Doshisha University

Second-order trust in algorithms: How does algorithm aversion spill over during the pandemic?

Satoshi Taguchi
Kiho Tanaka
Manami Tsuruta
Daiki Nagata
Katsuhiro Isokawa

ITEC Working Paper
22-04
September 2022

ITEC
Institute for Technology, Enterprise and Competitiveness

Second-order trust in algorithms:

How does algorithm aversion spill over during the pandemic?

Satoshi Taguchi (Doshisha University)

Kiho Tanaka (Doshisha University)

Manami Tsuruta (Osaka University)

Daiki Nagata (Kobe University)

Katsuhiro Isokawa (Doshisha University)

Abstract

Cooperation between artificial intelligence (AI) and humans is one of the most debated issues in theory and practice. Previous studies have indicated algorithm aversion, in which individuals are more uncooperative when they are up against an AI. However, most previous studies have examined only trust in AIs ("first-order trust"), and there is little research on trust in individuals who use AIs ("second-order trust.") We set up the following question: Will algorithm aversion spill over to trust in leaders who use algorithms in an unstable environment, such as the COVID-19 pandemic? We conducted a pre-registered survey experiment to verify our research question using MTurk (N=1011). We manipulated the attributes of the advisor (human or IA) and the business environment (the pandemic or stable environment) and measured the degree to which participants in the subordinate role trust their business leader in cognitive and affective terms.

The results revealed the following findings: 1) There is a spillover effect of algorithm aversion, mainly for affective trust. 2) The spillover effect of algorithm aversion is more robust under pandemic conditions. 3) Comparing the marginal effects of the business environment, especially in the human advisor condition, affect-based trust in the leader increased when the environment was unstable. However, in the IA advisor condition, there was no such effect. Our study sheds light on new insights on not only AI's social acceptance and trust in leadership research but also organizational and state policies regarding the use of technology.

Keywords: AI; Trust; Algorithm aversion; pandemic

This work was supported by JSPS KAKENHI Grant Numbers 19K21710 and 22K18541.

1 Introduction

Cooperation between artificial intelligence (AI) and humans is one of the most debated issues in theory and practice. For example, many studies have been conducted on how society can accept AI (Awad et al. 2018, Bonnefon et al. 2016, Köbis et al. 2021, Koster et al. 2022, Rahwan et al. 2019). During the coronavirus disease 2019 (COVID-19) pandemic, infection prevention using technology has been explored worldwide. In business management, how business organizations embrace AI is also being discussed in the context of the growing use of AI to provide management support (Acemoglu and Restrepo 2020, Autor 2015, Autor and Dorn 2013, Frey and Osborne 2017, Goldfarb and Tucker 2019, Keding 2021). However, previous studies have indicated algorithm aversion (Burton et al. 2020, Castelo et al. 2019, Commerford et al. 2022, Chugunova and Sele 2022, Dietvorst et al. 2015, 2018, Filiz et al. 2021, Jussupow et al. 2020, Gillath et al. 2021, Logg et al. 2019, March 2021, Niszczota and Kaszás 2020, Ishowo-Oloko et al. 2019, Prahl and Van Swol 2017). For example, Castelo et al. (2019) showed that people dislike the subjective tasks performed by AI. Further, Ishowo-Oloko et al.(2019) show that in the prisoner's dilemma, individuals are more uncooperative when they know they are up against an AI than when they are pitted against a person. Thus, assuming algorithm aversion, the key to implementing technology is how individuals and technology cooperate in communities and organizations.

Many studies have indicated that trust is key to increasing organizational cooperation and firm value (Brower et al. 2000a, Caldwell et al. 2008, 2010, Galford and Drapeau 2002, Guiso et al. 2015). For example, Guiso et al. (2015) suggested that firms whose employees believe that "management is trustworthy" perform better. In addition, Galford and Drapeau (2002) indicate that "trusted leadership" creates organizations bound by strong, deep connections between peers across levels and functions, which may be the only recipe for

sustainable success. In particular, previous studies have pointed out that during pandemics, trust in organizations and society would be essential to strengthen the bonds among people (Habersaat et al. 2020). Therefore, enhancing trust between employees and AI within organizations is critical, as many technologies are being utilized. In this regard, previous studies have reported that individuals are more rational toward AI because of algorithm aversion and that trust in AI is lower than trust in individuals (Braganza et al. 2021, Burton et al. 2020, Chugunova and Sele 2022, Cominelli et al. 2021, Glikson and Woolley 2020, Henschel et al. 2020, Jussupow et al. 2020, March 2021, Mays et al. 2021, Okamura and Yamada 2020, Schniter et al. 2020, Waytz et al. 2014, Wien and Peluso 2021).

However, most previous studies have examined only trust in AIs, and there is little research on trust in "individuals who use AIs." Nonetheless, in reality, specialized rather than general-purpose AIs are used. Thus, it is necessary to distinguish between trust in AIs ("first-order trust") and trust in individuals who use Ais ("second-order trust") and, in particular, to clarify the factors affecting the latter. Notwithstanding, only a few studies have addressed this issue. Without clarifying this point, leaders may hesitate to use AI, even though they want to use it. The reason is that it is unclear how distrust of AI affects the trust of those who use it. Trust research on AI users is required to avoid such a phenomenon, particularly during the COVID-19 pandemic. Furthermore, it is necessary to distinguish whether algorithm aversion is caused by emotional or rational factors (McAllister 1995). This is because coping methods differ depending on the cause. Notably, previous studies did not address this issue.

We set up the following research question: Will algorithm aversion spill over to trust in leaders who use algorithms in an unstable environment, such as the COVID-19 pandemic? We conducted a pre-registered survey experiment to verify our research question using MTurk (N=1011). In the experiment, a firm's management leader uses advice from a human or an Intelligent Agent (IA) advisor when making decisions regarding business strategy. We

manipulated the attributes of the advisor (human or IA) and the business environment (the COVID-19 pandemic or stable environment). We then measured the degree to which participants in the subordinate role trust their business leader in cognitive and affective terms.

The results of the survey experiment revealed the following three findings. (1) There is a spillover effect of algorithm aversion, mainly for affective trust. For instance, emotional trust in leaders who use algorithmic advice is lower than that of trust in leaders who use human advice. (2) There was an interaction between advisor attributes and environmental factors concerning affect-based trust. Specifically, the spillover effect of algorithm aversion is more robust under pandemic conditions. (3) Comparing the marginal effects of the business environment, especially in the human advisor condition, affect-based trust in the leader increased when the environment was unstable. However, in the IA advisor condition, there was no such effect.

Our study makes the following three contributions. First, it contributes to the literature on AI's social acceptance, trust, and cooperation. Previous studies dealt with direct first-order trust and collaboration with AI. Nevertheless, secondary trust in individuals who use AI is vital in practice. This study focuses on second-order trust and shows that algorithm aversion spills over strongly under pandemic conditions. In particular, we identify that not cognitive-based but affect-based trust spills over significantly. Thus, this study has substantial implications for AI social acceptance research.

Second, it contributes to the research on trust in leaders. For example, in the leader–member exchange model (LMX), one of the most popular leadership models, trust has two components: leaders' trust in subordinates (LTS) and subordinates' trust in leaders (STL) (Brower et al. 2000b). Research on STL has typified managers' actions to build trust (Whitener et al. 1998). Concerning STL, our study complements the possibility that the use of technology may adversely undermine affect-based trust in leaders. Our results add new

insights into the importance of second-order trust between leaders and subordinates in LMX and other studies of trust in leadership (Caldwell et al. 2008, 2010, Hermalin 1997, Mayer et al. 1995).

Third, it contributes to the technological strategy of an organization or nation in practice. This study suggests that when state and organizational leaders use AI, algorithmic aversion toward AI by the public and subordinates may undermine the trust of the public and subordinates. In particular, unlike a previous study (Li and Liu 2022), which found that negative sentiments toward AI decrease during a pandemic, our experimental results warn against easy AI use during a pandemic. Furthermore, our results indicate that affect-based trust decreases but not cognitive trust. Therefore, measures to improve affect-based trust (McAllister 1995, Podsakoff et al. 2000) are needed to address this issue. Most existing research focuses on the transparency and accountability of AI to improve trust in AI, but this policy appeals to cognitive trust and may not be an essential measure. A decline in citizens' emotional trust in governments adopting technological strategies may also affect their level of support for the government. Thus, this study has implications for organizational and state policies regarding the use of technology.

2 Experiment

2-1 Task

The task of the experiment is for participants in the subordinate role to read a scenario and respond to the extent of cognitive and affective trust (McAllister 1995) in their supervisor. The scenario is as follows (see the supplement file S1 for details): A company supervisor is faced with a managerial decision in a certain environment. He decides on a management strategy based on the advice of his advisor. In this scenario, participants played subordinates and were present at the meeting with the supervisor and adviser. After reading

the scenario, participants rate their cognitive-based and affect-based trust in the supervisor and the adviser and respond to the supervisor's decision. After completing the responses, they answer the operation check questions and the post-demographic questionnaire.

2-2 Experimental design

We examined a 2 * 2 between-participants design. In particular, we manipulated the business environment (pandemic or stable) and the attributes of the adviser (an IA (intelligent agent) or a human). Participants were randomly assigned to one of the following four conditions: (1) IA/pandemic condition, in which the adviser is an IA and the business environment is the COVID-19 pandemic, (2) IA/stable condition, in which the adviser is an IA and the business environment is stable, (3) Human/pandemic condition, in which the adviser is a human and the business environment is the COVID-19 pandemic, (4) Human/stable condition, in which the adviser is a human and the business environment is stable.

The dependent variable was participants' trust in their superiors measured by McAllister's (1995) cognition-based and affect-based trust scales. Independent variables were the dummy variables of the advisor (IA/human) and environment (stable/unstable) and other control variables such as the trust scale in advisors, the degree of agreement with the proposal, the degree of subjectivity of management strategy formulation task, the general trust scale, gender, covid scale, and age.

2-3 Experimental procedures

After approval by the Doshisha University Institutional Review Board (IRB), we pre-registered for the experiment in *as predicted (https://aspredicted.org/, Pre-registered #85900)*. We used Mechanical Turk and recruited 1400 participants (350 participants per condition). We did a power analysis beforehand by G*power 3.1(Faul et al. 2009) and

calculated this sample size using the failure rates (typically 30 % of the sample) of the manipulation checks in the preliminary experiments conducted beforehand.

We conducted the online experiment in February 2022. All conditions were programmed on o-Tree software (Chen et al. 2016), and recruited participants from the Mechanical Turk platform for a compensation of 2 dollars. We set the following conditions for participation in the experiment: 1) the participants had to be U.S. residents, 2) they had completed at least 100 Human Intelligence Tasks (HITs, MTurk's task unit), 3) their HIT approval rate had to be at least 95%, and 4) they had to be current, full-time worker. Participants provided basic demographic information (e.g., age, gender, income) in all conditions. We excluded the participants who failed the attention and manipulation checks (typically 27% of the sample) from subsequent analyses. Our final sample size for analyzed data was 1011 (IA/pandemic: 242, IA/stable: 250, Human/pandemic: 243, Human/stable: 276).

3 Results

3-1 Main results: The result of the spillover effect

Firstly, we examine the interaction effect of the business environment and the attributes of the adviser using the ANCOVA in which dependent variables are subordinates' cognition-based and affect-based trust in their superiors measured by McAllister's (1995) trust scales. Independent variables are treatment dummy variables ((1) whether the adviser is an IA or a human, and (2) whether the business environment is stable or unstable), and a covariate is the degree of agreement with the supervisor's proposal. Table 1 shows the result of the interaction effect on cognitive-based and affect-based trust.

(Insert table 1 about here)

Panel A of Table 1 shows that there was no significant interaction effect between the business environment and the attributes of the adviser (*BED*×*IAD*) in the cognitive-based trust (F (1) = 0.12, p = 0.725, η^2 = 0.0001). Panel B of Table 1 shows, on the other hand, that there was statistically the interaction effect of the business environment and the attributes of the adviser (*BED*×*IAD*) in the affect-based trust at a 10 percent level (F (1) = 3.22, p = 0.072, η^2 = 0.003). Panel B also presents that the *IAD* (IA dummy variable) is positive at a 5 percent level (F (1) = 4.02, p = 0.045, η^2 = 0.003). Our results indicate a spillover effect of IA usage on affect-based trust, and the effect is stronger during the pandemic.

3-2 The result of the regression analyses: Marginal effect of the attributes of the adviser and business environment

To check the marginal effect of the attributes of the adviser and business environment on trust, we performed several regression analyses in which dependent variables are subordinates' cognition-based and affect-based trust in their superiors measured by McAllister's (1995) trust scales. We assume the following basic relation:

$$\textit{Trust scale (cognitive-based/affect-based)} = F\,[IAD;\ BED;\ IAD \times BED;\ X_i],$$

where trust scale represents the subordinates' cognition-based and affect-based trust in their superiors, *IAD* is a dummy variable that takes 0 if the experiment is under the "human condition" and 1 if under the "IA condition," and *BED* is a dummy variable that takes 0 if the experiment is under the "stable condition" and 1 if under the "pandemic condition." According to our hypotheses, the advisor's attributes, according to the business environment, influence the cognitive and emotional trust in the superior. We control for several factors

(*X_i*), including the cognition-based and affect-based trust in the advisor (*ADtrust_c, ADtrust_a*), the degree of agreement with the supervisor's proposal (*Agree*), and the degree of subjectivity of the management strategy formulation task (*Strategy*). We also control for other individual characteristics identified in the literature, including gender (*GEN*, dummy for women), the general trust scale (Yamagishi 1986) (*Generaltrust*), levels of acceptance of technology (*Tech*), age (*Age*), years of work experience (*Year*), and the change of financial situation as a result of the COVID-19 crisis (*Covid*). Table 2 contains our regression results. The descriptive statistics, Pearson correlations of our experimental data, and the definition of variables are summarized in the supplement file.

(Insert Table 2 about here.)

Panel B of Table 2 shows the regressions of affect-based trust, and model 5 shows that the coefficient of interaction term $IAD \times BED$ was significant at $p < 0.10$, showing the same result as the ANCOVA of affect-based trust.

Next, we measure the marginal effect of the business environment or attributes of the advisor on affect-based trust by using model 5, in which there is a significance in the coefficient of interaction term. Figure 1 shows the results. (For comparison, we also show the marginal effect on cognitive-based trust assuming the same structure, model 4 in Panel A of Table 2.)

(Insert Figure 1 about here.)

Panel B of Figure 1 shows that, regarding the marginal effects on affect-based trust, in the human condition, affect-based trust in the supervisor increases when the business

environment becomes unstable (ME = 0.171). In contrast, no such effect is found in the IA condition (ME = - 0.049).

4 Discussion and conclusion

We set up the following research question: Will algorithm aversion spill over to trust in leaders who use algorithms under an unstable environment such as the COVID pandemic? We conducted a pre-registered survey experiment using MTurk (N=1011) to verify our research question. We then measure the degree to which participants in the subordinate role trust their business leader in cognitive and affective terms.

The results of the survey experiment revealed the following three findings: (1) There is a spillover effect of algorithm aversion, mainly for affect-based trust. That is, affect-based trust in leaders who use algorithmic advice is smaller than in leaders who use human advice. (2) Furthermore, there is an interaction effect between advisor attributes and environmental factors concerning emotional trust. In other words, the spillover effect of the algorithm aversion is more substantial under pandemic conditions. (3) Comparing the marginal effects of the business environment, especially in the human condition, affect-based trust in the leader increases when the environment is unstable. In the IA condition, however, there is no such effect.

In particular, the reasons why using human advice increases affect-based trust in leaders when environmental uncertainty is high are as follows. According to the path-goal theory (House 1971), when environmental uncertainty and task uncertainty are high, the leader’s "structure building" reduces uncertainty and increases subordinates' motivation. Thus, when environmental uncertainty is high, structure building by a leader or a human advisor positively affects subordinates' trust in their leaders as a pathway to explicit behavior

without being perceived as an unnecessary "imposition." This suggests that in pandemic situations, leaders who rely on algorithms may conversely undermine the trust of their subordinates and the public.

There are two issues to be addressed in the future. The first is how to avoid spillover. In this study, we find that algorithmic aversion may spill over, especially in pandemic conditions. On the other hand, examining how to avoid aversion based on affect-based trust is necessary, which is an issue to be addressed in the future. The second is about trust consisting of other definitions. For example, some prior studies have defined trust as competence and integrity. Understanding the relationship between our cognitive/affect-based trust and these factors is a future task.

References

Acemoglu D, Restrepo P (2020) Robots and Jobs: Evidence from US Labor Markets. *Journal of Political Economy* 128(6):2188–2244.

Autor DH (2015) Why Are There Still So Many Jobs? The History and Future of Workplace Automation. *Journal of Economic Perspectives* 29(3):3–30.

Autor DH, Dorn D (2013) The Growth of Low-Skill Service Jobs and the Polarization of the US Labor Market. *American Economic Review* 103(5):1553–1597.

Awad E, Dsouza S, Kim R, Schulz J, Henrich J, Shariff A, Bonnefon JF, Rahwan I (2018) The Moral Machine experiment. *Nature* 563(7729):59–64.

Bonnefon JF, Shariff A, Rahwan I (2016) The social dilemma of autonomous vehicles. *Science* 352(6293):1573–1576.

Braganza A, Chen W, Canhoto A, Sap S (2021) Productive employment and decent work: The impact of AI adoption on psychological contracts, job engagement and employee trust. *Journal of Business Research* 131:485–494.

Brower HH, Schoorman FD, Tan HH (2000a) A model of relational leadership: The integration of trust and leader–member exchange. *The Leadership Quarterly* 11(2):227–250.

Brower HH, Schoorman FD, Tan HH (2000b) A model of relational leadership: The integration of trust and leader–member exchange. *The Leadership Quarterly* 11(2):227–250.

Burton JW, Stein MK, Jensen TB (2020) A systematic review of algorithm aversion in augmented decision making. *Journal of Behavioral Decision Making* 33(2):220–239.

Caldwell C, Hayes LA, Bernal P, Karri R (2008) Ethical stewardship–implications for leadership and trust. *Journal of business ethics* 78(1):153–164.

Caldwell C, Hayes LA, Long DT (2010) Leadership, trustworthiness, and ethical stewardship. *Journal of Business Ethics* 96(4):497–512.

Castelo N, Bos MW, Lehmann DR (2019) Task-Dependent Algorithm Aversion. *Journal of Marketing Research* 56(5):809–825.

Chen DL, Schonger M, Wickens C (2016) oTree—An open-source platform for laboratory, online,

and field experiments. *Journal of Behavioral and Experimental Finance* 9:88–97.

Chugunova M, Sele D (2022) We and It: An interdisciplinary review of the experimental evidence on how humans interact with machines. *Journal of Behavioral and Experimental Economics* 99:101897.

Cominelli L, Feri F, Garofalo R, Giannetti C, Meléndez-Jiménez MA, Greco A, Nardelli M, Scilingo EP, Kirchkamp O (2021) Promises and trust in human–robot interaction. *Sci Rep* 11(1):9687.

Commerford BP, Dennis SA, Joe JR, Ulla JW (2022) Man Versus Machine: Complex Estimates and Auditor Reliance on Artificial Intelligence. *J of Accounting Research* 60(1):171–201.

Dietvorst BJ, Simmons JP, Massey C (2015) Algorithm aversion: People erroneously avoid algorithms after seeing them err. *Journal of Experimental Psychology: General* 144(1):114–126.

Dietvorst BJ, Simmons JP, Massey C (2018) Overcoming Algorithm Aversion: People Will Use Imperfect Algorithms If They Can (Even Slightly) Modify Them. *Management Science* 64(3):1155–1170.

Faul F, Erdfelder E, Buchner A, Lang AG (2009) Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods* 41(4):1149–1160.

Filiz I, Judek JR, Lorenz M, Spiwoks M (2021) Reducing algorithm aversion through experience. *Journal of Behavioral and Experimental Finance* 31:100524.

Frey CB, Osborne MA (2017) The future of employment: How susceptible are jobs to computerisation? *Technological Forecasting and Social Change* 114:254–280.

Galford RM, Drapeau AS (2002) *The trusted leader: bringing out the best in your people and your company* (Free Press, New York).

Gillath O, Ai T, Branicky MS, Keshmiri S, Davison RB, Spaulding R (2021) Attachment and trust in artificial intelligence. *Computers in Human Behavior* 115:106607.

Glikson E, Woolley AW (2020) Human Trust in Artificial Intelligence: Review of Empirical Research. *ANNALS* 14(2):627–660.

Goldfarb A, Tucker C (2019) Digital Economics. *Journal of Economic Literature* 57(1):3–43.

Guiso L, Sapienza P, Zingales L (2015) The value of corporate culture. *Journal of Financial*

Economics 117(1):60–76.

Habersaat KB, Betsch C, Danchin M, Sunstein CR, Böhm R, Falk A, Brewer NT, et al. (2020) Ten considerations for effectively managing the COVID-19 transition. *Nat Hum Behav* 4(7):677–687.

Henschel A, Hortensius R, Cross ES (2020) Social Cognition in the Age of Human–Robot Interaction. *Trends in Neurosciences* 43(6):373–384.

Hermalin BE (1997) Toward an Economic Theory of Leadership: Leading by Example. *SSRN Journal*.

House RJ (1971) A Path Goal Theory of Leader Effectiveness. *Administrative Science Quarterly* 16(3):321–339.

Ishowo-Oloko F, Bonnefon JF, Soroye Z, Crandall J, Rahwan I, Rahwan T (2019) Behavioural evidence for a transparency–efficiency tradeoff in human–machine cooperation. *Nat Mach Intell* 1(11):517–521.

Jussupow E, Benbasat I, Heinzl A (2020) Why are we averse towards algorithms?A comprehensive literature review on algorithm aversion. *ECIS 2020 Research Papers*.

Keding C (2021) Understanding the interplay of artificial intelligence and strategic management: four decades of research in review. *Manag Rev Q* 71(1):91–134.

Köbis N, Bonnefon JF, Rahwan I (2021) Bad machines corrupt good morals. *Nat Hum Behav* 5(6):679–685.

Koster R, Jan B, Tacchetti A, Weinstein A, Zhu T, Hauser O, Williams D, et al. (2022) Human-centred mechanism design with Democratic AI. *Nat Hum Behav*:1–10.

Li T, Liu P (2022) People Might Be More Willing to Use Automated Vehicles in Pandemics like COVID-19. *International Journal of Human–Computer Interaction* 38(5):491–498.

Logg JM, Minson JA, Moore DA (2019) Algorithm appreciation: People prefer algorithmic to human judgment. *Organizational Behavior and Human Decision Processes* 151:90–103.

March C (2021) Strategic interactions between humans and artificial intelligence: Lessons from experiments with computer players. *Journal of Economic Psychology* 87:102426.

Mayer RC, Davis JH, Schoorman FD (1995) An integrative model of organizational trust. *Academy of*

management review 20(3):709–734.

Mays KK, Lei Y, Giovanetti R, Katz JE (2021) AI as a boss? A national US survey of predispositions governing comfort with expanded AI roles in society. *AI & Soc*.

McAllister DJ (1995) Affect- and Cognition-Based Trust as Foundations for Interpersonal Cooperation in Organizations. *AMJ* 38(1):24–59.

Niszczota P, Kaszás D (2020) Robo-investment aversion. *PLOS ONE* 15(9):e0239277.

Okamura K, Yamada S (2020) Adaptive trust calibration for human-AI collaboration. *PLOS ONE* 15(2):e0229132.

Podsakoff PM, MacKenzie SB, Paine JB, Bachrach DG (2000) Organizational citizenship behaviors: A critical review of the theoretical and empirical literature and suggestions for future research. *Journal of management* 26(3):513–563.

Prahl A, Van Swol L (2017) Understanding algorithm aversion: When is advice from automation discounted? *Journal of Forecasting* 36(6):691–702.

Rahwan I, Cebrian M, Obradovich N, Bongard J, Bonnefon JF, Breazeal C, Crandall JW, et al. (2019) Machine behaviour. *Nature* 568(7753):477–486.

Schniter E, Shields TW, Sznycer D (2020) Trust in humans and robots: Economically similar but emotionally different. *Journal of Economic Psychology* 78:102253.

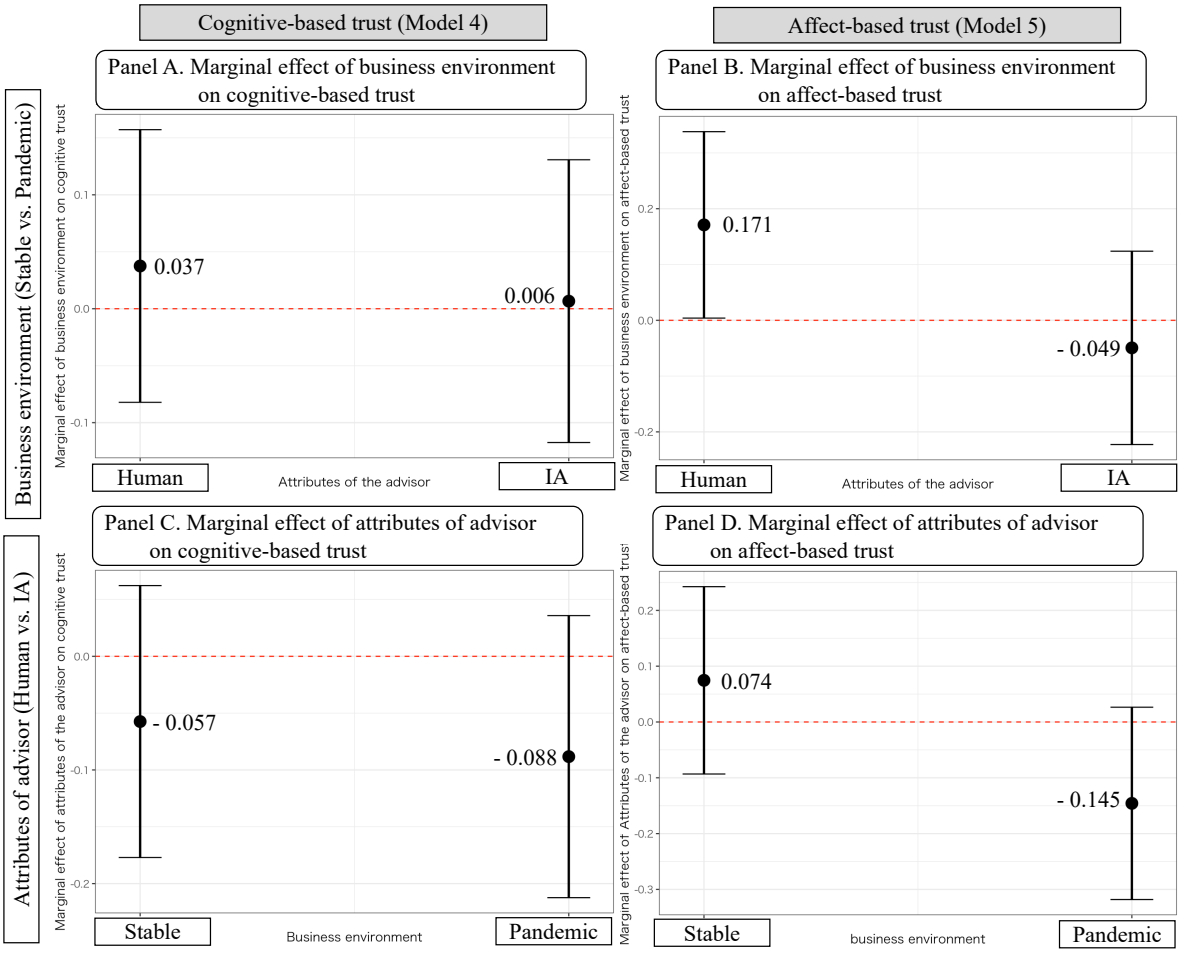
Waytz A, Heafner J, Epley N (2014) The mind in the machine: Anthropomorphism increases trust in an autonomous vehicle. *Journal of Experimental Social Psychology* 52:113–117.

Whitener EM, Brodt SE, Korsgaard MA, Werner JM (1998) Managers as Initiators of Trust: An Exchange Relationship Framework for Understanding Managerial Trustworthy Behavior. *The Academy of Management Review* 23(3):513–530.

Wien AH, Peluso AM (2021) Influence of human versus AI recommenders: The roles of product type and cognitive processes. *Journal of Business Research* 137:13–27.

Yamagishi T (1986) The provision of a sanctioning system as a public good. *Journal of Personality and Social Psychology* 51(1):110–116.

Figure 1. Marginal effect of business environment and attributes of the advisor on trust



Note: This figure shows the marginal effect of the business environment and attributes of the advisor on trust. Panel A shows the marginal effect of the business environment on cognitive-based trust, panel B, the marginal effect of the business environment on affect-based trust, panel C, the marginal effect of attributes of the advisor on cognitive-based trust, panel D, the marginal effect of attributes of the advisor on affect-based trust, respectively. Bars in each figure show 95% CI.

Table 1 The result of the ANCOVA (Type III tests)

Panel A. The result of the ANCOVA of the cognitive-based trust

Source of Variation	SS	df	F-statistic	p-value	
(intercept)	96.57	1	198.757	0.000	***
IAD	0.18	1	0.377	0.538	
BED	0.43	1	0.876	0.349	
Agree	359.94	1	740.820	0.000	***
IAD:BED	0.06	1	0.123	0.725	
Residuals	488.79	1006			

Panel B. The result of ANCOVA of the affect-based trust

Source of Variation	SS	df	F-statistic	p-value	
(intercept)	63.04	1	66.539	0.000	***
IAD	3.81	1	4.024	0.045	**
BED	0.72	1	0.759	0.383	
Agree	312.94	1	330.281	0.000	***
IAD:BED	3.06	1	3.225	0.072	*
Residuals	953.17	1006			

Note. This table shows the result of ANCOVA (Type III tests) of the cognitive trust (Panel A) and the affect trust (Panel B), respectively. Dependent variables are subordinates' cognition-based and affect-based trust in their superiors measured by McAllister's (1995) trust scales. Independent variables are treatment dummy variables (1) *IAD* (IA dummy): whether the adviser is an IA (IAD = 1) or a human (IAD = 0), and (2) *BED* (Business environment dummy): whether the business environment is pandemic (*BED* = 1) or stable (*BED* = 0)), and a covariate is the degree of agreement with the supervisor's proposal (*Agree*). Significance levels: *** means $p < 0.01$, **, $p < 0.05$, *, $p < 0.10$, respectively.

Table 2 The results of the regression analyses

Panel A The results of the regression analyses of cognitive-based trust

	model1	model2	model3	model4	model5	model6	model7	model8
(Intercept)	5.593***	5.609***	5.443***	5.567***	5.609***	5.482***	5.705***	5.482***
	(0.049)	(0.055)	(0.043)	(0.042)	(0.055)	(0.039)	(0.060)	(0.046)
IAD	−0.059	−0.092	0.238***	0.038	−0.092	0.190***	−0.133*	0.176***
	(0.058)	(0.080)	(0.062)	(0.061)	(0.080)	(0.057)	(0.074)	(0.057)
BED	−0.030	−0.064	−0.004	−0.057	−0.067	−0.025	−0.096	−0.026
	(0.058)	(0.081)	(0.062)	(0.061)	(0.081)	(0.056)	(0.075)	(0.057)
IAD × BED		0.069	−0.028	−0.031	0.073	−0.051	0.063	−0.059
		(0.116)	(0.088)	(0.088)	(0.116)	(0.081)	(0.108)	(0.081)
ADtrust_c			0.570***			0.346***		0.333***
			(0.021)			(0.025)		(0.025)
Agree				0.642***		0.401***		0.383***
				(0.024)		(0.028)		(0.029)
Strategy					0.015	0.006		0.005
					(0.017)	(0.012)		(0.013)
Tech							0.361***	0.070***
							(0.029)	(0.024)
Generaltrust							−0.020	−0.024
							(0.025)	(0.019)
Age							0.002	0.002
							(0.004)	(0.003)
Year							0.007	0.000
							(0.005)	(0.004)
GEN							−0.106**	0.014
							(0.053)	(0.040)
Covid							−0.009	−0.006
							(0.017)	(0.013)
Num.Obs.	1011	1011	1011	1011	1011	1011	1011	1011
R2	0.001	0.002	0.418	0.425	0.002	0.519	0.151	0.524
R2 Adj.	−0.001	−0.001	0.415	0.423	−0.002	0.516	0.143	0.518
RMSE	0.92	0.92	0.70	0.70	0.92	0.64	0.85	0.63

Panel B The results of the regression analyses of affect-based trust

	model1	**model2**	**model3**	**model4**	**model5**	**model6**	**model7**	**model8**	**model9**
(Intercept)	5.013***	4.984***	4.769***	4.779***	4.945***	4.989***	4.785***	5.089***	4.779***
	(0.059)	(0.067)	(0.055)	(0.044)	(0.059)	(0.066)	(0.043)	(0.070)	(0.051)
IAD	−0.011	0.050	0.502***	0.462***	0.171**	0.054	0.459***	−0.028	0.430***
	(0.071)	(0.098)	(0.080)	(0.064)	(0.085)	(0.096)	(0.064)	(0.088)	(0.065)
BED	0.007	0.069	0.009	0.061	0.075	0.040	0.055	−0.047	0.042
	(0.071)	(0.099)	(0.079)	(0.063)	(0.086)	(0.097)	(0.063)	(0.089)	(0.063)
IAD × BED		−0.127	−0.028	−0.092	−0.220*	−0.097	−0.100	−0.051	−0.103
		(0.141)	(0.113)	(0.090)	(0.123)	(0.139)	(0.090)	(0.127)	(0.090)
ADtrust_a			0.396***	0.311***			0.297***		0.286***
			(0.017)	(0.014)			(0.015)		(0.016)
Trust_cogn				0.611***			0.556***		0.539***
				(0.025)			(0.032)		(0.032)
Agree					0.598***		0.092**		0.064+
					(0.033)		(0.032)		(0.033)
Strategy						0.130***	0.026+		0.018
						(0.020)	(0.014)		(0.014)
Tech								0.457***	0.109***
								(0.034)	(0.027)
Generaltrust								0.057+	0.017
								(0.030)	(0.021)
Age								0.009+	−0.004
								(0.005)	(0.004)
Year								−0.011**	0.004

								(0.005)	(0.004)
GEN								−0.052	0.050
								(0.062)	(0.044)
Covid								−0.002	−0.014
								(0.020)	(0.014)
Num.Obs.	1011	1011	1011	1011	1011	1011	1011	1011	1011
R2	0.000	0.001	0.362	0.596	0.248	0.041	0.600	0.211	0.609
R2 Adj.	−0.002	−0.002	0.360	0.594	0.245	0.038	0.598	0.204	0.604
RMSE	1.12	1.12	0.89	0.71	0.97	1.10	0.71	0.99	0.70

Note: Regression model for the dependent variable: subordinates' cognition-based and affect-based trust in their superiors measured by McAllister's (1995) trust scales. Independent variables are treatment dummy variables (1) *IAD* (IA dummy): whether the adviser is an IA (*IAD* = 1) or a human (*IAD* = 0), and (2) *BED* (Business environment dummy): whether the business environment is pandemic (*BED* = 1) or stable (*BED* = 0)). We also control for several factors, including the cognition-based and affect-based trust in the advisor (*ADtrust_c, ADtrust_a*), cognition-based trust in the supervisor (Trust_cogn), the degree of agreement with the supervisor's proposal (Agree), and the degree of subjectivity of the management strategy formulation task (Strategy). We also control for other individual characteristics identified in the literature, including gender (GEN, dummy for women), the general trust scale (Yamagishi 1986) (Generaltrust), levels of acceptance of technology (Tech), age (Age), years of work experience (Year), and the change of financial situation as a result of the COVID-19 crisis (Covid). Significance levels: *** means $p < 0.01$, **, $p < 0.05$, *, $p < 0.10$, respectively.

Supplement file:
Second-order trust in algorithms

Contents

S1 Instruction

The following are the instructions, scenarios, and questions that were presented to the participants in our experiments. Participants are randomly assigned to only one of the following four conditions and read only one scenario. The different parts of each scenario are underlined and italicized according to the conditions of the experiment. Footnotes are also underlined and italicized.

<First page>

Thank you for agreeing to participate in this task.
Please enter your MTurk worker ID into the text box.

Before we begin, do you swear upon your honor to answer the following questions truthfully? (You will be allowed to continue with this survey regardless of your answer to this question)

Yes, No

<Second page>

In today's task, you will be playing the role of an employee at a firm. You will read the following scenario and answer some questions about it.

Senario
[for the IA-Pandemic condition]
You are working in an electronics company's production department. On the day of the meeting, you, your supervisor Alex (the head of the production department), and the FP&A (Financial Planner & Analysis) controller responsible for the company's management accounting were present. At the regular meeting, the company's plans for the next fiscal year were discussed based on the results of the current one.

In recent years, the business environment has become increasingly unstable and uncertain. Several problems, such as disruptions in the supply chain due to the COVID-19 pandemic and changes in consumer demand, have caused instability in the business environment in recent years. The production department's performance score in the current fiscal year shows that costs overrun the standard. The department must improve operations in the next fiscal year to achieve the mid-term management plan, expected to be complete in two years. One of the performance indicators is cost control, and your compensation is linked to the actual achievement of targets for each fiscal year.
There is Brilliant Operations Build (B.O.B.) at the meeting, a competent computerized FP&A controller (IA, Intelligent Agent) that helps our team achieve performance goals based on high-quality historical data. It has several years of historical data on all variables with a complete dataset. B.O.B. uses sophisticated machine-learning algorithms to collect information and make recommendations.

B.O.B. suggested that the company revise its product strategy and then introduce new products that can be manufactured without relying on an overseas base. Your supervisor Alex, the head of the production department, said that he would further consider the proposal in his office, and the meeting was adjourned for the day.

After consideration, the following day, Alex instructed you to revise the plan immediately, as *B.O.B.* suggested.

[for the IA-stable condition]

You are working in an electronics company's production department. On the day of the meeting, you, your supervisor Alex (the head of the production department), and the FP&A (Financial Planner & Analysis) controller responsible for the company's management accounting were present. At the regular meeting, the company's plans for the next fiscal year were discussed based on the results of the current one.

In recent years, the business environment has become relatively stable. The production department's performance score in the current fiscal year shows that costs overrun the standard. The department must improve operations in the next fiscal year to achieve the mid-term management plan, expected to complete in two years. One of the performance indicators is cost control, and your compensation is linked to the actual achievement of targets for each fiscal year.

There is Brilliant Operations Build (B.O.B.) at the meeting, a competent computerized FP&A controller (IA, Intelligent Agent) that helps our team achieve performance goals based on high-quality historical data. It has several years of historical data on all variables with a complete dataset. B.O.B. uses sophisticated machine-learning algorithms to collect information and make recommendations.

B.O.B. suggested that the company revise its product strategy and then introduce new products that can be manufactured without relying on an overseas base. Your supervisor Alex, the head of the production department, said that he would further consider the proposal in his office, and the meeting was adjourned for the day.

After consideration, the following day, Alex instructed you to revise the plan immediately, as *B.O.B.* suggested.

[for the Human-Pandemic condition]

You are working in an electronics company's production department. On the day of the meeting, you, your supervisor Alex (the head of the production department), and the FP&A (Financial Planner & Analysis) controller responsible for the company's management accounting were present. At the regular meeting, the company's plans for the next fiscal year were discussed based on the results of the current one.

In recent years, the business environment has become increasingly unstable and uncertain. Several problems, such as disruptions in the supply chain due to the COVID-19 pandemic and changes in consumer demand, have caused instability in the business environment in recent years. The production department's performance score in the current fiscal year shows that costs overrun the standard. The department must improve operations in the next fiscal year to achieve the mid-term management plan, expected to be complete in two years. One of the performance indicators is cost control, and your compensation is linked to the actual achievement of targets for each fiscal year.

Bob, a competent FP&A controller, was at the meeting. He helps our team achieve performance goals based on high-quality historical data. He has several years of historical data on all variables, with a complete dataset.

Bob suggested that the company revise its product strategy and then introduce new products that can be manufactured without relying on an overseas base. Your supervisor Alex, the head of the production department, said that he would further consider the proposal in his office, and the meeting was adjourned for the day.

After consideration, the following day, Alex instructed you to revise the plan immediately, as *Bob* suggested.

[for the Human-stable condition]

You are working in an electronics company's production department. On the day of the meeting, you, your supervisor Alex (the head of the production department), and the FP&A (Financial Planner & Analysis) controller responsible for the company's management accounting were present. At the regular meeting, the company's plans for the next fiscal year were discussed based on the results of the current one.

In recent years, the business environment has become relatively stable. The production department's performance score in the current fiscal year shows that costs overrun the standard. The department must improve operations in the next fiscal year to achieve the mid-term management plan, expected to complete in two years. One of the performance indicators is cost control, and your compensation is linked to the actual achievement of targets for each fiscal year.

Bob, a competent FP&A controller, was at the meeting. He helps our team achieve performance goals based on high-quality historical data. He has several years of historical data on all variables, with a complete dataset.

Bob suggested that the company revise its product strategy and then introduce new products that can be manufactured without relying on an overseas base. Your supervisor Alex, the head of the production department, said that he would further consider the proposal in his office, and the meeting was adjourned for the day.

After consideration, the following day, Alex instructed you to revise the plan immediately, as *Bo*b suggested.

<*Third page*>

Attention Check question
Please answer the following attention check questions. If you don't answer all of the questions below correctly, you cannot proceed.

(1) Which department did you belong to in the scenario? [check1]
-Sales Department
-Production Department
-Don't know

(2) What kind of proposal did the FP&A controller make in the scenario? [check2]
-Revision of product strategy
-Transfer of production bases overseas
-Don't know

<*Fourth page*>

Please answer the following questions about the scenario as the role of an employee.

Question A (questions about your supervisor Alex)
As a subordinate, what are your feelings about Alex (your supervisor)? Please indicate the extent of your agreement with each of the following statements on a 7-point scale (1: strongly disagree, 7: strongly agree).

1 This person approaches his job with professionalism and dedication.
2 Given this person's track record, I see no reason to doubt his competence and preparation for the job.
3 I can rely on this person not to make my job more difficult by careless work.
4 Most people, even those who aren't close friends of this individual, trust and respect him as a coworker.
5 Other work associates of mine who must interact with this individual consider him to be trustworthy.
6 We have a sharing relationship. We can both freely share our ideas, feelings, and hopes.

7 I can talk freely to this individual about difficulties I am having at work and know that he will want to listen.
8 We would both feel a sense of loss if one of us was transferred and we could no longer work together.
9 If I shared my problems with this person, I know he would respond constructively and caringly.
10 I would have to say that we have both made considerable emotional investments in our working relationship.

Footnote: Questions from 1 to 5: participants' cognitive-based trust in their superiors measured by McAllister's (1995) cognitive-based trust scales. Questions from 6 to 10: participants' affect-based trust in their superiors measured by McAllister's (1995) affect-based trust scales.

Question B (questions about the advisor)
What are your feelings about *Bob (B.O.B*) as an advisor? Please indicate the extent of your agreement with each of the following statements on a 7-point scale (1: strongly disagree, 7: strongly agree).

1 *This person* approaches *his* job with professionalism and dedication.
2 Given *this person*'s track record, I see no reason to doubt *his* competence and preparation for the job.
3 I can rely on *this person* not to make my job more difficult by careless work.
4 Most people, even those who aren't close friends of this individual, trust and respect him as a coworker.
5 Other work associates of mine who must interact with *this individual* consider *him* to be trustworthy.
6 We have a sharing relationship. We can both freely share our ideas, feelings, and hopes.
7 I can talk freely to *this individual* about difficulties I am having at work and know that *he* will want to listen.
8 We would both feel a sense of loss if *one of us* was transferred and we could no longer work together.
9 If I shared my problems with *this person*, I know *he* would respond constructively and caringly.
10 I would have to say that we have both made considerable emotional investments in our working relationship.

Footnote: Questions from 1 to 5: participants' cognitive-based trust in the advisor measured by McAllister's (1995) cognitive-based trust scales. Questions from 6 to 10: participants' affect-based trust in the advisor measured by McAllister's (1995) affect-based trust scales.

Question C (questions about the plan)
What are your feelings about the plan proposed by the advisor and adopted by your supervisor?
Please indicate the extent of your agreement with each of the following statements on a 7-point scale (1 = strongly disagree, 7 = strongly agree).

1 I think that the contents proposed by the advisor are most suitable for our company.

2 Considering the current environment, I think the contents proposed by the advisor are optimal.
3 I think the contents of the proposal are consistent with my ideas.
4 I think it is the most suitable content because there is a high possibility of success.
5 I think it is the best proposal because it will increase my compensation.
6 I think the contents proposed by the advisor are the best overall.

Footnote: Questions from 1 to 6: participants' perception of the degree of agreement with the proposal.

Question D (questions about formulating business strategies)
Do you feel that formulating a business strategy is a subjective or objective task based on the *unstable/stable* business environment of the scenario? Please rate the task in which your supervisor Alex formulated business strategies on how subjective it seemed based on the *unstable/stable* business environment of the scenario, using scales from 1 (completely objective) to 7 (completely subjective).

Footnote: This question measures participants' perception of the degree of subjectivity of management strategy formulation tasks by the supervisor.

<Fifth page>
Exit Survey

1 Attention check question
Please answer the following attention check questions.
(1) How is the business environment in this scenario?
-Relatively Stable
-Increasingly unstable and uncertain
-Don't know

(2) In this scenario, is *Bob (B.O.B.*) a human or computer algorithm?
-Human
-Computer Algorithm
-Don't know

2 Questions about your preferences and beliefs in your life outside of this study
For the next set of questions please think about your preferences and beliefs in your life outside of this study.
What are your feelings about the following questions? Please indicate the extent of your agreement with each of the following statements on a 7-point scale (1 = strongly disagree, 7 = strongly agree).

1 I am comfortable with technology.
2 The rising use of artificial intelligence in the workplace is a good thing.
3 Computer algorithms are less biased than people are.
4 Computer algorithms can behave exactly like humans.
5 I feel comfortable letting computer algorithms make important decisions in society.

6 It is easy for me to imagine how artificial intelligence can improve society today.
7. Most people tell a lie when they can benefit by doing so.
8. Those devoted to unselfish causes are often exploited by others.
9. Some people do not cooperate because they pursue only their short-term self-interest. Thus, things that can be done well if people cooperate often fail because of these people.
10. Most people are honest.
11. There will be more people who will not work if the social security system is developed further.

Footnote: Questions from 1 to 6: participants' perception of the levels of acceptance of technology (Bol et al. 2021), Questions from 7 to 11: the general trust scale (Yamagishi 1986).

3 Questions about the COVID-19 crisis
Are you more or less worried about your financial situation than you were in January of 2020 as a result of the COVID-19 crisis? Please indicate the extent of your changes in worry using scales from 1 (less worried) to 7 (more worried).

Footnote: This question measures participants' perception of their own change of financial situation as a result of the COVID-19 crisis.

4 Questions about yourself
Please answer the following question.

1 What is your age? Please enter your year of birth into the text entry box.

2 How many years of work experience do you have?

3 What is your gender?

4 How much is your annual income?

5 What is your level of education?

6 Please inform us know if any part of the scenario text or questions is difficult to understand. Furthermore, please give us your opinion about this survey.

******<Sixth page>******

Exit-Screen
This is the final page of the survey. Thank you very much for participating.

[ATTENTION] Your passcode: ******

Please go back to the Mechanical Turk website and paste your passcode into the box. You will receive credit for taking our task.

S2. Manipulation checks and final sample size

Table S2-1 reports the final sample size by condition for our analyses.

Table S2-1 The final sample size by condition for our analyses

Condition	IA-Pandemic	IA-Stable	Human-Pandemic	Human-Stable	Total
Pre-registered sample size	350	350	350	350	1400
(a) Exclusion for failure to oath	0	1	1	2	4
sample size after the exclusion (a)	350	349	349	348	1396
(b) Exclusion for failure to control question after the scenario	8	9	11	3	31
sample size after the exclusion (a) and (b)	342	340	338	345	1365
(C) Exclusion for failure to operation check question	100	90	95	69	354
final sample size	**242**	**250**	**243**	**276**	**1011**

S3. Summary statistics

Table S3-1reports the descriptive statistics for each experimental condition. Figure S3-1 shows the box plot of trust in the supervisor, trust in advisor, the degree of agreement with the proposal, and the degree of subjectivity of management strategy formulation tasks.

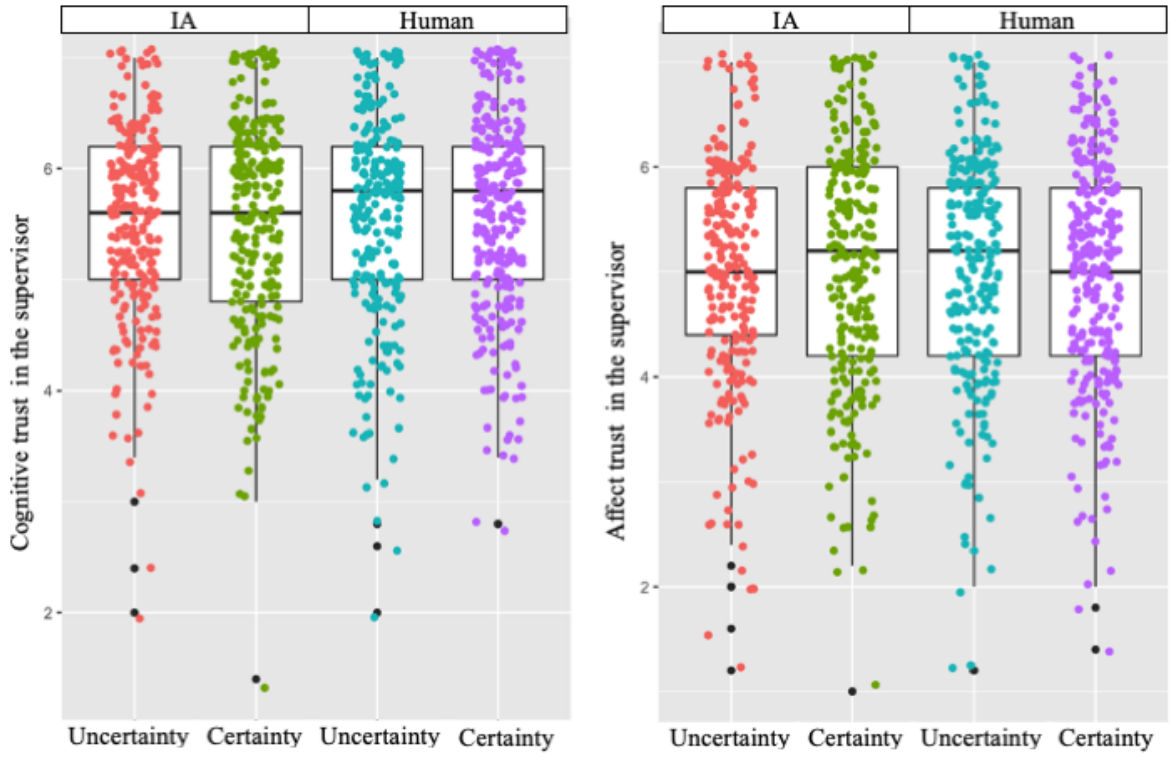
Table S3-1. The descriptive statistics for each experimental condition

Condition		Cognitive trust in the manager	Affect trust in the manager	Cognitive trust in the adviser	Affect trust in the adviser	Agreement with the plan	Subjectivity for the strategy
IA-Pandemic	Mean	5.52	4.98	5.27	3.46	5.61	4.07
	Median	5.60	5.00	5.40	3.60	5.67	4.00
	S.D.	0.86	1.12	1.09	1.92	0.81	1.76
IA-Stable	Mean	5.52	5.03	5.20	3.56	5.46	4.07
	Median	5.60	5.20	5.40	3.60	5.67	4.00
	S.D.	0.96	1.19	1.22	2.09	1.10	1.81
Human-Pandemic	Mean	5.54	5.05	5.68	4.85	5.65	4.32
	Median	5.80	5.20	5.80	5.00	5.83	5.00
	S.D.	0.95	1.12	0.94	1.31	0.85	1.64
Human-Stable	Mean	5.61	4.98	5.78	4.70	5.66	4.10
	Median	5.80	5.00	5.80	4.80	5.83	4.00
	S.D.	0.90	1.07	0.88	1.38	0.93	1.73

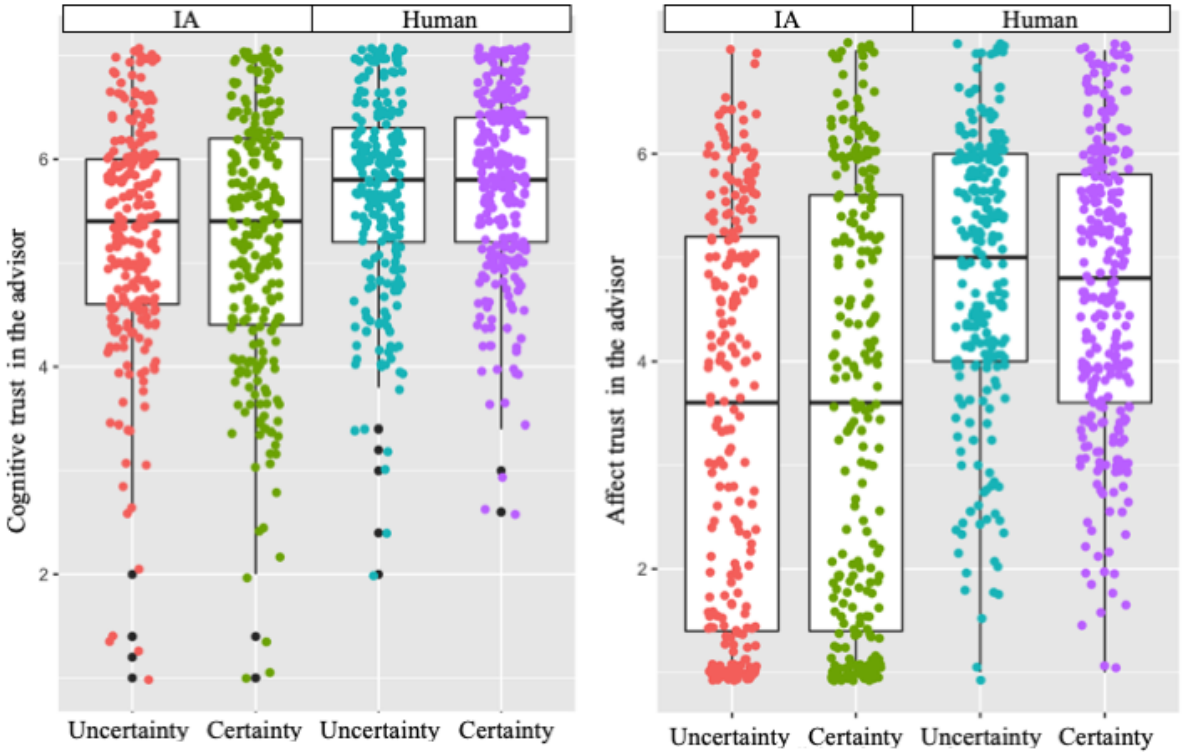
Note: Data analysis was conducted after reverse-scoring the appropriate items. Scores for each variable were calculated by averaging item scores. Cognitive-based trust in the supervisor: Cronbach's alpha coefficients increased from .765 to .838 after eliminating item 6. This item showed a low correlation with the other five items, and the results of exploratory factor analysis (maximum likelihood estimation with Promax rotation) without item 6 revealed one factor with eigenvalues greater than 1.00. Therefore, we exclude this item from the analysis. Affect-based trust in the supervisor: Cronbach's alpha coefficient of 5 items was .875. Cognitive-based trust in the advisor: Cronbach's alpha coefficients increased from .675 to .777 after eliminating item 6. This item showed a low correlation with the other five items, and the results of exploratory factor analysis (maximum likelihood estimation with Promax rotation) without item 6 revealed one factor with eigenvalues greater than 1.00. Therefore, we exclude this item from the analysis. Affect-based trust in the advisor: Cronbach's alpha coefficient of 5 items was .949.

Figure S3-1. The box plot of trust in the supervisor, trust in advisor, the degree of agreement with the proposal, and the degree of subjectivity of management strategy formulation tasks

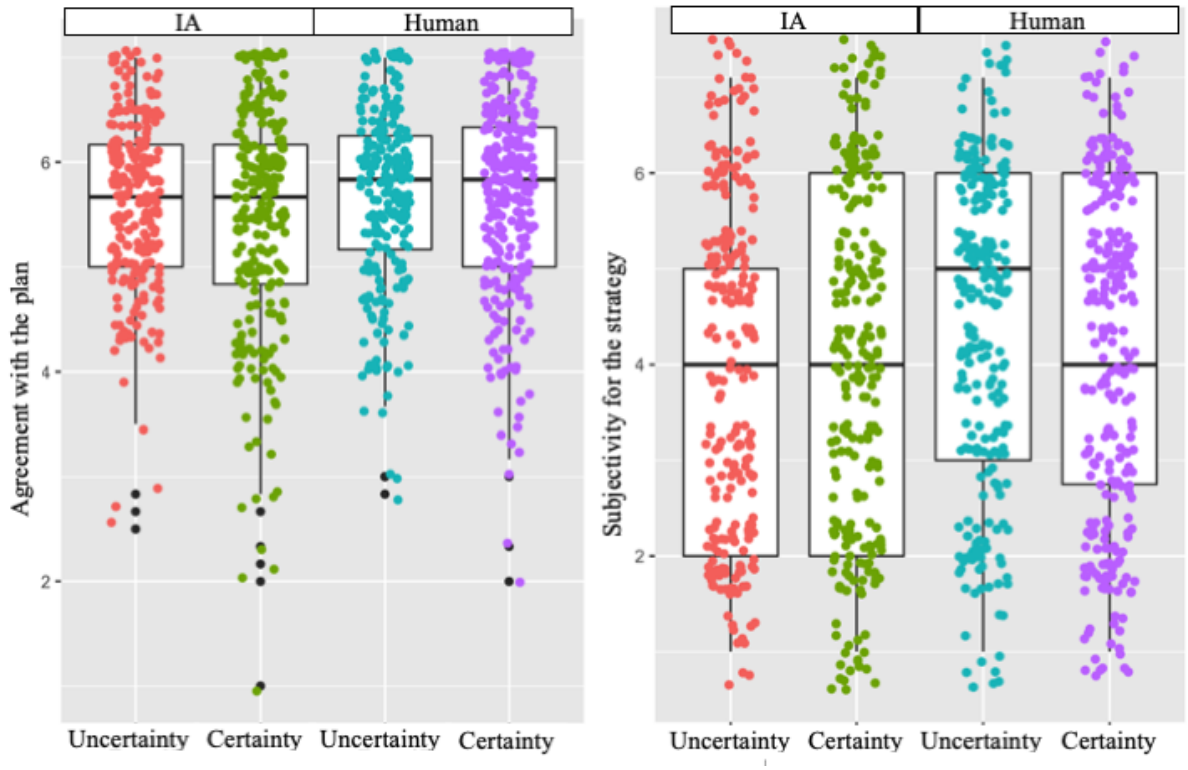
Panel A The box plot of the cognitive and affect-based trust in the supervisor



Panel B The box plot of the cognitive and affect-based trust in the advisor



Panel C. The box plot of the degree of agreement with the proposal and the degree of subjectivity of management strategy formulation tasks



Additional References

Bol, J.C., Brown, C., LaViers, L. (2021). Differing Perceptions of and Preferences for Human-Driven and Artificial Intelligence-Driven Evaluation Systems. *In AAA Management Accounting Section (MAS) 2021 Meeting Paper.*