Human Factors and Simulation, Vol. 83, 2023, 212–225
https://doi.org/10.54941/ahfe1003579

Stress and Motivation on Reliance Decisions With Automation

Mollie McGuire and Miroslav Bernkopf

Naval Postgraduate School, Monterey, CA, USA

ABSTRACT

The decision to rely on automation is crucial in high-stress environments where there is an element of uncertainty. It is equally vital in human-automation partnership that the human's expectations of automation reliability are appropriately calibrated. The current study examined the effects of stress and motivation on the decision to rely on automation. Participants were randomly assigned to a stress and motivation condition, and conducted a task two AI partners, one with high reliability and one with low reliability. While motivation had a stronger effect than stress, both motivation and stress affected reliance decisions with the high reliability AI. The low reliability AI was affected to a lesser degree if at all. Overall, the decision to not rely on the AI partner, especially with the higher in reliability was slower than the decision to rely on the AI partner, with the slowest decision times occurring in the high stress condition with motivated participants, suggesting more deliberate processing was utilized when deciding against the advice of the AI higher in reliability.

Keywords: Human-automation interaction, Trust in automation, Decision making, Stress, Motivation

INTRODUCTION

The human-autonomy partnership is only beneficial when the human makes appropriate reliance decisions. Over reliance or under reliance can pose a serious threat in high-risk contexts. Over reliance (i.e., being too reliant on automation) can result in failure to intervene when an error occurs, either due to overconfidence in the automation or inattention due to being "out-of-the-loop" (Endsley, 2017). Being overly reliant on automated navigation so that when an error occurs (e.g., GPS signal is lost) it is missed, which could have disastrous consequences, like a ship running aground (Lee & Sanquist, 2000). On the other end, under reliance (i.e., reluctance to use automation even when it outperforms human capabilities) can occur due to mistrust. This mistrust might be of automation in general or of a particular system that is prone to false alarms (Parasuraman & Riley, 1997). In order to have optimal human-automation partnerships, it is important to understand the human decision-making processes that underlie reliance behaviors. Some contexts, including environment, competing tasks, and time parameters can be more demanding, usurping cognitive resources, and therefore leaving fewer cognitive resources for reliance decisions. When there are fewer cognitive resources available, decisions are likely to be made based on intuitive judgments.

Additionally, high levels of stress can impact thinking and subsequently decision making strategies, shifting from more deliberative-based to intuitive-based decision making, that could result in poor reliance decisions. Rice and Keller examined the effect of time pressure on reliance on automation. They found that time pressure resulted in an increase in overall reliance due to time pressure rushing information processing, prompting a switch in strategy towards an intuitive-based decision-making approach. However, even though time pressure might lead to an increase in stress, it is not the same construct as stress. The switch to an intuitive-based decision-making strategy could have been due to limited time as opposed to stress. It is then important to understand the effect of stress on reliance in the absence of limited time. Stress without time pressure may lead to a similar shift in processing due to how the brain is affected by a stressor.

Stress and Decision Making

The prefrontal cortex (PFC) is the foundation for executive functioning, such as information integration and top-down attention allocation (Arnsten, 2009). Attention to, and the integration of relevant information is important for optimal, deliberative decision making. However, stress negatively affects the PFC (Arnsten). The decline in PFC resources can cause a shift to more intuitive information processing (Margittai et al., 2016). Margittai et al. reported that participants who were given hydrocortisone (cortisol agonist) scored worse on a task that measures analytical over intuitive processing, the cognitive reflection task (CRT).

In using analytical processing for decision making, cognitive resources are needed to (a) retrieve, from long-term memory the information relevant to that decision stored, (b) integrate and temporarily store that information into working memory, and (c) attend to relevant feedback to update the information needed for future decisions (Brand et al., 2006). Stress reallocating cognitive resources away from the PFC would diminish the ability to learn and integrate feedback for subsequent decisions. In terms of reliance decisions, if the expected reliability does not equal the actual reliability, enough cognitive resources must be available to notice the discrepancy and to update the reliability information in memory. Feedback in decision making is essential when results from prior decisions provide information for future decisions. However, stress can impede the capacity needed to notice and learn from the feedback (Starcke et al., 2008), resulting in continued reliance when the actual reliability is lower than the expected reliability.

Additionally, gender differences have been observed in decision making under stress (e.g., van den Bos et al., 2009; Lighthall et al., 2012; Mather & Lighthall, 2012). Van den Bos, Harteveld, & Stoop found that increased cortisol levels in men decreased performance on the Iowa Gambling Task. However, in women slightly elevated cortisol levels improved performance, but more elevated cortisol levels decreased performance. Thus, both genders had decreased performance with high cortisol levels, but small increases in cortisol made women perform better. In studies on risky decision making under stress, it has been shown that stress increases risky decision making in men but decreases risky decision making in women (Lighthall et al., 2009; Mather & Lighthall, 2012). More research is needed on gender differences in decision making under stress outside of the risk paradigm.

Motivation

Since stress has been shown to have an adverse effect on the PFC, it is interesting to also consider what may have a beneficial effect. Motivation has been found to enhance top-down attention allocation strategies (Locke & Braver, 2008). Locke and Braver used functional magnetic resonance imaging (fMRI) and had participants perform a task with positive, negative, or no incentives. They found that monetary incentives (e.g., rewards) were correlated with an increase in brain areas responsible for cognitive control. Additionally, they found that motivation resulted in faster reaction times without negatively affecting performance. Kouneiher et al. (2009) performed an fMRI study on PFC activation and motivation. They found an increase in the areas associated with cognitive control with incentives, especially higher incentives. While these studies show how motivation can increase brain activity and affect performance, there is a lack of research on whether motivation can overcome factors that may negatively affect cognition, such as stress.

Current Study

The current study investigated reliance decisions with automation under stress with and without motivation. The experiment was designed to study (a) whether there is an overall increase in reliance in the high stress versus low stress condition, (b) if stress affects the ability to incorporate feedback to inform subsequent reliance decisions, (c) if motivation is a moderator in the case of a deleterious effect of stress, (d) the extent that trust correlates with reliance in the high stress versus low stress condition, and (e) whether there are gender differences in reliance decisions under stress.

The following hypotheses were predicted:

1. It was predicted that there would be an increase in reliance in the high stress than low stress condition, due to the deleterious effect of stress on cognitive functions, resulting in a more intuitive-based decision-making approach.

2. It was predicted that motivation would lead to more accurate reliability assessments of the AI partners due to enhanced attention to feedback information.

3. It was predicted that there would be no change in trust, regardless of reliance, between the stress conditions.

METHOD

Participants and Design

Participants were recruited from the Naval Postgraduate School ($M_{age} = 32$), were mostly active duty military (75 active duty, 5 civilian), and male (63 male, 17 female). Due to COVID closures, the desired sample size was not achieved, and resulted in 80 total participants.

The design was a 2 (stress: high, low) x 2 (AI reliability: high, low) x 2 (motivation: incentives known, incentives unknown) mixed design with stress and motivation manipulated between subjects, and AI reliability manipulated within subjects. Participants were randomly assigned to a stress condition and a motivation condition. AI reliability consisted of two AI partners, one with high reliability and one with low reliability, to assess whether appropriate feedback was incorporated into subsequent reliance decisions among the stress and motivation conditions.

The main task was a pattern learning task that involved the participant selecting what number came next in an iterative sequence (i.e., 1, 2, or 3). Participants made an initial decision, then received advice from what was described as an artificial intelligence (AI) partner before making their final decision. Reliance on automation was only assessed on trials when there was a discrepancy between participant choice and AI partner choice; reliance was operationalized as the participant's final decision reflecting the AI choice instead of their own initial choice (van Dongen & van Maanen, 2013). See Bernkopf (2021) for details on pattern learning task.

Participants earned $0.40 for each correct response in the pattern learning task. Participants were told about the earnings in the motivation condition, and had them displayed throughout the pattern learning task, while participants in the non-motivation condition were neither informed of the earnings nor had them displayed. This study utilized gains instead of losses to avoid adding another later of stress that may be caused by incurring losses (Hochman & Yechiam, 2011). Motivation and trust experiments have commonly used incentive-based compensation (e.g., de Visser et al., 2016; Bland et al., 2017). Participants could earn anywhere from $0 to $50, paid with an Amazon gift card.

Stress Manipulation

Laboratory stress induction techniques that include uncontrollability and social-evaluative threat produce the largest increases in cortisol (Dickerson & Kemeny, 2004). The Trier Social Stress Test (TSST; Kirschbaum et al., 1993) is commonly used, contains both uncontrollability and social-evaluative threat, and has been shown to elicit the highest stress responses in a laboratory setting (Smyth et al., 2013). The TSST was used for the high stress condition, and the placebo TSST (p-TSST; Het et al., 2009) was used for the low stress condition. Both the TSST and pTSST have three phases, each phase consisting of five minutes.

Salivary cortisol and state assessment of State Trait Anxiety Inventory (STAI; Spielberger et al., 1970) were the primary indicators of stress used. Salivary cortisol was collected via passive drool and stored in a freezer in the lab at -20 degrees Celsius (Smyth et al., 2013). Additionally, participants wore a wristband that collected heart rate variability (HRV) and electrodermal activity (EDA) data continuously throughout the experiment.

AI Partners

There were two AI partners with different reliabilities, 90% (AI1; referred to as ALEX in the experiment) and 60% (AI2; referred to as SAML in the experiment). Participants were initially told that both AI partners had approximately 90% reliability so that the AI with reliability at 90% would be consistent with their expectations, and the AI with reliability at 60% would be inconsistent with expectations, and is generally considered unreliable (i.e., < 70%; Wickens & Dixon, 2007). Participants were given a cover story about the AI partners, that a software company conducting an evaluation of "pattern learning software before applying it to more complex tasks on naval ships" (modified from van Dongen & van Maanen, 2013). The AI partner switched every 10 trials on the main experimental task, and the starting AI partner was randomly assigned.

Trust Measures

Trust is commonly conceptualized as the perception that a trustee will act in a way that aids or meets the positive expectations of a trustor in situations where there is vulnerability and uncertainty (Hoff & Bashir, 2015; Lee & See, 2004). Trust was measured both objectively and subjectively. The objective measure of trust, the trust game, is commonly used to measure trust between humans where the investor is the trustor, and the investee is the trustee. The premise of the game is that the investor shares money with the investee, knowing that the money will be multiplied by a certain amount, and that the investor can share the money that they receive. The investor is then trusting the investee to share the money at the end (Berg et al., 1995). For the purposes of this experiment, each AI partner was an investee. The trust game was used to see (a) if reliance decisions correlated with the amount invested with each AI (reliance and trust correlation), and (b) if a subjective trust rating correlated with a traditionally used objective measure of trust.

Procedures

All testing took place in the afternoon, starting at noon due to concerns with awakening cortisol levels (Smyth et al., 2013). Participants were asked to abstain from exercise, smoking, eating, or drinking for an hour before the experiment due to potential interactions with cortisol production. Factors that affect cortisol were accounted for in the exit questionnaire (Smyth et al.).

Following an initial informed consent upon entry, the participants were fitted with a wristband to collect physiological data, afterwards they sat at a workstation where eye tracking was calibrated. The first task was a working memory test (i.e., Ospan; Unsworth et al., 2005), afterwards the participants rested for 10 minutes to mitigate any anxiety about the Ospan task and to have a sufficient rest period to establish a baseline for stress assessment. After the ten-minute rest, the first saliva sample was collected, and participants filled out the STAI. After the STAI, participants were given the cover story and became familiar with the main task through practice trials consisting of 40 trials total, 20 with each AI partner, alternating AI partner every ten trials. Instructions and practice trails were completed before stress induction so that learning the task was not affected by stress. The stress induction was next, either the TSST or p-TSST was administered, with the second and third saliva collection immediately before and after stress induction. After the post-stressor saliva collection, participants completed another STAI, and then the Cognitive Reflection Task (CRT). After the CRT, participants completed 100 trials of the pattern learning task, followed by the trust game with each AI partner, and finally the exit questionnaire that (a) asked participants subjective trust and subjective reliability questions, (b) collected demographic data, (c) assessed experiment engagement, and (d) accounted for potential confounds that inhibit cortisol production. The subjective assessments of trust and reliability were rated on a 7-point scale. Subjective trust was rated for each AI partner before subjective reliability in order to mitigate reliability assessments affecting trust ratings. Participants were then debriefed, which included a second consent request to use their data after they knew all the details of the study (one participant was lost due to not consenting post-experiment). Finally, participants were given an Amazon gift card totaling the amount they earned throughout the experiment, with a maximum amount of $50. This research complied with the American Psychological Association Code of Ethics and was approved by the Institutional Review Board at the Naval Postgraduate School.

RESULTS AND DISCUSSION

Analyses were 2 (Stress) x 2 (AI Reliability) x 2 (Motivation) repeated measures ANOVAs, unless indicated otherwise. There were not enough female participants to make comparisons between the groups; however, male participant results that differed from the combined results may be an indication of some gender differences in the data that need to be explored in future research. Therefore, male only data was reviewed if the male only results differ from the combined results.

Stress Manipulation Checks

Successful stress induction was assessed by analyzing cortisol and STAI score differences between the stress conditions. Both cortisol and the STAI showed an increase in the high but not low stress condition, indicating that the stress manipulation was successful. See Table 1 for the last two samples of cortisol and STAI differences between stress conditions.

Cortisol. The difference in cortisol between stress groups was analyzed using a 2 (stress) x 5 (cortisol) repeated measures using transformed cortisol data due to violations with normality. ANOVA. Overall, there was a significant interaction between stress and cortisol, $F(4, 75) = 10.58$, $p < .001$, $\eta_p^2 = .36$; the last two samples (samples 4 and 5 are when the cortisol should be raised in response to the stressor) were significantly higher for the high stress ($M = .26, .27, SD = .16, CI_{95}[.22, .30], [.23,.31]$, Sample 4 and 5 respectively) than low stress condition ($M = .16,.15, SD = .10, .07, CI_{95}[.12, .20], [.23, .31]$), S4: $t(78) = 3.77$, $p < .001$, d = .84; S5: $t(78) = 4.30$, $p <. 001$, d = .96.

STAI. The difference in pre- and post-stress STAI scores was compared between stress groups. There was a significant interaction between stress and STAI score, $F(1, 78) = 42.39, p < .001, \eta_p^2 = .35$; there was a significant increase in STAI score in the high stress condition ($M_{difference} = 9.72, SD = 9.47, CI_{95}[6.65, 12.79]$), $t(38) = 6.41, p < .001$, d = 1.02, but not in the low stress condition, $t(40) = -1.29, p = .21$.

Reliance and Reaction Times

Reliance on Automation. Reliance was operationalized as the percentage of time the participant chose the final answer consistent with the AI partner's choice on trials where there was initial disagreement (i.e., the participant's first choice and the AI's choice were different; average disagreement was 56%). Overall, there was a main effect of AI reliability; there was a higher rate of reliance with AI1 (AI higher in reliability) than AI2 (AI lower in reliability), $F(1, 76) = 104.09$, $p < .001$, $\eta_p^2 = .58$. While there were no other significant results overall, there are differences when the results are broken down by gender. See Table 1 for reliance by condition.

Gender. Both male and female data showed a main effect of AI reliability, ($p < .001$); female participants displayed no other significant or trending results. For male participants, there was a significant interaction between AI reliability and motivation, $F(1, 59) = 5.38, p = .024, \eta_p^2 = .08$; and a significant three-way interaction between AI reliability, stress, and motivation, $F(1, 59) = 4.28$, $p = .043$, $\eta_p^2 = .07$. There were differences between conditions with AI1, but not AI2. In the non-motivation condition, reliance on AI1 was higher in the low than high stress condition, t(29) = 2.23, $p = .034$, d = .81. There was no difference in the motivation condition. Reliance on AI1 in the high stress condition was higher in the motivation than non-motivation condition, t(27) = 3.96, $p < .001$, d = 1.47. Additionally, in the high stress condition, the difference between AI1 and AI2 was larger in the motivation ($M = .27$, $SD = .15$) than non-motivation condition ($M = .10$, $SD = .15$).

Overall, there was higher reliance on AI1 (i.e., the AI partner higher in reliability) than AI2. While the combined data did not show an effect of stress and reliance, when looking at the male data only, there were some interesting

Table 1. Reliance by condition.

	Low-Stress		High-Stress	
	No Motivation	Motivation	No Motivation	Motivation
AI 1	0.75 (.19) [.65, .84]	0.74 (.23) [.65, .83]	0.62 (.22) [.52, .71]	0.78 (.18) [.69, .88]
AI 2	0.58 (.19) [.50, .67]	0.56 (.19) [.48, .64]	0.48 (.18) [.39, .56]	0.55 (.19) [.47, .64]

Note. Means are reported with standard deviations in parentheses, and confidence intervals in brackets.

Table 2. Reliance by condition by gender.

	Low-Stress		High-Stress	
	No Motivation	Motivation	No Motivation	Motivation
		Male		
AI1	0.76 (.19)	0.73 (.26)	0.60 (.19)	0.83 (.11)
	[.66, .85]	[.64, .83]	[.50, .71]	[.73, .93]
AI2	0.60 (.19)	0.57 (.19)	0.51 (.12)	0.56 (.19)
	[.51, .69]	[.48, .66]	[.41, .61]	[.47, .66]
		Female		
AI1	0.69 (.27)	0.75 (.08)	0.65 (.27)	0.65 (.29)
	[.35, 1.02]	[.46, 1.04]	[.39, .91]	[.39, .97]
AI2	0.49 (.14)	0.53 (.13)	0.39 (.29)	0.53 (.22)
	[.22, .76]	[.30, .77]	[.18, .60]	[.32, .74]

Note. Means are reported with standard deviations in parentheses, and confidence intervals in brackets.

findings worth noting. Reliance with AI1 in the high stress condition with motivated participants, showed the highest rate of reliance, and the closest to the actual AI1 reliability of 90%. However, within the high stress condition with non-motivated participants, reliance on AI1 reflected levels closer to the less reliable AI, possibly indicating that stress decreased accurate performance detection when participants were not motivated. With AI2, reliance did not change much across the conditions, ranging from 51%-60%, aligning with the actual reliability of 60%.

Reaction Times. Reaction times were transformed using natural logarithm and are reported in milliseconds.

Overall final decision reaction times. There was a main effect of AI reliability; reaction times were faster with AI1 (M = 1100.62, SD = 315.55, CI_{95}[1028.66, 1168.74]) than AI2 (M = 1231.16, SD = 419.98, CI_{95}[1135.00, 1320.68]), $F(1, 76) = 41.29$, $p < .001$, $\eta_p^2 = .35$. The main effect held with both female and male participants; no other effects were found.

Reliance vs non-reliance reaction times. Reliance and non-reliance final choice reaction times between the two AI partners were analyzed. There was a main effect of reliance; reaction times were faster for reliance than non-reliance, $F(1, 69) = 15.11$, $p < .001$, $\eta_p^2 = .18$. There was an interaction between reliance and motivation, $F(1, 69) = 9.05$, $p = .004$, $\eta_p^2 = .12$. In the motivation condition reaction times were slower when participants chose not to rely on their AI partner (non-reliance) than when they relied on their AI partner (reliance), t(40) = 3.26, p = .002, d = .51 ; there was no difference in the non-motivation condition, p =.12. There was a three-way interaction with reliance, stress, and motivation, $F(1, 69) = 4.94$, $p = .03$, $\eta_p^2 = .07$. In the high stress, motivation condition, reaction times were slower with non-reliance than with reliance, t(19) = 3.11, p = .006, d = .70 ; there was no difference in the high stress, non-motivation condition, p = .86. Additionally, in the low stress, non-motivation condition, reaction times were slower with non-reliance than reliance, t(18) = 2.27, p = .036, d = .52; there was no difference in the low-stress motivation condition, p = .137. Finally, there was a three-way interaction with AI reliability, reliance, and motivation,

Table 3. Reliance reaction times by condition.

	Low-Stress		High-Stress	
	No Motivation	Motivation	No Motivation	Motivation
		AI1		
Reliance	1444.98 (562.29) [1245.80, 1644.16]	1405.18 (405.00) [1211.31, 1599.05]	1246.58 (338.84) [1052.71, 1440.45]	1276.61 (352.40) [1071.65, 1481.56]
Non-Reliance	1572.83 (722.06) [1259.98,1885.67]	1760.38 (776.72) [1455.88, 2064.88]	1276.30 (319.22) [971.80, 1580.80]	1880.42 (749.44) [1558.51, 2202.34]
		AI2		
Reliance	1444.15 (470.58) [1214.49, 1673.81]	1642.15 (638.07) [1418.62, 1865.69]	1348.25 (352.19) [1124.71, 1571.78]	1442.77 (442.60) [1206.45, 1679.09]
Non-Reliance	1559.46 (577.82) [1273.24, 1845.67]	1686.65 (803.98) [1408.07, 1965.23]	1272.93 (381.11) [994.34, 1551.51]	1634.99 (593.74) [1340.47, 1929.51]

Note. Means are reported in milliseconds with standard deviations in parentheses, and confidence intervals in brackets.

$F(1, 69) = 4.37$, $p = .04$, $\eta_p^2 = .06$. With AI1, in the motivation condition, reaction times were slower with non-reliance than reliance, t(35) = 4.83, $p < .001$, d = .81. See Table 3 for reliance reaction times among conditions.

Reaction times are often assessed in decision making to gain insight on cognitive effort, with slower reaction times indictive of more deliberate information processing due to increased cognitive effort (Kahneman, 2003; Payne & Bettman, 2004). Reaction times indicated participants were slower in their decision to *not* rely than to rely on their AI partner. Reaction times in the high stress, motivation condition were among the fastest when participants chose to rely on AI1, but were the slowest when participants chose *not* to rely on AI1. With AI1, in the high stress non-motivation condition, participants quickly made a reliance decision whether they chose to rely or not to rely, but in the motivation condition, non-reliance decisions were much slower than reliance decisions. The same pattern is found with AI2, but less pronounced. This indicates that making a decision inconsistent with an AI partner takes more time than a decision consistent with an AI partner, more prominent when that AI partner is highly reliable, but still found when the AI partner has low reliability.

Subjective Reliability

There was a main effect of AI reliability; reliability was rated higher for AI1 than AI2, $F(1, 76) = 73.41$, $p < .001$, $\eta_p^2 = .49$. There was a significant interaction with AI reliability and motivation, $F(1, 76) = 8.89$, $p = .004$, $\eta_p^2 = .11$. For AI1, reliability was rated higher in the motivation than non-motivation condition, $t(78) = 3.46$, $p = .001$; there was no difference for AI2. See Table 5 for subjective ratings of reliability among conditions.

The participant ratings of each AI partner provide some insight into their ability to attend to and remember AI reliability based on performance. Overall, participants reported accurately that AI1 was more reliable than AI2. Additionally, participants that were motivated rated the reliability of AI1 closer to its actual reliability compared to those that were not motivated. The ability to attend to, remember, and accurately update reliability expectations is important in the ongoing relationship between human and automation.

Trust in Automation

Objective and subjective measures of trust were analyzed in relation to reliance, and to each other.

Partner Investment. There was a main effect of AI reliability; participants invested more in AI1 than AI2, $F(1, 73) = 5.44$, $p = .022$, $\eta_p^2 = .07$.

Percent invested and reliance. Overall, investment amount and reliance were not significantly correlated, $r = .10$, $p = .21$. However, there was a significant correlation between investment and reliance with AI1, $r = .20$, $p = .04$, but not with AI2, $r = -.03$, $p = .41$. Correlations were also examined between the stress conditions. Investment and reliance with AI1 were correlated in the high, $r = .29$, $p = .04$, but not low stress condition, $r = .12$, $p = .23$. There were no significant correlations between the motivation conditions.

Percent invested and subjective reliability. The correlation between subjective reliability and investment was significant with both AI partners, AI1, $r = .59$, $p < .001$, AI2, $r = .50$, $p < .001$. The correlation between investment and subjective reliability with both AI partners was stronger in the high stress, (AI1: $r = .70$, AI2: $r = .55$, $p < .001$) than low stress condition, (AI1: $r = .46$, $p = .003$, AI2: $r = .37$, $p = .017$). When examined between motivation conditions, all correlations were significant at $p < .001$, with the exception of the non-motivation condition with AI2, but was still significant at $p = .004$.

Subjective Trust. There was a main effect of AI reliability; participants reported higher trust in AI1 than AI2, $F(1, 76) = 77.01$, $p < .001$, $\eta_p^2 = .50$. There was a main effect of stress; participants reported higher trust overall regardless of AI in the high than low stress condition, $F(1, 76) = 7.24$, $p = .009$, $\eta_p^2 = .09$. Additionally, there was an interaction with AI reliability and motivation, $F(1, 76) = 7.33$, $p = .008$, $\eta_p^2 = .09$. Participants reported higher trust with AI1 in the motivation than the non-motivation condition, $t(78) = 2.78$, $p = .007$; there was no difference in trust ratings with AI2 between motivation conditions, $p = .36$.

Subjective trust and subjective reliability were highly correlated with both AI partners, AI1, $r = .79$, $p < .001$, AI2 $r = .76$, $p < .001$. Correlations were significant among all conditions. Finally, the correlation between investment and subjective trust was significant for both AI partners; AI1, $r = .49$, $p < .001$, and AI2, $r = .47$, $p < .001$. Correlations were significant among all conditions.

Trust is a large factor in determining reliance, therefore it was important to assess trust in this experiment with three objectives: (a) how trust is affected by stress, motivation, and AI reliability, (b) how trust and reliance correlated under stress and motivation conditions, and (c) how useful an objective measure of human-human trust is when assessing human-automation trust. Trust was hypothesized to not be affected by stress, and while there was not a formal hypothesis on whether motivation would influence trust, it was not expected that motivation would have an effect on trust.

Trust between Conditions. Overall, there was greater trust with AI1 than AI2 as evidenced by investment and subject trust. In both trust measures, motivation appears to be the primary influencing factor, with higher trust reported for AI1 in the motivation than non-motivation condition. However, stress did appear to increase overall trust with both AI partners. Additionally, investment and subjective trust were correlated across all conditions, and with both AIs.

Trust and Reliance. There was a significant correlation between investment and reliance with AI1, but not AI2. With AI1, the correlation held for the high but not the low stress condition, which is the opposite of what was predicted. Both objective (i.e., investment amount) and subjective measures of trust were correlated with each other, but inconsistent in their correlation with reliance.

LIMITATIONS AND FUTURE RESEARCH

There were some limitations to note in this study, and with these limitations there are opportunities for future research. First, no direct gender comparisons could be made due to an insufficient number of female participants. Gender differences in reliance decisions in general, and then reliance decisions with and without stress and motivation requires more inquiry. Additionally, the population was mostly military officers, and therefore has limited generalizability beyond the sample demographic; different results may be found in a sample more representative of the general population. Future research should include a sample closer to the general population and examine the cortisol levels in comparison.

CONCLUSION

The current study investigated the effect of stress and motivation on reliance decisions with autonomous partners that differed in their reliability, one with high reliability and one with low reliability. Motivation had a stronger effect than stress. However, there was evidence that the combination of high stress and motivation affected reliance decisions. Reaction times were slower in the high stress condition when participants were motivated, so that participants took longer to decide *not* to rely on the highly reliable AI partner. This suggests more deliberative decision making when deciding against the advice of a reliable autonomous partner. Subjective reliability indicated the importance of motivation, with participants in the motivation condition reporting reliability closer to the actual reliability of AI1 compared to the non-motivation condition.

Decision making and ratings with each AI partner show that AI2 was perceived more closely to its actual reliability across conditions. However, AI1 was perceived the closest to its actual reliability in the motivation condition. Motivation has been shown to increase activity in the PFC. While it cannot be determined if motivation in this experiment increased PFC processing, it was evident that motivation did enhance decision making with a reliable autonomous partner as seen in the ability to report more accurate AI reliability and the use of slower, more deliberate decision making.

Advancements in technology are leading to more human-automation partnerships across all areas in the military. However, those systems are only useful when relied on appropriately. Understanding human decision making with automation in contexts filled with stress and uncertainty is vital to successful and efficient hybrid partnerships in the military. Overall, the effect of motivation on reliance decisions with automation should be investigated further, both with and without stress, as motivation was the stronger manipulation in the present experiment.

ACKNOWLEDGMENTS

This work was sponsored by the Department of the Navy, Office of Naval Research, Consortium for Robotics Unmanned Systems Education and Research at the Naval Postgraduate School.

The rationale and methods portion of this work was published in the proceedings of the 52th Hawaii International Conference on System Sciences (HICSS), 2019, pp. 237-243.

Some of the data in this paper are reported in the corresponding thesis from [the second author]; citation in references.

REFERENCES

Arnsten, A. F. (2009). Stress signalling pathways that impair prefrontal cortex structure and function. Nature reviews neuroscience, 10(6), 410–422. https://doi.org/10.1038/nrn2648

Berg, J., Dickhaut, J., & McCabe, K. (1995). Trust, reciprocity, and social history. Games and economic behavior, 10(1), 122–142. https://doi.org/10.1006/game.1995.1027

Bernkopf, M. (2021). Effects of stress and motivation on human reliance on automated decision-support systems (Master's thesis, Monterey, CA; Naval Postgraduate School).

Bland, A. R., Roiser, J. P., Mehta, M. A., Schei, T., Sahakian, B. J., Robbins, T. W., & Elliott, R. (2017). Cooperative behavior in the ultimatum game and prisoner's dilemma depends on players' contributions. Frontiers in psychology, 8, 1017. https://doi.org/10.3389/fpsyg.2017.01017

Brand, M., Labudda, K., & Markowitsch, H. J. (2006). Neuropsychological correlates of decision-making in ambiguous and risky situations. Neural Networks, 19(8), 1266–1276. https://doi.org/10.1016/j.neunet.2006.03.001

de Visser, E. J., Monfort, S. S., McKendrick, R., Smith, M. A., McKnight, P. E., Krueger, F., & Parasuraman, R. (2016). Almost human: Anthropomorphism increases trust resilience in cognitive agents. Journal of Experimental Psychology: Applied, 22(3), 331. https://doi.org/10.1037/xap0000092

Dickerson, S. S., & Kemeny, M. E. (2004). Acute stressors and cortisol responses: a theoretical integration and synthesis of laboratory research. Psychological bulletin, 130(3), 355. https://doi.org/10.1037/0033-2909.130.3.355

Endsley, M. R. (2017). From here to autonomy: lessons learned from human–automation research. Human factors, 59(1), 5–27. https://doi.org/10.1177/0018720816681350

Het, S., Rohleder, N., Schoofs, D., Kirschbaum, C., & Wolf, O. T. (2009). Neuroendocrine and psychometric evaluation of a placebo version of the 'Trier Social Stress Test'. Psychoneuroendocrinology, 34(7), 1075-1086.https://doi.org/10.1016/j.psyneuen.2009.02.008

Hochman, G., & Yechiam, E. (2011). Loss aversion in the eye and in the heart: The autonomic nervous system's responses to losses. Journal of behavioral decision making, 24(2), 140–156. https://doi.org/10.1002/bdm.692

Hoff, K. A., & Bashir, M. (2015). Trust in automation: Integrating empirical evidence on factors that influence trust. Human factors, 57(3), 407–434. https://doi.org/10.1177/0018720814547570

Kahneman, D. (2003). A perspective on judgment and choice: mapping bounded rationality. American psychologist, 58(9), 697. https://doi.org/10.1037/0003-066X.58.9.697

Kirschbaum, C., Pirke, K. M., & Hellhammer, D. H. (1993). The 'Trier Social Stress Test'–a tool for investigating psychobiological stress responses in a laboratory setting. Neuropsychobiology, 28(1-2), 76–81. https://doi.org/10.1159/000119004

Kouneiher, F., Charron, S., & Koechlin, E. (2009). Motivation and cognitive control in the human prefrontal cortex. Nature neuroscience, 12(7), 939–945. https://doi.org/10.1038/nn.2321

Lee, J. D., & Sanquist, T. F. (2000). Augmenting the operator function model with cognitive operations: Assessing the cognitive demands of technological innovation in ship navigation. IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans, 30(3), 273–285. doi: 10.1109/3468.844353.

Lee, J. D., & See, K. A. (2004). Trust in automation: Designing for appropriate reliance. Human factors, 46(1), 50–80. https://doi.org/10.1518/hfes.46.1.50_30392

Lighthall, N. R., Sakaki, M., Vasunilashorn, S., Nga, L., Somayajula, S., Chen, E. Y.,... & Mather, M. (2012). Gender differences in reward-related decision processing under stress. Social cognitive and affective neuroscience, 7(4), 476–484. https://doi.org/10.1093/scan/nsr026

Locke, H. S., & Braver, T. S. (2008). Motivational influences on cognitive control: behavior, brain activation, and individual differences. Cognitive, Affective, & Behavioral Neuroscience, 8(1), 99–112. https://doi.org/10.3758/CABN.8.1.99

Margittai, Z., Nave, G., Strombach, T., van Wingerden, M., Schwabe, L., & Kalenscher, T. (2016). Exogenous cortisol causes a shift from deliberative to intuitive thinking. Psychoneuroendocrinology, 64, 131–135.https://doi.org/10.1016/j.psyneuen.2015.11.018

Mather, M., & Lighthall, N. R. (2012). Risk and reward are processed differently in decisions made under stress. Current directions in psychological science, 21(1), 36–41. https://doi.org/10.1177/0963721411429452

Parasuraman, R., & Riley, V. (1997). Humans and Automation: Use, Misuse, Disuse, Abuse. Human Factors, 39(2), 230–253. https://doi.org/10.1518/001872097778543886

Payne, J. W., & Bettman, J. R. (2004). Walking with the scarecrow: The information-processing approach to decision research. Blackwell handbook of judgment and decision making, 110–132.

Rice, S., & Keller, D. (2009). Automation reliance under time pressure. Cognitive Technology.

Smyth, N., Hucklebridge, F., Thorn, L., Evans, P., & Clow, A. (2013). Salivary cortisol as a biomarker in social science research. Social and Personality Psychology Compass, 7(9), 605–625. https://doi.org/10.1111/spc3.12057

Spielberger, C. D., Gorsuch, R. L., & Lushene, R. E. (1970). STAI. Manual for the State-Trait Anxiety Inventory (Self Evaluation Questionnaire). Palo Alto California: Consulting Psychologist.

Starcke, K., Wolf, O. T., Markowitsch, H. J., & Brand, M. (2008). Anticipatory stress influences decision making under explicit risk conditions. Behavioral neuroscience, 122(6), 1352. https://doi.org/10.1037/a0013281

Unsworth, N., Heitz, R. P., Schrock, J. C., & Engle, R. W. (2005). An automated version of the operation span task. Behavior research methods, 37(3), 498–505. https://doi.org/10.3758/BF03192720

Van den Bos, R., Harteveld, M., & Stoop, H. (2009). Stress and decision-making in humans: performance is related to cortisol reactivity, albeit differently in men and women. Psychoneuroendocrinology, 34(10), 1449–1458.https://doi.org/10.1016/j.psyneuen.2009.04.016

van Dongen, K., & van Maanen, P. P. (2013). A framework for explaining reliance on decision aids. International Journal of Human-Computer Studies, 71(4), 410–424. https://doi.org/10.1016/j.ijhcs.2012.10.018

Wickens, C. D., & Dixon, S. R. (2007). The benefits of imperfect diagnostic automation: A synthesis of the literature. Theoretical Issues in Ergonomics Science, 8(3), 201–212. https://doi.org/10.1080/14639220500370105