Cognition, Technology & Work
https://doi.org/10.1007/s10111-022-00717-z

ORIGINAL ARTICLE

Task priority reduces an adverse effect of task load on automation trust in a dynamic multitasking environment

Tetsuya Sato[1] · Samia Islam[1] · Jeremiah D. Still[1] · Mark W. Scerbo[1] · Yusuke Yamani[1]

Received: 8 April 2022 / Accepted: 21 September 2022

Abstract
The present study examined how task priority influences operators' scanning patterns and trust ratings toward imperfect automation. Previous research demonstrated that participants display lower trust and fixate less frequently toward a visual display for the secondary task assisted with imperfect automation when the primary task demanded more attention. One account for this phenomenon is that the increased primary task demand induced the participants to prioritize the primary task than the secondary task. The present study asked participants to perform a tracking task, system monitoring task, and resource management task simultaneously using the Multi-Attribute Task Battery (MATB) II. Automation assisted the system monitoring task with 70% reliability. Task load was manipulated via difficulty of the tracking task. Participants were explicitly instructed to either prioritize the tracking task over all other tasks (tracking priority condition) or reduce tracking performance (equal priority condition). The results demonstrate the effects of task load on attention distribution, task performance and trust ratings. Furthermore, participants under the equal priority condition reported lower performance-based trust when the tracking task required more frequent manual input (tracking condition), while no effect of task load was observed under the tracking priority condition. Task priority can modulate automation trust by eliminating the adverse effect of task load in a dynamic multitasking environment.

Keywords Multitasking · Human–automation trust · Eye movement · Attention distribution

1 Introduction

Many professional tasks such as controlling an aircraft (e.g., Billings 1997), a robotic arm (e.g., Li et al. 2014), and an air traffic control system (e.g., Loft et al. 2016) require operators to perform multiple concurrent tasks. A human operator, as often conceptualized as a limited-capacity information processor (Neisser 1980), is assumed to allocate attentional resources to meet the demand of each task by systematically adjusting resource allocation policy (Kahneman 1973; Wickens et al. 2015; Yamani and Horrey 2018). However, modern applied tasks often impose heavy computational and processing demands, necessitating the use of automation to successfully and efficiently execute its mission. Automation is often defined as a technological system that performs functions that can or cannot be accomplished by human operators (Bainbridge 1983; Parasuraman et al. 2000) and hypothesized to systematically reduce the attentional demand of the required task at various human information-processing stages including sensory processing, perception/working memory, decision making, and response selection (Yamani and Horrey 2018).

The proliferating use of automation has shifted the operator's role from actively controlling the system to passively monitoring system behavior (Bainbridge 1983). Unfortunately, research has demonstrated that humans are particularly poor at monitoring performance for a period of time (e.g., vigilance decrement; Mackworth 1948; Molloy and Parasuraman 1996; Warm et al. 2008; McCarley and Yamani 2021). To aid with this task, practitioners developed alerted-monitor systems to present the state of an automated system at every moment and help direct the operators' attention to system errors. However, alerted-monitor systems can produce signaling errors (i.e., false alarms and miss events) due to the system's threshold setting (Getty et al. 1995). Such signaling errors from an automated system could influence an operator's trust toward the automated system (e.g.,

✉ Tetsuya Sato
tsato003@odu.edu

[1] Department of Psychology, Old Dominion University, 325B Mills Godwin Building, Norfolk, VA 23529, USA

Published online: 27 September 2022

Chancey et al. 2017), delaying human response (Breznitz 1984; Getty et al. 1995; Sorkin 1988) and increasing workload (Dixon and Wickens 2006).

Trust is a critical factor for successful human–automation interaction (Hoff and Bashir 2015; Lee and Moray 1992; Lee et al. 2021; Long et al. 2022; Lyons and Stokes 2012; Muir 1994; Muir and Moray 1996; Schaefer et al. 2016; Parasuraman and Riley 1997; Yamani et al. 2020). Human–automation trust refers to "an attitude that an agent will help achieve an individual's goals in a situation characterized by uncertainty and vulnerability" (Lee and See 2004, pp. 51). Though human–automation trust has been linked with operators' strategies for using automation (Chancey et al. 2017; Karpinsky et al. 2018), the psychological mechanism that underlies human–automation trust is unknown. Established upon previous frameworks of interpersonal trust (Rempel et al. 1985; Barber 1983) and Muir's work (Muir 1987, 1994; Muir and Moray 1996), human–automation trust has been theorized to arise from three separate informational sources, including performance (i.e., what the automation is doing), process (i.e., how the automation works), and purpose (i.e., the designer's intent for developing the automation; Lee and Moray 1992). Empirical evidence suggests that the three dimensions capture different trajectories of human-automation trust (Chancey et al. 2017; Karpinsky et al. 2018; Long et al. 2020; Sato et al. 2020). For example, in a study that manipulated perceived risk and task load, only performance-based but not process- or purpose-based trust was reliably modulated when the participants interacted with a novel alerted-monitor system for the first time (Sato et al. 2020).

Several researchers have examined the influence of task load on human–automation trust in multitasking environments (Bailey and Scerbo 2007; Karpinsky et al. 2018; Sato et al. 2020). For example, Karpinsky et al. (2018) examined the effect of task load on human–automation trust in a low-fidelity simulator. They asked undergraduate participants to concurrently perform a tracking task manually and a system monitoring task assisted by an imperfect signaling system with 70% reliability in the Multi-Attribute Task Battery (MATB-II; Santiago-Espada et al. 2011). Task load was manipulated via difficulty levels of the tracking task. Results indicated that the participants rated lower performance- and process-based trust toward the signaling system under the high task load condition than the low task load condition. Karpinsky et al.'s (2018) analysis of trust revealed that the participants' ratings reflected their perception of the automation's behavior and mechanism. A more recent study demonstrated that this adverse effect of workload on performance-based trust only arises when operators perceive high risk (Sato et al. 2020). Additionally, analysis of eye movement showed less fixation on a task monitored by the signaling system (i.e., system monitoring task) under the high task load condition than the low task load condition, suggesting that attention allocation is a critical factor that influences human–automation trust in a dynamic multitasking environment assisted with imperfect automation. Specifically, their results are consistent with the view that human–automation trust depends on the extent to which attentional resources are allocated to scan the behavior of the automation. Yet, the causal relationships among attention allocation and trust remain largely unknown.

What factors potentially influenced the participants' trust and their scanning strategies in high task load conditions? A possible account for lower trust rating and fewer fixations on the automated task in the high task load condition is that operators placed a higher priority on the tracking task. The high task load condition demanded more frequent manual input with more force than the low task load condition in the tracking task, which could have encouraged the participants to attend to the tracking task more. This change in task priority might have caused participants to reduce their sampling of behaviors of the signaling system. Consequently, participants will not have enough information to assess the capability of the signaling system, lowering trust toward the signaling system. Task priority is conceptualized as the value of a task (Gutzwiller et al. 2014; Gutzwiller and Stizman 2017; Wickens et al. 2016). Freed (2002) suggested that task priority is influenced by various information sources including urgency, importance, task duration, and interruption cost. Several previous studies have attempted to directly examine the effect of task priority in multitasking environments and provided mixed results (Gilbert and Wickens 2017; Gopher et al. 1982; Gutzwiller et al. 2014; Gutzwiller and Sitzman 2017; Wickens et al. 2016). For example, Gopher et al. (1982) successfully manipulated task priority by providing continuous feedback on the participant's tracking performance and instructing them to prioritize the tracking task at a certain level (i.e., 30%, 50%, and 70%). Additionally, the researchers presented a desired performance line which denotes the target performance level of the tracking task and serves as an index of the participant's tracking performance. However, more recent works demonstrated a minimal effect of task priority (Gilbert and Wickens 2017; Gutzwiller et al. 2014; Gutzwiller and Sitzman 2017; Wickens et al. 2016). In these recent studies, participants were verbally instructed to prioritize the tracking task or prioritize all tasks equally without specifying the target performance level, which could be responsible for the lack of the reliable effect of task priority manipulation. Gutzwiller and Sitzman (2017) suggested that the lack of effect is not due to the task load, but it was due to the participants sequentially performing the task. Additionally, Gilbert and Wickens (2017) suggested that the magnitude of the effect depends on the participant's evaluation of the task's priority. Based on Yamani and Horrey's (2018) theoretical model of human–automation interaction, we speculated that participants were not able to update their resource allocation policy, since they could not evaluate their performance without objective target performance level.

The present study examined the effect of task priority on attention allocation and trust toward an imperfect signaling system in a simulated environment using the MATB-II. The participants performed three concurrent tasks (i.e., tracking task, system monitoring task, and resource management task) in the MATB environment. We measured attention allocation via eye movements, specifically percent dwell time (PDT) on area of interests (AOIs) within the displays for the three tasks. Previous research used PDT as a measure of attentional resources in applied settings (Horrey et al. 2006; Schriver et al. 2017; Wickens et al. 2003) which is applicable to the current experiment. We used the trust questionnaires developed in Jian et al. (2000) and in Chancey et al. (2017) to measure trust in automation. In brief, these two questionnaires have been implicated to measure two separate constructs as the one generated by Jian et al. (2000) was determined empirically while the other generated by Chancey et al. (2017) was theoretically grounded based on Lee and See's (2004) triadic model of automation trust. In addition to task load as in Karpinsky et al. (2018), task priority was manipulated based on Gopher et al.'s (1982) work. In Gopher et al.'s (1982) study, participants in the tracking priority condition were encouraged to improve their tracking performance by 20 percent more over their own baseline performance level, while those in the equal priority condition were encouraged to perform the tracking task 20 percent less than the baseline. The present study goes beyond prior works (Gilbert and Wickens 2017; Gutzwiller et al. 2014; Gutzwiller and Sitzman 2017; Wickens et al. 2016) by demonstrating an effect of task priority using Gopher et al.'s (1982) method. We predicted that participants would display lower trust and fixate less frequently toward an imperfect signaling system in the high task load condition than the low task load condition, as observed in the previous studies (Karpinsky et al. 2018; Sato et al. 2020). Furthermore, we predicted that the effect of task load on attention allocation and trust would be diminished when participants equally prioritized the tracking task in high task load condition. Specifically, participants in high task load condition would present similar trust ratings and eye movements to previous studies (Karpinsky et al. 2018; Sato et al. 2020) when the tracking task is prioritized. However, trust ratings and eye movements would be comparable between the task load conditions when participants equally prioritized all the tasks.

2 Methods

2.1 Participants

Forty participants (31 females and 9 males; $M = 21.05$ years, $SD = 6.25$) were recruited from Old Dominion University (ODU). All participants had a normal or corrected-to-normal vision and normal color perception. Participants were compensated with research credits for their participation. This research complied with the American Psychological Association Code of Ethics and was approved by the College of Sciences Institutional Review Board at ODU. Informed consent was obtained from each participant.

2.2 Apparatus

A Samsung T24C550 23.6″ LED monitor (1920 × 1080) with a frame rate of 75 Hz was used for the study. The monitor was placed 80 cm away from the chin rest. MATB-II (Santiago-Espada et al. 2011) was run on Windows 7 (Dell OptiPlex 9020). EyeLink II (SR Research, Mississauga, Ontario, Canada) was used to record the participant's eye movement with a sampling rate of 250 Hz. The experiment took place in a quiet room with dimmed light.

2.3 MATB-II tasks

MATB-II (Santiago-Espada et al. 2011) is a software developed by NASA Langley Research Center, Hampton, VA, designed to assess human performance in a simulated environment that hosts flight-related tasks. Participants in the present study performed the tracking task, system monitoring task, and resource management task. Figure 1 presents a sample display of the MATB-II task.

2.3.1 Tracking task

In the compensatory tracking task, participants controlled the joystick to keep the moving circular target within the dotted square. The circular target depicts the direction in which the aircraft moves, while the dotted square reflects the designated route. In the experimental session, the circular target deviated from the dotted square by setting the frequency of the force function to either 0.12 Hz or 0.06 Hz (i.e., high or low task load condition, respectively). In the practice session, the frequency of the force function was set to 0.09 Hz. The program computed the root mean squared error (RMSE) by sampling the participant's input in the XY dimension at 20 Hz. The average RMSE was computed for each block to assess tracking performance.

2.3.2 System monitoring task

Participants monitored the four vertical gauges and corrected the vertical fluctuating pointer at the lower or upper extremity. The four vertical gauges represent the temperature and pressure of the aircraft's two engines. The rectangular box (i.e., signaling system) above the gauges presents the engine's state (i.e., normal or warning). The

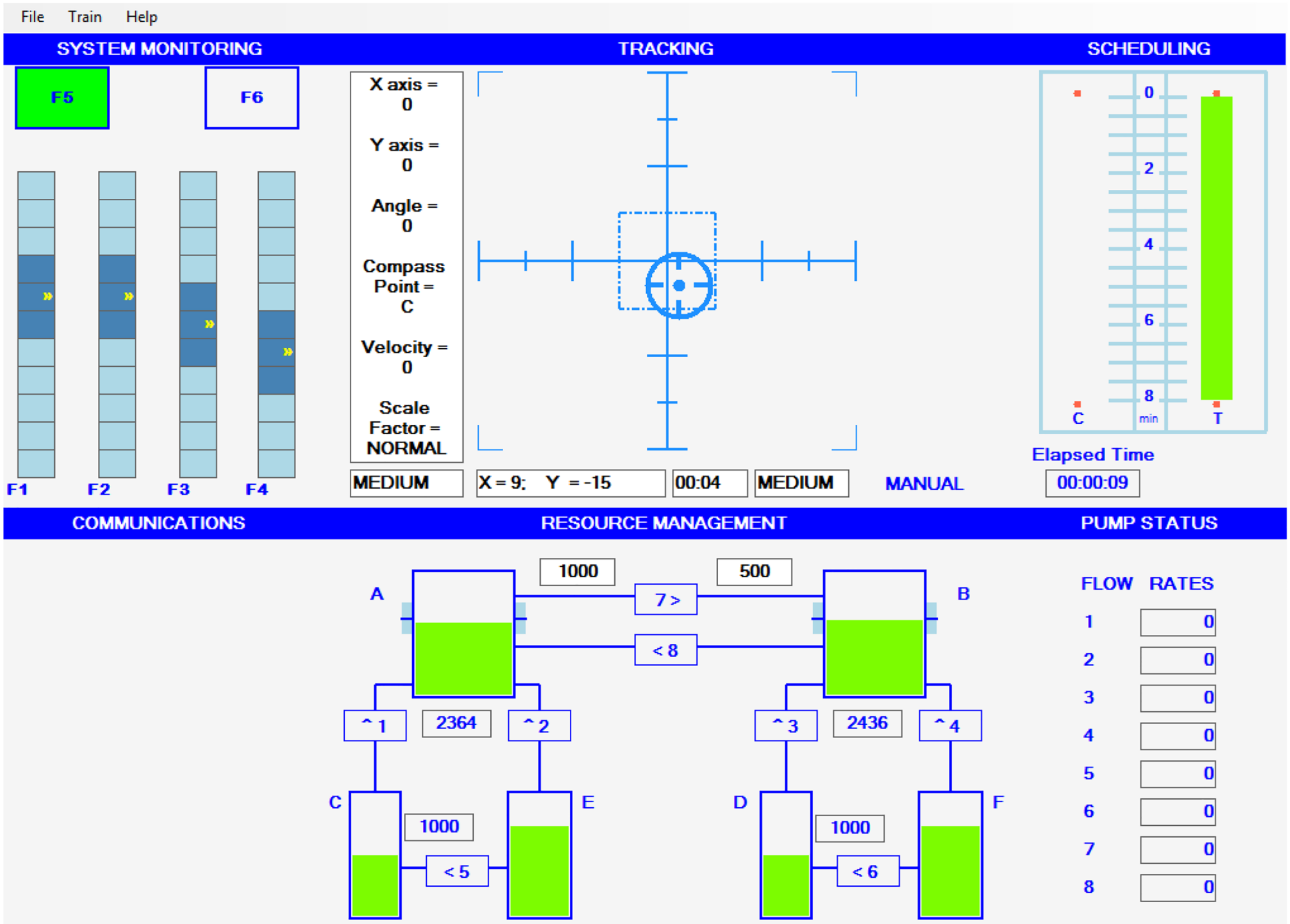

Fig. 1 Sample display of MATB-II. System monitoring (top left), tracking task (top center), and resource management task (bottom center)

engine is in "normal" state when a green rectangular box is illuminated. A green rectangular box illuminates when the vertical pointer fluctuates between the center of the vertical gauge. The engine is in a "warning" state when a red rectangular box is illuminated. A red rectangular box illuminates when the vertical pointer hits the extremity of the vertical gauge. For each block, 28 hit events and 12 false alarm (FA) events occurred randomly (70% reliability). The present study excluded miss events, because task performance and trust did not differ between miss and FA event in the Karpinsky et al. (2018) study. During a hit event, the vertical fluctuating pointer hits either extremity of the vertical gauge turning the green rectangular box off and illuminating the red rectangular box. In this case, participants were asked to respond to the signaling system and correct the vertical fluctuating pointer using a mouse. Specifically, participants clicked the red rectangular box, green rectangular box, and the corresponding gauge labeled F1–F4. During a FA event, the green rectangular box turns off, and the red rectangular box illuminates even though the vertical fluctuating pointer did not hit either extremity of the vertical gauge. In this case, participants were asked to respond only to the signaling system.

2.3.3 Resource management task

Participants maintained fuel in Tank A and Tank B, located next to letters A and B, respectively. The depletion rate of Tank A was set to 1,000 units per minute, while Tank B was set to 500 units per minute. When the tank's volume is below 2,500 units, participants transferred fuel from lower supply tanks, located next to letter C–F. Fuel can be transferred using a mouse to click the corresponding pumps, labeled with numbers from 1 to 8. The flow rate of the pump was set to 900 units per minute. Each block included eight pump failure events where a pump deactivates for 10 s. The pump presents three different states represented by colors. A green pump indicates that the pump is activating. A white pump indicates that the pump is deactivated but can be activated anytime. A red pump indicates a pump failure event.

2.4 Design

The present study employed a 2 × 2 mixed design with Task Priority (equal vs. tracking) as a between-subjects factor and Task Load (low vs. high) as a within-subjects factor. Dependent variables were subjective workload, trust, attention allocation, tracking performance, system monitoring performance, and resource management performance.

2.5 Dependent variables

2.5.1 Subjective workload

A modified version of NASA-TLX (Hart and Staveland 1988) was administered to measure subjective workload without pair-wise comparison (Hart 2006). The questionnaire consisted of 6 items, each representing 6 subscales (mental demand, physical demand, temporal demand, performance, effort, and frustration), on a 21-point gradient scale ranging from very low to very high (minimum score = 6, maximum score = 126).

2.5.2 Trust

Chancey et al.'s (2017) and Jian et al.'s (2000) trust questionnaires were administered to measure human–automation trust (see Appendix A and B). Chancey et al.'s (2017) trust questionnaire included 13 items on a 12-point Likert scale ranging from (1) not descriptive to (12) very descriptive (minimum score = 13, maximum score = 156). The items were categorized into one of three subscales (i.e., performance, process, and purpose). Jian et al.'s (2000) trust questionnaire included 12 items on a 7-point Likert scale ranging from (1) not at all to (7) extremely (minimum score = 12, maximum score = 84).

2.5.3 Attention allocation

For each area of interest (AOI), percentage dwell time (PDT) were computed by calculating the proportion of time that the participants fixated on an AOI. AOI is defined as the areas within which the participants' fixations was analyzed to examine PDT. The AOI was defined for each of the tracking, system monitoring and resource management displays.

2.5.4 MATB-II performance

The mean RMSE for each block measured tracking performance. System monitoring performance was measured by the mean error rate and response time (RT) for their first response in each block separately for hit and FA events. Error rates are the proportion of events that participants executed incorrectly. RT is the time interval between the onset of an event and the participant's initial response. Resource management performance was assessed by the mean volumes for Tank A and Tank B.

2.6 Procedures

Participants completed an informed consent and demographics form. Then, participants were screened for color perception and visual acuity using the Ishihara color blindness test and the Snellen chart. Participants were randomly assigned to either the tracking or equal priority condition. Following Gopher et al.'s (1982) procedure, participants in the equal priority condition were asked to prioritize the tracking task at a priority level of 30%. That is, participants performed the tracking task at a level better than the lowest 30% of their own baseline level performance. Alternatively, participants in the tracking priority condition were asked to prioritize the tracking task at a priority level of 70%. In the practice session, participants performed the MATB-II task separately for a total of 9 min (part-task training) and simultaneously for 3 min (whole-task training). Upon completion of the practice session, participants received their average RMSE reflecting their baseline performance of the tracking task during the whole-task training and a target value unique to each participant based on their own baseline performance. In the equal priority condition, target value was computed by adding one standard deviation to average RMSE. In the tracking priority condition, target value was computed by subtracting one standard deviation from the average RMSE. Participants were instructed to aim for the target value during the experimental session. In the experimental session, participants completed two 20-min blocks, which differed in the difficulty of the tracking task. The two blocks were counterbalanced to reduce order effects. After each block, participants completed two human–automation trust questionnaires (Chancey et al. 2017; Jian et al. 2000) and the NASA-TLX (Hart and Staveland 1988). Participants were provided with research credit for the efforts.

2.7 Statistical analysis

Bayesian analyses were employed instead of null-hypothesis significance tests (NHSTs). Unlike the p value in NHSTs, Bayesian analyses provide evidence for or against the effect of interest. Specifically, in the default Bayesian framework (Rouder and Morey, 2012), Bayes factor, denoted as $B_{10,}$ represents a likelihood ratio between statistical evidence for a model including an effect of interest to that excluding the effect. Thus, its magnitude provides direct information about the strength of statistical evidence for or against the presence of an effect (Wetzels et al. 2011). Bayes factors were interpreted following Jeffrey's (1961) descriptive term.

3 Results

A 2 × 2 mixed Bayesian analysis of variance (ANOVA) was employed with Task Load (low vs. high) as a within-subject factor and Task Priority (equal vs. tracking) as a between-subject factor for each dependent variable. To examine whether the manipulation of Task Priority was successful, Bayesian paired-samples t tests were employed to compare the participant's RMSE for each block with the assigned target value computed from the participant's RMSE in the practice session. Two participants were removed from the analysis since their system monitoring performance was below the inclusion criteria (performance accuracy of 50%). Additionally, three participants were removed from the analysis due to technical issues with the eye tracker. One participant withdrew from the study, because the participant felt sick. Thus, a total of 34 participants (27 females and 7 males; $M = 21.06$ years, SD = 5.83) were included in the analysis.

3.1 Manipulation check

Equal priority condition. When participants were instructed to reduce tracking performance, tracking RMSE was close to the target value in the high task load condition [paired-samples $t(16) = -1.88$, $B_{10} = 1.04$, $d = 0.55$]. However, tracking RMSE was decisively lower than the target value in the low task load condition [paired-samples $t(16) = 6.78$, $B_{10} = 4.60 \times 10^3$, $d = 1.77$]. Figure 2 presents the average tracking RMSE and the target value for each condition.

Tracking priority condition. When participants were instructed to prioritize the tracking task over the other two tasks, tracking RMSE decisively exceeded the target value in the high task load condition [paired-samples $t(16) = -7.90$ $B_{10} = 2.62 \times 10^4$, $d = 2.04$], suggesting that participants' path deviated from the center more than required by the priority instruction. On the other hand, tracking RMSE was substantially below the target value in the low task load condition [paired-samples $t(16) = 3.18$ $B_{10} = 8.43$, $d = 0.66$].

3.2 Subjective workload

Participants' subjective workload was decisively higher for the high task load condition than the low task load condition, demonstrating successful manipulation of Task Load [$M = 77.00$ vs. 66.00 for the high and low task load condition, respectively; $F(1, 32) = 20.70$, $B_{10} = 376.00$, $\eta^2_G = 0.10$]. However, data gave no substantial evidence for the main effect of Task Priority [$F < 1$, $B_{10} = 1/2.80$] and the interaction effect [$F < 1$, $B_{10} = 1/2.19$].

3.3 Chancey et al.'s (2017) trust scale

The three subscales in Chancey et al.'s (2017) trust scale were analyzed separately. Figures 3, 4, and 5 present the mean trust ratings for performance-, process-, and purpose-based trust, respectively. Participants showed substantially lower performance-based trust in the high task load condition than the low task load condition [$M = 44.18$ vs.

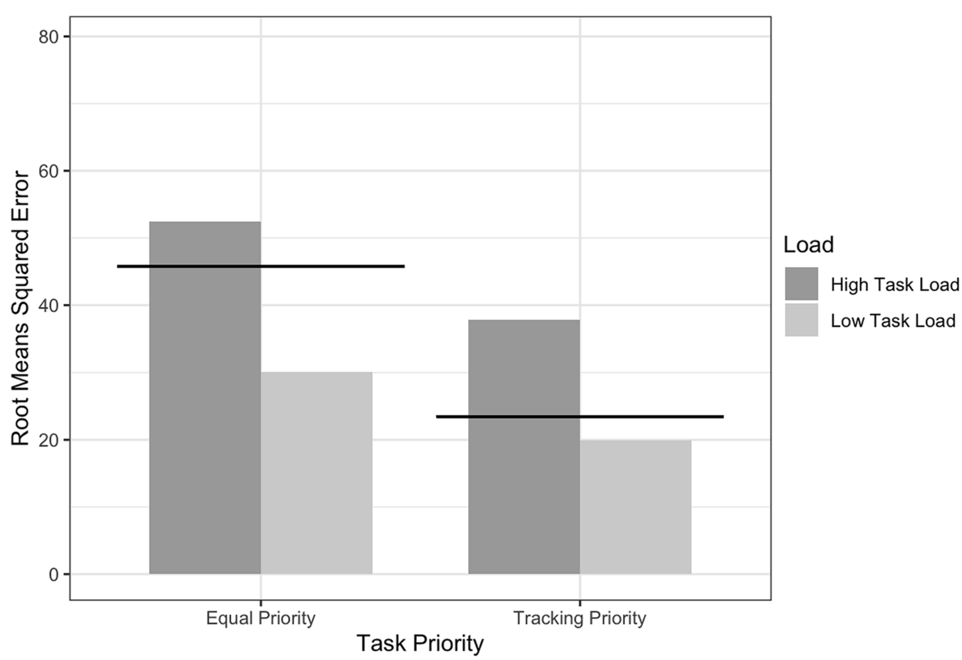

Fig. 2 Mean tracking RMSE as a function of the task priority conditions and task load. Horizontal bar represents the mean target value

41.62, respectively; $F(1, 32) = 6.50$, $B_{10} = 3.24$, $\eta^2_G = 0.01$], replicating the result of the previous study (Karpinsky et al. 2018). The main effect of Task Load was qualified by the two-way interaction [$F(1, 32) = 6.50$, $B_{10} = 3.46$, $\eta^2_G = 0.01$], indicating that the effect was stronger in the equal priority condition [$M = 40.41$ vs. 45.53; paired-samples $t(16) = -3.56$, $B_{10} = 16.64$, $d = 0.39$] than in the tracking priority condition [M = 42.82 vs. 42.82; paired-samples $t(16) = 0$, $B_{10} = 1/4.00$]. Data gave no evidence for the main effect of Task Priority [$F < 1$, $B_{10} = 1/1.67$]. Between Task Load conditions, data pattern for process-based trust ratings was similar to performance-based trust ratings [$F(1, 32) = 11.58$, $B_{10} = 22.86$, $\eta^2_G = 0.03$]. However, data indicated no substantial evidence for the main effect of Task Priority [$F < 1$, $B_{10} = 1/1.40$] and the interaction effect [$F < 1$, $B_{10} = 1/2.61$]. Finally, purpose-based trust did not substantially vary between the conditions [$1/1.59 < B_{10} < 1.62$].

3.4 Jian et al.'s (2000) trust scale

Data indicated no substantial evidence for any of the effects [$1/2.02 < B_{10} < 1/1.10$].

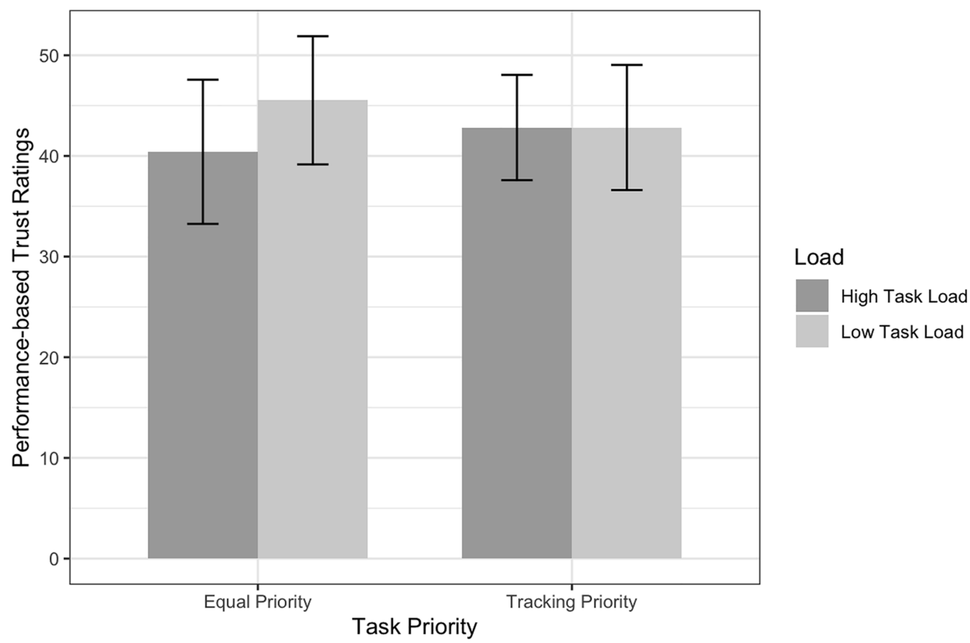

Fig. 3 Mean scores for performance-based trust as a function of the task priority conditions and task load. Error bars represent 95% confidence intervals

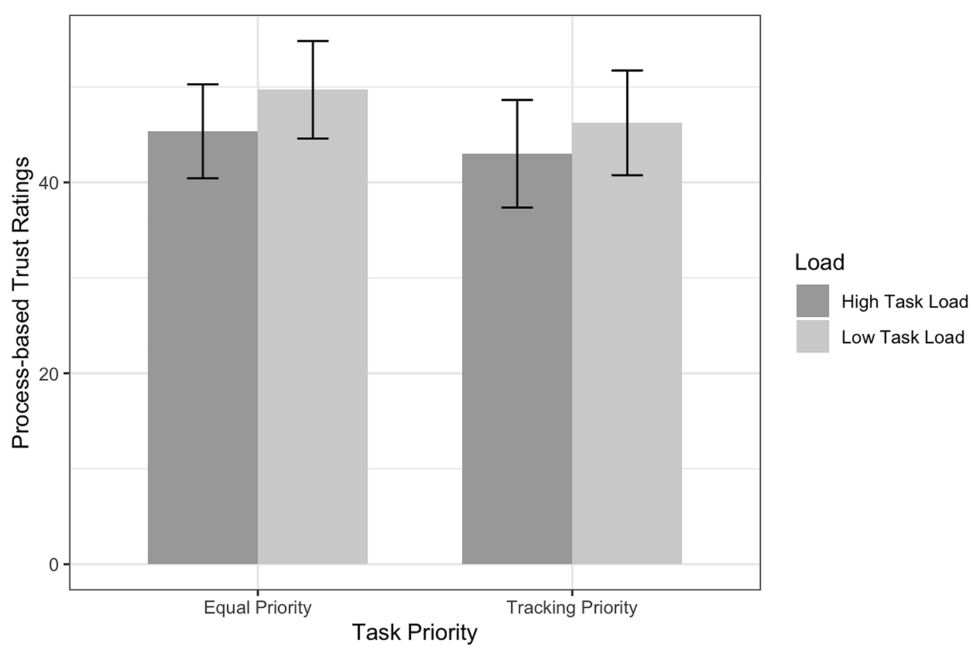

Fig. 4 Mean scores for process-based trust as a function of the task priority conditions and task load. Error bars represent 95% confidence intervals

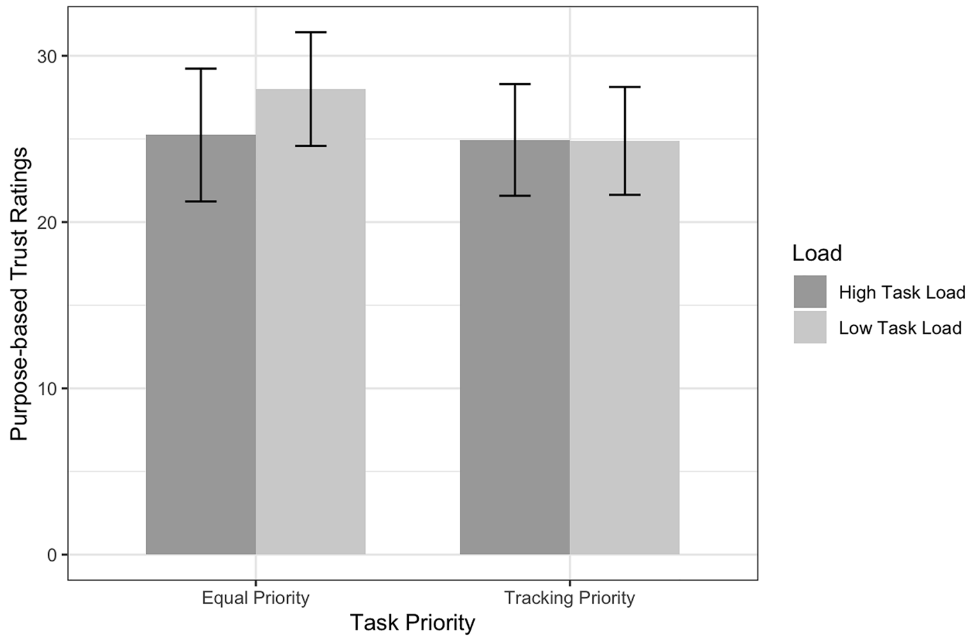

Fig. 5 Mean scores for purpose-based trust as a function of the task priority conditions and task load. Error bars represent 95% confidence intervals

3.5 Attention allocation

PDT on the tracking, system monitoring, and resource management tasks were analyzed separately. Data gave strong evidence that the participants made more frequent fixations on the tracking task when the tracking task required more frequent corrections [$M = 0.46$ vs. 0.32 for the high task load and low task load, respectively; $F(1, 32) = 45.16$, $B_{10} = 6.85 \times 10^4$, $\eta^2_G = 0.16$]. Furthermore, when the tracking task required more frequent input, participants spent less time fixating the system monitoring task [$M = 0.11$ vs. 0.13 for the high task load and low task load, respectively; $F(1, 32) = 11.05$, $B_{10} = 15.22$, $\eta^2_G = 0.04$] and the resource management task [$M = 0.36$ vs. 0.45 for the high task load condition and low task load condition, respectively; $F(1, 32) = 34.52$, $B_{10} = 7.30 \times 10^3$, $\eta^2_G = 0.09$].

When participants prioritized the tracking task, as expected, participants fixated on the tracking task more frequently [$M = 0.51$ vs. 0.27 for the tracking priority and the equal priority condition conditions, respectively; $F(1, 32) = 22.85$, $B_{10} = 466.52$, $\eta^2_G = 0.38$]. In turn, they fixated the system monitoring task less in the tracking priority condition than the equal priority condition [$M = 0.10$ vs. 0.14; $F(1, 32) = 7.94$, $B_{10} = 5.45$, $\eta^2_G = 0.1$], and the same pattern of attention allocation was obtained in the resource management task [$M = 0.31$ vs. 0.50 for the tracking priority and the equal priority conditions, respectively; $F(1, 32) = 15.89$, $B_{10} = 53.37$, $\eta^2_G = 0.31$]. Finally, data indicated no substantial evidence for the interaction effect [$1/1.42 < B_{10} < 1/2.96$]. Figures 6 and 7 present the PDT on the system monitoring display and tracking display, respectively.

3.6 Tracking performance

The RMSE was decisively greater in the high task load condition compared to the low task load condition [$M = 45.14$ vs. 24.99; $F(1, 32) = 230.11$, $B_{10} = 2.90 \times 10^{14}$, $\eta^2_G = 0.53$], suggesting that the cursor deviated more from the target when the tracking task required more frequent corrections. Additionally, RMSE was higher when participants were asked to reduce tracking performance [$M = 41.28$ vs. 28.84 for the equal priority and the tracking priority conditions, respectively; $F(1, 32) = 16.16$, $B_{10} = 71.06$, $\eta^2_G = 0.30$]. Data indicated no evidence for the interaction effect [$F(1, 32) = 2.88$, $B_{10} = 1/1.12$, $\eta^2_G = 0.01$].

3.7 System monitoring performance

3.7.1 RTs

There was strong evidence that participants in the low task load condition compared to the high task load condition responded faster to both FA [$M = 3.09$ vs. 3.62 s; $F(1, 32) = 16.23$, $B_{10} = 63.52$, $\eta^2_G = 0.06$] and hit events [$M = 2.83$ vs. 3.34 s; $F(1, 32) = 36.35$, $B_{10} = 8.97 \times 10^3$, $\eta^2_G = 0.11$]. Furthermore, when tracking task was underprioritized, participants decisively responded faster to FA [$M = 2.48$ vs. 4.24 s; $F(1, 32) = 23.02$, $B_{10} = 430.06$, $\eta^2_G = 0.39$] and hit events [$M = 2.42$ vs. 3.75 s; $F(1, 32) = 31.10$, $B_{10} = 3.04 \times 10^3$, $\eta^2_G = 0.46$], compared to the tracking priority condition. Data gave no evidence for the interaction effect in both events [$1.17 < B_{10} < 2.18$].

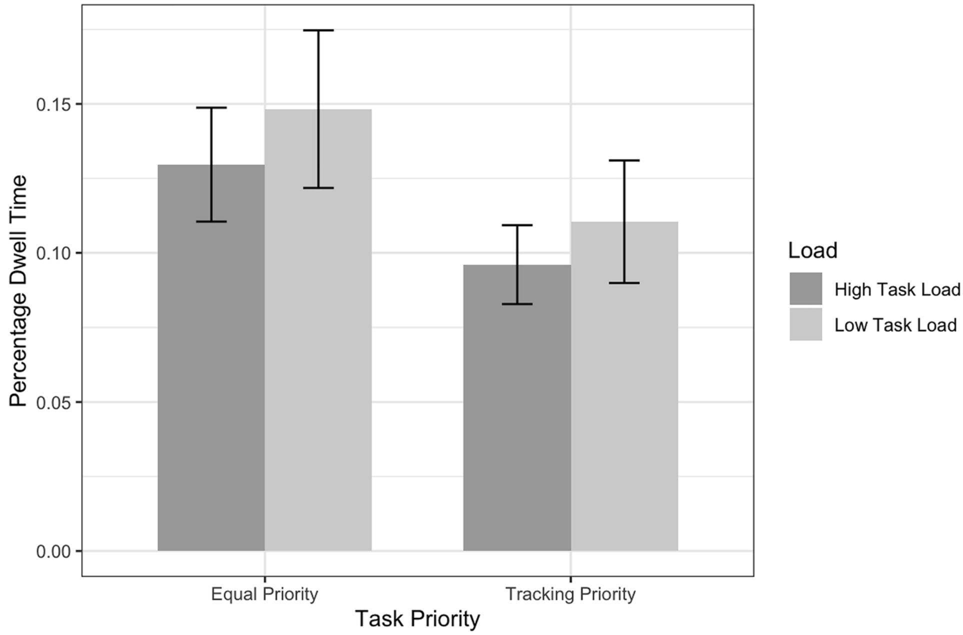

Fig. 6 Mean PDT on the system monitoring display as a function of the task priority conditions and task load. Error bars represent 95% confidence intervals

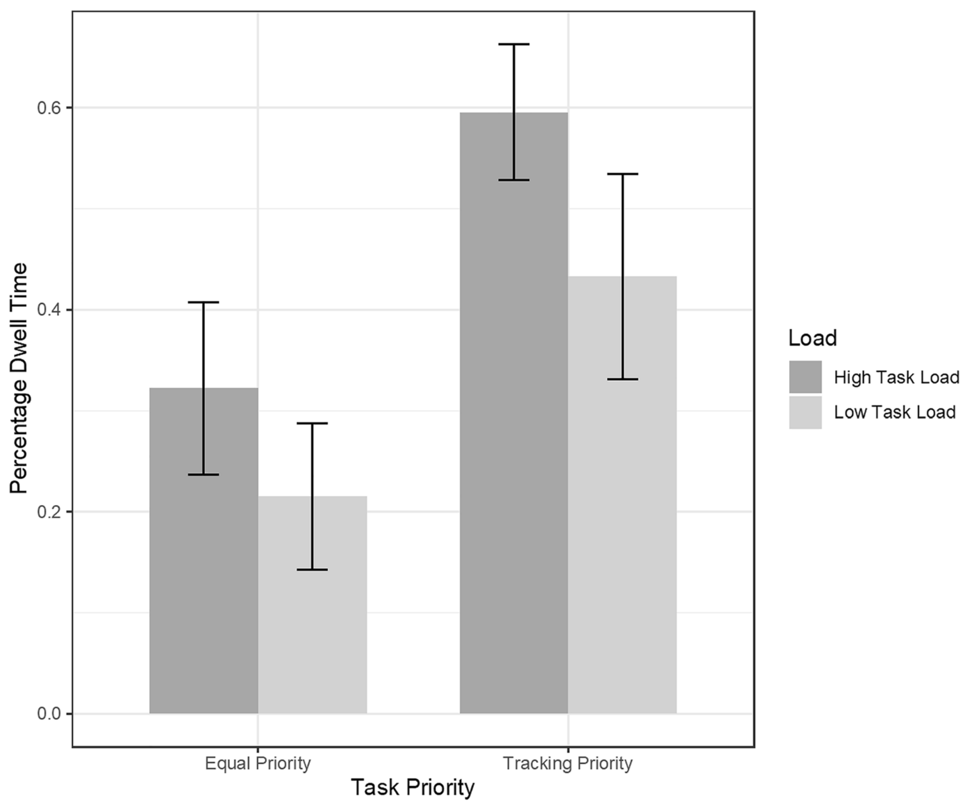

Fig. 7 Mean PDT on the tracking display as a function of the task priority conditions and task load. Error bars represent 95% confidence intervals

3.7.2 Error rate

For hit events, participants made substantially more errors in the high task load condition than the low task load condition [$M = 0.09$ vs. 0.06; $F(1, 32) = 6.87$, $B_{10} = 3.46$, $\eta^2_G = 0.03$] and in the tracking priority condition than the equal priority condition [$M = 0.11$ vs. 0.03; $F(1, 32) = 9.70$, $B_{10} = 7.40$, $\eta^2_G = 0.19$]. However, for FA events, data gave substantial evidence against the main effect of Task Load [$F < 1$, $B_{10} = 1/3.37$]. The remaining effects were not substantial [$1/1.05 < B_{10} < 2.18$].

3.8 Resource management performance

Data provided substantial evidence against the main effect of Task Load on the amount of fuel in Tank A [$B_{10}=1/4.05$] and Tank B [$B_{10}=1/3.34$], indicating that resource management performance did not differ between Task Load conditions, and the interaction effect for Tank A [$B_{10}=1/3.23$]. No remaining effects were substantial [$1/2.28<B_{10}<1.16$].

4 Discussion

Previous works demonstrated that participants exhibit lower trust toward imperfect automation on the system monitoring task while frequently fixating on the tracking task (Karpinsky et al. 2018; Sato et al. 2020). We speculate that participants indicate lower trust toward the automation, because higher priority is set to the tracking task to match their attentional demand of the tracking task, causing them to misperceive the signaling system's performance. The present study examined whether task priority modulates the effect of task load on attention allocation and trust toward imperfect automation. Participants concurrently performed both the tracking task and the resource management task manually while completing the system monitoring task with assistance from a 70% reliable signaling system. Following Gopher et al.'s (1982) procedure, participants prioritized the tracking task by aiming for the objective target value based on their baseline tracking performance, resulting in a successful manipulation of task priority.

The present study replicated the effects of task load as found in Karpinsky et al. (2018). In the high than low task load condition, participants reported higher subjective workload levels, spent less time scanning the system monitoring, and reported lower performance-based and process-based trust. Participants were asked to control the moving cursor with more frequent disturbances which required more frequent monitoring of the target for the tracking task, where participants experienced higher levels of workload (cf. Vanderhaegen et al. 2020). As noted above, operators develop performance-based trust from the current and historical behaviors of automation observable to them. Operators may develop process-based trust from the appropriateness of the algorithm and regulatory mechanisms of the automation's behaviors. Then, purpose-based trust refers to trust base on understanding of the intention of the automation designers. Additionally, when the tracking task required frequent input, the participant's PDT on the tracking task elevated while the participant's PDT on the system monitoring task decreased. This reciprocal relationship on PDT illustrates tradeoffs between the two tasks as the attentional demand of the tracking task varied. Together, these findings suggest that participants distributed more attentional resources to the tracking task to cope with a greater task demand, while reducing sampling of the signaling system's behavior. Consequently, less information about the system's behavior could have induced misperception of the system's reliability and lowered performance-based and process-based trust. Within Lee and See's (2004) theoretical framework, reduced trust toward the signaling system could be attributed to a mismatch between the participant's perception of the signaling system's behavior and the actual capability, and its related regulatory mechanisms, of the signaling system. We observed differences in the participants' system monitoring performance between task load conditions, which were not observed in Karpinsky et al.'s (2018) study. These differences could be attributed to the presence of the resource management task unlike in Karpinsky et al. (2018), presumably degrading system monitoring performance due to the added attention demand.

Using Gopher et al.'s (1982) technique, we provided a specific target value for the tracking task based on the participants' own baseline performance. Participants were asked to prioritize the tracking task at a level of 30% in the equal priority condition (worse performance than their baseline), while 70% in the tracking priority condition (better performance than their baseline). The current results show that task priority can modulate the effect of task load on automation trust. Contrary to our expectation, participants in the high task load condition reported lower performance-based trust in the equal priority condition, but prioritizing the tracking task over the other tasks in the tracking priority condition eliminated the effect of task load on automation trust. One possible explanation is that task priority influenced the mobility of the attention (Yamani and Horrey 2018). Setting higher priority to perform the tracking task modulated the effect of task load on the attentional resource capacity, blocking mobilization of attentional resources to the signaling system (Young and Stanton 2002). More limited resources allocated to the signaling system then could have degraded information-processing critical for the development of performance- and process-based trust. That is, it is possible that the participants did not possess sufficient attentional resources to allocate to accurately observe and monitor behaviors of the automation and consider the regulating algorithms and characteristics (e.g., reliability). Lastly, none of these effects were observed on trust scores using Jian et al.'s (2000) questionnaire. This discrepancy between Jian et al. (2000) and Chancey et al. (2017) questionnaires may represent the fact that Jian et al.'s questionnaire is empirically developed while Chancey et al.'s questionnaire is theoretically driven. On one hand, Jian et al. (2000) questionnaire is based on no pervasive theory but instead the results of a three-phased study involving elicitation and comparison of words related to trust and distrust. Chancey et al. (2017) questionnaire, on the other, adapted a trust questionnaire on trust during human–computer interaction developed by Madsen and Gregor (2000) to map onto Lee and See's (2004) human-automation trust. Our recent preliminary work (Long et al. 2020; Yamani et al. in preparation) using multi-level confirmatory factor analysis demonstrated that these two questionnaires measure two separate constructs, supporting this interpretation.

There exist at least four caveats when interpreting the data. First, the present study recruited undergraduate students who were not familiar with the MATB task. It calls for future research whether the current results generalize to trained experts with a better mental model of the task and the automation compared to novices. Second, the present study did not manipulate perceived risk even though risk is a critical factor influencing automation trust (Chancey et al. 2017; Sato et al. 2020). Elevated levels of perceived risk may further increase the effects of task load and task priority, especially in real-world flight environments. Third, due to the technical constraints, we were not able to provide moment-to-moment feedback on their tracking performance as in Gopher et al.'s (1982). The effect of task priority might be stronger than found in the current study, which requires additional research. Lastly, a more rigorous and advanced analysis of eye movement and trust data may reveal the underlying processes responsible for the current findings. The application of gaze transition entropy analysis (Krejtz et al. 2015) would allow quantifying randomness of gaze distribution and characterize complex transitions among multiple AOIs. These advanced analytic techniques may reveal specific eye movement processes critical for the development of automation trust.

In conclusion, the current study directly examined the effect of task priority on eye movements and trust toward an imperfect signaling system in a simulated dynamic multitasking environment. The results not only replicate the adverse effect of tracking task demand on trust but also demonstrate that this effect was eliminated when participants were instructed to prioritize the tracking task over the other two concurrent tasks. The results imply that operators may allocate their attentional resources to different tasks based on their perceived task demand. Additionally, verbal instruction to prioritize the tracking task can override resource allocation strategies, impacting their trust toward the signaling system. Practically, task priority should be considered when developing training programs involving human-automation interaction and trust in a multitasking environment. Implementing task priority in training programs can potentially control an operator's trust to prevent disuse or misuse of automation (Parasuraman and Riley 1997).

There are two important avenues for future research. First, future research should consider examining the interaction between attentional allocation and working memory resource consumption (cf. Baddeley and Hitch, 1974). According to Wickens (2002), signal processing can demand the same or different resources, impacting operator performance. For instance, if visual spatial working memory resources are consumed by the primary tracking task signaling fuel pump status with a verbal representation might facilitate dual-task monitoring as it is drawing from an orthogonal pool of resources (Iani and Wickens 2007; Wickens 2002). Research examining the interaction between attention allocation and working memory resource consumption might lead to effective design recommendations. Second, future research should focus on the effects of task priority on trust toward and interactions with multiple automated systems. Advanced Air Mobility (AAM) is an emerging technology that allows transportation of people and goods in urban and rural areas via fully automated aerial vehicles (National Academics of Sciences, Engineering, and Medicine 2020; Chancey et al. 2021). Human operators responsible for AAM operations will likely need to monitor and, when necessary, manually intervene multiple automated aerial vehicles in the AAM platform. This new transportation technology in an integrated National Airspace System will impose numerous research questions including development of training programs involving task priority as a key element for AAM operators and the mechanism of their trust development toward a set of automated aerial vehicles.

5 Conclusion

The current results replicate the previous finding that operators exhibit less trust toward imperfect automation assisting with the secondary task when the primary task demands more attention (e.g., Karpinsky et al. 2018). Additionally, the results indicate that this attenuation effect is reduced or eliminated by explicitly prioritizing the primary task than the secondary task assisted by automation. That is, when their attention was more constrained due to increased task priority, their trust ratings did not vary with different levels of task load even though their workload increased. In practice, automation designers should consider task priority and operators' attention distribution when designing training programs for appropriate human–automation trust in dynamic multitasking workspace.

Appendix A

Scale items from the Chancey et al. (2017) trust questionnaire. (The numbers indicate the order that the items were presented to the participants when administered).

Performance

2. For me to perform well, I can rely on the automated aid to function.

4. The automated aid's advice reliably helps me perform well.

5. The automated aid's advice consistently helps me perform well.

12. The automated aid always provides the advice I require to help me perform well.

13. The automated aid adequately analyzes the system consistently, to help me perform well.

Process

3. It is easy to follow what the automated aid does to help me perform well.

6. I understand how the automated aid will help me perform well.

8. Although I may not know exactly how the automated aid works, I know how to use it to perform well.

10. To help me perform well, I recognize what I should do to get the advice I need from the automated aid the next time I use it.

11. I will be able to perform well the next time I use the automated aid because I understand how it behaves.

Purpose

1. Even when the automated aid gives me unusual advice, I am certain that the aid's advice will help me to perform well.

7. Even if I have no reason to expect that the automated aid will function properly, I still feel certain that it will help me to perform well.

9. To help me perform well, I believe advice from the automated aid even when I don't know for certain that it is correct.

Appendix B

Scale items from the Jian et al. (2000) trust questionnaire.

1. The system is deceptive.
2. The system behaves in an underhanded manner.
3. I am suspicious of the system's intent, action, or outputs.
4. I am wary of the system.
5. The system's actions will have a harmful or injurious outcome.
6. I am confident in the system.
7. The system provides security.
8. The system has integrity.
9. The system is dependable.
10. The system is reliable.
11. I can trust the system.
12. I am familiar with the system.

Author contributions TS and YY developed the experimental design and the experimental protocol. TS performed the data analysis and wrote the original manuscript. All authors reviewed and edited the manuscript.

Declarations

Competing interests The authors declare no competing interests.

References

Baddeley AD, Hitch G (1974) Working memory. Psychol Learn Motiv 8:47–89

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

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

Barber B (1983) The logic and limits of trust. Rutgers University Press, New Brunswick

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

Breznitz S (1984) Cry wolf: the psychology of false alarms. Erlbaum, Hillsdale

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 57:947–958

Chancey ET, Politowicz, MS, Le Vie L (2021). Enabling advanced air mobility operations through appropriate trust in human-autonomy teaming: foundational research approaches and applications. In: AIAA Scitech 2021 Forum, p 0880

Comstock JR, Arnegard RJ (1992) The multi-attribute task battery for human operator workload and strategic behavior research (NASA Tech. Memorandum 104174). NASA Langley Research Center, Hampton

Dixon SR, Wickens CD (2006) Automation reliability in unmanned aerial vehicle control: a reliance-compliance model of automation dependence in high workload. Hum Factors 48:474–486

Freed M (2000) Reactive prioritization. In: Proceedings of the international workshop on planning and scheduling in space, San Francisco, 2000

Getty DJ, Swets JA, Pickett RM, Gonthier D (1995) System operator response to warnings of danger: a laboratory investigation of the effects of the predictive value of a warning on human response time. J Exp Psychol Appl 1:19–33

Gilbert KM, Wickens CD (2017) Experimental evaluation of STOM in a business setting. In: Proceedings of the human factors and ergonomics society annual meeting, vol 61. SAGE publications, Los Angeles, p 767–771

Gopher D, Brickner M, Navon D (1982) Different difficulty manipulations interact differently with task emphasis: evidence for multiple resources. J Exp Psychol Hum Percept Perform 8:146–157

Gutzwiller RS, Wickens CD, Clegg BA (2014) Workload overload modeling: an experiment with MATB II to inform a computational model of task management. In: Proceedings of the human factors and ergonomics society annual meeting, vol 58. SAGE publications, Los Angeles, p 849–853

Gutzwiller RS, Sitzman DM (2017) Examining task priority effects in multi-task management. In: Proceedings of the human factors and ergonomics society annual meeting, vol 61. SAGE publications, Los Angeles, p 762–766

Hart SG (2006) NASA-task load index (NASA-TLX); 20 years later. In: Proceedings of the human factors and ergonomics society annual meeting, vol 50. SAGE publications, Los Angeles, p 904–908
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 57:407–434
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
Iani C, Wickens CD (2007) Factors affecting task management in aviation. Hum Factors 49:16–24
Jeffreys H (1961) Theory of probability, 3rd edn. University Press, Oxford
Jian J, Bisantz AM, Drury CG (2000) Foundations for an empirically determined scale of trust in automated systems. Int J Cogn Ergon 4:53–71
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
Lee JD, Moray N (1992) Trust, control strategies and allocation of function in human machine systems. Ergonomics 35:1243–1270
Lee JD, See KA (2004) Trust in automation: designing for appropriate reliance. Hum Factors 46:50–80
Li H, Wickens CD, Sarter N, Sebok A (2014) Stages and levels of automation in support of space teleoperations. Hum Factors 56:1050–1061
Loft S, Chapman M, Smith RE (2016) Reducing prospective memory error and costs in simulated air traffic control: external aids, extending practice, and removing perceived memory requirements. J Exp Psychol Appl 22:272–284
Long S, Sato T, Millner N, Mirabelli J, Loranger R, Yamani Y (2020) Empirically and theoretically driven scales on automation trust: a multi-level confirmatory factor analysis. In: Proceedings of the human factors and ergonomics society annual meeting, vol 64. SAGE publications, Los Angeles, p 1829–1832
Lyons JB, Stokes CK (2012) Human–human reliance in the context of automation. Hum Factors 54:112–121
Mackworth NH (1948) The breakdown of vigilance during prolonged visual search. Quart J Exp Psychol 1:6–21
Molloy R, Parasuraman R (1996) Monitoring an automated system for a single failure: vigilance and task complexity effects. Hum Factors 38:311–322
Muir BM (1987) Trust between humans and machines, and the design of decision aids. Int J Man Mach Stud 27:527–539
Muir BM (1994) Trust in automation: part I. Theoretical issues in the study of trust and human intervention in automated systems. Ergonomics 37:1905–1922
Muir BM, Moray N (1996) Trust in automation. Part II. Experimental studies of trust and human intervention in a process control simulation. Ergonomics 39:429–460
Parasuraman R, Riley V (1997) Humans and automation: use, misuse, disuse, abuse. Hum Factors 39:230–253
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
Rempel JK, Holmes JG, Zanna MP (1985) Trust in close relationships. J Pers Soc Psychol 49:95–112
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
Sato T, Yamani Y, Liechty M, Chancey ET (2020) Automation trust increases under high-workload multitasking scenarios involving risk. Cogn Technol Work 22:399–407
Schaefer KE, Chen JYC, Szalma JL, Hancock PA (2016) A meta-analysis of factors influencing the development of trust in automation: implications for understanding autonomy in future systems. Hum Factors 58:377–400
Schriver AT, Morrow DG, Wickens CD, Talleur DA (2017) Expertise differences in attentional strategies related to pilot decision making. Decision making in aviation. Routledge, London, pp 371–386
Sorkin RD (1988) Why are people turning off our alarms? J Acoust Soc Am 84:1107–1108
Warm JS, Parasuraman R, Matthews G (2008) Vigilance requires hard mental work and is stressful. Hum Factors 50:433–441
Wetzels R, Matzke D, Lee MD, Rouder JN, Iverson GJ, Wagenmakers EJ (2011) Statistical evidence in experimental psychology: an empirical comparison using 855 *t* tests. Perspect Psychol Sci 6:291–298
Wickens CD (2002) Multiple resources and performance prediction. Theor Issues Ergon Sci 3:159–177
Wickens CD, Alexander AL (2009) Attentional tunneling and task management in synthetic vision displays. Int J Aviat Psychol 19:182–199
Wickens CD, Goh J, Helleburg J, Horrey WJ, Talleur DA (2003) Attentional models of multi-task pilot performance using advanced display technology. Hum Factors 45:360–380
Wickens CD, Hollands JG, Banbury S, Parasuraman R (2015) Engineering psychology and human performance. Psychology Press
Wickens CD, Gutzwiller RS, Vieane A, Clegg BA, Sebok A, Janes J (2016) Time sharing between robotics and process control: validating a model of attention switching. Hum Factors 58:322–343
Yamani Y, Horrey WJ (2018) A theoretical model of human-automation interaction grounded in resource allocation policy during automated driving. Int J Hum Factors Ergonom 5:225–239
Yamani Y, Long SK, Itoh M (2020) Human–automation trust to technologies for naïve users amidst and following the COVID-19 pandemic. Hum Factors 62:1087–1094
Young MS, Stanton NA (2002) Malleable attentional resources theory: a new explanation for the effects of mental underload on performance. Hum Factors 44:365–375
Vanderhaegen F, Wolff M, Mollard R (2020) Non-conscious errors in the control of dynamic events synchronized with heartbeats: a new challenge for human reliability study. Saf Sci 129:1–11