ORIGINAL ARTICLE

Automation trust increases under high-workload multitasking scenarios involving risk

Tetsuya Sato[1] · Yusuke Yamani[1] · Molly Liechty[2] · Eric T. Chancey[2]

Abstract

Trust is a critical construct that influences human–automation interaction in multitasking workspaces involving imperfect automation. Karpinsky et al. (Appl Ergon, 70, 194–201, 2018) investigated whether trust affects operators' attention allocation in high-load scenarios using the multi-attribute task battery II (MATB). Results suggested that task load reduces trust towards imperfect automation, then reducing visual attention allocation to the monitoring task aided by the automation. Participants also reported reduced levels of trust in high-load conditions. However, it is possible that the participants in high-load conditions did not trust the system because their poor task performance did not have expressly adverse consequences (i.e., risk). The current experiments aimed to replicate and extend Karpinsky et al. (2018) by asking forty participants to concurrently perform a tracking task and system monitoring task in the MATB II with or without risk. The reliability of the automated aid supporting the system monitoring task was 70%. The study employed a 2 × 2 split-plot design with task load (easy vs. difficult) via magnitude of errors in the tracking task as a within-participant factor and risk (high vs. low) as a between-participant factor. Participants in the high-risk group received an instruction that poor performance would result in a repeat of the experiment, whereas participants in the low-risk group did not receive this instruction. Results showed that trust was comparable between the high- and the low-load conditions, but the high risk elevated trust in the high-load condition. This implies that operators display greater levels of trust when a multitasking environment demands greater attention and they perceive risk of receiving expressly adverse consequence, regardless of the true reliability of automated systems.

Keywords Human–machine systems · Trust · Risk · Attention allocation · Human–automation interaction

1 Introduction

Many modern professional tasks such as piloting an aircraft (e.g., Billings 1997), robotic arm control during space tele-operation (e.g., Li et al. 2014), and technology-aided military operations (e.g., Chen and Terrence 2009) require that operators concurrently perform multiple perceptual–cognitive tasks in information-rich environments.

✉ Yusuke Yamani
  yyamani@odu.edu

[1] Department of Psychology, Old Dominion University, 236B Mills Godwin Building, Norfolk, VA 23529, USA

[2] Leidos, Dayton, OH, USA

Human information-processing models conceptualize human operators as limited-capacity information processors, with a stream of discrete psychological stages including sensory detection, perception, interpretation, decision-making, and response selection and execution, supported by attentional resources (Wickens et al. 2013; Yamani and Horrey 2018). Attentional resources can refer to a single pool of energy that activates the information-processing stages and determines the system's ability to process information (Gopher 1993; Kahneman 1973). Information-rich multitasking workspaces, therefore, demand operators effectively allocate their attention to multiple information sources. When multiple tasks compete for operators' limited attentional resources, workload may increase and overall operator performance becomes compromised due to information overload (Tsang and Wilson 1997; Hancock and Warm 1989; Young et al. 2015).

In such attention-demanding multitasking workspaces, automated systems may be designed and implemented to

help reduce operator workload and improve human and system-level performance. Automated systems are devices or computers that replace or partially support the human information-processing functions (e.g., Bainbridge 1983) such as information acquisition, information analysis, decision selection and action implementation at varying levels of support (Parasuraman et al. 2000). With the recent advancement and proliferation of automation, the human has shifted from an active controller to a passive monitor of highly automated systems (e.g., Hoogendoom et al. 2014; Metzger and Parasuraman 2001; Parasuraman et al. 1996; Sheridan 1970). Unfortunately, humans are poor monitors (see Warm et al. 2008). To help the human accomplish this task, alerted-monitor systems have also become prevalent. Alerted-monitor systems consist of the human monitor and a sensor-based signaling system (Sorkin and Woods 1985). Signaling systems provide the human with information about the state of a highly automated system, and direct attention of the human to system faults and failures. Unfortunately, however, these systems are not perfect and can produce errors due to mechanical constraints and threshold settings. Interestingly, automated systems including a sensor-based signaling system can negatively influence behaviors of the human operator (Wickens and Dixons 2007; Parasuraman and Riley 1997).

1.1 Trust and attention allocation

One critical contributor to human–automation interaction, particularly as it relates to the alerted-monitor paradigm, is trust (Bliss et al. 1995; Chancey et al. 2017; Meyer 2001; Rice 2009; Sheridan 2019a, b; Vanderhaegen 2017). Lee and See (2004) define human–automation trust as "an attitude that an agent will help achieve an individual's goals in a situation characterized by uncertainty and vulnerability." From this theoretical perspective, trust is determined by three goal-oriented informational bases: performance, which describes what the automation is doing; process, which describes how the automation is performing a task; purpose, which describes why the automation was developed (Lee and See 2004). The stability of trust depends on the degree to which these informational sources are represented, and ultimately affects the appropriateness of trust and subsequent behavioral responses (i.e., trust calibration). Research has shown that trust may play a critical role in determining operators' attention allocation and monitoring strategies (Bailey and Scerbo 2007; Bliss and Dunn 2000; Karpinsky et al. 2018; Molloy and Parasuraman 1996).

To illustrate, Karpinsky et al. (2018) recently investigated the effects of task load on attention allocation strategies and human–automation trust in a multitasking environment using the multi-attribute task battery (MATB; Comstock and Arnegard 1992). In their experiments, participants were asked to concurrently perform a manual tracking task and a system monitoring task assisted by an imperfect signaling system in a low-fidelity flight-simulation environment. Participants' eye movements were used as a measure of attention allocation (e.g., Horrey et al. 2006), and their subjective trust rating and task performance were analyzed.

Participants in Karpinsky et al. (2018) reported substantially lower levels of performance- and process-based trust, but not purpose-based trust, toward the imperfect signaling system supporting the system monitoring task, when the tracking task required more frequent manual corrections. This occurred even though the reliability of automation remained the same (70%) between the high- and low-load conditions. The researchers reasoned that participants who were unfamiliar with the signaling system and MATB tasks likely calibrated their trust based on their perception of the behaviors of the signaling system. Yet, when the tracking task demanded more attention, participants allocated less attention to monitor and therefore generated misperception of the signaling system's behavior, which also showed a lower level of the performance and process dimensions of trust.

It remains unclear, however, whether participants in high-load conditions reported less trust towards the system because poor task performance did not have expressly adverse consequences (i.e., risk). That is, it is possible that the participants intentionally neglect to closely monitor the system in the high-load conditions demanding more attention to the tracking task because they did not perceive risk associated with their task performance. Risk is frequently cited as a key factor in determining trust and trust-related behaviors in human–automation interaction (e.g., Chancey et al. 2017; Lyons and Stokes 2012; Mayer et al. 1995; Hoff and Bashir 2015; Riley 1994). Risk can be defined as "the extent to which there is uncertainty about whether potentially significant and/or disappointing outcomes of decisions will be realized" (Sitkin and Pablo 1992, p. 10). Corritore et al. (2003) argue that environments that involve risk and vulnerability are the only conditions where trust evolves (also see de Vries et al. 2003; Luhmann 1979, 1988). Though risk as a condition where operators face potential harm or exposure is necessary for trust formation (Corritore et al. 2003; de Vries et al. 2003; Hoff and Bashir 2015; Lee and See 2004), there exist only a few studies directly looking at the effects of risk on trust in the literature of human–automation interaction (e.g., Chancey et al. 2017; Lyons and Stokes 2012; Lewandowsky et al. 2000; see Vanderhaegen 2017 on how risk relates to the use of automated technology in cyber-physical and human systems), and none have investigated its impact on trust on the three dimensions and visual attention when interacting with imperfect signaling system.

To investigate the effects of risk on automation trust in an attention-demanding flight environment, we employed a

similar experimental design as Karpinsky et al. (2018), yet added a second condition in which we manipulated the risk involved with performing poorly on the experimental tasks. We first aimed to replicate the results of Karpinsky et al. (2018) that participants would trust the imperfect signaling system less in the high-load condition than the low-load condition. Furthermore, we hypothesized that this effect of load on automation trust would diminish under high risk. Finally, we hypothesized that the effect of load would be smaller in the high-risk group than the low-risk group, showing the interaction effect.

2 Method

The current experiment asked participants to concurrently perform the tracking task and system monitoring task in the MATB-II (Santiago-Espada et al. 2011). An imperfect signaling system with reliability of 70% assisted the participants in the system monitoring task. The study employed a 2×2 split-plot design with task load (high vs. low) via magnitude of errors in the tracking task as a within-participant factor and risk (high vs. low) as a between-participant factor. Participants in the high-risk group received an instruction that poor performance would result in a repeat of the experiment, whereas participants in the low-risk group did not receive this instruction.

2.1 Participants

Forty participants (26 females; $M = 22.4$ years, SD = 7.7 years) were recruited from the community of Old Dominion University, Norfolk, Virginia. Participants were screened for normal color perception, using the Ishihara color blindness test (1989), and normal or corrected-to-normal vision (near acuity = 20/23.4, SD = 5.3, far acuity = 20/20.8, SD = 4.6). Those who score over 19 on the motion sickness susceptibility questionnaire (MSSQ; Golding 1998) were not eligible for the experiment. Participants were compensated with class credit for participation.

2.2 Apparatus

Stimuli were presented on a Samsung T24C550 23.6″ LED monitor with a resolution of 1920×1080 pixel and a frame rate of 75 Hz. The MATB-II was hosted on a Dell Opti-Plex 9020 on Windows 7. Eye movements were recorded at a sampling rate of 250 Hz via Eyelink II (SR Research, Mississauga, Ontario, Canada) controlled by a Dell Precision 390 running on Windows XP. Participants viewed the monitor at the distance of approximately 80 cm on a chin rest. The experiment was conducted in a quiet room with dimmed lights.

2.3 Flight-simulation task

The MATB-II is a computer-based flight simulator designed to evaluate human performance in up to four tasks regularly and simultaneously performed by aircraft pilots in flight (tracking, system monitoring, fuel management, and communication). Participants in the current study performed the tracking and system monitoring tasks (e.g., Karpinsky et al. 2018).

2.3.1 Tracking task

The compensatory tracking task required participants to keep the moving circular target within a dotted square at the center of the window using a joystick. The circular target represents the aircraft's deviation from the designated route. The tracking task is located at the top center of the MATB display (Fig. 1). The frequency of the force function that determines the deviation at every refresh was set to either 0.12 Hz or 0.06 Hz (high or low load, respectively). The computer sampled the participant's input in *X* and *Y* dimensions at 20 Hz, with the root mean square (RMS) deviation between the target and the center point computed. The mean of RMS over the duration of each block was calculated.

2.3.2 System monitoring task

The system monitoring task is located at the top left of the display (Fig. 1). The system monitoring task required participants to monitor the temperature and pressure levels of the left and right engine, indicated by four vertical gauges, and respond if a vertically fluctuating pointer for any one of the four gauges (F1–F4) deviates from the middle point of the gauge. Two boxes above the four gauges signify the states of the automated system. The green box (F5) indicates a "normal" state, while the red box (F6) indicates a "warning" state. Each 20-min block consisted of 14 hit (Hit) and 6 false alarm (FA) events, presented in a random order and random time interval (mean = 29.2 s; range = 13–53 s). This study did not employ "miss" events because a previous study found no difference in human performance and trust between environments involving FA events or Miss events (Karpinsky et al. 2018). In Hit events, the signaling system accurately detected system malfunction and notifies participants by illuminating the "warning" box (F6). In this case, participants were required to (1) turn off the warning by pressing the red box (F6), (2) turn the state of the signaling system back to normal by pressing the green box (F5), and (3) correct the malfunctioning gauge by selecting the corresponding gauge (F1–F4). In FA events, the signaling system
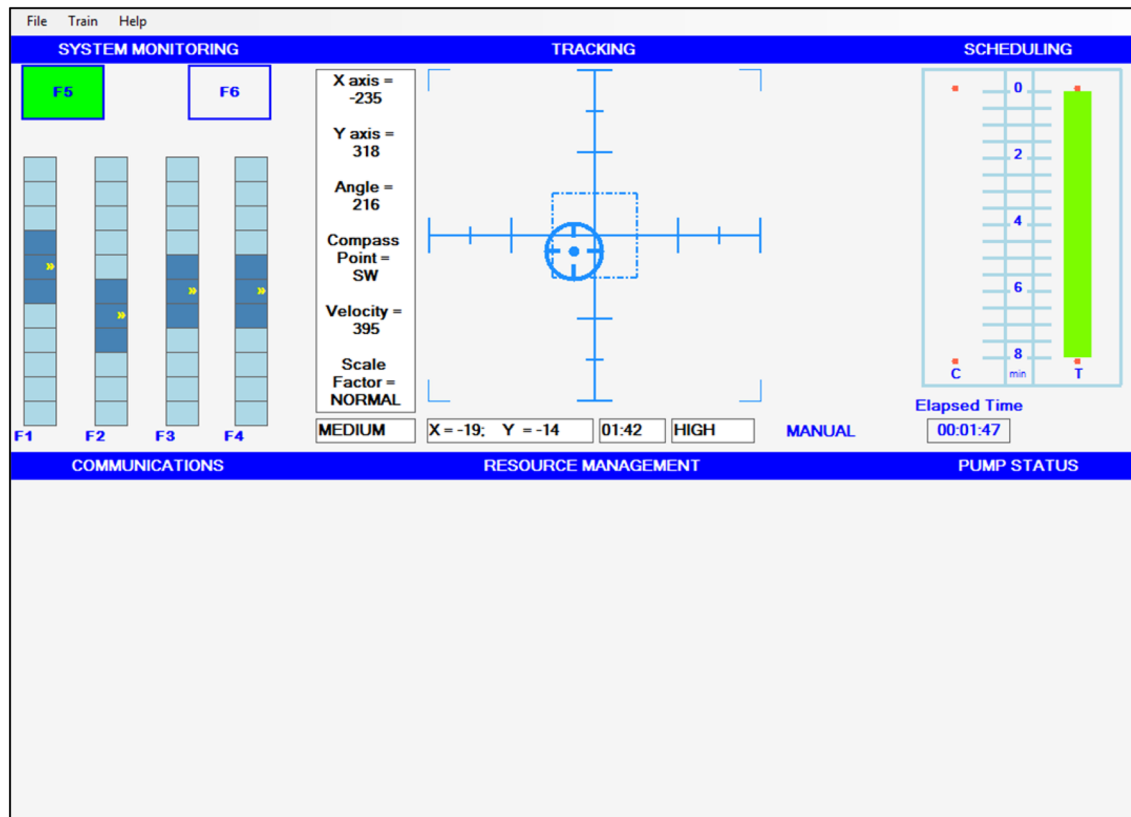
Fig. 1 A sample display of the MATB task. System monitoring task (top left) and tracking task (top center)

illuminated the "warning" box even though the gauges were functioning normally. In this case, participants were asked to turn the state of the signaling system back to normal by pressing the green box (F5).

2.4 Dependent variables

Subjective workload, trust, visual attention allocation, perceived risk, system monitoring performance, and tracking performance were the dependent measures of the experiment.

2.4.1 Subjective workload

Subjective workload was measured with the NASA-Task Load Index (TLX; Hart and Staveland 1988), which was comprised of 6 items on a 21-point gradient scale ranging from very low to very high (minimum score = 6, maximum score = 126). The 6 items assessed mental demand, physical demand, temporal demand, effort, performance, and frustration level.

2.4.2 Trust

Trust was measured via the trust questionnaire used by Chancey et al. (2017), which consisted of three factors (i.e., performance, process, and purpose) and 13 items on a 12-point Likert scale ranging from (1) not descriptive to (12) very descriptive (minimum score = 13, maximum score = 156).

2.4.3 Perceived risk

Perceived risk was measured via a modified version of the risk questionnaire by Simon et al. (1999; cf. Chancey et al. 2017). The questionnaire consisted of five items on a 12-point Likert scale ranging from (1) not descriptive to (12) very descriptive (minimum score = 5, maximum score = 60).

2.4.4 Visual attention allocation

Visual attention allocation was quantified by percent dwell time (PDT) and saccade probabilities. For both eye movement measures, two display areas covering the tracking task and the system monitoring task were defined as the areas of

interest (AOIs). PDT is the proportion of time that fixation existed within a particular AOI, and PDTs for the tracking task and the system monitoring task were computed. Additionally, we calculated the proportion of saccades that started from the tracking task and ended at the system monitoring task, and those that started from the system monitoring and ended at the tracking task separately.

2.4.5 MATB-II performance

For tracking performance, the mean of RMS for each block was calculated. For system monitoring performance, error rates and reaction times (RTs) were calculated for each trial and the means were calculated separately for the Hit and FA events. Specifically, error rates indicate the proportion of events with at least one incorrect response and RTs indicate the time interval between the onset of an event (e.g., deviation of the fluctuating pointer of the target gauge and illumination of the "warning" box for the Hit event, and illumination of the "warning" box for the FA event) and participants' first response to the event, regardless of its accuracy.

2.5 Procedure

Participants completed an informed consent form, a general demographics form, and the visual screening process for color perception and visual acuity, followed by the MSSQ (Golding 1998) and simulator sickness questionnaire (SSQ) pre-test (Kennedy et al. 2009). Participants were then randomly assigned to either high- or low-risk group. Participants in the high-risk group received instructions indicating that failure to perform well would result in longer participation time in the experiment, but this consequence was not enforced (cf., Chancey et al. 2017). Participants in the low-risk group did not receive this instruction. Participants were told that the signaling system was imperfect prior to the practice session, but the exact reliability level was not given. In the practice session, participants familiarized themselves with the tracking and system monitoring tasks by performing a 5-min block including Hits and FAs. Once participants were adequately familiarized with the tasks, participants were asked to complete the perceived risk questionnaire. In the experimental session, participants completed two 20-min experimental trials under high- and low-load conditions in a counterbalanced order across participants. Halfway through each experimental trial, participants were presented with the trust questionnaire. Following this, participants completed the remainder of the trial. At the end of each trial, participants completed the NASA-TLX. Participants were allowed to rest between the two experimental trials. Upon the completion of the experimental session, the participants filled out the SSQ post-test and were debriefed the purpose of the
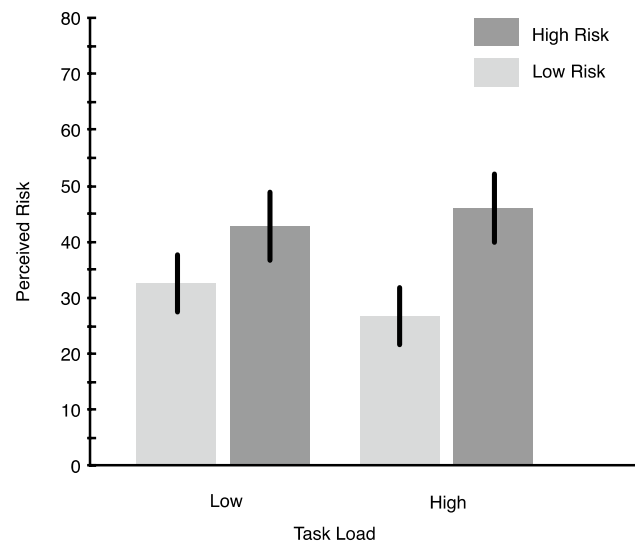

Fig. 2 Mean perceived risk scores for the low- and high-load conditions for the high-risk and low-risk groups. Error bars represent 95% between-subject confidence intervals (CIs)

experiment and risk manipulation. Participants received a credit for their participation and exited the lab.

2.6 Statistical analysis

Default Bayesian tests (Rouder and Morey 2012) were employed instead of null-hypothesis significance tests (NHSTs). Bayesian analysis allows providing evidence for or against the null hypothesis, while the NHSTs do not. Bayes factors are the measure of evidence for an effect of interest, reported as B_{10}. Bayes factors indicate ratios of likelihood that the observed data arise from a model including an effect of interest to that excluding the effect. Another benefit of using Bayesian tests over the NHSTs is that Bayes factors measure the strength of evidence unlike the p values of the NHSTs. That is, Bayes factors of 100, for example, indicates that it is 100 times more likely that observed data have arisen from a model with the effect of interest to the null model, while the p values do not provide such information. Finally, we follow the terminologies introduced by Jeffreys (1961) to describe the magnitude of each effect.

3 Results

Subjective workload, trust, visual attention, and performance data were subject to separate 2×2 split-plot Bayesian analyses of variance (ANOVAs) with load (high vs. low) as a within-subject factor and risk (high vs. low) as a between-subject factor. Due to technical issues with the eye tracker, data from two participants were excluded from the analyses. As a manipulation check, a t test was performed to compare

perceived risk levels between the group that received the instruction and the group that did not. Participants in the high-risk group reported decisively higher perceived risk than the low-risk group [$M = 44.36$ vs. 30.02; independent-samples $t(36) = 4.23$, $B_{10} = 152.63$]. Figure 2 presents mean perceived risk scores.

3.1 Subjective workload

Unexpectedly, data showed substantial evidence that operators reported comparable levels of subjective workload between the high- and low-load conditions [$M = 49.8$ vs. 51.7 for the low- and high-load conditions respectively; $F < 1$, $B_{10} = 1/3.31$]. The data were indifferent to the remaining effects [$B_{10} = 1/2.05$ for the effect of risk and $B_{10} = 1/1.92$ for the interaction effect].

3.2 Trust

Scores on each of the three subscales, performance, process, and purpose dimensions of trust of the trust questionnaire were analyzed separately. Figures 3, 4, and 5 present mean trust scores for the performance-, process-, and purpose information dimensions, respectively. First, and most importantly, data indicate substantial evidence for the two-way interaction [$F(36, 7) = 7.04$, $\eta^2_G = 0.02$, $B_{10} = 4.60$], suggesting that operators tended to rate performance dimension of trust lower when the task load was higher under low perceived risk [$M = 42.84$ vs. 46.31 for the high- and low-load conditions; paired-samples $t(18) = 1.65$, $B_{10} = 0.75$] but the opposite under high-risk group [$M = 48.47$ vs. 45.21 for the high- and low-load conditions; paired-samples $t(18) = 2.28$,
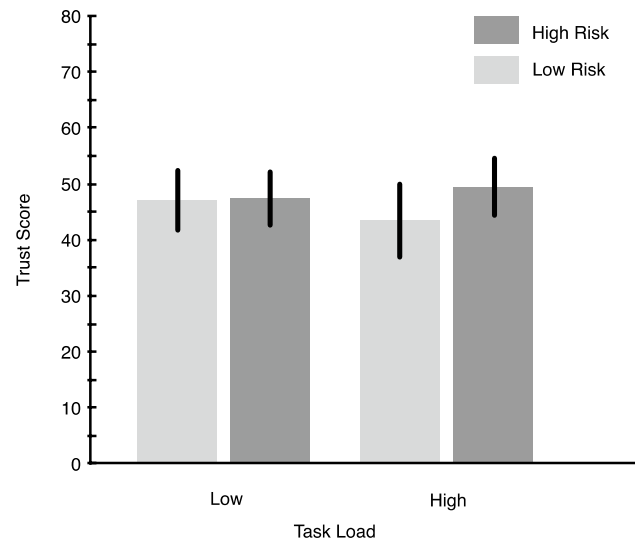

Fig. 4 Mean process-based trust scores for the low- and high-load conditions for the high-risk and low-risk groups

$B_{10} = 1.88$]. The data provided substantial evidence against the main effect of load and anecdotal evidence against the main effect of risk [both Fs < 1, $B_{10} = 1/4.11$ for load and $B_{10} = 1/1.73$]. A similar data pattern emerged for the process dimension of trust, but the data showed only anecdotal evidence for the interaction effect [$F(1, 36) = 5.32$, $\eta^2_G = 0.02$, $B_{10} = 2.31$]. Again, the data substantially supported the absence of the main effect of Load and were indifferent to the effect of Risk [both Fs < 1, $B_{10} = 1/3.70$ for load and $B_{10} = 1/1.69$]. Finally, none of the effects were substantial for the purpose dimensions of trust [$1/2.99 < B_{10} < 1/1.24$].
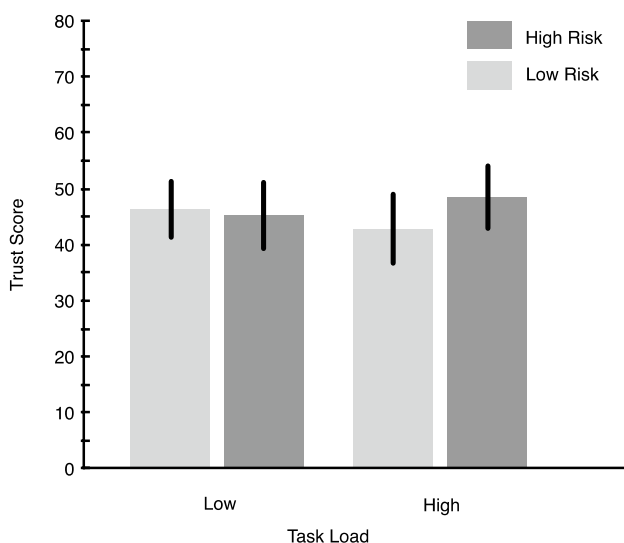

Fig. 3 Mean performance-based trust scores for the low- and high-load conditions for the high-risk and low-risk groups
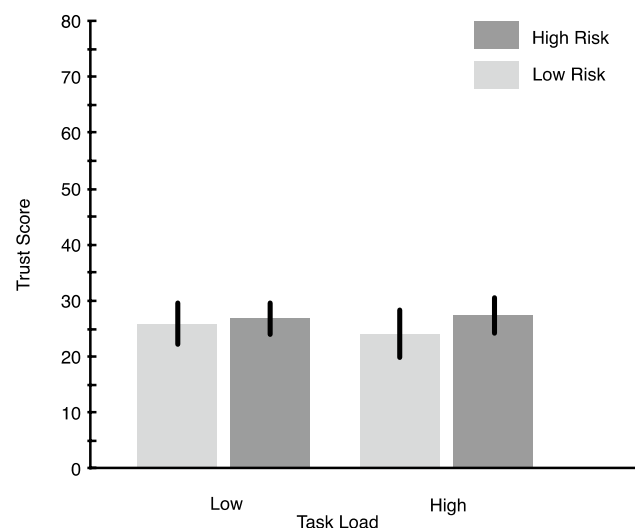

Fig. 5 Mean purpose-based trust scores for the low- and high-load conditions for the high-risk and low-risk groups

3.3 Visual attention directed to system monitoring

The analysis of PDT within the system monitoring display and frequency of saccades starting from the tracking display to the system monitoring display did not detect substantial differences between the experimental conditions [$1/3.32 < B_{10} < 1.09$]. However, the high-load condition produced greater probabilities that operators made saccades starting from the system monitoring display and ending at the tracking display [$M = 0.46$ vs. 0.41, for the high- and the low-load conditions, respectively; $F(1, 36) = 6.67$, $\eta^2_G = 0.02$, $B_{10} = 3.66$]. Data did not indicate the presence of the other effects [$F(1,36) = 2.30$, $\eta^2_G = 0.05$, $B_{10} = 1.00$ for the main effect of risk; $F < 1$, $B_{10} = 1/3.12$, for the interaction effect].

3.4 RTs

For the FA events, data gave no substantial evidence for or against all of the effects [$1/2.99 < B_{10} < 1/1.11$]. For the Hit events, data produced substantial evidence against the effect of load [$F < 1$, $B_{10} = 1/3.62$], indicating that operators in the high- and low-load conditions responded at comparable speed. The other effects were not substantial [$B_{10} = 1/1.51$ for the main effect of risk and $B_{10} = 1/1.99$ for the interaction effect].

3.5 Error rates

For the FA events, data showed substantial evidence against the effect of load on error rates [$F < 1$, $B_{10} = 1/3.38$]. No other effects were substantial [$B_{10} = 1/1.38$ for the main effect of risk and $B_{10} = 1/259$ for the interaction effect]. For the Hit events, the high-load condition produced numerically greater error rates than the low-load condition [$M = 0.26$ vs. 0.19] but the data did not indicate substantial evidence [$F(1, 36) = 4.62$, $\eta^2_G = 0.02$, $B_{10} = 1.66$]. Data indicated no substantial evidence for or against the effect of risk [$B_{10} = 1/2.22$] and substantial evidence against the interaction effect [$B_{10} = 1/3.01$].

4 Discussion

A previous flight simulator experiment using the MATB paradigm showed that when operators concurrently performed the tracking task and the system monitoring task aided by the imperfect signaling system, operators reported less trust when the tracking task demanded more manual correction, despite the identical reliability of the signaling system (cf. Karpinsky et al. 2018). It is possible that operators in the high-load condition failed to allocate their attention to monitor the behavior of the signaling system, and a resulting misperception of the system's performance led to declined trust toward the system. We hypothesized that operators neglected to allocate sufficient attention to the system monitoring task because poor task performance did not have expressly adverse consequences. Using the experiment design of Karpinsky et al. (2018), our participants in the high-risk group received an instruction that poor performance would result in a repeat of the experiment, whereas participants in the low-risk group did not receive this instruction. We expected that operators continue to trust less toward the imperfect signaling system due to elevated attention demands by the tracking task in the high-load condition compared to the low-load condition, regardless of the presence of risk.

The current experiment generally did not replicate the results of Karpinsky et al. (2018), showing no evidence that operators' performance and trust levels decline in the high-load condition compared to the low-load condition. This failed manipulation of the tracking task difficulty was also apparent from subjective workload levels measured by the NASA-TLX. However, analysis of eye movements indicated that operators in the high-load condition were more likely to return from the system monitoring task to the tracking task given that they fixated within the system monitoring task when compared to those in the low-load condition. This suggests that the load manipulation was still effective to a measurable and reliable extent to modulate operator's attention allocation strategies. The risk manipulation contributed to larger variability between participants, which might have reduced the power of the current experiment. Alternatively, "task load" was manipulated by the magnitude of manual corrections that operators must make which is not attention-demanding per se but largely involves motor responses. Future experiments should specifically examine effects of operators' cognitive load on attention allocation strategies and automation trust.

Unexpectedly, in the absence of the main effect of task load, operators in the high-risk group reported greater trust toward the imperfect signaling system than those in the low-risk group. Specifically, the operators showed these trends on the performance- and process dimensions of trust, but not on the purpose dimension. In a situation where performance failure can lead to adverse consequences, operators may display more trust in an imperfect automated system based on its performance (i.e., what it is doing) and perhaps process (i.e., how it is performing a task). It is noteworthy that greater risk elevated the performance-based trust without measurable differences in operator performance and attention allocation. Theoretically, this point accords with the view that trust is characterized as an affective evaluation of beliefs (Lee and See 2004) instead of behavioral responses (Chancey et al. 2017). The fact that their trust ratings varied in spite of the reliability of the signaling system regardless of perceived

risk levels also supports this view. Furthermore, for the participants who were not particularly familiar with the MATB task and the signaling system, it appears that operators' trust evolves first from performance, then process, and finally purpose. More research is needed to investigate how the effect of risk impacts trajectories of trust development within the framework of three information bases of human–automation trust (Lee and See 2004) as operators gain more experience with imperfect signaling system.

Future research should use expert pilots to examine the generalizability of the current finding to real-world piloting of modern aircrafts. Though the current experiment used undergraduate students as participants, the MATB is an experimental paradigm without modeling a specific aircraft system where the environment is novel regardless of their expertise. However, novices and experts may form their trust towards a specific signaling system or the controlling system as a whole, and further research is necessary for examining this possibility. Additionally, performance and trust elicited by the current paradigm might depend on participant's experience with video gaming. Future research may incorporate a game experience questionnaire (Green and Bavelier 2003) to examine a potential link between video gaming and performance in flight simulation.

Practically, automation designers should consider the potential impact of risk on users' trust toward automated decision aids with varying sensor thresholds. Though previous research suggests that operators misuse automated systems that issue frequent errors (Parasuraman and Riley 1997), operators especially in high-load environments can display higher levels of trust towards imperfect automated systems when exposed to high levels of risk with clear behavioral consequences. Training programs may incorporate elements of risk to control and elevate user trust in such systems, which can influence subsequent behavioral responses in multitasking workspaces (Chancey et al. 2017; Karpinsky et al. 2018). However, it is unclear if the effect of risk on trust generalizes to various workspaces where multiple tasks impose more severe attention demands, and future research should explore this possibility.

References

Bailey NR, Scerbo MW (2007) Automation-induced complacency for monitoring highly reliable systems: the role of task complexity, system experience, and operator trust. Theory Issues Ergon Sci 8:321–348

Bainbridge L (1983) Ironies of automation. Automatica 19:775–779

Billings CE (1997) Aviation automation: the search for a human-centered approach. Lawrence Erlbaum Associates Publishers, Mahwah

Bliss JP, Dunn MC (2000) Behavioral implications of alarm mistrust as a function of task workload. Ergonomics 43:1283–1300

Bliss JP, Gilson RD, Deaton JE (1995) Human probability matching behaviour in response to alarms of varying reliability. Ergonomics 38:2300–3212

Chancey ET, Bliss JP, Yamani Y, Handley HAH (2017) Trust and the compliance-reliance paradigm: the effects of risk, error bias, and reliability on trust and dependence. Hum Factors 59:333–345

Chen JYC, Terrence PI (2009) Effects of imperfect automation and individual differences on concurrent performance of military and robotics tasks in a simulated multitasking environment. Ergonomics 58:907–920

Comstock JR, Arnegard RJ (1992) The multi-attribute task battery for human operator workload and strategic behavior research. NASA Langley Research Center, Hampton

Corritore CL, Kracher B, Wiedenbeck S (2003) On-line trust: concepts, evolving themes, a model. Int J Hum Comput Stud 58(6):737–758

de Vries P, Midden C, Bouwhuis D (2003) The effects of errors on system trust, self-confidence, and the allocation of control in route planning. Int J Hum Comput Stud 58(6):719–735

Golding JF (1998) Motion sickness susceptibility questionnaire revised and its relationship to other forms of sickness. Brain Res Bull 47:507–516

Gopher D (1993) The skill of attentional control: acquisition and execution of attention strategies. In: Meyer DE, Kornblum S (eds) Attention and performance XIV. MIT Press, Cambridge, pp 299–322

Green CS, Bavelier D (2003) Action video game modifies visual selective attention. Nature 423(6939):534

Hancock PA, Warm JS (1989) A dynamic model of stress and sustained attention. Hum Factors 31:519–537

Hart SG, Staveland LE (1988) Development of NASA-TLX (Task Load Index): results of empirical and theoretical research. Adv Psychol 52:139–183

Hoff KA, Bashir M (2015) Trust in automation: integrating empirical evidence on factors that influence trust. Hum Factors 57(3):407–434

Hoogendoom R, van Arerm B, Hoogendoom S (2014) Automated driving, traffic flow efficiency, and human factors: literature review. Transp Res Rec 2442:113–120

Horrey WJ, Wickens CD, Consalus KP (2006) Modeling drivers' visual attention allocation while interacting with in-vehicle technologies. J Exp Psychol Appl 12:67–78

Jeffreys H (1961) Theory of probability, 3rd edn. Oxford University Press, New York

Kahneman D (1973) Attention and effort. Prentice-Hall, Englewood Cliffs

Karpinsky ND, Chancey ET, Palmer DB, Yamani Y (2018) Automation trust and attention allocation in multitasking workspace. Appl Ergon 70:194–201

Kennedy RS, Lane NE, Berbaum KS, Lilienthal MG (2009) Simulator sickness questionnaire: an enhanced method for quantifying simulator sickness. Int J Aviat Psychol 3:203–220

Lee JD, See KA (2004) Trust in automation: designing for appropriate reliance. Hum Factors 46:50–80

Lewandowsky S, Mundy M, Tan G (2000) The dynamics of trust: comparing humans to automation. J Exp Psychol Appl 6(2):104

Li H, Wickens CD, Sarter N, Sebok A (2014) Stages and levels of automation in support of space teleoperations. Hum Factors 56:1050–1061

Luhmann N (1979) Trust and power: two works. Wiley, Hoboken

Luhmann N (1988) Familiarity, confidence, trust: Problems and alternatives. In: Gambetta D (ed) Trust: making and breaking cooperative relations. Basil Blackwell, New York, pp 94–108

Lyons JB, Stokes CK (2012) Human–human reliance in the context of automation. Hum Factors 54:112–121

Mayer RC, Davis JH, Schoorman FD (1995) An integrative model of organizational trust. Acad Manag Rev 20:709–734

Metzger U, Parasuraman R (2001) The role of the air traffic controller in future air traffic management. An empirical study of active control versus passive monitoring. Hum Factors 43:519–528

Meyer J (2001) Effects of warning validity and proximity on responses to warnings. Hum Factors 43:563–572

Molloy R, Parasuraman R (1996) Monitoring an automated system for a single failure. Vigilance and task complexity effects. Hum Factors 38:311–322

Parasuraman R, Riley V (1997) Humans and automation: use, misuse, disuse, abuse. Hum Factors 39:230–253

Parasuraman R, Mouloua M, Molloy R, Hilburn B (1996) Monitoring of automated systems. In: Parasuraman R, Mouloua M (eds) Automation and human performance: theory and applications. Erlbaum, Hillsdale, NJ, pp 91–115

Parasuraman R, Sheridan TB, Wickens CD (2000) A model for types and levels of human interaction with automation. IEEE Trans Syst Man Cybern Part A Syst Hum 30:286–297

Rice S (2009) Examining single- and multiple-process theories of trust in automation. J Gen Psychol 13:303–319

Riley V (1994) A theory of operator reliance on automation. In: Mouloua M, Parasuraman R (eds) Human performance in automated systems: current research and trends. Erlbaum, Hillsdale, pp 8–14

Rouder JN, Morey RD (2012) Default Bayes factors for model selection in regression. Multivar Behav Res 47:877–903

Santiago-Espada Y, Myer RR, Latorella KA, Comstock JR (2011) The multi-attribute task battery II (MATB-II) software for human performance and workload research: a user's guide (NASA/TM-2011-217164). National Aeronautics and Space Administration, Langley Research Center, Hampton

Sheridan TB (1970) On how often the supervisor should sample. IEEE Trans Syst Sci Cybern 6:140–145

Sheridan TB (2019) Extending three existing models to analysis of trust in automation: signal detection, statistical parameter estimation, and model-based control. Hum Factors. https://doi.org/10.1177/0018720819829951

Sheridan TB (2019b) Individual differences in attributes of trust in automation: measurement and application to system design. Front Psychol 10:1117

Simon M, Houghton SM, Aquino K (1999) Cognitive biases, risk perception, and venture formation: how individuals decide to start companies. J Bus Ventur 15:113–134

Sitkin SB, Pablo AM (1992) Reconceptualizing the determinants of risk behavior. Acad Manag Rev 17:9–38

Sorkin RD, Woods DD (1985) Systems with human monitors: a signal detection analysis. Hum Comput Interact 1:49–75

Tsang PS, Wilson G (1997) Mental workload. In: Salvendy G (ed) Handbook of human factors and ergonomics, 2nd edn. Wiley, New York, pp 243–268

Vanderhaegen F (2017) Towards increased systems resilience: new challenges based on dissonance control for human reliability in Cyber-Physical & Human Systems. Annu Rev Control 44:316–322

Warm JS, Parasuraman R, Matthews G (2008) Vigilance requires hard mental work and is stressful. Hum Factors 50:433–441

Wickens CD, Dixons SR (2007) The benefits of imperfect diagnostic automation: a synthesis of the literature. Theor Issues Ergon Sci 8:201–212

Wickens CD, Hollands JG, Banbury S, Parasuraman R (2013) Engineering psychology and human performance, 4th edn. Pearson, Boston

Yamani Y, Horrey WJ (2018) A theoretical model of human–automation interaction grounded in resource allocation policy during automated driving. Int J Hum Factors Ergon 5:225–239

Young MS, Brookhuis KA, Wickens CD, Hancock PA (2015) State of science: mental workload in ergonomics. Ergonomics 58:1–17