ORIGINAL ARTICLE

The effect of cognitive workload on decision authority assignment in human–robot collaboration

Michail Karakikes[1] · Dimitris Nathanael[2]

Abstract

Human–robot collaboration in dynamic industrial environments warrants robot flexibility and shifting between tasks. Adaptive robot behavior unavoidably carries decision-making needs regarding task allocation and scheduling. Such decisions can be made either by the human team members or autonomously, by the robot's controlling algorithm. Human authority may help preserve situational awareness but increases mental demands due to increased responsibilities. Conversely, granting authority to the robot can offload the operator, at the cost of reduced intervention readiness. This paper aims to investigate the question of decision authority assignment in a human–robot team, in terms of performance, perceived workload and subjective preference. We hypothesized that the answer is influenced by the cognitive workload imposed on the human operator by the work process. An experiment with 21 participants was conducted, in which decision authority and induced workload through a secondary task were varied between trials. Results confirmed that operators can support the robot better when decision authority is allocated according to their workload. However, operator decision authority (a) may cause inferior performance at any secondary tasks performed in parallel with robot supervision and (b) increases perceived workload. Subjective preference was found to be evenly divided between the two levels of decision authority, and unaffected by task difficulty. In brief, if human–robot team performance is a priority, humans should be granted decision authority when their overall workload allows it. In high-workload conditions, system decision-making algorithms should be developed. Nonetheless, process designers should be mindful of the interpersonal differences between operators who are destined to collaborate with robots.

Keywords Human–robot collaboration · Decision authority · Task allocation · Adaptive automation · Secondary task

1 Introduction

The deployment of robots in production and testing processes enables possibilities of human–robot collaboration (HRC) to achieve complex tasks with demanding requirements, foregoing the costs of developing dedicated, specialized machinery. However, care must be taken in the design of the process, so that human operators can stay safe, avoid errors, be productive and accept to work alongside robots. The fact that a task can be automated does not necessarily imply that it ought to be. Automation should be considered within the broader team context in which it is integrated.

The level of autonomy (LoA) describes the degree of independence in robotic action inference, and it can be defined for both physical and cognitive functions, such as information acquisition, analysis, and decision/action selection (Parasuraman et al. 2000). Many researchers suggest that the LoA ought to be dynamic, rather than predefined, to allow adaptation to the unpredictability of process-related factors, such as temporal demand variations, or the state of the operator (e.g. fatigue) (Sheridan and Parasuraman 2005). Adaptive automation unavoidably creates challenges regarding the shifting between autonomy levels, also termed "modes" of the robot, an important one being how is the change triggered. In other words, a "task manager" (system or human operator) of the process needs to decide whether to allocate a task to the automation or the operator and when to do so (Wickens et al. 2015). Human decision authority (HDA) assigns a leading, executive role to the operator and

✉ Dimitris Nathanael
dnathan@central.ntua.gr

Michail Karakikes
mkarakikes@gmail.com

1 Department of Mechanical and Process Engineering, ETH Zurich, Zurich, Switzerland

2 School of Mechanical Engineering, National Technical University of Athens, Athens, Greece

subordinate role to the robot regarding task allocation and scheduling, while robot decision authority (RDA) refers to the inverse.

Classic human factors research suggests that, in HRC scenarios, the human should be kept in the decision loop, to avoid phenomena of complacency, loss of skill and loss of situational awareness (Wickens et al. 2015). However, some empirical human–robot interaction studies indicate that human control and decision-making are not always desirable in terms of process efficiency and subjective satisfaction (Baraglia et al. 2016; Gombolay et al. 2015a, b; Scerbo 1996; Tausch and Kluge 2020). This can be partially explained by the increased workload that follows human decision authority and responsibility in supervising autonomous systems (Harris et al. 1993). This workload partially depends on the information exchange interface and the temporal demand imposed on the decision. As suggested by models of effort–performance relationship, there are levels of attentional effort that can cause attention overload or underload, each associated with different problems (Eberts and Salvendy 1986). Specifically, depending on work demands and operator competencies, attentional overload may be associated with operator stress, sensory overload and increased errors while attentional underload with sensory deprivation, inattentiveness (i.e. decreased vigilance) and boredom (Johannsen 1979). Furthermore, especially in cases of divided attention between competing tasks, attentional or cognitive tunneling may occur (Wickens 2005) where operators' perception is tunneled only to a salient part of the perceptual space thus leading to decreased situational awareness at the perception level (Endsley 1995). Such scenarios of cognitive dissonance during an operator's effort to control dynamic events can become non-conscious and result in erroneous behavior that remains undetected, posing safety risks to the process (Vanderhaegen et al. 2020).

Researchers have proposed to use human workload or performance monitoring to automatically adapt the level of autonomy (Scerbo 1996; Sheridan and Parasuraman 2005; Wickens et al. 2015). One of the findings is that system suggestions versus mandates increase mental workload because they need to be evaluated by the operator (Kaber and Riley 1999). However, system-invoked changes in LoA may impair the operator's situational awareness due to their unpredictability, raising questions about user acceptance (Sheridan and Parasuraman 2005). Ultimately, in designing the control of dynamic autonomy, there is a trade-off between unpredictability and cognitive workload for decision-making (Sheridan and Parasuraman 2005).

In operator-led adaptive automation, information evaluation and action selection are not the sole contributors to the operator's mental workload. The concept of "flow" in human psychology discusses the immersion of oneself into an activity (Csikszentmihalyi 1990). Work performed in a state of "flow" can be more productive and pleasant, and this is achievable, among others, by allowing focused, uninterrupted work to develop flexibly when needed, at a suitable combination of challenge and skill levels. Despite the effort to preserve "flow", events such as interruptions and task switching will inevitably occur in a multitasking context where the operator supervises a robot while also performing another task in proximity. Research has demonstrated that switching work tasks only when the previous task has been completed is advantageous to limit the division of attention between tasks, while a large time window for the decision to switch causes more cognitive effort to be carried over to the next task (Leroy 2009). Moreover, a perhaps counterintuitive outcome is that external interruptions from a work task could be less disruptive than self-initiated interruptions (at least when the former are appropriately timed) probably because of the reduced need for decision-making (Katidioti et al. 2016). Such findings, despite originating from different scientific fields, are quite relevant to HRC scenarios where the robot makes assistance requests to the operator who decides when to intervene.

To make robots more supportive in decision-making, research ought to explore how robots can respect and facilitate the—inherently unobservable—human state of "flow" at work, not only through properly timing interruptive assistance requests, but also by allowing the human to operate at a cognitively comfortable level of task hierarchy. As described by Sheridan and Stassen (1979), in man–machine work, there is a trade-off between the lower 'servo-level' versus the higher 'management level' of control over a task, the former referring to the executory, perceptual and mechanical subtasks and the latter referring to the executive, planning subtasks. While they individually contribute to the perceived mental workload of the operator, concurrent operation at both levels can be the major factor. Hence, in work design, it is important to consider the effect of the selected decision-making scheme in the disentanglement of these two cognitive levels and the induced frequency of interlevel shifting. For instance, HDA along with high physical task requirements may lead to an increased need for the operator to alternate between the executive and executory levels, resulting in excessive overload and diminished performance.

Typical approaches to quantifying mental effort include subjective ratings, the secondary task method, user modeling and physiological measures (Eberts and Salvendy 1986). In the secondary task method, the primary task of interest is performed simultaneously with a secondary task and performance is compared to a benchmark (i.e. the secondary task completed individually), to estimate mental reserve capacity from the primary task and thereby to quantify its associated workload (Ogden et al. 1979). Another function of the secondary task can be to induce additional workload, i.e. to stress or overload

the operator when the primary task is insufficiently loading. In subjective measures, participants explicitly state their perceived workload in questionnaires. A well-established tool is the NASA Task Load Index (NASA-TLX), which has been proven to be reliably sensitive and a good predictor of perceived workload (Hart 2006; Rubio et al. 2004). In the present study, a secondary task is used as a stressor and the NASA-TLX as a means for assessing perceived mental workload.

The present paper aims to investigate the effect of decision authority schemes in HRC tasks, on performance, overall perceived workload, and subjective preference criteria, at different levels of induced cognitive workload. For this purpose, an experimental HRC task was designed based on a prospective industrial scenario, where decision authority on task allocation and scheduling could be given either to the human or to the robot. It is hypothesized that, in low to moderate induced workload conditions, increased human control could be beneficial, resulting in better situational awareness and thus performance, whereas in high-induced workload conditions, it may be more advantageous to offload the human by shifting task allocation responsibility to the robot. In other words, human control is expected to promote vigilance and reduce attentional tunneling, which can be desirable in mentally undemanding scenarios, but it could lead to attentional resource saturation and erratic behavior in high-workload settings.

2 Methods

2.1 Case study description

An industrial case study was used as a platform for the investigation, namely the collaborative human–robot assembly of a medical autoinjector device in small batches, in the context of a testing laboratory. In such environments akin to a medical device laboratory, work is inherently flexible because of the unique procedures and protocols that follow the various daily tasks (e.g. assemblies, stress tests, function tests for different devices). Operators can benefit from parallelizing work using automation, but the low "production" volumes of a laboratory versus a production line impose the need for adaptability of said automation to render it usable and cost-effective. The automation should be capable of performing multiple tasks, switching between them easily and altering the way it performs them if needed (e.g. adding, removing or changing the sequence of steps). In the present case study, a Panda collaborative robot inserts a pre-filled syringe inside the autoinjector device, respecting a maximum force specification (see Fig. 1). Upon assembly, a considerable sample of devices undergoes function testing on a specialized test setup to ensure the quality of injection parameters. The robot performs the device assembly autonomously, while the ensuing function testing can be performed either by the robot or manually by the operator, in adaptation to the current state and goals of the human–robot team.

Fig. 1 Setup for autonomous assembly and testing of autoinjector devices. From left to right: **a** robotic arm, **b** transparent autoinjector device tray, **c** 2-finger pneumatic gripper for device stabilization during assembly, **d** white pre-filled syringe tray, **e** blue function testing platform

Device assembly and testing were defined as the primary task, and the experimental scenario was enriched with the addition of a secondary parallel task assigned exclusively to the operator. A Trail Making Test (TMT) was selected as a secondary task to substitute for any cognitive task required in the context of laboratory work which typically demands high attention for short periods of time during processing of a single item (in this case, one Trail), followed by more relaxed temporal demand between items (Sánchez-Cubillo et al. 2009). In practice, the operator simultaneously supervises the assembly and performs the unrelated secondary task, with potential short interruptions to manually test devices (see Fig. 2).

2.2 Participants

In total, twenty-one participants (*M* 29.1 years, SD 9.1 years, 8 women) completed the experiment, all professionally active in the pharmaceutical industry (11 scientists, 7 operators, 3 apprentices). They were divided into two groups (Group E: *N* = 10, *M* 30.2 years, SD 10.9 years, and Group D: *N* = 11, *M* 28.2 years, SD 7.7 years, 4 women), approximately balanced in age, gender, and occupation, each receiving a different level of secondary task difficulty ("easy" or "difficult").

2.3 Experimental design

A dual-task scenario was created for this study, in which the tasks to be completed in parallel were:

1. Assembly and testing of devices (primary task)
2. Computerized Trail Making Test (secondary task)

Device assembly was automated, performed exclusively by the robot, whereas device testing, which was needed only for a subsample of devices, could be performed by either the human operator or robotic agent. Moreover, the human operator was tasked with identifying and correcting machine errors that could randomly occur in the automatic part of the process.

The Trail Making Test was used as the secondary task (i.e. induced workload) to simulate any sensorimotor task that may be required in a laboratory, being a widely used tool for investigating cognitive performance, which depends on visuo-spatial working memory and executive function. The difficulty of the test has been proven to depend on the type of the test, total Trail length and visual interference (Gaudino et al. 1995). Multiple Trails were completed by each participant throughout the experiment.

A multifactorial experimental design was used, using two factors, each with two levels: (a) decision authority for subtask allocation within the primary task and (b) secondary task difficulty (i.e. induced workload level). Decision authority was a within-subject factor, with both levels completed by all participants in a counterbalanced order, whereas secondary task difficulty was a between-subject factor, where two participant groups were formed, each experiencing one of the two levels.

Variation of factor "decision authority" was realized by interchanging the authority between the human and the robot in the following decisions: (a) whether a device undergoing assembly will be tested, and if so (b) which of the two agents will physically perform the test. Under "human decision authority" (HDA), the human operator acquires an executive role to make the two aforementioned decisions during the assembly of each device, throughout the experiment and informs the robot through a virtual button interface, whereas under "robot decision authority" (RDA) the robot's

Fig. 2 Experiment execution by participant. From left to right: **a** participant on secondary task (laptop computer), **b** robot assembling a device

controlling algorithm makes these decisions and informs the human only in case a manual test is decided, and human action is needed. The actual timing of the potential human action was always selected by the human.

Under the HDA condition, the human decision of which device to test was based on a protocol given to participants, according to which every second assembled device ought to be tested (i.e. the second, fourth, and so forth). Adherence to this arbitrary protocol was used to simulate any supervisory task over the automation which requires situational awareness of the process' progress, as well as mental manipulation. The second decision of assigning the actual testing of devices to self or robot was used to introduce a need for temporal prediction and tactical planning considering the status of both the primary and secondary task. Choosing to conduct a test manually, would accelerate the assembly process because the robot would start assembling the next device in parallel, at the cost of time loss from the secondary task, to a degree depending on test timing selection. Under the RDA condition, the automation makes these decisions following a test protocol unknown to the participant, while automatic task allocation is programmed with the aim to keep assembly productivity above a threshold (in practice, the same number of manual test assignments was preserved across all participants to reduce variability). The RDA condition imposed little requirements for the participant to act at the executive level and allowed focus at the executory level for manual device testing and secondary task completion.

Secondary task difficulty was varied using a different type of the TMT test. The "easy" level had 30 targets and was based on "Part A" Trails, entailing a sequence of numbers (i.e. 1-2-3 and so forth), whereas the "difficult" level had 34 targets and was based on "Part B" Trails, entailing a sequence of alternating numbers and letters (i.e. 1-A-2-B or A-1-B-2 and so forth), with the latter requiring the use of further cognitive abilities (Sánchez-Cubillo et al. 2009).

In the working hypothesis, participants facing the "difficult" secondary task are expected to perform sub-optimally when they are allocated decision-making authority (HDA) for the primary task, because of attentional overload. In other words, they might not allocate and/or schedule device tests in a time-effective way, and/or they might be slower at the secondary task compared to the RDA scenario. Conversely, participants facing the "easy" secondary task are expected to perform superiorly under HDA surpassing the benchmark RDA performance. On the other hand, robot decision authority in the "easy" secondary task could lead to attentional underload and a subsequent decrease of engagement and performance in the primary task, which will require only rare moments of attention. Robot requests for assistance, such as manual device tests or error correction may be overlooked by the participant.

Independent variables:

1. Decision authority: ("HDA", "RDA"); within-subject
2. Secondary task difficulty: ("easy", "difficult"); between-subject

Dependent variables:

1. Primary task performance (assembly and testing)
2. Secondary task performance (TMT)
3. Overall perceived workload
4. Subjective preference of decision authority

2.4 Measurement setup

The Panda collaborative robot by Franka Emika (™) was programmed to perform device assembly and testing (under request), in combination with external, pneumatically controlled fixtures. For the secondary task, a variation of the original version of the TMT was developed, using PEBL (Mueller and Piper 2014) and ran on a laptop computer using the laptop's trackpad as the input method. Information exchange between participant and robot for device testing decisions occurred through a graphical user interface on one corner of the laptop screen, visible at the same time as the secondary task. In HDA, the interface consisted of virtual buttons used by the participant to inform the robot about testing task need and allocation. In RDA, the graphical window would display a flashing text whenever the robot requested a manual test from the participant.

Experiments were video recorded to extract performance measures of the assembly task. Overall perceived workload data were captured using the NASA-TLX tool on the laptop, after the completion of each of the two trials. Subjective preference of level of autonomy was measured at the end of each session through a single-item bipolar questionnaire, using a seven-point Likert scale: "Which one of the two collaboration scenarios that you experienced would you prefer to work with?".

2.5 Experimental procedure

First, the participants received a verbal introduction to the experiment, which included videos of the robot's operation, one manual device test practice, and one practice round of the secondary task. Decision authority was set to a certain level (HDA or RDA, counterbalanced) and the trial started, in which participants were instructed to supervise the assembly of 12 devices, ensure the testing of 6 specific devices, all while completing as many rounds of the secondary task as possible. Participants could choose when to start each TMT Trail, which was accompanied by a digital timer, providing feedback on performance in that Trail. According to the guideline they received, participants aimed to prioritize first the assembly/testing of devices, second the Trail completion

time, and third the number of Trails completed. In the scenario created by these rules, the optimal task scheduling strategy (under both RDA and HDA) was to perform any manual device tests (i) before the robot completes the assembly of all devices, (ii) in-between TMT Trails, and (iii) based on personal preference and fatigue from the TMT. Regarding task allocation, the optimal strategy in HDA was to perform all device tests manually, since automated testing was visibly slower. Under RDA, task allocation was automated; the number of manual tests was predetermined (half of the total number of required device tests) and functioned as a benchmark. Participants' (and ultimately team) performance was essentially determined by their successful deduction of the aforementioned tactics, as well as their ability to adhere to them, under circumstances of temporal pressure and division of attention.

The duration of one full experimental trial was determined by the total assembly time (i.e. the trial ended when the device assembly and testing were finished) and was approximately 15 min on average. Participants completed the NASA-TLX ratings of the condition they had just experienced upon its completion. This was followed by a short 5-min break, and the introduction of the second experimental condition with remaining decision authority condition. Subsequently, the second trial was completed in approximately another 15 min, and another round of NASA-TLX ratings was performed upon trial completion, with the option to view/modify previous ratings. The NASA-TLX pairwise comparisons for relative weighting of the factors were answered only once, at the end of the second trial. Finally, a questionnaire was administered about the subjective preference between the two conditions the participant experienced. The complete duration of the experiment including introductions was approximately 1 h.

2.6 Measured values

The evaluation metrics extracted from the experiment are discussed in this section. They consist of both objective performance measures (capturing efficiency and effectiveness) and subjective measures.

2.6.1 Primary task performance

2.6.1.1 Assembly completion time The total time required for the human–robot team to finish all device assemblies and necessary tests. This measure includes the effects of (a) the assistance given to the robot by the human in the form of manual tests, helping to improve efficiency and (b) the response time to robotic faults which require human intervention to resume the process.

2.6.1.2 Number of manual tests The number of tests that a participant chooses to complete manually while working in parallel with the robot. Due to parallelization of work, a higher number of manual tests results in higher process efficiency (inversely correlated to assembly completion time).

2.6.2 Secondary task performance

2.6.2.1 Trail completion time For each Trail, this is the time between the first mouse click in a Trail and the moment the final target is clicked on. Participants were instructed that this was a priority over the number of Trails completed. This variable is intended as a measure of the focus allocated to the secondary task and abstractly represents the effectiveness the participant would exhibit in any equivalent secondary task.

2.6.2.2 Number of completed Trails The total number of Trails that the participant is able to complete until the assembly and tests are finished. If a Trail is incomplete at the time of assembly finish, the completion percentage is counted. Participants were instructed to prioritize this less than Trail completion time. This metric is intended to capture the success of the temporal strategy employed by the participant, and thereby it is thought as the equivalent efficiency in any equivalent secondary task.

2.6.3 Subjective measures
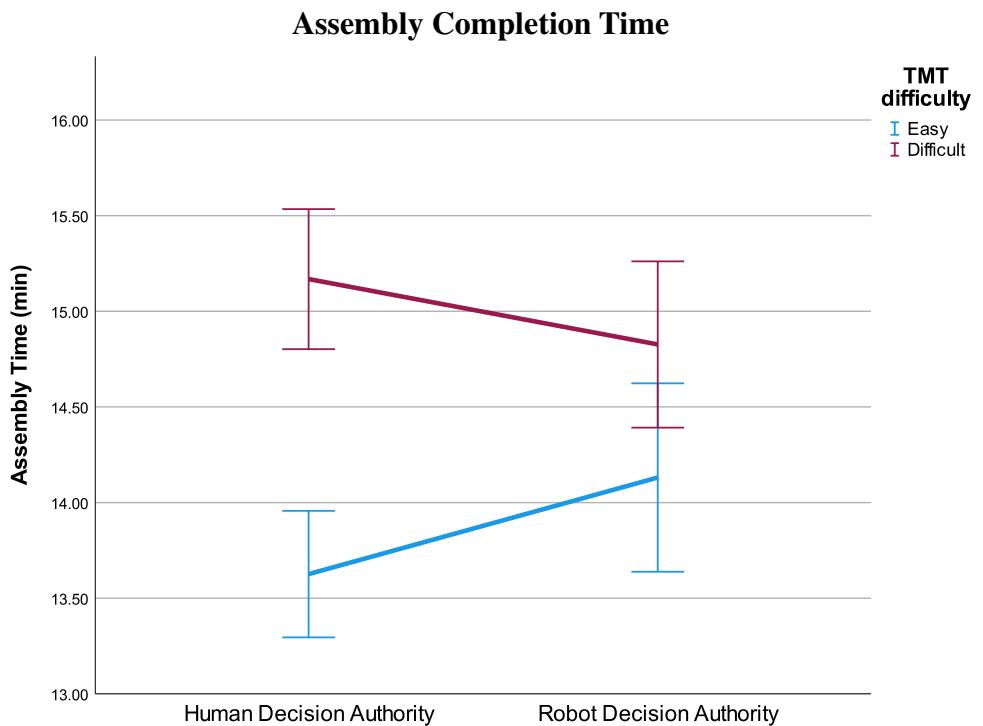
2.6.3.1 NASA-TLX score The overall perceived workload assessment of participants for each condition they experienced.

2.6.3.2 Subjective preference score of autonomy level The rating of preference between the two conditions of decision authority, in a seven-point Likert scale (1: strong preference for RDA, 7: strong preference for HDA).

3 Results

Mixed-model ANOVA was performed to analyze the interaction between decision authority and secondary task difficulty, for the variables that fulfilled the test's requirements. Moreover, Wilcoxon signed-rank tests (non-parametric) were used to establish statistical significance for within-subject comparisons (i.e. between the two decision authority levels) for all the variables, unless otherwise stated. In the latter tests, the participant group in the "easy" condition (hereafter named Group E) and the group in the "difficult" condition (Group D) were analyzed separately, to find the effects of decision authority at each given secondary task difficulty level.

Fig. 3 Mean assembly completion time. Error bars: 95% CI
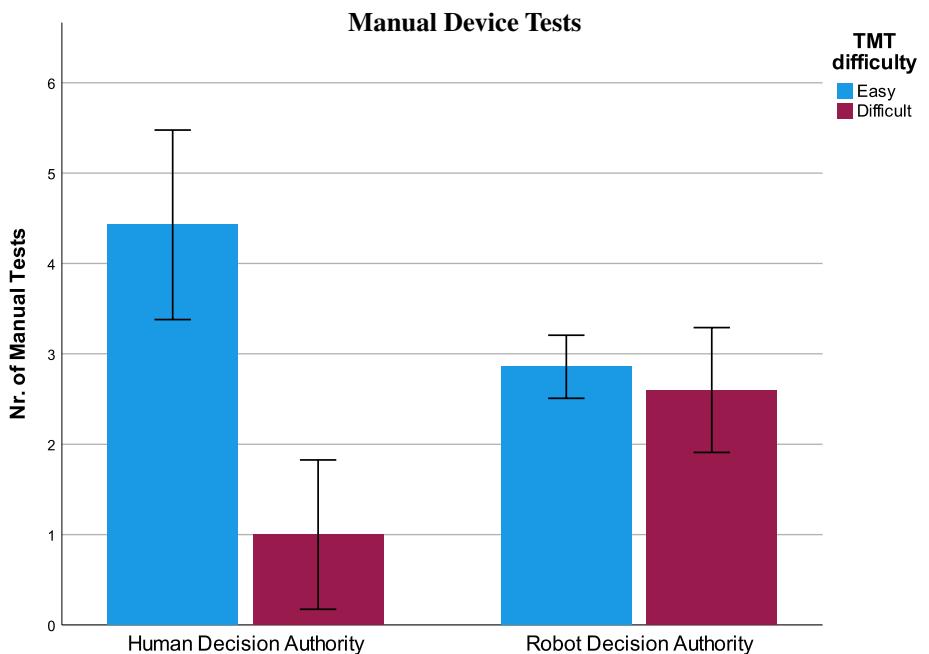

3.1 Primary task performance

The ANOVA results regarding assembly completion time suggest that there exists interaction between the factors decision authority and secondary task difficulty ($F(1,15) = 4.754$, $p = 0.046$, $\eta_p = 0.241$), exhibited by the reversing trend shown in Fig. 3. Priorly, the data were assessed with the Shapiro–Wilk test and were found to be normally distributed

in all four conditions (lowest $p = 0.153$). Three outliers were excluded after boxplot inspection, and another participant was excluded due to incomplete data capture. ANOVA on other metrics could not be performed due to test assumption violations.

Wilcoxon tests revealed no simple main effects of decision authority for the variable assembly completion time (Group E: $Z = 1.521$, $p = 0.128$ and Group D: $Z = 1.070$,

Fig. 4 Mean number of manual device tests performed. Error bars: 95% CI

$p = 0.285$). Nevertheless, regarding the number of manual tests performed (see Fig. 4), participants in Group E performed significantly more manual tests under human decision authority ($Z = 2.232$, $p = 0.026$), whereas Group D performed significantly fewer manual tests under human decision authority ($Z = 2.422$, $p = 0.015$), compared to RDA. Results are summarized in Table 1.

3.2 Secondary task performance

Regarding the number of completed Trails, the ANOVA suggests that there is interaction between decision authority and secondary task difficulty ($F(1,18) = 5.850$, $p = 0.026$, $\eta_p = 0.245$). Priorly, the data were tested for normality with the Shapiro–Wilk test and were found to be normally distributed in all four conditions (lowest $p = 0.064$). One outlier was excluded. No ANOVA was performed for the mean Trail completion time because test assumptions were violated.
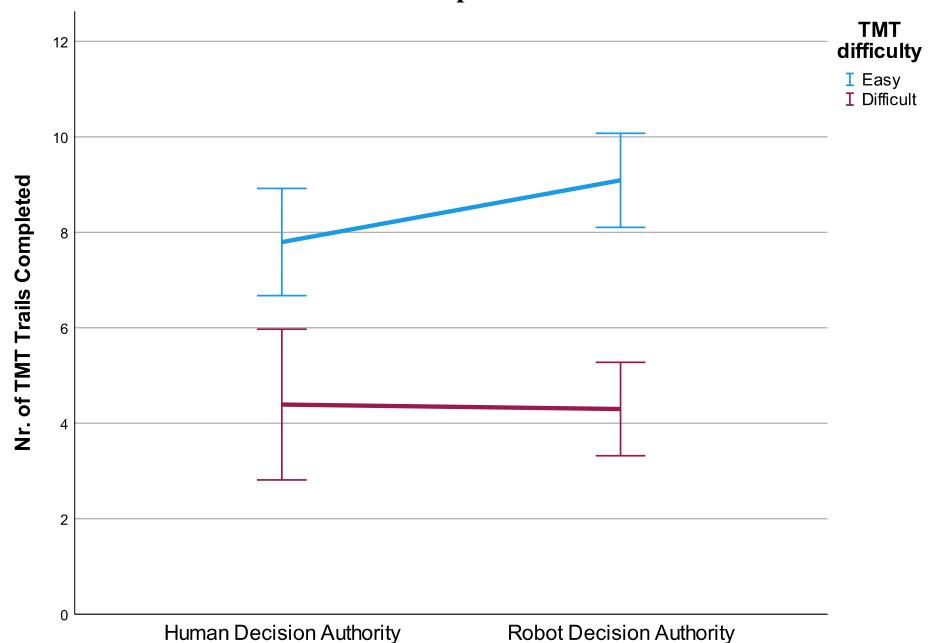
Using a paired-sample t test, a significant difference in the number of completed Trails was found only in Group E,

where more Trails were completed under RDA than HDA ($t(8) = 2.809$, $p = 0.023$) (see Fig. 5). The mean Trail completion time was not significantly different between decision authority levels, for both groups, although it appeared to have a decreasing tendency under RDA (see Fig. 6). TMT performance results are summarized in Table 2.

3.3 Subjective measures

Main effects from the ANOVA on NASA-TLX data for the overall perceived workload revealed that HDA had a significantly increased TLX score compared to RDA ($F(1,18) = 5.690$, $p = 0.028$, $\eta_p = 0.240$) (see Fig. 7). No significant interaction was found between factors secondary task difficulty and decision authority. The data were found to be normally distributed in all four conditions, using the Shapiro–Wilk test. One outlier was excluded from the data. NASA-TLX results are summarized in Table 3.

Correlation between the TLX score and performance metrics was analyzed using the Pearson coefficient. Only a moderate correlation was found between the TLX score and the number of completed TMT Trails ($r = -0.347$), as well as mean Trail completion time ($r = 0.318$). In other words, when participants completed fewer TMT Trails, at a slower pace per Trail, they perceived the whole scenario as more demanding. Assembly completion time and manual device tests were uncorrelated to NASA-TLX.

Concerning subjective preference, no significant difference was found between HDA and RDA in either of the two groups. Five participants in Group D stated they preferred RDA (rating less than 4) versus six who were in favor of

Table 1 Results of primary task performance in M (SD) format ($N = 17$)

TMT difficulty	Assembly completion time (min)		Manual tests (#)	
	HDA	RDA	HDA	RDA
Easy	13.63 (0.36)	14.13 (0.53)	4.43 (1.13)*	2.86 (0.38)*
Difficult	15.17 (0.51)	14.83 (0.61)	1.00 (1.15)*	2.60 (0.97)*

HDA human decision authority, *RDA* robot decision authority, *$p < 0.05$

Fig. 5 Mean number of completed TMT Trails. Error bars: 95% CI

Number of Completed TMT Trails

Fig. 6 Mean Trail completion time (s). Error bars: 95% CI

TMT Trail Completion Time
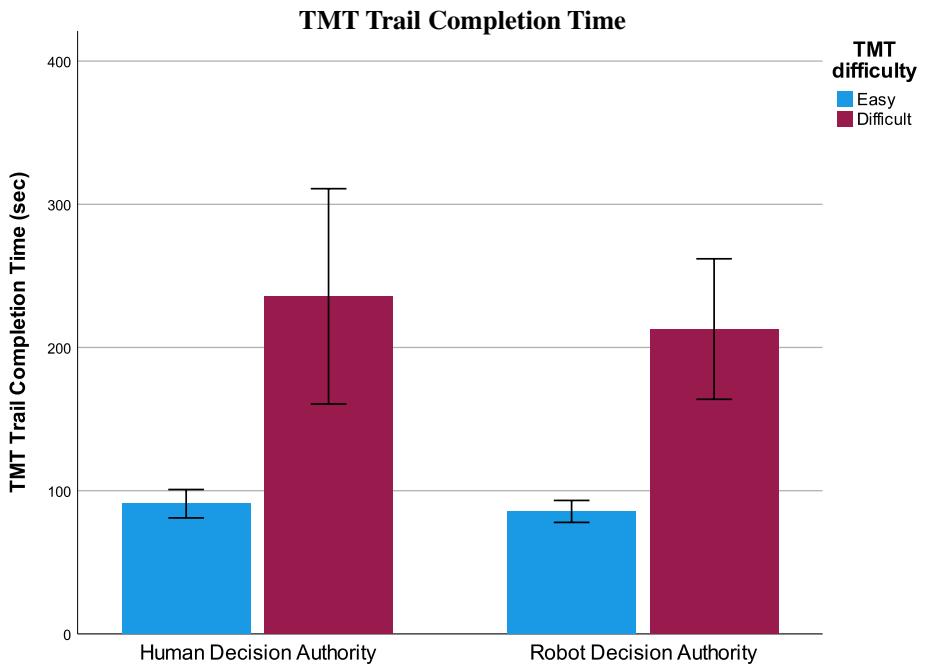
Table 2 Results of secondary task performance in M (SD) format ($N = 20$)

TMT difficulty	Completed Trails (#)		Trail completion time (s)	
	HDA	RDA	HDA	RDA
Easy	7.80 (1.46)*	9.09 (1.28)*	90.91 (12.87)	85.52 (9.97)
Difficult	4.39 (2.35)	4.30 (1.46)	235.73 (111.93)	212.89 (73.04)

HDA human decision authority, *RDA* robot decision authority, *$p < 0.05$

Fig. 7 NASA-TLX score
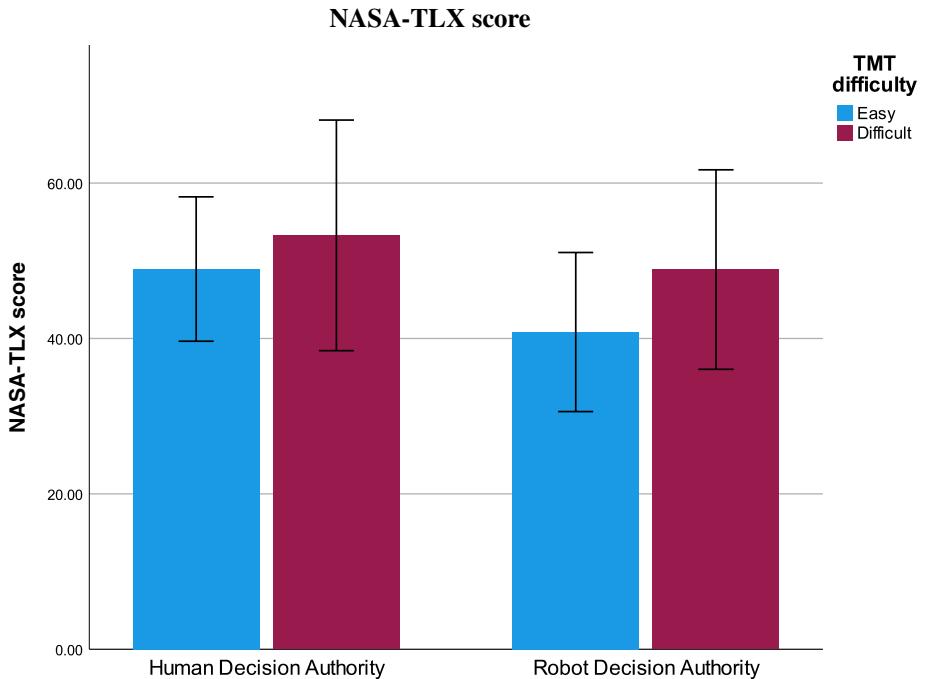
NASA-TLX score

Table 3 NASA-TLX score results in M (SD) format ($N=20$)

TMT difficulty	NASA-TLX score	
	HDA	RDA
Easy	48.94 (12.08)	40.83 (13.32)
Difficult	53.27 (22.10)	48.86 (19.11)

HDA human decision authority, *RDA* robot decision authority

HDA. In Group E, both decision authority levels were preferred by five participants each (see Fig. 8).

To study the correlation of preference (unpaired variable) with the other dependent variables (paired), new variables had to be constructed, by subtracting HDA variables from RDA variables. The transformed variables represented the difference of performance of each participant between the two levels of decision authority (e.g. how many more TMT Trails a participant completed under HDA versus RDA). The Spearman's rank order coefficient indicated a moderate correlation between subjective preference and both secondary task metrics, namely number of completed TMT Trails ($\rho=0.414$) and mean Trail completion time ($\rho=-0.434$). Lastly, subjective preference was found to be correlated with the NASA-TLX score ($\rho=-0.464$). In short, participants preferred the decision authority scenario in which they completed more TMT Trails, quicker per Trail on average, and with less perceived effort.

4 Discussion

This study examined the effect of decision authority assignment in HRC tasks under different induced operator workload levels. An experiment based on an industrial dual-task scenario was conducted, in which the HRC task (primary task) was the semi-automated assembly of a medical device. The HRC task was complemented with a secondary cognitive task for the human operator to test task performance and subjective preference criteria at different levels of induced cognitive workload. Decision authority for task allocation and scheduling was given to either the operator or the robot's algorithm. It was hypothesized that, in light-to-moderate induced workload conditions, increased human control would result in better situational awareness and thus performance, whereas in high-induced workload conditions, it may be more advantageous to offload the human by shifting task allocation responsibility to the robot. The results of the study partly support the above hypothesis.

The existence of interaction between the factors decision authority and secondary task difficulty, established by the ANOVA for the metric assembly completion time, is evidence that increasing the difficulty of the secondary task alters the effect of decision authority, in the direction

initially hypothesized: lower secondary task difficulty makes HDA beneficial for primary task performance, whereas RDA yields better results under higher secondary task difficulty. However, in secondary task performance (number of completed Trails), the discovered interaction had the opposite tendency: RDA was more beneficial in the easier secondary task group (i.e. Group E), but not in the difficult secondary task group (i.e. Group D).

In the easier secondary task group, participants fared better in the primary task under the HDA condition (i.e. more manual tests), suggesting that they were able to utilize their increased level of control to benefit the automation. However, this came at a cost of declined performance in the secondary task, for which the RDA condition proved to be more beneficial, possibly due to increased focus when not charged with decision-making responsibilities.

Conversely, in the difficult secondary task group, primary task performance was better with RDA (i.e. more manual tests). No difference was found in secondary task performance metrics between the two levels of decision authority, suggesting that participants were equally focused on the secondary task in both conditions. This could explain their poorer performance in assisting the automation in the primary task under HDA, in which they executed significantly fewer manual tests, thus increasing the assembly completion time. It is possible that participants, in the presence of a more demanding secondary task, committed a higher portion of their cognitive resources to it, negatively affecting their ability to strategize effectively, and/or adjust and follow their plans for manual device tests.

These findings hint at the existence of a trade-off in performance between the primary and secondary tasks which could indicate that in this experiment, participants approached their cognitive resource limitations.

No significant difference was found in the mean Trail completion time variable in the TMT, which serves as an indicator of how focused participants were able to remain on the secondary task, without undesired interruptions for manual tests in the primary task. Participants of each group remained equally focused, under both decision authority levels, suggesting that the quality of work they would perform in any secondary task would not be significantly affected by the level of decision authority in the primary task.

NASA-TLX results confirmed the assumption that HDA causes higher overall perceived workload to participants than RDA, because of the added decision-making responsibilities and the required shifting between cognitive levels (executive and executory). Moreover, TLX scores were found to be mildly correlated to secondary task performance, but not to primary task performance, potentially because the secondary task demanded more of the participant's attention, mental effort, and time. Further, performance in the secondary task

Subjective Preference Rating
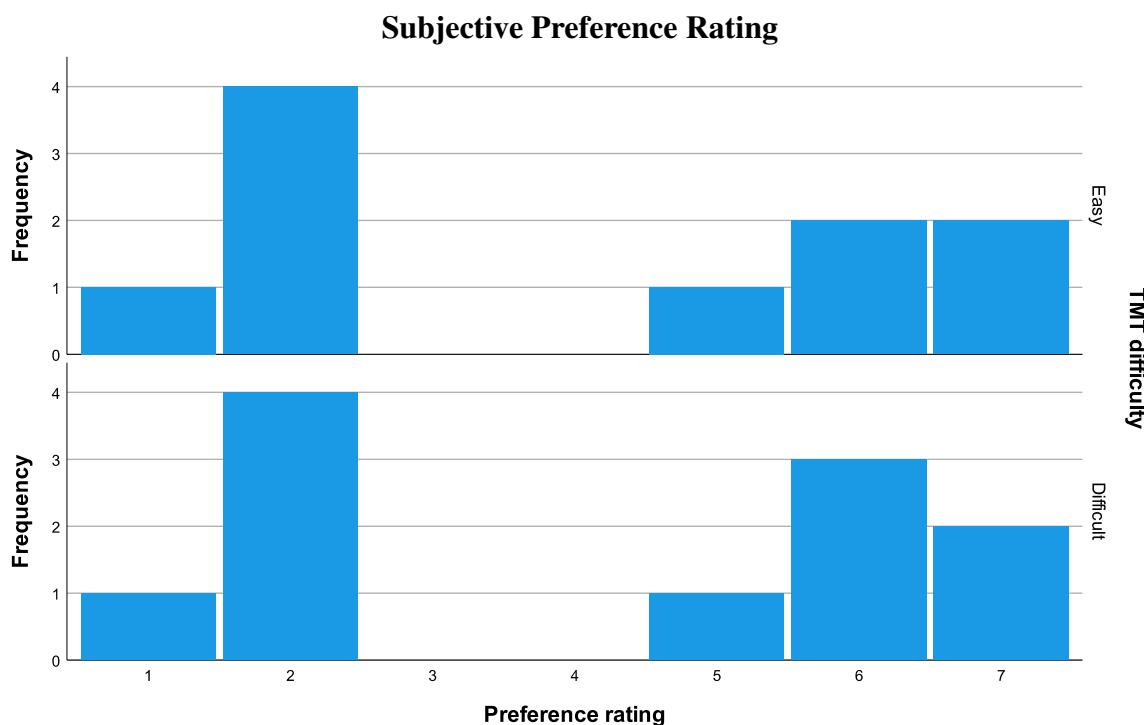
Fig. 8 Subjective preference rating. 1 = (preference for robot decision authority) and 7 = (preference for human decision authority)

was more comprehensive and easily comparable between experiment trials (e.g. live timer on TMT screen).

Interestingly, in spite of the various differences in performance and workload produced by decision authority and secondary task difficulty, these did not translate to subjective preference towards any of the conditions. The data clearly suggest that preference was evenly divided between the two authority levels, independently to difficulty. Nevertheless, preference was moderately linked to—already correlated—variables of perceived workload and secondary task performance. Overall, evidence indicates that interpersonal effects and predisposition, possibly related to performance/perceived workload, are the determining factors of preference. Some participants felt more comfortable focusing on the secondary task when they possessed decision-making authority, often due to a lack of self-trust in noticing and remembering the robot's requests under RDA. Other participants experienced the RDA scenario as more comfortable because it reduced the need for continuous assembly process monitoring.

5 Conclusion and outlook

In an era of expanding automation, robots are increasingly being introduced in workspaces beyond mass production lines, towards lower-volume processes, forming teams with human operators. Their flexibility makes them attractive to dynamic environments with varying demands and tasks sought to be automated. To maximize resource utilization, robots may need to change tasks repeatedly throughout a workday, while fluently collaborating with teammates.

Such flexibility requires frequent, ad hoc decision-making, which poses an additional responsibility for the team's members. In the present work, the question of decision authority assignment between human operator and robot is investigated, regarding performance, overall perceived workload, and subjective preference. Traditionally, human factors research recommends that operators should be granted control when collaborating with semi-autonomous systems, but some empirical studies have shown that this is not always efficient or even preferred by participants. We hypothesized that the answer is dependent on the cognitive workload imposed on the operator by the work process (i.e. induced workload).

In a study involving 21 participants, it was found that, in lower induced workload conditions (easier secondary tasks), human decision authority can increase the productivity of the primary, semi-automated task by engaging the operator in the loop, but it may also lead to a deterioration of performance in any secondary tasks performed in parallel due to division of attentional resources. On the contrary, in higher induced workload conditions (difficult secondary tasks), human authority in decisions can lead to cognitive overload,

causing inferior performance in the primary task, without corresponding improvement in secondary task performance, compared to robot decision authority.

The experimental HRC study was held to confirm the hypothesis that human decision-making is more suitable to low workload scenarios while high human workload increases the need for autonomous decision-making. This was found to hold true, at least in the case study under question. Specifically, under low induced workload, there seems to be a trade-off in performance between the primary and secondary task—human decisions do benefit the former but tend to neglect the latter. Under high-induced workload, such a trade-off is not evident; human decision authority decreases primary task performance, with no respective increase in secondary task performance, hinting at a diminishing returns scenario.

The above findings add to the ongoing discussion in the literature about adaptive automation in the field of human–robot interaction, by providing evidence on the importance of cognitive workload in assigning the authority to invoke automation to the operator or the robotic system itself. In accordance with classic Human Factors research, collaborative robots should best be operated at an autonomy level which facilitates optimal arousal, subject to operator skill. It is suggested that pertinent HRC studies ought to consider the level of mental workload imposed on subjects, as it can have varying effects on team efficiency and limit the generalizability of the results.

In summary, to allocate decision authority appropriately in a human–robot collaborative process, designers need to consider the potential cognitive workload that will be imposed on an operator by tasks either inherent to the process or external, which may be performed in parallel with robot supervision. A detailed cognitive task analysis for the operator may be required before selecting a decision authority scheme. However, in dynamic collaborative settings it is perhaps preferable if the decision authority is not predesigned but rather an adjustable option. The criteria of selection should include the case-specific relative importance of performance of the tasks involved, as well as the subjective preference of the operator.

Some limitations of this study should be mentioned. First, the instructions given to participants regarding task prioritization may have had a varying effect among participants resulting in divergent task prioritization strategies among them. Second, besides increased difficulty, each Trail of the difficult secondary task also had longer completion time than the easy version, which may have biased participant task-shifting decisions for time-management reasons, potentially confounding the analysis for cognitive workload effect. Third, due to the relatively short duration of the trials (30 min) no substantial skill building was possible. It is possible that after the learning phase,

humans utilize the features of each autonomy level and environmental cues more effectively, so that performance trends and even their subjective preference are altered. Lastly, the presented experiment was conducted using a specific secondary task, with particular temporal features and cognitive subsystem requirements (e.g. visual perception channel, symbolic information processing). For results to be more generalizable, a wider variety of tasks of different natures could be examined, using the principles of cognitive task analysis.

Declarations

Competing interests The authors declare no competing interests.

Ethical Standards Obtention of consent from study participants for data collection and publication complied with the ethics guidelines of ETH Zurich.

References

Baraglia J, Cakmak M, Nagai Y, Rao R, Asada M (2016) Initiative in robot assistance during collaborative task execution. In: Paper presented at the 2016 11th ACM/IEEE international conference on human-robot interaction (HRI)

Csikszentmihalyi M (1990) Flow: the psychology of optimal experience, vol 1990. Harper & Row, New York

Eberts R, Salvendy G (1986) The contributions of cognitive engineering to the safe design and operation of CAM and robotics. J Occup Accid 8(1–2):49–67

Endsley MR (1995) A taxonomy of situation awareness errors. Hum Fact Aviat Oper 3(2):287–292

Gaudino EA, Geisler MW, Squires NK (1995) Construct validity in the trail making test: what makes part B harder? J Clin Exp Neuropsychol 17(4):529–535. https://doi.org/10.1080/01688639508405143

Gombolay MC, Gutierrez RA, Clarke SG, Sturla GF, Shah JA (2015a) Decision-making authority, team efficiency and human worker satisfaction in mixed human–robot teams. Auton Robot 39(3):293–312

Gombolay MC, Huang C, Shah J (2015b) Coordination of human–robot teaming with human task preferences. In: Paper presented at the 2015b AAAI fall symposium series

Harris C, Hancock P, Arthur E (1993) The effect of taskload projection on automation use, performance, and workload. In: Paper presented at the proceedings of the seventh international symposium on aviation psychology

Hart SG (2006) Nasa-task load index (NASA-TLX); 20 years later. Proc Hum Fact Ergon Soc Annu Meet 50(9):904–908. https://doi.org/10.1177/154193120605000909

Johannsen G (1979) Workload and workload measurement. In: Moray N (ed) Mental workload: its theory and measurement. Springer, US, Boston, pp 3–11

Kaber DB, Riley JM (1999) Adaptive automation of a dynamic control task based on secondary task workload measurement. Int J Cogn Ergon 3(3):169–187

Katidioti I, Borst JP, van Vugt MK, Taatgen NA (2016) Interrupt me: external interruptions are less disruptive than self-interruptions. Comput Hum Behav 63:906–915. https://doi.org/10.1016/j.chb.2016.06.037

Leroy S (2009) Why is it so hard to do my work? The challenge of attention residue when switching between work tasks. Organ Behav Hum Decis Process 109(2):168–181. https://doi.org/10.1016/j.obhdp.2009.04.002

Mueller ST, Piper BJ (2014) The psychology experiment building language (PEBL) and PEBL test battery. J Neurosci Methods 222:250–259. https://doi.org/10.1016/j.jneumeth.2013.10.024

Ogden GD, Levine JM, Eisner EJ (1979) Measurement of workload by secondary tasks. Hum Fact 21(5):529–548. https://doi.org/10.1177/001872087902100502

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(3):286–297

Rubio S, Díaz E, Martín J, Puente JM (2004) Evaluation of subjective mental workload: a comparison of SWAT, NASA-TLX, and workload profile methods. Appl Psychol 53(1):61–86

Sánchez-Cubillo I, Periáñez JA, Adrover-Roig D, Rodríguez-Sánchez JM, Ríos-Lago M, Tirapu J, Barceló F (2009) Construct validity of the Trail Making Test: role of task-switching, working memory, inhibition/interference control, and visuomotor abilities. J Int Neuropsychol Soc 15(3):438–450. https://doi.org/10.1017/S1355617709090626

Scerbo MW (1996) Theoretical perspectives on adaptive automation. Automation and human performance: theory and applications. Lawrence Erlbaum Associates Inc, Hillsdale, pp 37–63

Sheridan TB, Parasuraman R (2005) Human–automation Interaction. Rev Hum Fact Ergon 1(1):89–129. https://doi.org/10.1518/155723405783703082

Sheridan TB, Stassen HG (1979) Definitions, models and measures of human workload. In: Moray N (ed) Mental workload: its theory and measurement. Springer, US, Boston, pp 219–233

Tausch A, Kluge A (2020) The best task allocation process is to decide on one's own: effects of the allocation agent in human–robot interaction on perceived work characteristics and satisfaction. Cogn Technol Work. https://doi.org/10.1007/s10111-020-00656-7

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 129:104814. https://doi.org/10.1016/j.ssci.2020.104814Wickens

Wickens CD (2005) Attentional tunneling and task management. In: 2005 international symposium on aviation psychology, p 812

Wickens CD, Hollands JG, Banbury S, Parasuraman R (2015) Automation and human performance. In: Engineering psychology and human performance. Psychology Press, pp 377–404