarding algorithmic decisions in favor of one's own decision or another's decision (Mahmud et al. 2022).

Algorithmic systems, however, do not always elicit a negative reaction from people; rather, these reactions are often influenced by a wide variety of factors (Mahmud et al. 2022, Berger et al., 2021, Dietvorst et al. 2015, Kawaguchi, 2021). Where some people prefer algorithmic advice, others do not (Mahmud et al. 2022, McBride et al. 2012). The situations in which people trust algorithmic advice more than human advice is referred to as “algorithm appreciation” (Hou & Jung, 2021). The two schools of thought, algorithm aversion and algorithm appreciation, predict varying outcomes when humans receive decision inputs from algorithms. The researchers also propose different methods to increase the acceptability of algorithmic decisions, which are often in contradiction with one another (Hou & Jung, 2021). To maximize the benefits of algorithmic decisions, it is imperative to understand how algorithm aversion or algorithm appreciation arises.

The purpose of this thesis is to reconcile the conflicting literature on algorithm aversion and algorithm appreciation. Two main factors, perceived risk and control, are examined as influential factors in affecting humans’ trust towards automated systems. I conducted an online experiment using a 2x2 design with 120 adult participants in order to examine how the control and risk environment of an algorithm’s decision-making process affects human trust towards algorithmic decision making. Results indicate that humans are less trusting, or more averse, of automated systems in situations with higher perceived risk and lower human control.

Literature Review

Human Machine Communication

As machines become increasingly integrated in decision making systems with humans, they become increasingly accepted as a part of the collaborative decision making process (Chui & Malhotra, 2018). The Computer as Social Actors (CASA) paradigm postulates that humans apply the same—or similar but modified—social heuristics to computers as they do to other humans (Gambino et al., 2020; Reeves & Nass, 1996). In a CASA experiment, the computer replaces a human in a specified social phenomenon. This method has found that people use politeness, gender stereotypes, and principles of reciprocal disclosure with computers (Reeves & Nass, 1996; Nass et al. 1994; Moon, 2000). While this doesn't directly apply in an organizational context, the idea that humans can identify with machines as teammates or colleagues provides a meaningful context behind increasingly anthropocentric reactions to computers.

Algorithmic Aversion and Algorithmic Appreciation

While algorithms have been seen to be more reliable than human judgment in many situations (Dawes, Faust, & Meehl, 1989), there is a growing library of debate around the attitudes surrounding algorithmic judgment -- in other words, "algorithm appreciation" versus "algorithm aversion."

Meehl (1954) is thought to be the first scholar to mention psychological distrust of algorithms (Logg et al., 2018). In his book, *Clinical versus Statistical Prediction,* Meehl contends that algorithms outperform humans at predicting outcomes. However, at the time, human experts within the clinical field held the opinion that a linear model would not be able to outperform human judgment.

Since the publication of Meehl's book, many researchers have empirically tested the mistrust of algorithms; however, these studies have yielded mixed results. Various studies suggest that humans rely on human recommendations in subjective decision making (Sinha and Swearingen, 2001) Researchers have also found that people trust human judgment over algorithmic judgment after seeing an algorithm make an error (Dietvorst et al., 2015, Dzindolet et al., 2002). On the other hand, studies have shown that participants trusted algorithmic advice instead of human advice when considering logic problems (Dijkstra, Liebrand, & Timminga, 1998). This inclination towards algorithmic advice endured even after participants saw the algorithmic system make an error. (Dijkstra, 1999).

Prior literature suggests that experienced professionals will heavily discount algorithmic advice when making judgements in their fields of expertise (Jacobson, Dobbs-Marsh, Liberman, & Minson, 2011, Meehl, 1954). Logg et al. (2018) found that experienced judges, those who made regular forecasts in the related field, distrusted algorithms more than laypersons. This indicates that on an individual basis, humans may be inclined to trust algorithmic advice based on personal knowledge and experience of the specific situation.

Risk

Trust, as defined by Lee and See (2004), is an expectation that an agent can help an individual achieve their goals in a situation characterized by uncertainty and vulnerability (Satterfield et al. 2017). The complexity of autonomous systems results in uncertainty when interacting with automation, which means that all interactions with such systems involve some level of risk. Thus, as humans try to maximize the benefits from autonomous systems and minimize the costs, they need to calibrate their levels of trust appropriately.

According to the framework provided by Hoff and Bashir (2014), perceptions of risk are critical to development of trust in automated systems. Additionally, Perkins et al. (2010) found that participants did not trust GPS suggestions as much in a riskier environment. Satterfield et al. (2017) also found that behavioral trust was lower with increased risk (in this case, how much money a participant stood to lose). Interestingly, however, Lyons and Stokes (2012) found that participants in a high-risk condition relied more on an automated aid than a human assistant. In essence, perceived risk heightens the cost of uncertainty in situations involving automation, which in turn lessens the trust of the operator towards the system. This leads to the first hypothesis.

H_1: The greater the level of risk, the lower the level of human trust in automated systems.

Control

The concept of locus of control pertains to an individual's belief in their ability to influence the events of their life (internal locus of control) or that the environment or fate dictates the events of their life (external locus of control) (Judge et al., 1998, Mahmud et al. 2022). Previous literature suggests that individuals' desire for control may inhibit them from accepting automated decision aid (Mahmud et al. 2022, Bigman & Gray 2018). Shaffer et al. (2013) found that in medical diagnostic decisions, greater internal locus of control was associated with more algorithmic aversion. In other words, one's need to feel in control may lessen their ability to trust automated systems in the decision-making process.

Kawaguchi (2021) found that workers entrusted with a higher degree of task delegation, and therefore given a greater level of control over the task, are less willing to trust algorithmic aids. However, Kawaguchi (2021) also found that workers are more willing to trust algorithmic advice when their forecasts are integrated with the algorithm. This means that people may not necessarily want total control over the decision making process. Previous literature has found that negative sentiment and distrust for algorithmic aids can be reduced by giving people even slight control over the process, such as the ability to modify algorithmic inputs (Dietvorst et al. 2015, Mahmud et al. 2022). To summarize, the ability to make modifications to the system's actions and decisions can foster a greater sense of trust between humans and automation. This leads to the second hypothesis.

H_2: The greater the level of human control, the higher the level of human trust in automated systems.

Methods

To test the hypotheses, an online experiment was conducted using a 2x2 design.

Pilot Study

A pilot study was conducted with 55 UC Santa Barbara undergraduate students. These participants were recruited using SONA and were compensated for their time with course credit. Due to observations of the pilot study, several questions were removed after reliability testing to ensure a Cronbach's Alpha score of greater than 0.7. One scale item was removed from the Attitudes Toward Artificial Intelligence Scale (Schepman, A. & Rodman, P. 2020), which read "Artificial Intelligence is used to spy on people." One scale item was removed from the Checklist for Trust Between People and Automation Scale (Jian et al., 2000), which read "I am wary of the system." Several categories from the Digital Skills Measurement Scale (Van Deursen et al. 2014). were removed due to excessive length of the survey so that only the Operational and Social categories were measured.

Sample

Participants were 120 adults from the United States, aged 18 or above (M = 18 - 34 years old). Participants were recruited using Prolific and were compensated for their time at a rate of $21.80 per hour. 60.8% of participants identified as female, 38.3% of participants identified as male, and 0.8% of participants identified as nonbinary. The majority of participants were Caucasian or White (87.5%, N = 105). Most of the participants' highest level of education was a Bachelor's degree (44.2%, N = 53), followed by some college with no degree (16.7%, N = 20).

Procedure

Participants from Prolific were directed to fill out the survey using a Qualtrics link. They were given a consent form informing them of the purpose, compensation, and privacy safeguards of the study. On average, the survey took approximately 30 minutes and 35 seconds to complete.

Measures

Manipulation Check

The participants were given a manipulation check after reading every scenario, which was a 5-point likert scale with 4 items asking participants how much they agree with various statements (1 = strongly disagree, 5 = strongly agree). The statements were "This situation is risky," "Using this system is risky," "I have control over this system's actions," and "I have control over the outcome of the scenario."

Checklist for Trust Between People and Automation. Participants' trust of the given automated systems was measured using a pre-established 5-point Likert with 12 items (Jian et al., 2000). Participants were asked how much they agree with various statements regarding the automated system (scale ranging from 1 to 7, 1 = not at all, 5 = extremely). One of the items labeled "I am wary of the system" was removed following a reliability analysis from the pilot study, with a resulting Cronbach's Alpha of 0.81.

General Attitudes Towards Artificial Intelligence Scale. Participants first answered the General Attitudes Towards Artificial Intelligence Scale (Schepman, A. & Rodman, P. 2020). There are 20 items, with positive and negative statements of AI. For example, "I am interested in using artificially intelligent systems in my daily life" or "People like me will suffer if Artificial Intelligence is used more and more." After a reliability analysis, the item "Artificial Intelligence was used to spy on people" was removed in order to achieve Cronbach's Alpha > 0.7. An attention check question was also added to this portion of the survey to ensure reliable data. The reliability analysis of the final participant group showed a Cronbach's Alpha of 0.916. Schepman and Rodman suggested against aggregating the scores of each item to create an overall scale mean due to a lack of unidimensionality. Therefore, I combined disagreement and agreement from the "strongly" and "somewhat" levels. retained the neutral category, and plotted the frequencies of each category in order to aid visualization (Appendix, Fig. 1)

Digital Skills Measurement Scale. After completing the four scenarios, participants were asked to complete the Digital Skills Measurement Scale (Van Deursen et al. 2014). This measured participants' digital skills using a pre-established 5-point Likert scale with a total of 16 items (scale ranging from 1 to 5, 1 = not at all true of me, 5 = very true of me). Only the Social and Operational categories were included in the final survey following suggestions from the pilot test. For example, in the operational category, The overall mean of the scale was 3.63 with a standard deviation of 0.21.

Risk. The participants were then given four scenarios and asked to imagine themselves interacting with the given automated system, each scenario with varying levels of risk. Two scenarios were High Risk conditions: vehicle driving systems and piloting systems. Two scenarios were Low Risk conditions: spell check systems and movie recommendation systems.

Control. The scenarios given to each participant also differed in their Control conditions. Malone's (2017) control continuum proposes that the roles of AI systems can change based on differing levels of control, ranging from tool, assistant, peer, and manager (Gibbs et al. 2021). The Control conditions were designed based on Malone's concept, with the High Control condition allowing for a high amount of human input, thus allowing the autonomous system to act as an assistant. The Medium Control condition allowed for some input by the human operator, but only a limited amount, thus treating the autonomous system as a peer. The Low Control condition did not allow any input from humans and gave the algorithmic system complete autonomy over the decision making process, thus treating the autonomous system as a manager.

For example, in this experiment, High Control groups were given the four Risk scenarios (vehicle, piloting, spell check, and movie recommendation systems), but were able to exercise a high amount of control over the decision making process. On the other hand, Low Control groups were given the four Risk scenarios, but could not control the decision making due to autonomous control by the system. A visual representation of the condition formatting used for this experiment can be found in the appendix (Fig. 2).

Data Analysis

The data was entered into SPSS and all personally identifiable information was removed. An independent sample t-test was performed to evaluate the effects of perceived risk on human trust, and a one-way ANOVA test was performed to evaluate the effects of control on human trust. Additionally, an interaction analysis was performed to evaluate the interaction effects between perceived risk and control.

Results

Manipulation Check

In order to confirm that the participants perceived the risk and control conditions as intended, a manipulation check was included in the experiment. A one-sample t-test verified that participants perceived High Risk and Low Risk scenarios differently (High Risk $M = 2.66$, Low Risk $M = 1.28$, $p < 0.001$). Additionally, three one-sample t-tests were performed in order to verify that participants perceived the Low Control, Medium Control, and High Control conditions differently (Low $M = 2.13$, Medium $M = 3.87$, High $M = 4.16$, $p < 0.001$). The analysis of these manipulation checks affirmed that participants perceived the risk and control conditions as intended.

Risk (H1)

The Checklist for Trust between People and Automation (Jian et al. 2000) was used to measure trust levels in participants; thus, a higher trust score indicates a higher level of trust, and vice versa. An independent sample t-test was used to compare trust scores between High Risk and Low Risk conditions. The results of this test revealed a significant difference in the trust scores between the High Risk condition and the Low Risk condition ($p < 0.001$, High Risk $t = 62.82$, Low Risk $t = 85.46$). The trust scores were higher in the Low Risk condition ($M = 3.85$, $SD = 0.49$) compared to the High Risk condition ($M = 3.35$, $SD = 0.58$). This indicates that overall, participants trusted automated systems more in situations of low risk. Thus, H1 was supported.

Control (H2)

A one-way ANOVA test was performed to compare the effect of High Control, Medium Control, and Low Control conditions on trust scores. Results found a significant difference in trust scores between at least two Control conditions ($F = 8.26$, $p < 0.001$). Tukey's HSD Test for multiple comparisons found that the mean value of trust scores was significantly different between the Low Control and High Control groups ($p < 0.001$), as well as a significant difference between the Low Control and Medium Control groups ($p = 0.006$). There was no statistically significant difference between the Medium Control and High Control groups ($p = 0.36$). This indicates a trend for low trust scores to be associated with a lower level of control. Thus, H2 was supported.

Post Hoc Analysis

General Attitudes Towards Artificial Intelligence and Perceived Control

Several secondary tests, unrelated to the original hypotheses, were also conducted. Four mediation analyses were conducted to analyze the mediating relationship between general attitudes towards Artificial Intelligence and perceived risk, general attitudes towards Artificial Intelligence and perceived control, digital skills and perceived risk, and digital skills and perceived control. From the statistical analyses, only one statistically significant mediating effect was revealed, as the effect of perceived control on trust towards Artificial Intelligence was mediated via the general attitudes towards Artificial Intelligence. This was tested using Hayes's PROCESS MACRO for SPSS (v 4.1). Perceived control was the independent variable, human trust was the dependent variable, and general attitudes towards

Artificial Intelligence was the mediating variable. The MACRO was run using 5,000 bootstrap samples and 95% confidence intervals. There was evidence that general attitudes towards Artificial Intelligence positively mediated the relationship between control and human trust (CI = [0.013, 0.044], $p = 0.003$).

Discussion

This study investigated the effect differing levels of control and risk have on human trust in an automated system. The data from this study contributes to a developing model of factors affecting trust in automation. It was hypothesized that trust levels would be higher in scenarios with higher levels of control as well as lower levels of risk. Overall, the results confirmed the hypothesis that human trust in automated systems is higher in scenarios where humans have greater control in the decision making process. Results also confirmed the hypothesis that human trust in automated systems is higher in scenarios where there is lower perceived risk. Furthermore, a post-hoc analysis found that general attitudes towards Artificial Intelligence positively mediated the relationship between control and human trust, but not control and perceived risk. Ultimately, this experimental study illuminated the ways in which human trust in automated systems may fluctuate based on the situational factors of perceived risk and control.

Theoretical Implications

Current literature regarding algorithm aversion and algorithm appreciation often conflict in their approach to building human trust in automated systems. My findings suggest that regarding perceived risk and control as two important factors in influencing human trust, which may help reconcile the two diverging schools of thought.

Previous scholars have suggested that algorithms may outperform humans when predicting outcomes (Meehl, 1954), yet there are varying results about human inclination to trust automated systems. Dijkstra (1999) found that participants relied on algorithmic advice even after seeing the algorithmic system make an error, directly contradicting the results of Dietvorst et al. (2015). However, these mixed results may be attributable to the difference in perceived risk of the different scenarios given to the participants of each study. Dijkstra (1999) gave participants a scenario regarding a criminal law case and asked to give a verdict of guilty or not guilty, which is a relatively high-risk situation, given the potential risk of punishment towards the defendant in the given case study. On the other hand, Dietvorst et al. (2015) gave participants estimation tasks and $1 bonus rewards if the estimate they selected was accurate. Therefore, the cost of an error of an automated system's decision is greater in Dijkstra's (1999) study as opposed to Dietvorst et al.'s (2015) study. My study suggests that the perceived risk of the situation may have influenced algorithm aversion and algorithm appreciation of participants.

Importantly, the results are consistent with those of Dietvorset et al. (2015) as well as Logg et al. (2019). In those studies as well as this one, participants were more trusting, or demonstrated algorithm appreciation, when they were able to have more control over the decision-making process. Thus, while total autonomous control will often be viewed with more algorithm aversion, the ability to modify the algorithmic input may reduce algorithm aversion and enhance algorithm appreciation.

Practical Implications

Notably, there was no significant difference in trust scores between the Medium Control and High Control groups, even though the manipulation check showed an effective perception of control differences between the two conditions. This may indicate that although trust scores increase with higher control, the rate at which trust scores increase may decrease even with higher control. Therefore, varying the amount of control that a user has over algorithmic systems may be effective to a lesser extent the more control that is given to the user. This corroborates findings by Dietvorst et al. (2016), which suggested that people are less sensitive to the amount that they can modify the algorithmic input, as long as they can participate in the decision making process (Burton et al., 2019).

Thus, even the illusion of autonomy may increase human trust in algorithmic systems and help users overcome algorithm aversion. The CASA paradigm suggests that humans and machines can collaborate as teammates or colleagues, but there still lies a barrier of trust due to the inherent black box nature of algorithmic decision making (Gambino et al., 2020; Reeves & Nass, 1996). The mediating effect of general attitudes towards Artificial Intelligence may be significant in this process. Positive preconceptions of Artificial Intelligence enhanced trust between humans and algorithms in situations where users had more control over the automated system. Perhaps an increase in the usage of automated systems that provide human users with the ability to make inputs could increase trust overall with automated systems, creating a positive feedback loop.

The rich literature surrounding algorithm aversion and algorithm appreciation often focuses on the idea "human versus machine," rather than addressing the idea of a combination of the two (Burton et al., 2019). Designers of algorithmic systems can incorporate these findings into a broader framework of understanding for human trust towards algorithmic systems. For example, designers may opt to create more interactive decision making processes in order to mitigate decreased trust and increase use of automated systems.

Additionally, these findings may be helpful for operators of algorithmic systems to be knowledgeable about the situational factors that influence human trust as use of algorithmic systems proliferates. My findings suggest operators may under-rely on automation in situations with higher perceived risk and lower control, or conversely, over-rely on automation in situations with lower perceived risk and higher control. In order to maximize the benefits provided by automation, operators should calibrate their trust appropriately.

Limitations and Future Research

The current study is only a preliminary investigation into reconciling algorithm aversion and algorithm appreciation, so it has a few limitations worth mentioning in this paper. Firstly, all respondents were from the United States, and most were white (87.5%), as well as female (60.8%), and had attained a Bachelor's degree (44.2%). The lack of diversity suggests that these analyses may not extend towards minority populations, and future research should be done before generalizing the results to the entire population.

Additionally, the order of scenarios were not randomized. Participants only viewed two high risk scenarios (driving an automated car and piloting an automated plane) and two low risk scenarios (using automated spell check and automated movie suggestions), in that order. Thus, this suggests that there may be a possible order effect between conditions due to the lack of randomization.

The experiment also did not include any information about the accuracy or reliability of the automated system. Future research should investigate the accuracy performance of automated systems as a potential mediating variable between perceived risk and human trust. It could be useful in determining how low or high accuracy may impact perceptions of risk, and therefore levels of human trust in the algorithmic systems.

Conclusion

This study aimed to reconcile previous findings regarding algorithm aversion and algorithm appreciation through analyses of perceived risk and control factors. Although algorithmic systems have grown in ubiquity, recent research has faced a lack of clarity that clouds the understanding of human trust towards algorithmic systems. The results of this study have implications for how humans calibrate their trust levels in high risk situations, as well as situations with differing levels of control in automated systems. This paper has found that people that interact with algorithmic systems in situations with high perceived risk tend to have less trust in the system. Additionally, it has been found that people that have less control over automated systems tend to have less trust in the system. Overall, this pattern of results suggests a need for integration of human and algorithmic input into the decision making process, as well as a careful consideration of the perceived risk in deployment of algorithmic systems, in order to provide a more trusting human-algorithm relationship. Given the rapid development of digitalization and datafication into daily life, it is imperative to build a cooperative relationship between humans and the algorithms that can augment our lives.

References

Abdelnour, S., Hasselbladh, H., & Kallinikos, J. (2017). Agency and institutions in organization studies. *Organization Studies, 38*, 1775–1792. https://doi.org/10.1177/0170840617708007

Berger, B., Adam, M., Rühr, A., & Benlian, A. (2021). “Watch me improve - Algorithm aversion and demonstating the ability to learn.” *Business and Information Systems Engineering 63*(1): 55-68.

Burton, J.W., Stein, M., Jensen, T.B. (2018). A systematic review of algorithm aversion in augmented decision making. *Journal of Behavioral Decision Making.*

Carter, L., McBride, M., Ntuen, C. (2012). The impact of personality on nurses’ bias towards automated decision aid acceptance. *International Journal of Information Systems and Change Management, 6*(2).

Chui, M., & Malhotra, S. (2018). *Notes from the AI frontier: AI adoption advances, but foundational barriers remain.* https://www.mckinsey.com/featured-insights/artificial- intelligence/ai-adoption-advances-but-foundational-barriers-remain

Dana, J., & Thomas, R. (2006). In defense of clinical judgment...and mechanical prediction. *Journal of Behavioral Decision Making, 19*(5), 413–428.

Dawes, R. M. (1979). The robust beauty of improper linear models in decision making. *American Psychologist, 34*(7), 571. https://doi.org/10.1037/0003-066x.34.7.571.

Dawes, R. M., Faust, D., & Meehl, P. E. (1989). Clinical versus actuarial judgment. *Science, 243*(4899), 1668–1674. https://doi.org/10.1126/science.2648573

DeSanctis, G., & Poole, M. S. (1994). Capturing the complexity in advanced technol- ogy use: Adaptive structuration theory. *Organization Science, 5,* 121–147. https://doi. org/10.1287/orsc.5.2.121

Dietvorst, B. J., Simmons, J. P., & Massey, C. (2015). Algorithm aversion: People erroneously avoid algorithms after seeing them err. *Journal of Experimental Psychology: General, 144*(1), 114.

Dietvorst, B. J., Simmons, J. P., & Massey, C. (2016). Overcoming algorithm aversion: People will use imperfect algorithms if they can (even slightly) modify them. *Management Science.*

Dijkstra, J. J. (1999). User agreement with incorrect expert system advice. *Behaviour & Information Technology, 18*(6), 399–411. https://doi.org/10.1080/ 014492999118832.

Dijkstra, J. J., Liebrand, W. B., & Timminga, E. (1998). Persuasiveness of expert systems. *Behaviour & Information Technology, 17*(3), 155–163. https://doi.org/10.1080/ 014492998119526.

Dzindolet, M. T., Pierce, L. G., Beck, H. P., & Dawe, L. A. (2002). The perceived utility of human and automated aids in a visual detection task. *Human Factors: The Journal of the Human Factors and Ergonomics Society, 44*(1), 79–94. https://doi.org/10.1518/ 0018720024494856.

Gambino, A., Fox, J., & Ratan, R. A. (2020). Building a stronger CASA: Extending the computers are social actors paradigm. *Human-Machine Communication, 1,* 71–86. https:// doi.org/10.30658/hmc.1.5

Gibbs, J., Kirkwood, G., Fang, C., & Wilkenfeld, J. N. (2021). Negotiating Agency and Control: Theorizing Human-Machine Communication from a Structurational Perspective. *Human-Machine Communication, 2*. https://doi.org/10.30658/hmc.2.8

Giddens, A. (1979). Central problems in social theory: Action, structure and contradiction in social analysis. *University of California Press.*

Giddens, A. (1984). The constitution of society. University of California Press.

Gino, F., & Moore, D. A. (2007). Effects of task difficulty on use of advice. *Journal of Behavioral Decision Making, 20*(1), 21–35. https://doi.org/10.1002/bdm.539.

Hastie, R., & Dawes, R. M. (2001). Rational choice in an uncertain world: The psychology of judgment and decision making. *Sage Publications.*

Jacobson, J., Dobbs-Marsh, J., Liberman, V., & Minson, J. A. (2011). Predicting civil jury verdicts: How attorneys use (and misuse) a second opinion. *Journal of Empirical Legal Studies, 8*(1), 99–119.

Jian, J., Bisantz, A.M., Drury, C.G. (2010). Foundations for an Empirically Determined Scale of Trust in Automated Systems. International Journal of Cognitive Ergonomics, 4(1), 53-71, https://doi.org/10.1207/S15327566IJCE0401_04.

Jung, M. F. & Hou, Y. T. (2021). Who is the Expert? Reconciling Algorithm Aversion and Algorithm Appreciation in AI-Supported Decision Making. *Proceedings of the ACM on Human-Computer Interaction, 5*, 1-22. https://doi.org/10.1145/3479864

Kawaguchi, K. (2020). When Will Workers Follow an Algorithm? A Field Experiment with a Retail Business. https://doi.org/10.1287/mnsc.2020.3599

Latour, B. (2005). Reassembling the Social: An Introduction to Actor-Network Theory. *Oxford University Press.*

Lee, J. & See, K. (2004). Trust in Automation: Designing for Appropriate Reliance. *Human Factors: The Journal of the Human Factors and Ergonomics Society, 46*(1). https://doi.org/10.1518/hfes.46.1.50_30392

Logg, J. M., Minson, J.A., & Moore, D. A. (2018). Algorithm Appreciation: People Prefer Algorithmic to Human Judgement. *Harvard Business School,* 17-086.

Mahmud, H., A.K.M. N.I., Ahmed, S.I., & Smolander, K. (2022). "Wht influences algorithmic decision-making? A systematic leterature review on algorithm aversion." *Technological Forecasting & Social Change, 175*(1-26).

Malone, T. W. (2018). *Superminds: The surprising power of people and computers thinking together.* Little, Brown, and Company.

Meehl, P. E. (1954). Clinical versus statistical prediction: A theoretical analysis and a review of the evidence. Minneapolis, MN: University of Minnesota Press. Meehl, P. E. (1957). When shall we use our heads instead of the formula? *Journal of Counseling Psychology, 4*(4), 268–273. https://doi.org/10.1037/h0047554.

Moon, Y., "Intimate exchanges: Using computers to elicit self-disclosure from consumers," *Journal of Consumer Research, 26*(4), pp. 323 - 339, 2000.

Moore, D. A., & Healy, P. J. (2008). The trouble with overconfidence. *Psychological Review, 115*(2), 502.

Moore, D. A., Tenney, E. R., & Haran, U. (2015). Overprecision in judgment. *The Wiley Blackwell Handbook of Judgment and Decision Making,* 182–209.

Nass, C., Steuer, J., & Tauber, E.R. (1994). "Computers are social actors," in P*roceedings of the SIGCHI Conference on Human Factors in Computing Systems*.

Neff, G., & Nagy, P. (2016). Automation, algorithms, and politics| Talking to Bots: Symbiotic agency and the case of Tay. *International Journal of Communication, 10*, 4915–4931. https://ijoc.org/index.php/ijoc/article/view/6277

Pondy, L. R., & Mitroff, I. I. (1979). Beyond open systems models of organizations. In B. M. Staw (Ed.), *Research in Organizational Behavior, 1*, 3–39.

Raghuram, S., Hill, N. S., Gibbs, J. L., & Maruping, L. M. (2019). Virtual work: Bridging research clusters. *Academy of Management Annals, 13*, 308–341. https://doi.org/10.5465/ annals.2017.0020

Reeves, B., & Nass, C. (1996). Media equation: How people treat computers, television, and new media like real people and places. *Cambridge University Press.*

Schepman, A. & Rodway, P. (2020). Initial validation of the general attitudes towards Artificial Intelligence Scale. *Computers in Human Behavior Reports, 1*, https://doi.org/10.1016/j.chbr.2020.100014

Shaffer, V. A., Probst, C. A., Merkle, E. C., Arkes, H. R., & Medow, M. A. (2013). Why do patients derogate physicians who use a computer-based diagnostic support system? *Medical Decision Making, 33*, 108 –118. http://dx.doi.org/10.1177/0272989X12453501

Sinha, R. R., & Swearingen, K. (2001). Comparing Recommendations Made by Online Systems and Friends. *Personalisation and recommender systems in digital libraries.*

Soll, J. B., & Mannes, A. E. (2011). Judgmental aggregation strategies depend on whether the self is involved. *International Journal of Forecasting, 27*(1), 81–102.

Van Deursen, AJ.A.M., Helsper, E. J., & Eynon, R. (2014). *Measuring Digital Skills: From Digital Skill to Tangible Outcomes project report.*

Yaniv, I., & Kleinberger, E. (2000). Advice taking in decision making: Egocentric discounting and reputation formation. *Organizational Behavior and Human Decision Processes, 83*(2), 260–281.

Appendix

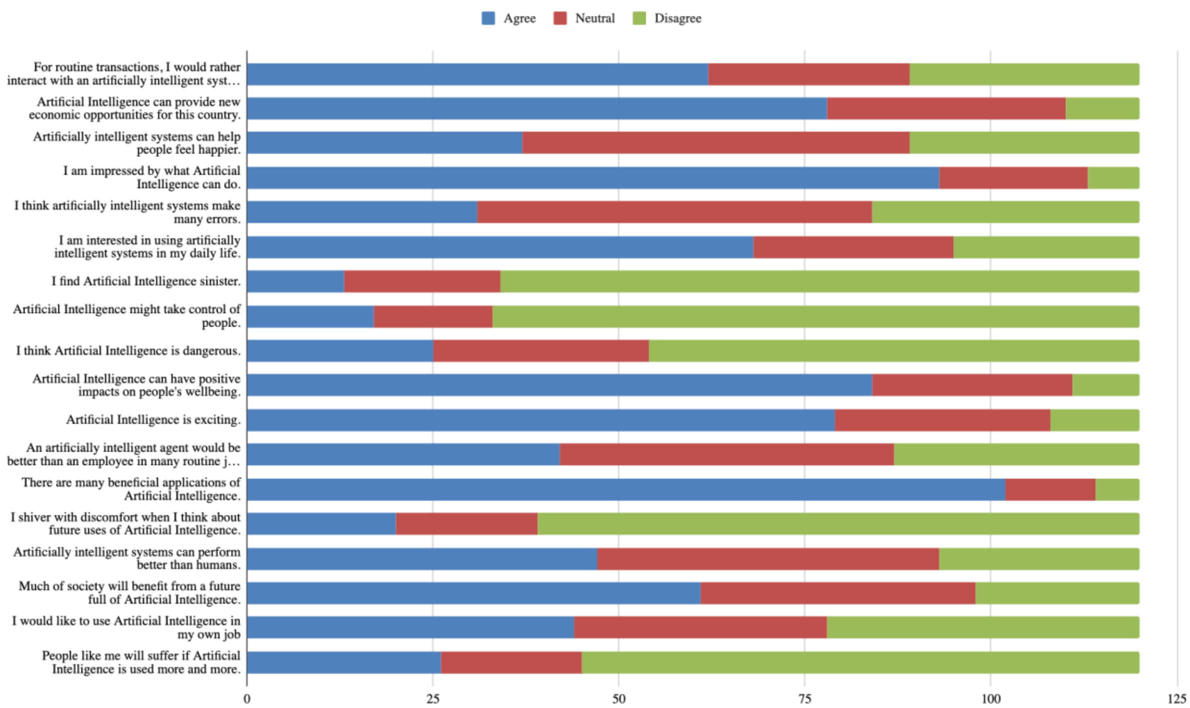

Figure 1. Frequencies of responses to statements in the General Attitudes to Artificial Intelligence questionnaire. Disagreement and agreement combine the “somewhat” and “strongly” categories. Agreement is presented by the blue bars, neutrality is presented by the red bars, and disagreement is presented by the green bars.

	High Control *Assistant*	**Medium Control** *Peer*	**Low Control** *Manager*
High Risk *Driverless Vehicles*	High Control High Risk	Medium Control High Risk	Low Control High Risk
High Risk *Automatic Piloting*	High Control High Risk	Medium Control High Risk	Low Control High Risk
Low Risk *Spellcheck*	High Control Low Risk	Medium Control Low Risk	Low Control Low Risk
Low Risk *Movie Recommendations*	High Control Low Risk	Medium Control Low Risk	Low Control Low Risk

Figure 2. A visual representation of the scenario conditions given to participants in the survey. Participants were put into one of three control groups, High Control, Medium Control, and Low Control. All participants viewed each of the four risk scenarios within their control conditions.