Human Factors and Systems Interaction, Vol. 192, 2025, 58–67
https://doi.org/10.54941/ahfe1006628

The Impact of Cultural Values on Human-AI Collaboration in a Decision-Making Task

Riley Schwanz, Tiffany G. Lui, Elizabeth Fox, Gene Alarcon, August Capiola, James Bliss, and George Reis

Air Force Research Laboratory, Dayton, OH 45324, USA

ABSTRACT

As artificial intelligence (AI) becomes increasingly part of everyday life including transportation, manufacturing, etc., it is important to understand how humans utilize AI to achieve an effective human-AI collaboration. Furthermore, it may be possible that one's interaction with AI is influenced by differences in their cultural values. Currently, literature on cognition differences in cultural values that go beyond the Eastern and Western comparison is lacking. Consequently, the current study examined how behavior and performance in a decision-making (DM) task are influenced by differences in individual cultural values and the presence of an AI decision aide. To examine cultural values, we used Hofstede's (1984) cultural dimensions: power distance, masculinity, long-term orientation, uncertainty avoidance, and collectivism. Participants completed a DM task consisting of local shapes (e.g., squares and diamonds) encompassed by a larger global shape. They were asked to determine if there were more squares or diamonds from the local shapes, while ignoring the global shape. The global shape matched (GC) or mismatched (GI) the local answer, or the global shape was absent (LO). Participants' DM was aided by high (80%) or low (60%) accuracy AI. Results showed higher accuracy and faster response times in GC and GI compared to LO. Eye tracking data indicated fewer fixations and longer dwell times in GC and GI compared to LO. Taken together, this may indicate that the global shape, whether it matched or mismatched the correct answer, reduced perceptual demand by acting as a boundary to constrain visual attention. In relation to cultural dimensions, increases in collectivism and long-term orientation predicted decreases in performance only when there was no AI while increases in power distance predicted increases in performance when there was no AI and when AI was highly accurate. Overall, performance may be influenced by cultural values and an AI decision aide.

Keywords: Culture, Artificial intelligence, Performance, Decision-making

INTRODUCTION

Many of the well-established findings in cognition research have been based predominantly on WEIRD (white, educated, industrialized, rich, democratic) samples (Gutchess et al., 2023). However, other research demonstrates that cognitive differences exist in relation to culture, indicating that not all cognitive principles are generalizable regardless of individual and/or cultural differences. For example, Nisbett et al. (2001) proposed the General Holistic-Analytic Model to explain differences in cognitive processing between Eastern and Western cultures. According to this model, Eastern cultures are more collectivistic and engage in holistic processing by placing greater emphasis on group features whereas Western cultures are more individualistic and engage in analytic processing by placing greater emphasis on individual features.

These cultural differences are further supported by eye tracking metrics including fixation count and fixation dwell time. For instance, Cenek et al. (2020) examined global and local attention in an Eastern culture (Taiwanese) and Western culture (Czechs) by having them view various scenes consisting of focal (local) objects against different backgrounds. Although both groups had more fixations and longer dwell times on the focal objects than the background, Czechs showed more fixations on the focal objects compared to Taiwanese individuals. Meanwhile, Taiwanese individuals spent longer fixating on the background compared to the Czechs. Findings from Cenek et al. (2020) indicate that there may be cultural differences in how people process the same stimuli.

It is important to note that cognition researchers who have studied cultural differences have operationalized culture mainly by geographical region (i.e., Eastern vs. Western; e.g., Cenek et al., 2020). Although geographical region is likely an important factor that distinguishes culture, there may be other cultural differences that should be examined to further understand the extent to which there are cognition differences at an individual level. Hofstede's five-dimensional measure of cultural values, which includes power distance, uncertainty avoidance, collectivism, long-term orientation, and masculinity, might be useful to assess cultural differences at the individual level.

As artificial intelligence (AI) becomes increasingly integrated in transportation, manufacturing, etc., it is important that humans and AI work together effectively, which requires a certain degree of compliance from the human counterpart (Love et al., 2023). Researchers have found relationships between Hofstede's five cultural values and attitudes in trust in AI. People higher in uncertainty avoidance (i.e., low tolerance for uncertainty) and people lower in collectivism are more likely to rely on AI (Chien et al., 2016). In the context of education, Viberg et al. (2023) found that teachers who are higher in long-term orientation (i.e., persistence, perseverance) showed higher trust in AI while those higher in masculinity show more concerns with using AI. Overall, there is evidence that cultural differences affect individuals' attitudes towards AI.

We sought to further investigate the relationship between individual cultural values and the extent to which people benefited from AI in the context of cognition. Participants performed a decision-making (DM) task with or without an AI aide. We implemented eye-tracking measures such as fixation count and dwell time to examine the extent to which differences exist across cultural values, providing further support for past research (e.g., Cenek et al., 2020) indicating that different cultures can influence eye movement patterns and performance.

METHOD

One hundred and three participants (62 males, 39 females, 2 did not disclose) from a midsize Midwestern university participated in the study in exchange for monetary compensation. Their ages ranged from 18 to 41 years with a mean of 25.03 and a standard deviation of 4.04. The distribution of ethnicities was as follows: 81% South Asian, 9% North American, 2% African American, 2% Middle Eastern, 2% African, 1% South/Central American, and two chose not to disclose. All participants were required to have 1) normal or corrected to normal vision, 2) normal motor control, and 3) no history of epilepsy or seizures. A3 (Trial Type) x 3 (AI Block) within-subjects design was used. The three trial types included: Local Only (LO), Global Congruent (GC), and Global Incongruent (GI). The three AI blocks consisted of: high accuracy AI (80% accurate), low accuracy AI (60% accurate), and no AI. For the AI confidence level, we sampled from a lognormal distribution centered around 70–90% confident with a standard deviation of 3.5.

Perceptual Decision-Making Task

The perceptual DM task was based on Navon's (1977) perceptual stimuli task and was further expanded. Participants viewed a display consisting of local and global shapes. The local stimuli consisted of a set of shapes: diamonds and squares. The global stimulus was a larger shape (either a diamond or square) that encompassed the local stimuli. The goal of the task was to determine if there were more squares or more diamonds from the local stimuli, while ignoring the global shape. Participants responded using a mouse by clicking the left mouse button if they believed there were more squares or clicking the right mouse button if they believed there were more diamonds. After each trial, participants were asked to rate their confidence in their answer using a confidence interval scale with 100% at each anchor for either square or diamond with 50% anchored in the middle of the scale. The LO trial type consisted of only the local stimuli. In GC, the local stimuli were encompassed by the global shape that matched with the correct answer to the local stimuli. In GI, the global shape did not match with the correct answer to the local stimuli. In addition, participants completed the task with or without the aid of an AI, which was depicted as a blue circle on the scale displaying how confident the AI was in determining if there were more squares or diamonds (see Figure 1).

Participants completed a demographics questionnaire and Yoo et al.'s (2011) updated scales to assess cultural differences at the individual level based on Hofstede's (1984) cultural dimensions. Then, participants were calibrated to the eye-tracker at less than 1.5° of angular error using the iMotions software. Next, participants completed a practice session of the perceptual DM task and were provided feedback on correct decisions, misses, and incorrect decisions. Participants then performed the experiment session consisting of the three AI blocks. Prior to the AI blocks, participants completed an AI training session in which they observed how the AI was making its decision on the confidence interval scale. Participants were then given feedback after each trial. In sum, participants completed a total of 1155 trials. Each trial was presented onscreen for 3000 ms followed by a display for participants to make their confidence judgment. In sum, the study was a single 2-hour session and participants were compensated at a rate of $10 per 30 mins. They also received a $10 bonus if their accuracy was greater than or equal to 80%.

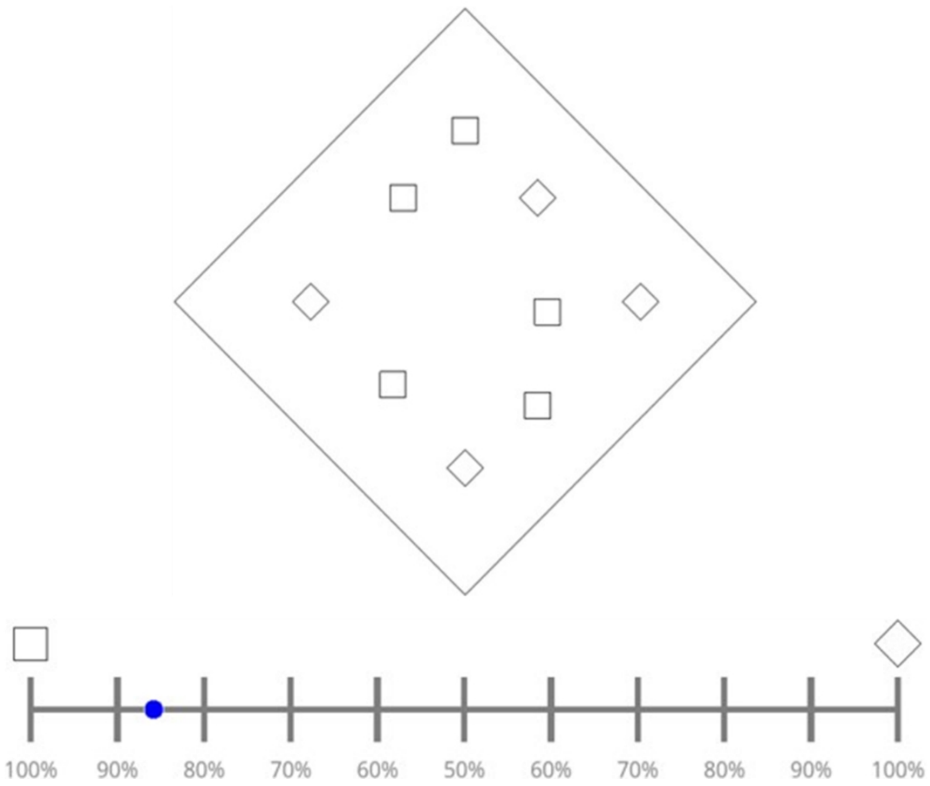

Figure 1: A global incongruent trial.

RESULTS

We used the lme4 (Bates et al., 2015) package in R (R Core Team, 2022) to examine the extent to which AI block and trial type moderated the relationship between individual cultural dimensions and outcomes (i.e., accuracy, response time, fixation count, and dwell time) with participant as a random effect.

Performance

There was a significant difference in accuracy between each of the three trial types: GC = 94%, GI = 93%, LO = 92% ($R^2 = .03$, $F(2,111111) = 28$, $p <.001$). GC ($b = .01$, $SE = .003$, $p <.001$) and GI ($b = .01$, $SE = .003$, $p <.001$) resulted in significantly higher accuracy compared to LO. There was no significant difference in accuracy between the three AI blocks: high accuracy AI = .93, low accuracy AI = .93, no AI = .93.

There was not a significant difference in RT between GC (1.54 sec) and GI (1.54 sec). However, GC ($b = -.01$, $SE = .003$, $p < .01$) and GI ($b = -.01$, $SE = .003$, $p < .001$) resulted in significantly faster RT compared to LO (1.56 sec; $R^2 = .19$, $F(2,111111) = 6.19$, $p = .002$). There was not a significant difference in RT between the low accuracy AI (1.55 sec) and no AI (1.56 sec). However, high accuracy AI resulted in significantly faster RT (1.53 sec; $R^2 = .03$, $F(2,111111) = 44.6$, $p < .001$) compared to low accuracy AI ($b = -.02$, $SE = .003$, $p < .001$) and no AI ($b = -.02$, $SE = .003$, $p < .001$).

Oculomotor Activity

Table 1 displays the mean number of fixations and dwell times (ms) per trial to the local stimuli area of interest (AOI) for each trial type and AI block. The number of fixations refers to the numbers of times that a participant fixated on an area in the visual display while dwell time refers to the duration of each of those fixations.

For fixation count, there was a significant main effect of trial type ($R^2 = .01$, $F(2, 544) = 19.91$, $p < .001$). The LO trial type showed significantly more fixations than the GC trial type ($b = -0.25$, $SE = 0.046$, $p < .001$) and GI trial type ($b = -0.25$, $SE = 0.046$, $p < .001$). However, there was no significant main effect of AI block and the Trial Type x AI Block interaction was not significant.

For dwell time, results revealed a significant main effect of trial type ($R^2 = .002$, $F(2, 66681) = 51.96$, $p < .001$). Post-hoc analyses showed that the LO trial type had significantly shorter dwell times compared to the GC trial type ($b = 25.79$, $SE = 2.80$, $p < .001$) and GI trial type ($b = 27.38$, $SE = 2.80$, $p < .001$). There was also a significant main effect of AI block ($R^2 = .002$, $F(2, 66684) = 19.75$, $p < .001$). The high accuracy AI block had shorter dwell times compared to the low accuracy AI block ($b = -15.69$, $SE = 2.24$, $p < .001$) and no AI block ($b = -14.36$, $SE = 2.25$, $p < .001$). Finally, the Trial Type x AI Block interaction was not significant.

Table 1: Mean number of fixations and dwell times per trial for each trial type and AI block. Standard error of the mean is in parentheses.

AI Block	High Accuracy AI			Low Accuracy AI			No AI		
Trial	GC	GI	LO	GC	GI	LO	GC	GI	LO
Fixations	3.69	3.70	3.91	3.69	3.67	3.91	3.65	3.65	3.90
	(0.02)	(0.02)	(0.03)	(0.02)	(0.02)	(0.03)	(0.02)	(0.02)	(0.03)
Dwell	414.65	415.50	394.62	427.65	432.36	401.46	430.05	429.95	405.51
Time (ms)	(2.71)	(2.78)	(4.45)	(2.70)	(2.83)	(4.36)	(2.82)	(2.84)	(4.47)

Predicting Performance and Eye Tracking Metrics With Hofstede's (1984) Cultural Dimensions

Trial type and AI block resulted in significant slopes in terms of collectivism predicting accuracy. For GC ($b = -.02$, $SE = .007$, $p = .003$) and GI ($b = -.01$, $SE = .003$, $p = .04$) in the no AI block, increases in collectivism predicted decreased accuracy. No significant slopes were found for the relationship between collectivism and response time. Trial type and AI block were significant moderators in terms of collectivism predicting number of fixations and dwell time. However, no significant slopes were found for the relationship between collectivism and number of fixations or dwell time.

Trial type and AI block resulted in significant slopes in terms of long-term orientation predicting accuracy and response time. For GC ($b = -.03$, $SE = .01$, $p = .008$), GI ($b = -.04$, $SE = .01$, $p < .001$), and LO ($b = -.03$, $SE = .01$, $p = .01$) in the no AI block, increases in long-term orientation predicted decreased accuracy. For all trial types and in all AI blocks, increases in long-term orientation predicted slower response time (e.g., no AI block = GC: $b = .12$, $SE = .05$, $p = .01$; GI: $b = .12$, $SE = .05$, $p < .01$; LO: $b = .16$, $SE = .05$, $p = .002$). Trial type and AI block were significant moderators in terms of orientation predicting number of fixations and dwell time. Specifically for the LO trial type and no AI block, increases in long-term orientation significantly predicted more fixations ($b = 0.65$, $SE = 0.30$, $p = .033$). No significant slopes were found for the relationship between long-term orientation and dwell time.

Trial type and block resulted in significant slopes in terms of power distance predicting response time. For GC ($b = -.08$, $SE = .03$, $p = .02$), GI ($b = -.07$, $SE = .03$, $p = .03$), and LO ($b = -.08$, $SE = .03$, $p = .02$) in the no AI block and for GC ($b = -.08$, $SE = .03$, $p = .02$), GI ($b = -.07$, $SE = .03$, $p = .04$), and LO ($b = -.07$, $SE = .03$, $p = .03$) in the high accuracy AI block, increases in power distance predicted faster response time. No significant slopes were found for the relationship between power distance and accuracy. Trial type and AI block were significant moderators in terms of power distance predicting number of fixations and dwell time. For the GC trial type ($b = -0.49$, $SE = 0.18$, $p = .010$), GI trial type ($b = -0.44$, $SE = 0.18$, $p = .020$), and LO trial type ($b = -0.47$, $SE = 0.19$, $p = .014$), increases in power distance significantly predicted fewer fixations in the high accuracy AI block. Similarly, for the GC trial type ($b = -0.46$, $SE = 0.18$, $p = .015$), GI trial type ($b = -0.44$, $SE = 0.18$, $p = .018$), and LO trial type ($b = -0.52$, $SE = 0.19$, $p = .007$), increases in power distance significantly predicted fewer fixations in the no AI block. For the LO trial type and low accuracy AI block, increases in power distance also significantly predicted fewer fixations ($b = -0.40$, $SE = 0.18$, $p = .035$). No significant slopes were found for the relationship between power distance and dwell time.

No significant slopes were found for the relationship between masculinity and accuracy or response time. Trial type and AI block were significant moderators in terms of masculinity predicting number of fixations and dwell time. For the GC trial type ($b = -0.29$, $SE = 0.12$, $p = .019$), GI trial type ($b = -0.31$, $SE = 0.12$, $p = .014$), and LO trial type ($b = -0.33$, $SE = 0.12$, $p = .009$), increases in masculinity predicted fewer fixations in the high accuracy AI block. Similarly, for the GC trial type ($b = -0.25$, $SE = 0.12$, $p = .046$), GI trial type ($b = -0.26$, $SE = 0.12$, $p = .038$), and LO trial type ($b = -0.27$, $SE = 0.12$, $p = .033$), increases in masculinity predicted fewer fixations in the no AI block. For dwell time, increases in masculinity predicted longer dwell time for the GI trial type and no AI block ($b = 37.85$, $SE = 18.50$, $p = .045$).

Trial type and AI block resulted in significant slopes in terms of uncertainty avoidance predicting response time. For GC (b = .10, SE = .05, p <.05), GI (b = .10, SE = .05, p <.05), and LO (b = .10, SE = .05, p <.05) in the no AI block, increases in uncertainty avoidance predicted slower response time. No significant slopes were found for the relationship between uncertainty avoidance and accuracy. Trial type and AI block were significant moderators in terms of uncertainty avoidance predicting number of fixations and dwell time. No significant slopes were found for the relationship between uncertainty avoidance and number of fixations. For the GC trial type ($b = -89.69$, $SE = 42.30$, $p = .038$), GI trial type ($b = -106.80$, $SE = 42.31$, $p = .014$), and LO trial type ($b = -87.68$, $SE = 43.15$, $p = .046$), increases in uncertainty avoidance predicted shorter dwell times for the low accuracy AI block. Similarly, for the GC trial type ($b = -108.71$, $SE = 42.30$, $p = .012$), GI trial type ($b = -102.38$, $SE = 42.30$, $p = .018$), and LO trial type ($b = -92.41$, $SE = 42.30$, $p = .035$), increases in uncertainty avoidance predicted shower dwell times for the no AI block.

DISCUSSION

Our purpose was to understand the extent to which individual cultural values influenced perceptual bias for local and global stimuli and to examine the relationship between cultural values and the benefit from AI. Overall, participants exhibited poorer performance in the LO trial type (slowest RT and lowest accuracy) compared to GC and GI. This finding is paralleled by LO showing more fixations and shorter dwell times than GC and GI trial types. Regarding the AI, participants had faster RT in the high accuracy AI block compared to the no and low accuracy AI block, which is paralleled by shorter dwell times in the high accuracy AI block compared to the no and low accuracy AI block. Our results raise theoretical implications relating to the influence of global stimuli, individual cultural differences, and the presence and accuracy of AI on performance and oculomotor activity.

The first theoretical issue raised by our results is related to the influence of global stimuli on people's performance in a visual search task. Participants had greater accuracy, faster RT, fewer fixations, and longer dwell times in GC and GI compared to LO. These findings indicate that the global shape, whether congruent or incongruent, reduces the level of perceptual demand placed on people's visual attention compared to the unbounded local only area. A direction for future research is to investigate the presence of a control condition that includes a benign global shape (i.e., a circle) compared to LO.

The second theoretical issue raised by our results is related to the relationship between individual cultural values and AI, and the influence of this relationship on performance.

Viberg et al. (2023) found that higher collectivism related to greater perceived concerns when using AI. In contrast, we found that higher levels of collectivism (i.e., people who are group-oriented and likely portray more global processing) led to decreased performance only when there was no AI, suggesting that participants were using the AI to their benefit regardless of its accuracy. One explanation for the contrast in findings is the importance of context. Viberg et al. (2023) studied teachers' attitudes for AI adoption in the classroom over the long-term while we were concerned with the benefit from AI in a short-term task. People likely have less concerns using AI as an aid to make more efficient decisions in the short-term compared to implementing AI to take over a substantial portion of an employee's roles and responsibilities over the long-term.

Viberg et al. (2023) found that people higher in long-term orientation perceived greater benefits from using AI. In line with past research, we found that people higher in long-term orientation showed decreased performance only when the AI was absent. These findings suggest that people higher in long-term orientation (i.e., people who value planning and gathering more information when making decisions) used the AI to their benefit regardless of its accuracy.

Chi et al. (2023) found that higher power distance related to higher expectations for AI and lower tolerance for unreliable AI. In line with these findings, we found that higher levels of power distance led to increased performance when the AI was either absent or high in accuracy. This finding suggests that people higher in power distance (i.e., people who recognize power hierarchies, value authority, and are achievement-oriented) might have viewed the AI as an authority figure, which they relied on to make more accurate and efficient decisions. However, people higher in power distance did not experience increased performance in the low accuracy AI block because they likely viewed the AI as an unreliable authority figure, which hindered their ability to make quick and accurate decisions.

Overall, eye tracking metrics (e.g., fixation count and dwell time) supported the conclusions regarding cultural impacts on visual task performance for collectivism, long-term orientation, and power distance. A direction for future research is to investigate additional objective and subjective underlying dimensions of culture (e.g., sociocultural norms, religion, socioeconomic status, etc.) and to analyze the interactive effect between multiple cultural dimensions on performance and DM.

LIMITATIONS

It is important to note limitations that may have affected the generalizations drawn from our results. We were unable to analyze group differences between cultures due to an imbalance in our sample. However, recent researchers have suggested the importance of analyzing the effect of individual cultural differences (e.g., Yoo et al., 2011), and we found evidence supporting that individual cultural differences have a significant impact on benefit from AI, oculomotor activity, and performance.

CONCLUSION

The purpose of our study was to 1) examine the influence of individual cultural differences on perceptual bias, and 2) examine the relationship between individual cultural differences and the benefit from AI in a visual search task. We found that global stimuli played an important role in oculomotor activity and performance, and we found that individual cultural differences including collectivism, long-term orientation, and power distance had a significant impact on oculomotor activity and performance with implications for reliance on AI-enabled decision-support systems. The findings of our study contribute to the existing literature by expanding Navon's (1977) visual search task, examining the influence of global stimuli on performance, and investigating the relationships between individual cultural differences and AI. In sum, we found evidence that people's performance in a visual search task is beneficially influenced by the presence of global stimuli and impacted by people's individual cultural values and the extent to which they benefited from the AI.

ACKNOWLEDGMENT

The views expressed are those of the author and do not necessarily reflect the official policy or position of the Department of the Air Force, the Department of Defense, or the U.S. government.

This research was supported in part by an appointment to the Department of Defense (DOD) Research Participation Program administered by the Oak Ridge Institute for Science and Education (ORISE) through an interagency agreement between the U.S. Department of Energy (DOE) and the DOD. ORISE is managed by ORAU under DOE contract number DE-SC0014664. All opinions expressed in this paper are the author's and do not necessarily reflect the policies and views of DOD, DOE, or ORAU/ORISE.

Distribution A. Approved for public release; distribution unlimited. AFRL-2025-0442; Cleared 27 Jan 2025.

REFERENCES

Bates, D., Mächler, M. and Bolker, B. (2015) 'Fitting linear mixed-effects models usinglme4', *Journal of Statistical Software*, 67 (1), doi: 10.18637/jss.v067.i01.

Čeněk, J., Tsai. J. and Šašinka, Č. (2020) 'Cultural variations in global and local attention and eye-movement patterns during the perception of complex visual scenes: Comparison of Czech and Taiwanese university students', *PLOS ONE*, 15 (11), doi: 10.1371/journal.pone.0242501.

Chi, O. *et al.* (2023) 'Customers' acceptance of Artificially Intelligent Service Robots: The influence of trust and culture', *International Journal of Information Management*, 70, S.102623, doi: 10.1016/j.ijinfomgt.2023.102623.

Chien, S. *et al.* (2016) 'Relation between trust attitudes toward automation, Hofstede's cultural dimensions, and big five personality traits', *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 60 (1), S.841–845, doi: 10.1177/1541931213601192.

Gutchess, A. and Rajaram, S. (2023) 'Consideration of culture in cognition: How we can enrich methodology and theory', *Psychonomic Bulletin & Review*, 30 (3), S.914–931, doi: 10.3758/s13423-022-02227-5.

Hofstede, G. (1984) *Culture's Consequences: International Differences in Work-Related Values*, (Vol. 5). Sage.

Love, J. *et al.* (2023) 'Trust in Human-bot Teaming: Applications of the Judge Advisor System', *In Proceedings of the Annual Meeting of the Cognitive Science Society*, (Vol. 45, No. 45).

Navon, David. (1977) 'Forest before trees: The precedence of global features in visual perception', *Cognitive Psychology*, 9(3), 353–383.

Nisbett, R. *et al*. (2001) 'Culture and systems of thought: Holistic versus analytic cognition', *Psychological Review*, 108 (2), S.291–310, doi: 10.1037//0033-295x.108.2.291.

Team, R. (2020) 'RA language and environment for statistical computing, R Foundation for Statistical', *Computing*.

Viberg, O. *et al*. (2024) 'What explains teachers' trust in AI in education across six countries?', *International Journal of Artificial Intelligence in Education*, doi: 10.1007/s40593-024-00433-x.

Yoo, B., Naveen, D. and Lenartowicz, T. (2011) 'Measuring Hofstede's five dimensions of cultural values at the individual level: Development and validation of CVSCALE', *Journal of International Consumer Marketing*, 23(3–4), pp. 193–210.