Using Tools to Help Us Think:

Actual But also Believed Reliability Modulates Cognitive Offloading

Patrick P. Weis & Eva Wiese

George Mason University, Fairfax, VA, USA

Précis: When do people offload cognitive tasks onto devices in their environment? We found that both the device's actual reliability and erroneous beliefs about the device's reliability influence cognitive offloading. These results emphasize the relevance of factors beyond feedback-related performance optimization when offloading cognition.

Running head: Reliability and Cognitive Offloading

Manuscript type: Research Article

Word count: 5102

Corresponding author:
Patrick Weis
Email: pweis@gmu.edu

ABSTRACT

Objective: A *distributed cognitive system* is a system in which cognitive processes are distributed between brain-based internal and environment-based external resources. In the current experiment, we examined the influence of metacognitive processes on external resource use (i.e., *cognitive offloading*) in such systems. **Background:** High-tech working environments oftentimes represent distributed cognitive systems. Since cognitive offloading can both support and harm performance, depending on the specific circumstances, it is essential to understand when and why people offload their cognition. **Methods:** An extension of the mental rotation paradigm was used. It allowed participants to rotate stimuli either internally as in the original paradigm or with a rotation knob that afforded rotating stimuli externally on a computer screen. Two parameters were manipulated: the knob's actual reliability (AR) and an instruction altering participants' beliefs about the knob's reliability (believed reliability; BR). Cognitive offloading proportion and perceived knob utility were measured. **Results:** Participants were able to quickly and dynamically adjust their cognitive offloading proportion and subjective utility assessments in response to AR, suggesting a high level of offloading proficiency. However, when BR instructions were presented that falsely described the knob's reliability to be lower than it actually was, participants reduced cognitive offloading substantially. **Conclusion**: How much people offload their cognition is not solely based on utility maximization but is additionally affected by possibly erroneous pre-existing beliefs. **Application:** To support users in efficiently operating in a distributed cognitive system, an external resource's utility should be made transparent and pre-existing beliefs should be adjusted prior to interaction.

Keywords: Human systems integration, Situated Cognition; Metacognition; Distributed Cognition; HCI.

INTRODUCTION

Opportunities to outsource thought have become abundant. During the industrial revolution, the availability of machines that replaced or supported *physical* labor increased dramatically. Nowadays, we are in the middle of a similar revolution as we experience an extensive rise in machines that replace or support *mental* labor: computers. Computers can increasingly be used to for unpopular tasks, freeing our mental resources for what is more relevant (Storm & Stone, 2015). This rise in computer's abilities is partly due to a better understanding of how humans incorporate the environment into the cognitive loop, leading to better design choices during the creation of computer-based systems that afford the outsourcing of brain-based processing. A prominent everyday example where such understanding is implemented can be found in wayfinding support: modern GPS-based navigation systems are designed to match the external representation to the internal cognitive map, aiming for intuitive human-centric use (Huang, Tsai, & Huang, 2012). More generally, environments in which cognitive processes are distributed between brain-based, internal and environment-based, external resources have been termed socio-technical or distributed cognitive systems (Hollan, Hutchins, & Kirsh, 2000; Hutchins, 1995).

However, despite the positive impact of cognitive engineering and increased computational capacities on creating external resources that afford outsourcing thought, there remain instances where outsourcing thought, also called *cognitive offloading* (Risko & Gilbert, 2016; for a recent review), is not advisable. In tasks focusing on efficiency, cognitive offloading is contraindicated when the external resource is simply slower or less accurate than the internal resource. Such an inefficient external resource could, for example, be an unreliable decision aid (on average, decision aids have been found to be inefficient if their reliability is below 70%; Wickens & Dixon, 2007) or a reliable externally stored information that is however inefficient to access (e.g., because the interface does not abide Fitt's law and incorporates small buttons to access relevant information; Experiment 2 in Gray, Sims, Fu, & Schoelles, 2006). There is a multitude of other possible reasons not to offload cognition besides short-term efficiency: for example, in tasks focusing on flexibility, cognitive offloading can be contraindicated because it hinders the establishment of domain-specific knowledge that could be transferred to similar problems (O'Hara & Payne, 1998). In conclusion, outsourcing thought oftentimes comes at a cost that might be higher than the benefit.

Unfortunately, people's offloading behavior is not always well calibrated to these costs. Automation-induced complacency describes the phenomenon that people tend to over-rely on automation, thereby sometimes missing erroneous automation behavior and sometimes following erroneous advice from the automation (Parasuraman, Molloy, & Singh, 1993; Parasuraman & Riley, 1997). One might argue that such errors could be warranted, given the benefit of being relieved from the cognitive-resource-draining system monitoring. However, in safety-critical environments complacent offloading behavior can contribute to catastrophes that are hardly justifiable with decreased monitoring costs (e.g. airplane accidents; National Transportation Safety Board, 1994). Similarly, suboptimal offloading behavior has been reported when people were asked to remember letters while given the opportunity to write the letters down if necessary (Risko & Dunn, 2015): people used pen and paper in more than 40% of the cases when two letters had to be remembered, and in around 90% of the cases when ten letters had to be remembered. This pattern is surprising when compared to people's task performance without the opportunity to offload memory: without pen and paper, recall performance for two letters was excellent (i.e. above 97%) whereas it was extremely poor (i.e., below 1% accuracy) for ten letters. Participants seem to offload cognitive resources unnecessarily often when internal processing is efficient (i.e., two letters), and do not fully make use of external resources when it is highly useful (i.e., ten letters), which makes it impossible to justify participant's offloading behavior in terms of short-term performance optimization.

Understanding the reasons behind inefficient and possibly harmful offloading choices is imperative to remediate such badly calibrated behavior. One possible reason relates to erroneous metacognitive judgments about the utility of one's internal (i.e., brain-based) and external (e.g., pen and paper) resources. Decisions regarding the use of external resources might be, in addition to lower-level cognitive processes based on higher-level metacognitive processes. For example, the use of a GPS-based navigation system might be dependent on spatial navigation skills (i.e., a lower-level cognitive process) but also be influenced by explicit beliefs about the navigation system's efficacy (i.e., a higher-level metacognitive process). This idea has been put forward by the *Metacognitive Model of Cognitive Offloading* (Risko & Gilbert, 2016). The influence of higher-level metacognitive factors on cognitive offloading is also backed by correlational data from a follow-up experiment to the memory study reported above: when participants who preferred using pen and paper to remember two letters over using internal memory were asked why they chose this external strategy, they argued that the external strategy was associated with higher accuracy, which was a misjudgment (in reality, both strategies yielded similar accuracy; Risko & Dunn, 2015). Thus, the use of external resources is likely dependent on possibly erroneous higher-order metacognitive judgments regarding the resources' utility.

In the current study, we employed an experimental design to further examine the impact of metacognitive judgments about an external resource on the inclination to actually use that resource. Specifically, we measured how a rotation device's actual and believed reliability affected cognitive offloading proportion (i.e., knob recruitment) during an object rotation task. We expected both factors to affect cognitive offloading proportion independently. The rationale is that actual reliability should influence cognitive offloading via lower-level cognitive processes like performance monitoring while believed reliability should influence cognitive offloading via higher-level metacognitive processes, i.e. beliefs about the external resource's utility. In the present study, reliability beliefs were manipulated via instruction, thus representing rather superficial beliefs that should act like expectations and be less integrated than intrinsically formed beliefs. Nevertheless, we would argue such superficial beliefs to influence cognitive offloading by the same mechanisms as intrinsically formed metacognitive beliefs (compare Risko & Gilbert, 2016; Figure 3).

In particular, we predicted negative beliefs regarding the knob's utility (i.e., *incongruent* condition) to reduce cognitive offloading proportion as well as usefulness ratings as compared to a *congruent* (i.e., belief consistent with actual reliability) or *naïve* condition (i.e., no belief instruction). Whereas previous studies have used post-hoc questionnaires to assess influences of pre-existing beliefs on decisions to offload cognition (e.g., Dunn & Risko, 2016; Risko & Dunn, 2015), pre-existing beliefs were manipulated experimentally via instruction in the current experiment, which allows causal rather than correlational inferences regarding the role of metacognitive processes in cognitive offloading. For exploratory purposes, we also measured knob utility assessments (i.e., usefulness ratings) to compare them between reliability and belief conditions.

METHODS & MATERIALS

Participants

In total, 126 undergraduate students participated in the experiment. Four participants were excluded due to extremely poor task performance (i.e. answering incorrectly in more than 30% of the trials), resulting in a final sample size of 122 participants (77 females; mean age: 20.9; range: 18 – 47; 109 right handed). Participants were randomly assigned to one of the three experimental conditions (41 *naïve*, 42 *congruent,* 39 *incongruent*). All participants were recruited from the psychology undergraduate student pool at George Mason University and reimbursed via research participation credits. To motivate participants to perform well, the three participants with the best performance in the rotation task were rewarded with Amazon vouchers (1st place: 15$; 2nd place: 10$; 3rd place: 5$). All participants were at least 18 years old and had normal or corrected to normal vision. This research complied with the APA's code of ethics and was approved by the local Ethics Committee at George Mason University. Participants provided informed consent prior to participation.

Apparatus

Stimuli were presented at a distance of about 100 cm on an ASUS VB198T-P 19-inch monitor set to a resolution of 1280 × 1024 pixels and a refresh rate of 60 Hz using MATLAB version R2015b (The Mathworks, Inc., Natick, MA, United States) and the Psychophysics Toolbox (Brainard, 1997; Pelli, 1997). Button press responses were recorded using a USB-connected standard keyboard. The rotation knob consisted of a potentiometer (SpinTrak Rotary Control; Ultimarc, London, UK) sampled at 1000 Hz. One full rotation of the rotation knob corresponded to one full rotation of the working stimulus on the screen.

Stimuli

For the rotation task, twenty different 2D stimuli were in MATLAB using a script provided by Collin & McMullen (2002) that followed the Attneave procedure (Attneave & Arnoult, 1956; for a detailed description). The stimuli used in the current study differed from each other only with regard to the edge parameter, ranging from three to twenty-one edges (see **Figure 1**).

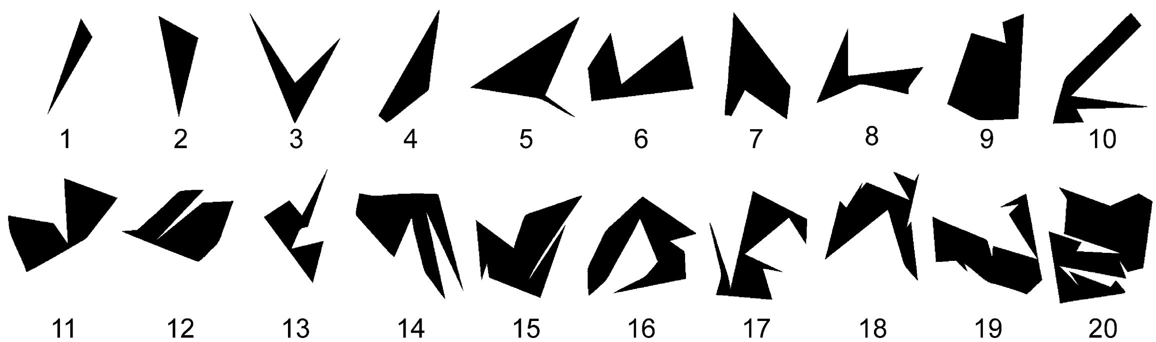

Fig. 1. Stimuli used for the extended rotation task: Twenty stimuli were created using the Attneave procedure (see *Stimuli*).

Task

An extension of the classic mental rotation paradigm (Shepard & Metzler, 1971; see **Figure 2a**) was used because it provides a moderately challenging cognitive task and allows implementation of a novel external resource that minimizes differences between participants due to prior experience and affords internal brain-based and external computer-based strategies.

At the beginning of each trial, a base stimulus is presented on the left and a working stimulus on the right side of the screen (see **Figure 2b**). The working stimulus represents either the base stimulus rotated clockwise by 60 or 120 degrees (*same handedness*), or the mirror image of the base stimulus rotated clockwise by 60 or 120 degrees (*different handedness*). Base and working stimulus appear on the screen at the same time and participants have up to five seconds to indicate the working stimulus' handedness via button press. Participants can either rotate one of the two stimuli internally or use the rotation knob to rotate the working stimulus externally on the screen to inform their answer. Importantly, rotating the knob would fail to rotate the stimulus in a systematic fashion (i.e., *Reliability* manipulation): knob reliability varied between 50% and 100% in increments of 10%, and was blocked throughout the experiment, with 40 rotation trials per block and reliability (i.e., in the 50% block, the knob would not rotate the working stimulus in 20 out of 40 trials). At the beginning of each block, a message on the screen informed participants about the knob reliability in the upcoming block (i.e., *belief* manipulation): in the *naive* condition, participants were only told that the knob might not work all the time, without inducing an explicit bias. In the *congruent* condition, participants were informed about the rotation knob's actual reliability, whereas in the *incongruent* condition, participants were wrongly informed about knob reliability (the provided reliability information was 30% lower than the actual reliability). Importantly, the actual reliability was comparable across all three conditions; only participants' expectations regarding reliability were varied.

It should be noted that the current design does not follow the typical "Choice/No Choice Paradigm" frequently employed in studies researching cognitive offloading (Risko & Gilbert, 2016, p. 678; Siegler & Lemaire, 1997). In such a design, participants are either forced to solve a task internally, forced to solve a task externally, or able to choose between internal and external strategies. Here, the main interest lies in participant's choice behavior and forced conditions are therefore omitted.

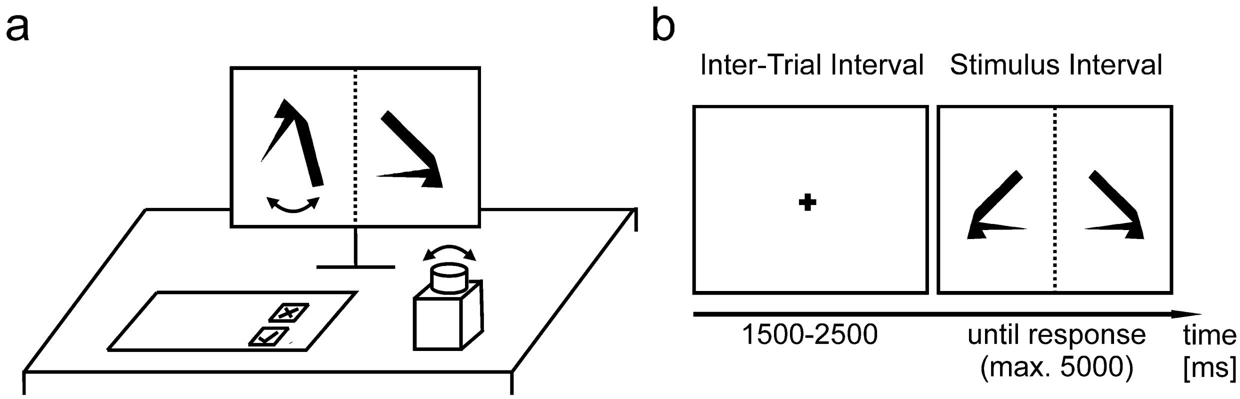

Fig. 2. *Extended rotation paradigm*: (a) The experimental set-up contained a computer screen, a standard keyboard, and a rotation knob. (b) Participant's task was to determine whether the base stimulus has the same handedness as the working stimulus. Participants could solve the task by mentally rotating one of the stimuli or by using the knob to rotate the working stimulus on the screen (for details, see *Task*). Stimuli and devices are not drawn to scale.

Procedure

At the beginning of each experimental session, participants were welcomed and seated in front of a computer screen. After providing informed consent, participants performed a computer version of the *rotary pursuit task* (i.e. exploratory measure of visuo-motor coordination; Melton, 1947; Mueller & Piper, 2014), and then solved 240 rotation problems as the main task of the experiment. The session concluded with a demographic survey. The study took 30 minutes to complete.

The rotation task follows a 6 x 2 x 2 x 3 mixed design with the within-participants factors *Reliability* (50%, 60%, 70%, 80%, 90%, 100%), *Handedness* (same, different), and *Angle* (60°, 120°), and the between-participants factor *Belief* (naive, congruent, incongruent). Trials were presented in blocks of 40, and each reliability condition was assigned to a specific block. The distribution of the unreliable trials was randomized within a block, and all stimuli were presented as working stimuli twice, once rotated by 60° and once by 120°. The order in which the different reliability blocks were presented was partially counter-balanced using a Latin square approach (Cochran & Cox, 1950).

Participants were allowed to take breaks every twenty trials. During the break, a message on the screen showed the amount of points gained during the last twenty trials to indicate their performance (100% of trials correct: 5 points; >= 90% of trials correct: 2 points; >= 70% of trials correct: 1 point). The three participants with the overall highest scores were awarded Amazon vouchers. To measure participant's metacognitive evaluations of the utility of the external resource, we prompted them twice during the experiment to evaluate the usefulness of the rotation knob on a 10-point scale (0: not at all; 9: very much): the first prompt was presented after finishing block one (i.e. after participants had encountered only one reliability condition), and the second prompt was presented at the end of the experiment (i.e., after all reliability conditions had been encountered).

Analysis

All trials with missing answers or RT values above or below 3 SD of the individual mean of the respective angle condition and trials with RT values below 150ms were excluded from analysis (0.8% of trials in total). To determine if participants used the external resource, we created a binary variable on a trial-by-trial basis that indicated if the participants turned the stimulus on the screen for more than 3° (i.e., external resource used) or less than 3° (i.e., external resource not used). The statistical approaches are described in the results section preceding the respective results. Effect sizes are reported as generalized eta squared (η_G^2). Generalized eta-square enables comparison between between-participants and within-participants designs (Bakeman, 2005; Olejnik & Algina, 2003).

RESULTS

Performance

Neither reaction time (F(2, 119) = 1.49, p = .229, η_G^2 = .016) nor accuracy (F(2, 119) = .12, p = .883, η_G^2 = .001) differed between belief conditions, suggesting comparable overall performance across groups. The ANOVA results are summarized in **Table S1 and S2**. P-values are reported Greenhouse-Geisser-corrected where applicable.

Cognitive offloading proportion

To analyze the influence of actual and believed reliability on cognitive offloading proportion (i.e., proportion in which participants used the knob to turn the stimulus for more than 3°), we conducted a 6 x 2 x 2 x 3 mixed ANOVA with the within-participants factors *Reliability* (50%, 60%, 70%, 80%, 90%, 100%), *Handedness* (same, different), *Angle* (60°, 120°) and the between-participants factor *Belief* (naive, congruent, incongruent). The ANOVA was followed up with non-parametric post-hoc Wilcoxon rank sum tests to account for deviations from normality in the DV's distributions.

Both actual knob *Reliability* (F(5, 595) = 23.69, p < .001, η_G^2 = .042), and *Beliefs* regarding the knob's reliability (F(2, 119) = 3.49, p = .034, η_G^2 = .035) had a significant impact on the extent to which participants used the rotation knob (i.e., cognitive offloading proportion) during the task. The *Reliability* x *Belief* interaction did not reach the level of significance (F(10, 595) = 1.64, p = .115, η_G^2 = .005). As expected, but of minor interest for the purposes of this study, *Angle* (F(1, 119) = 71.62, p < .001, η_G^2 = .004, M(60°) = 64.3%, M(120°) = 68.6%) and *Handedness* (F(1, 119) = 5.85, p = .017, η_G^2 = .0002, M(congruent) = 66.9%, M(incongruent) = 66.0%)) also affected cognitive offloading proportion. The interaction between *Reliability*, *Angle*, and *Handedness* was close to significance but also of minor interest to the main purposes of this study

($F(5, 595) = 2.15$, $p = .058$, $\eta_G^2 = .0003$). No other effects reached statistical significance (all $F < 2.2$, all $p > .1$, all $\eta_G^2 < .006$, **see Table 1**). The effect of actual and believed reliability on participants' external resource use is shown in **Figure 3.**

Post-hoc two-sided Wilcoxon rank sum tests (Hollander & Wolfe, 1973) showed that it had no influence on overall cognitive offloading proportion whether participants were correctly informed about the actual reliabilities of the external resource or had to deduce the reliabilities during the block (*congruent vs. naïve*, $W = 901$, $p = .719$, *M*(congruent) = 72.56, *M*(naïve) = 70.54), which suggests that participants promptly picked up on the actual knob reliability in the naïve condition and adjusted their cognitive offloading proportion accordingly. However, if participants were given incongruent information stating lower knob reliability, two single-sided Wilcoxon rank sum tests confirmed that participants used the external resource significantly less often than when given no information (i.e., *naïve* vs. *incongruent*, $W = 1005.5$, $p = .036$, *M*(naïve) = 70.54, *M*(incongruent) = 55.71) or when given congruent information (i.e. *congruent – incongruent*, $W = 1051.5$, $p = .036$) about the external resource's reliability. The data shows that correct utility beliefs, in contrast to incorrect utility beliefs, did not alter cognitive offloading proportion. All p-values were corrected for multiple comparisons using the Bonferroni-Hochberg method (BH; Benjamini & Hochberg, 1995).

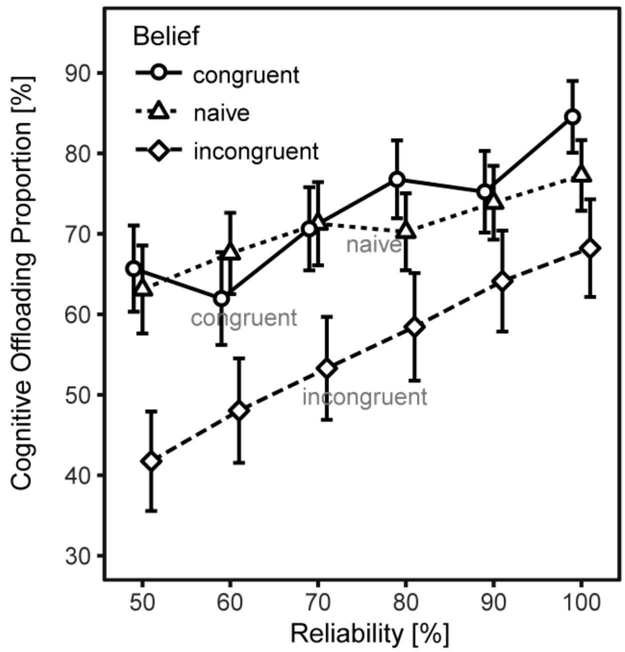

Fig. 3. *Cognitive offloading proportion as a function of actual and believed reliability*. Participant's cognitive offloading behavior depends on both actual (x-axis) and believed (line types) reliabilities. Error bars depict SEM.

Table 1
ANOVA results for cognitive offloading proportion

	DF1	DF2	F	p	η_G^2
Belief *	2	119	3.49	0.0338	0.0422
Reliability ***	5	595	23.69	< 0.0001	0.0355
Angle ***	1	119	71.62	< 0.0001	0.0035
Handedness *	1	119	5.85	0.0171	0.0002
Reliability x Belief	10	595	1.64	0.1150	0.0051
Belief x Angle	2	119	1.19	0.3090	0.0001
Belief x Handedness	2	119	1.96	0.1460	0.0001
Reliability x Angle	5	595	1.09	0.3630	0.0002
Reliability x Handedness	5	595	1.84	0.1150	0.0003
Angle x Handedness	1	119	0.09	0.7580	0.0000
Belief x Reliability x Angle	10	595	0.84	0.5810	0.0002
Belief x Reliability x Handedness	10	595	0.67	0.7290	0.0002
Belief x Angle x Handedness	2	119	0.99	0.3760	0.0001
Reliability x Angle x Handedness	5	595	2.15	0.0577	0.0003
Reliability x Belief x Angle x Hand.	10	595	1.27	0.2460	0.0004

Notes. *** $p < 0.001$, * $p < 0.05$; Handedness describes the stimulus', not the participant's handedness.

Stability of cognitive offloading proportion over time

Even though the naïve condition indicates that participants are in principle able to quickly calibrate their external resource use according to the actual reliability, the incongruent condition indicates that false expectations about the knob's reliability can significantly modulate cognitive offloading proportions. To assess the stability of this belief-induced offloading modulation, we conducted an exploratory follow-up analysis that investigated how participants adjusted their external resource use over time. We created a *Time* variable representing the within-block progression in steps of ten trials each (i.e. a value of 1 represents the average of trials 1-10, etc.) and conducted a mixed ANOVA with the within-participants factors *Reliability* and the between-participants factor *Belief*. We used orthogonal polynomial instead of treatment contrasts for the time factor to investigate the nature of changes over time. We did not include further factors in the analysis since those were not balanced within the 10-trial segments.

If participants in the false belief condition indeed adjusted their cognitive offloading proportion over time, *Belief* and *Time* should interact in their influence on external resource use. Though this was the case, the interaction between *Belief* and *Time* was further moderated by *Reliability* (i.e. 3-way interaction *Belief* x *Reliability* x *Time*, $F(30, 2142) = 1.56$, $p = 0.027$, $\eta_G^2 = 0.003$). The polynomial contrasts for *Time* revealed that the linear component ($F(10, 2142) = 3.75$, $p < .0001$), but not the quadratic ($F(10, 2142) = .52$, $p = .879$) or cubic ($F(10, 2142) = .43$, $p = .934$) component interacted with the relationship between *Belief* and *Reliability*. When further inspecting the offloading pattern, Wilcoxon-signed rank tests (Hollander & Wolfe, 1973; the *V* statistic resembles the sum of positive ranks) suggested that participants in the incongruent *Belief* condition adjusted their external resource use between the first ten and the last ten trials (i.e. between Time 1 and Time 4) only for low reliabilities (i.e.; 50%, $V = 110.5$, $p = .099$; 60%, $V = 74.5$, $p = .099$; 70%, $V = 76.5$, $p = .099$), but not for high reliabilities (80%, $V = 107$, $p = .164$; 90%, $V = 135$, $p = .832$; 100%, $V = 107$, $p = .832$). All six p-values are corrected for multiple comparisons using the BH-procedure. Thus, participants with incongruent beliefs appear to partly readjust their offloading behavior over time in low but not in high reliability conditions, an interpretation that is backed by the highly significant linear term of the three-way interaction. The offloading pattern is illustrated in **Figure 4.** The ANOVA results are summarized in the supplementary material, **Table S3**.

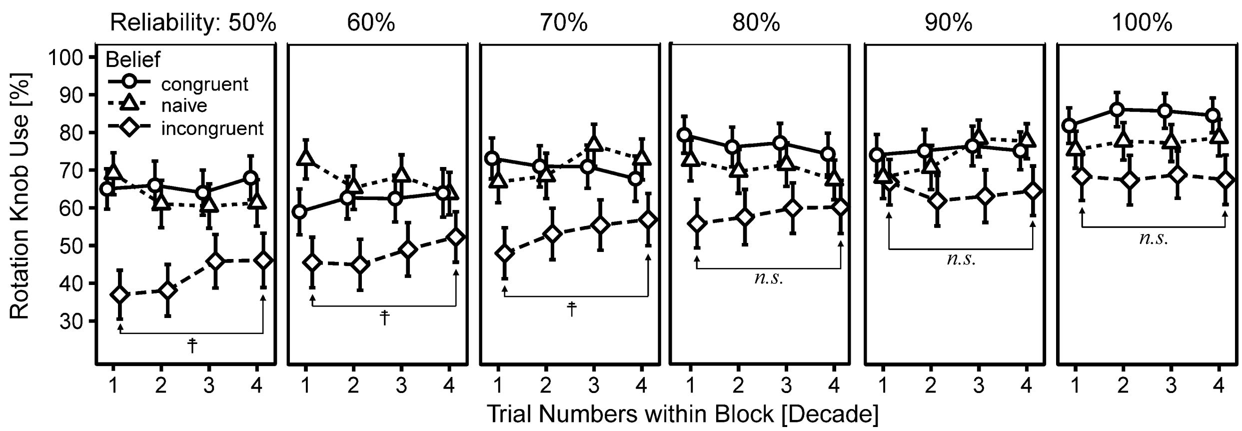

Fig. 4. *Exploration of the stability of false beliefs*: As indicated by post-hoc pairwise comparisons (lines with arrows), for low reliabilities (50%, 60%, 70%), participants with incongruent beliefs seem to converge towards naïve behavior over time whereas for higher reliabilities (80%, 90%, 100%), no such convergence seems to happen. This interpretation is backed by a significant linear component of the three-way interaction between Belief, Reliability, and Time (see text for details). † $p < .1$ after correction for multiple comparisons; n.s. $p > .1$

Knob utility ratings

Metacognitive beliefs regarding the knob's usefulness were analyzed using a 6 x 3 ANOVA with the between-participants factors *Reliability* and *Belief*, respectively. The ANOVA exclusively used the usefulness ratings obtained after the first block. This procedure enabled comparing usefulness ratings of different reliabilities and beliefs simultaneously while rendering *Reliability* a between-participants factor. Since the order in which the different reliability conditions were presented was counter-balanced, the procedure yielded an equal amount of information for the six reliability levels.

We expected the belief manipulation to alter evaluations of the external resource's usefulness. In contrast, the main effect of *Belief* on usefulness evaluations was not significant ($F(2,103) = .63$, $p = .550$, $\eta_G^2 = .012$). However, the effect of *Reliability* was significant ($F(5,103) = 5.10$, $p < .001$, $\eta_G^2 = .199$), with higher usefulness ratings when actual knob reliability was high compared to when it was low; see **Figure 5.** Interestingly, the knot (i.e., the kink in a bilinear function) seen in **Figure 5** occurs at the same reliability that has been identified as 'crossover point' between beneficial and disadvantageous automation (Wickens & Dixon, 2007). Specifically, Wickens and Dixon (2007) found that automation with reliabilities below 70% was, on average, worse than no automation at all. Although we do not argue the 70% reliability knot to be a generalizable characteristic of external resources, such a knot is present in our data as supported by two one-sided post-hoc t-tests (i.e., *60% Reliability* vs. *70% Reliability*, $t = 1.88$, $p = .034$, $M(50\%) = 5.9$, $M(60\%) = 7.3$, and *70%* vs. *80%*, $t = 0.87$, $p = .804$, $M(80\%) = 6.8$). ANOVA results are summarized in **Table 2.** One participant had to be excluded from usefulness rating analyses due to missing data.

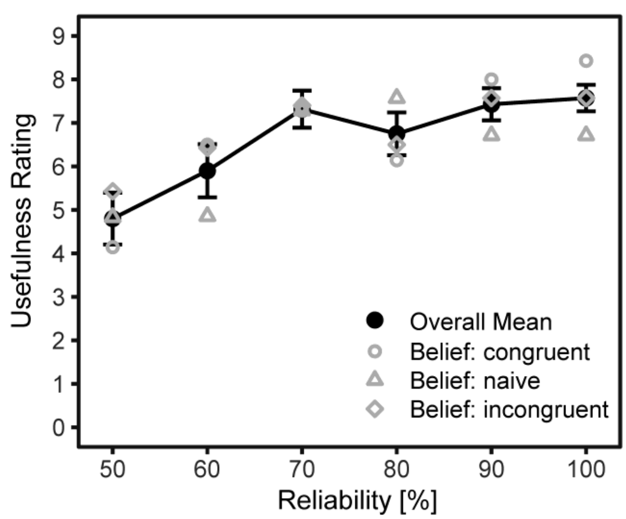

Fig. 5. *External resource Usefulness Evaluation*: Only Reliablity, not Beliefs about reliability altered usefulness evaluations (see Figure 3 for offloading behavior; see Table 2 for ANOVA results). Usefulness was rated on a 10-point scale ranging from 0 to 9. Error bars depict SEM.

Table 2.
ANOVA results for knob usefulness ratings

	DF1	DF2	F	p	η_G^2
Belief	2	103	0.63	0.5304	0.0122
Reliability ***	5	103	5.10	0.0003	0.1986
Belief x Reliability	10	103	0.75	0.6727	0.0682

Notes. *** $p < 0.001$

DISCUSSION

In the current experiment, an adaptation of the mental rotation paradigm (Shepard & Metzler, 1971) was employed to explore how human problem solvers decide when to use external and when to rely on internal resources. We manipulated actual and believed reliability of an external resource, a rotation knob and measured how frequently participants tried to use the knob as well as how useful they perceived the knob to be. Results show that participants were less likely to recruit the external resource when its actual reliability was low (versus high) but also when they believed that the reliability was low (versus high). Whether participants were correctly informed about the reliability of the external resource (i.e., congruent condition) or told that it might sometimes not work properly (i.e., naïve condition) did not differentially affect cognitive offloading, suggesting that participants' reliability assessments based on experience with the system have been well calibrated. Negative beliefs about the resource's reliability (i.e., incongruent condition), however, significantly reduced offloading as compared to the other two conditions, suggesting notable influences of false beliefs on cognitive offloading. The effect of false beliefs was declining over time for lower knob reliabilities but stable for higher knob reliabilities, suggesting at least partial readjustment over time. However, further evidence is needed to make conclusive statements about the effects of false beliefs over time. Lastly, and unexpectedly, explicit assessments of the external resource's usefulness were only affected by actual but not believed reliability, suggesting that reliability and belief manipulations influence offloading through different mechanisms.

The results highlight the importance of higher-level metacognitive judgments in cognitive offloading and thereby confirm the general assumption behind the Metacognitive Model of Cognitive Offloading, which states that "selecting between offloading and relying on internal processes is influenced by metacognitive evaluations of our (internal) mental capacities and the capacities of our extended mental systems encompassing body and world" (Risko & Gilbert, 2016, p. 684). Importantly, the present study demonstrates that induced beliefs about the extended mental system can *cause* sustainable changes in cognitive offloading proportion, even when beliefs are in harsh contrast to reality (i.e., 30% discrepancy between actual and believed reliability), which adds to the correlational findings postulating the influence of metacognitive judgments on cognitive offloading (e.g., Dunn & Risko, 2016; Risko & Dunn, 2015). The results are also well consistent with studies showing that offloading frequency is dependent on the external resource's utility (Gray & Fu, 2004; Gray, Sims, Fu, & Schoelles, 2006; O'Hara & Payne, 1998; Risko et al., 2014; Walsh & Anderson, 2009), which was manipulated via reliability in the present study.

Contrary to our expectations, belief-dependent changes in cognitive offloading proportion were not reflected in the ratings of the knob's usefulness. Though we had no strong hypotheses, we expected the belief manipulation to influence people's explicit theories about knob utility, which should then affect both cognitive offloading and eventually knob usefulness assessments. Such a causal chain would have been in line with what has been termed theory- or information-based judgments in memory research (Koriat, 1997; Koriat & Helstrup, 2007) and well compatible with in the Metacognitive Model of Cognitive Offloading. Also, metacognitive judgments have already been associated with offloading behavior: judgments of internal utility were found to correlate with offloading independently from actual internal utility (Gilbert, 2015; Risko & Dunn, 2015) and judgments of an external resource's utility (i.e., a display from which information had to be retrieved) were correlated with offloading independently from the external resource's actual utility (Dunn & Risko, 2016).

So why is the belief manipulation only affecting knob use, not perceived knob usefulness? We speculate that theory-based metacognitive judgments can influence offloading behavior independently from any ongoing experience-driven monitoring effort (the latter would drive what has been termed experience-based judgments in memory research; Koriat, 1997; Koriat & Helstrup, 2007). While experience might affect offloading via experience-based usefulness evaluations (which can happen without awareness; Cary & Reder, 2002), beliefs might affect offloading differently, without being 'translated' into the utility domain, for example via trust in the external resource and subsequent adjustments in attentional resource allocation. Concordantly, the *Integrated Model of Complacency and Automation Bias* (Parasuraman & Manzey, 2010, Fig. 6) assumes different pathways for person-related parameters (e.g., beliefs) and system-related parameters (e.g., reliability) in influencing attentional resource allocation when interacting with automation, ultimately leading to possibly inefficient distributed processing. Though we deem the knob usefulness ratings interesting enough to report, we want to emphasize that our speculations are based on an exploratory null finding and that further research is needed to disentangle the mechanisms by which theorizing and experiencing affect cognitive offloading.

From an applied perspective, our findings help understand and improve user behavior in tech-infused environments that afford cognitive offloading. It should be kept in mind that cognitive offloading is desirable in some (e.g. when outsourcing memory onto a cockpit; Hutchins, 1995) but not in other (e.g. when overrelying on a vehicle's autopilot; National Transportation Safety Board, 1994; Parasuraman & Riley, 1997) cases. It thus seems critical for users to learn and choose the most beneficial offloading behavior, depending on the system and the particular circumstances. Regarding objective system parameters, the presented data confirms previous findings (Gray & Fu, 2004; Gray, Sims, Fu, & Schoelles, 2006; O'Hara & Payne, 1998; Risko et al., 2014; Walsh & Anderson, 2009), demonstrating that users can automatically extract relevant information (e.g., an external resource's reliability) and adapt cognitive offloading accordingly. In fact, naive participants were so proficient in extracting reliabilities in the present study that their offloading proportion was nearly identical to the one from participants that were correctly informed about the external resource's reliability. Our results thereby confirm that by increasing a user's experience with a system, optimal behavior becomes more likely.

However, merely increasing exposure time is oftentimes not enough to inform optimal behavior. It is crucial *how* that time is being used. In the domain of automated decision aids, it has proven helpful to increase the 'quality' of the time spent with a system by implicitly incentivizing participants to increase monitoring behavior rather than being 'blindly compliant' with the system. This has been, for example, done by varying the external resource's reliability (higher variance leads to increased monitoring; Parasuraman et al., 1993) or exposure to external resource failure during a training session (more failures lead to increased monitoring; Bahner, Hüper, & Manzey, 2008). The present results add another possible intervention to improve offloading behavior: helping participants to form correct beliefs concerning an external resource's performance. Providing performance information and thus altering pre-existing beliefs can help novel users inform their initial offloading choices and experienced but inefficient users to remediate their offloading behavior. Such an approach could not only be useful to remediate erroneous beliefs about an external resource but also erroneous beliefs about internal resources like overconfidence in the own abilities (which correlates with cognitive offloading independently from actual ability; Gilbert, 2015). Whereas experience-based adjustments of cognitive offloading strategies take time, theory-based belief adjustments are fast and would thus be especially useful when exposure to the respective system is short or when the system is too complex to allow extracting its performance parameters via experience.

Although our study provides insights into belief-based interventions that could aid users readjust their cognitive offloading proportion, there is substantial need to carve out the details of such interventions (see also Risko & Gilbert, 2016, p. 685). It would also be useful to increase the understanding of the mechanisms by which belief manipulation affects offloading. In particular, it would be relevant to examine if the effect is mediated by trust in the external resource or changes in attentional resource allocation or monitoring behavior (compare to Parasuraman & Manzey, 2010, Fig. 6). Future efforts also need to clarify if belief manipulations in domains not related to utility have equally strong effects on cognitive offloading, examine if belief manipulations are equally powerful when beliefs are induced outside a highly trustworthy surrounding like a university-based laboratory, and more closely investigate the time-course of induced beliefs' effects on cognitive offloading.

ACKNOWLEDGMENTS

We thank Sharon Guanilo for collecting the data and Alan Anticevic for providing his updates of the Matlab script written by Collin & McMullen (2002) that allowed us to create the Attneave stimuli. The associated R analysis script and data files can be freely accessed online through the Open Science Framework at https://osf.io/98u3r/.

KEY POINTS

- Many everyday environments increasingly allow us to offload our cognitive processing onto digital devices. However, offloading cognitive processing can be both beneficial and detrimental to our overall performance, emphasizing the relevance of an individual's decision whether to solve a certain cognitive task internally or externally.
- We manipulated the actual and believed reliability of a rotation device. Participants were able to calibrate their offloading frequency according to the device's reliability. However, participants also calibrated their offloading frequency according to erroneous beliefs about its reliability.
- The influence of pre-existing beliefs demonstrates a substantial role of metacognitive processes on cognitive offloading decisions, implying opportunities to guide and remediate cognitive offloading behavior.

REFERENCES

Attneave, F., & Arnoult, M. D. (1956). The quantitative study of shape and pattern perception. *Psychological Bulletin*, *53*(6), 452.

Bahner, J. E., Hüper, A.-D., & Manzey, D. (2008). Misuse of automated decision aids: Complacency, automation bias and the impact of training experience. *International Journal of Human-Computer Studies*, *66*(9), 688–699. https://doi.org/10.1016/j.ijhcs.2008.06.001

Bakeman, R. (2005). Recommended effect size statistics for repeated measures designs. *Behavior Research Methods*, *37*(3), 379–384.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, 289–300.

Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision*, *10*, 433–436. https://doi.org/10.1163/156856897x00357

Cary, M., & Reder, L. M. (2002). Metacognition in strategy selection. In *Metacognition* (pp. 63–77). Boston, MA: Springer.

Cochran, W. G., & Cox, G. M. (1950). *Experimental designs*. Oxford, England: Wiley.

Collin, C. A., & McMullen, P. A. (2002). Using Matlab to generate families of similar Attneave shapes. *Behavior Research Methods, Instruments, & Computers*, *34*(1), 55–68.

Dunn, T. L., & Risko, E. F. (2016). Toward a Metacognitive Account of Cognitive Offloading. *Cognitive Science*, *40*(5), 1080–1127. https://doi.org/10.1111/cogs.12273

Gilbert, S. J. (2015). Strategic use of reminders: Influence of both domain-general and task-specific metacognitive confidence, independent of objective memory ability. *Consciousness and Cognition*, *33*, 245–260. https://doi.org/10.1016/j.concog.2015.01.006

Gray, W. D., & Fu, W.-T. (2004). Soft constraints in interactive behavior: the case of ignoring perfect knowledge in-the-world for imperfect knowledge in-the-head. *Cognitive Science*, *28*(3), 359–382. https://doi.org/10.1016/j.cogsci.2003.12.001

Gray, W. D., Sims, C. R., Fu, W.-T., & Schoelles, M. J. (2006). The soft constraints hypothesis: A rational analysis approach to resource allocation for interactive behavior. *Psychological Review*, *113*(3), 461–482. https://doi.org/10.1037/0033-295X.113.3.461

Hollan, J., Hutchins, E., & Kirsh, D. (2000). Distributed cognition: toward a new foundation for human-computer interaction research. *ACM Transactions on Computer-Human Interaction (TOCHI)*, *7*(2), 174–196.

Hollander, M., & Wolfe, D. A. (1973). *Nonparametric statistical methods.* New York, NY: John Wiley and Sons.

Huang, J. Y., Tsai, C.-H., & Huang, S.-T. (2012). The next generation of GPS navigation systems. *Communications of the ACM*, *55*(3), 84–93.

Hutchins, E. (1995). How a Cockpit Remembers Its Speeds. *Cognitive Science*, *19*(3), 265–288. https://doi.org/10.1207/s15516709cog1903_1

Koriat, A. (1997). Monitoring one's own knowledge during study: A cue-utilization approach to judgments of learning. *Journal of Experimental Psychology: General*, *126*(4), 349.

Koriat, A., & Helstrup, T. (2007). Metacognitive aspects of memory. In S. Magnussen & T. Helstrup (Eds.), *Everyday Memory* (1st ed., p. 26). New York, NY: Taylor & Francis.

Melton, A. W. (1947). *Apparatus Tests* (Research Report). ARMY AIR FORCES WASHINGTON D C AVIATION PSYCHOLOGY PROGRAM.

Mueller, S. T., & Piper, B. J. (2014). The psychology experiment building language (PEBL) and PEBL test battery. *Journal of Neuroscience Methods*, *222*, 250–259.

National Transportation Safety Board. (1994). *Stall and loss of control on final approach* (Aircraft accident report No. 6308A).

O'Hara, K. P., & Payne, S. J. (1998). The Effects of Operator Implementation Cost on Planfulness of Problem Solving and Learning. *Cognitive Psychology*, *35*(1), 34–70. https://doi.org/10.1006/cogp.1997.0676

Olejnik, S., & Algina, J. (2003). Generalized eta and omega squared statistics: measures of effect size for some common research designs. *Psychological Methods*, *8*(4), 434.

Parasuraman, R., & Manzey, D. H. (2010). Complacency and Bias in Human Use of Automation: An Attentional Integration. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *52*(3), 381–410. https://doi.org/10.1177/0018720810376055

Parasuraman, R., Molloy, R., & Singh, I. L. (1993). Performance consequences of automation-induced'complacency'. *The International Journal of Aviation Psychology*, *3*(1), 1–23.

Parasuraman, R., & Riley, V. (1997). Humans and automation: Use, misuse, disuse, abuse. *Human Factors*, *39*(2), 230–253.

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*, *10*(4), 437–442. https://doi.org/10.1163/156856897X00366

Risko, E. F., & Dunn, T. L. (2015). Storing information in-the-world: Metacognition and cognitive offloading in a short-term memory task. *Consciousness and Cognition*, *36*, 61–74.

Risko, E. F., & Gilbert, S. J. (2016). Cognitive Offloading. *Trends in Cognitive Sciences*, *20*(9), 676–688. https://doi.org/10.1016/j.tics.2016.07.002

Risko, E. F., Medimorec, S., Chisholm, J., & Kingstone, A. (2014). Rotating with rotated text: a natural behavior approach to investigating cognitive offloading. *Cognitive Science*, *38*(3), 537–564.

Shepard, R. N., & Metzler, J. (1971). Mental rotation of three-dimensional objects. *Science*, *171*(3972), 701–703.

Siegler, R. S., & Lemaire, P. (1997). Older and younger adults' strategy choices in multiplication: Testing predictions of ASCM using the choice/no-choice method. *Journal of Experimental Psychology: General*, *126*(1), 71.

Silver, D., Huang, A., Maddison, C. J., Guez, A., Sifre, L., van den Driessche, G., … Hassabis, D. (2016). Mastering the game of Go with deep neural networks and tree search. *Nature*, *529*(7587), 484–489. https://doi.org/10.1038/nature16961

Storm, B. C., & Stone, S. M. (2015). Saving-Enhanced Memory The Benefits of Saving on the Learning and Remembering of New Information. *Psychological Science*, *26*(2), 182–188. https://doi.org/10.1177/0956797614559285

Walsh, M. M., & Anderson, J. R. (2009). The strategic nature of changing your mind. *Cognitive Psychology*, *58*(3), 416–440.

Ward, A. F., Duke, K., Gneezy, A., & Bos, M. W. (2017). Brain Drain: The Mere Presence of One's Own Smartphone Reduces Available Cognitive Capacity. *Journal of the Association for Consumer Research*, *2*(2), 140–154. https://doi.org/10.1086/691462

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

BIOGRAPHIES

Patrick Weis is a PhD student in the Human Factors and Applied Cognition Program at George Mason University. He received an MS in Neuroscience from the University of Tuebingen in 2014.

Eva Wiese is an Assistant Professor in the Human Factors and Applied Cognition Program at George Mason University. She received her PhD in Neuroscience from the Ludwig-Maximilian University of Munich, Germany in 2013.