THINKING OUTSIDE THE BOX: WHEN DO HUMAN PROBLEM SOLVERS OFFLOAD COGNITION?

by

Patrick P. Weis

A Dissertation
Submitted to the
Graduate Faculty
of
George Mason University
in Partial Fulfillment of
The Requirements for the Degree
of
Doctor of Philosophy
Psychology

Committee:

____ Director

____

____

____ Department Chairperson

____ Program Director

____ Dean, College of Humanities and Social Sciences

Date: ___________________________________ Fall Semester 2019
George Mason University
Fairfax, VA

Thinking Outside the Box: When do Human Problem Solvers Offload Cognition?

A Dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy at George Mason University

by

Patrick P. Weis
Master of Science
Eberhard-Karls-University Tuebingen, 2014
Bachelor of Science
Otto-von-Guericke-University Magdeburg, 2012

Director: Eva Wiese, Professor
Department of Psychology

Fall Semester 2019
George Mason University
Fairfax, VA

Copyright 2019 Patrick P. Weis
All Rights Reserved

ACKNOWLEDGEMENTS

I am grateful to have had many humble beings supporting my academic endeavors either directly or using some—most valuable—detours. The most direct support was given throughout numerous meetings at numerous places accompanied by a wide variety of caffeinated or alcoholinated drinks with either Eva Wiese, Aziz Abubshait, or both. While the former support was frequently aimed at improving scientific quality, the latter support was frequently aimed at decreasing the PhD's daily sufferings. Another valuable source of support—that undoubtfully raised the quality of my research—was the thriving R and statistics community at GMU: the MRES group led by Dr. Patrick McKnight including the R missionaries Jake Quartuccio, Sam Monfort, and Aziz Abubshait as well as the fantastic stats practitioners and teachers Dr. José Cortina and Dr. Jeffrey Stuewig. I am also most grateful for the constant support and valuable brainstorming meetings with Dr. Craig McDonald. Another source of direct support were the undergraduate students I was able to work with and who were the backbone of the data collection behind many of my dissertation's studies: Chanelle Roegler, Sahar Miller, Antonn Chasse, Sharon Guanilo-Cueva, and Caitlin Mahoney. The more indirect support was given by my former academic advisors Dr. Toemme Noesselt and Dr. Cornelia Herbert, who raised me to where I was at the beginning of my PhD. It was also given by Susan Ridley who always had answers to the administrative issues I ran into at GMU—and there were plenty! I also need to mention my Dad, who likely sparked my scientific interest and might thus be the cause for all my academic pleasures and miseries. Lastly—and most importantly, since I would not have endeavored without it—I am grateful for the consistent emotional support from a wide variety of family members and friends, including but not limited to my Mum, my Uncle Peter, my Aunt Ruth, my former partner Raissa, my current partner Jana, my friends Jenny, Bing, Sutirtha, Tobias, Christopher, Felix, Matthias, Grazyna, Johanna, Donnie, Katja, and Daniel, my colleagues and friends Ali, Stephanie, Noushin, Nicholas, Molly, and Spencer, and my most caring landlady Pat with her partner Jeff.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF EQUATIONS

ABSTRACT

THINKING OUTSIDE THE BOX: WHEN DO HUMAN PROBLEM SOLVERS OFFLOAD COGNITION?

Patrick P. Weis, Ph.D.

George Mason University, 2019

Dissertation Director: Dr. Eva Wiese

The common main objectives across the three studies presented in the present dissertation lie in improving the understanding of parameters that influence a human problem solver's decision to use environment-based external instead of brain-based internal resources. In Study 1, it was shown that a human problem solver's inclination to offload cognition depends on both monitoring the external resource's actual performance as well as on pre-existing beliefs about the external resource's performability, indicating comparable importance of both factors. In Study 2, it was shown that a human problem solver's inclination to offload cognition also depends on the performability of the internal brain-based resources that are relevant for the task at hand. Good internal performability decreased cognitive offloading frequency even when the task at hand looked comparably difficult, thereby suggesting that internal performability can supersede task appearance in determining offloading propensity. Both Study 1 and Study 2 showed the human problem

solver’s sensitivity to performance parameters for determining cognitive offloading frequency. In Study 3, this sensitivity to performance parameters was investigated further. Specifically, it was investigated whether human problem solvers are able to choose between internal and external processing in a way that serves their current performance goals—in Study 3 accuracy or speed—or whether they prefer different heuristics like minimizing mental effort or maximizing certain performance metrics like speed irrespectively of the current goal. Results of Study 3 confirmed the former, indicating the human problem solver’s capability of choosing between internal and external resources to achieve current performance goals.

INTRODUCTION

Why is Patrick always using the navigation app on his smartphone while Peter is relying on the navigational capabilities of his brain? In my dissertation, I employ experimental paradigms in laboratory settings to investigate people's c*ognitive offloading* behavior. Cognitive offloading describes the use of external resources beyond our skulls to support cognitive tasks like spatial navigation, arithmetic, or memory. My main objective is to help identifying key parameters that influence our decision to use external resources instead of brain-based internal resources. That way, I hope to increase the understanding of human-technology-interaction, ultimately contributing to a healthier and more humane use of technology. The prominence of Nicholas Carr's essay *Is Google making us stupid?* (Carr, 2008) illustrates both societal importance and the unresolved nature of my dissertation's topic. As for now, we are lacking guidelines for both humane design and healthy habits when including external resources into our cognitive operations. Exploring and understanding the reasons behind people's decision to use external resources provides a first step towards better design principles and interventions to remediate technology use.

Cognitive offloading – How we use the environment to help us think

Undoubtedly, the brain plays a crucial role in our ability to cognize. It affords information storage and retrieval. For example, it enables us to store our neighbors' names and to retrieve them whenever we see them. It affords the creation of meaningful sentences, making plans for our future, internal simulations like mental rotation, and much more. Still: without technological support, many humans struggle to solve cognitive tasks involving arithmetic (Osiurak, Navarro, Reynaud, & Thomas, 2018; Matthew M. Walsh & Anderson, 2009), spatial navigation (Fenech, Drews, & Bakdash, 2010), or prospective memory (Cherkaoui & Gilbert, 2017; Gilbert, 2015) efficiently. Thus, if one is aiming to understand how people cognize in their everyday environments, looking at their brains is not enough. It is imperative to look at their bodies, their environments, and their interactions with their environments as well (Clark, 2004; Clark & Chalmers, 1998; Risko & Gilbert, 2016; Scaife & Rogers, 1996; Wilson, 2002). The benefit of such a systemic perspective is captured by the concept of an *epistemic action* (Kirsh & Maglio, 1994). An epistemic action is an action undertaken to advance a cognitive process. An epistemic action is contrasted by an action that is undertaken to alter the structure of the physical environment. For example, reordering Scrabble tiles is thought to support the cognitive process of creating meaningful words out of the letters depicted on the tiles (Maglio, Matlock, Raphaely, Chernicky, & Kirsh, 1999), an action quite distinct from putting tiles onto a garden path to afford comfortable walking. Without a systemic perspective that expands beyond the skull, one would struggle to understand human cognizing in the Scrabble case as well as in much of everyday problem solving.

Investigating human cognition from a holistic perspective that includes brain, body, and environment is nothing particularly novel. Far earlier than the current debates about a holistic focus in cognitive science (Clark, 1999), in his *Activity Theory*, Leont'ev (1981) already took a holistic approach when trying to capture an *activity* with all of its relevant aspects, thus trying to unify thought, action, and environment. In particular, he was interested in the "integrated, goal-oriented configuration of internal and external resources" (p. 50), which he called a *functional organ* [1], when solving cognitive problems. Such configurations could for example encompass a blind man's stick that helps him navigating to a friend's house or a notebook that supports working memory when solving a math equation[2]. However, though conceptually promising, researching functional organs has been and still currently is challenging. Cognitive science research traditionally used to restrict the kind of human-environment interactions necessary for researching functional organs - partly because confounds are harder to control in such a setting - and favored simpler non-interactive paradigms instead (Risko & Gilbert, 2016).

As of today, several research streams of cognitive science, i.e. *Embodied Cognition* (Wilson, 2002), *Situated Cognition* (Kirsh, 2009), *Distributed Cognition* (Hollan, Hutchins, & Kirsh, 2000), and *Extended Cognition* (Clark & Chalmers, 1998)

[1] Originally, the term was introduced by Hegel and Marx (Nardi, 1996). From Marx' perspective, a fundamental goal of any individual is to create the functional organs needed for individual and societal growth.

[2] As a side remark, it might be of interest to the reader that activity theory claims external resource use to have consequences far beyond immediate performance. It is claimed that external resources can have aesthetic and moral qualities as well, as can be seen in the following anecdote (Nardi, 1996, p. 295): a musician was upset about being painted without his Cello because he conceived him and the Cello as one entity, feeling incomplete without his Cello. The musician perceived the Cello, an external musical resource, as a "continuation of the human soul", which could possibly be true for other cognition-related external resources as well. Activity Theory thus postulates that symbols like words but also tools like computers are a crucial part of an activity and that they not only influence what we do but also how we do it, and ultimately also who we are. Thus, to fully answer this dissertation's main question about the reasons behind external resource use, one might need to consider identity-related consequences as well.

adopted this holistic perspective, each with a slightly different agenda. Traditionally, Activity Theory has been quite dissimilar to Cognitive Science. Activity Theory's focus is on, as might be expected, activities, i.e. on doing things. In contrast, cognitive science's focus has long been exclusively on sequential mental information processing and neural representations (Shapiro, 2010). When meeting a cognitive scientist, an activity theorist might well have called him or her a reductionist, accusing him of focusing too narrowly on sequential brain-based information processing when trying to explain and predict human behavior. With the current developments in Embodied, Situated, Distributed, and Extended Cognition, such accusations are becoming increasingly void. In the following section, these developments, specifically focusing on the determinants of cognitive offloading, are summarized.

Determinants of cognitive offloading

This section introduces theoretical considerations and empirical evidence concerning a cognizer's decision to include the environment into a thought process rather than to exclusively rely on brain-based resources. The focus is on aspects currently debated in cognitive science outlets.

Optimizing performance

Being a well-known proponent of performance optimizing theories, the *Soft Constraints Theory* (Gray, Sims, Fu, & Schoelles, 2006) postulates that people always strive to maximize speed when having to decide between internal and external processing. Following that rationale, if presented with an arithmetic task, one is supposed to use a calculator only if it is faster than solving the task with mental arithmetic. The

Soft Constraints Theory focuses on speed as performance metric whereas other approaches could include metrics like accuracy. The common denominator of performance-based theories is that brain-based and environment-based resources are seen as interchangeable. There is no innate preference for either of them, a principle that has been termed *cognitive impartiality* (for a review, see Risko & Gilbert, 2016). Preferences for an internal or external processing mode only emerge if there are performance differences, turning the favor towards the more efficient processing mode. There is profound evidence that, in tightly controlled settings, participants are indeed using external resources to optimize or at least improve performance (Dunn & Risko, 2016: Experiment 5; Gray et al., 2006; Risko, Medimorec, Chisholm, & Kingstone, 2014; Matthew M. Walsh & Anderson, 2009). It has also been shown that people with lower cognitive skills tend to use smartphone-based search engines more frequently than people with higher cognitive skills, which would support the adaptive use of external resources on a societal level (Barr, Pennycook, Stolz, & Fugelsang, 2015). However, it should be noted that performance optimization can be tricky if tasks are extensive in time or cognitive demand, in which case simple but imperfect heuristics might be used in lieu of performance optimization strategies (see e.g., *local optima* in Fu & Gray, 2006, and the concept *bounds of rationality* Simon, 1956). Also, as will be mentioned in more detail in the following, performance optimization might not always be the first choice in complex environments. Performance optimization is unlikely to be the sole reason to recruit external resources for support thought.

Unburden the brain

A direct competitor to the principle of performance optimization would be if people strive to free as many neural resources as possible, thus offloading cognitive processing onto the environment whenever possible. Intuitively, such a preference for environment-based processing might sound plausible because mental resources might induce a substantial amount of metabolic costs and can interfere with other, possibly more relevant, brain-based cognitive processing (Storm & Stone, 2015). Originally, such a preference for external resources has been proposed in the memory domain by Ballard *et al.* (1997) and been termed *Minimal Memory Theory* since (e.g., in Gray *et al.*, 2006). It should however be noted that in the original paper (Ballard et al., 1997), it was not proposed that people strive to unburden the brain at all costs: Ballard *et al*. (1997) already showed that after increasing the accessibility costs for the external resource, it was used less frequently. In other words, the more time-consuming it was to access an external resource, the more people preferred relying on their brain-based memory. Analogous observations have been made by Gray *et al*. (2006). More generally however, it has been observed that people, all else being equal, prefer less effortful cognitive strategies over more effortful alternative strategies in a wide range of behavioral decision tasks (Kool, McGuire, Rosen, & Botvinick, 2010).

Whether in accordance with the original manuscript by Ballard *et al*. (1997) or not, the core of Minimal Memory Theory, as the concept is used today (e.g. in Gray *et al.*, 2006), is a bias favoring environment-based processing over brain-based cognizing. Thus, Minimal Memory Theory is directly conflicting with the cognitive impartiality assumption underlying the Soft Constraints Theory. However, empirical findings currently question the existence of a bias favoring external resources based on mental effort, favoring an alternative explanation for the bias: people might engage in metacognitive evaluations of internal and external strategies and oftentimes misjudge their internal abilities or the usefulness of the external resource, which mediates the biased use of external strategies (Gilbert et al., 2019). This influence of metacognitive influence on cognitive offloading is explored in the next section.

Follow metacognitive evaluations

Recently, researchers observed external resource use that was not consistent with performance optimization (Dunn & Risko, 2016; Gilbert et al., 2019; Risko & Dunn, 2015). For example, participants frequently[3] chose to exclusively rely on internal resources when asked to remember an array consisting of ten letters despite extremely bad internal performance (Risko & Dunn, 2015). Analogously, participants frequently chose to rely on an external resource – pen and paper – when asked to remember an array consisting of two letters despite extremely good internal performance. Both findings illustrate that participants frequently behaved inefficiently. In a follow-up experiment, the arrays of letters were presented once more and participants were to rate expected accuracy and effort when remembering the letters either internally or with support of pen and paper (Risko & Dunn, 2015: Experiment 2). It was found that participants deemed tasks with two letters to be harder to solve internally than was actually the case, which

[3] i.e., in more than 10% of all trials, despite an accuracy of close to 0 when exclusively relying on internal ressources

possibly was the reason for the higher offloading rate reported in the original experiment (Risko & Dunn, 2015: Experiment 1). Similarly, participants massively overestimated how well they can handle ten letters internally. While this follow-up experiment only provides correlational rather than causal evidence, it nevertheless lets is appear likely for metacognitive evaluations to influence offloading. Further correlational evidence was provided by Dunn & Risko (2016), which lead to the proposal of the *Metacognitive Theory* of cognitive offloading. Additionally, Gilbert (2015) showed that metacognitive confidence seems to influence offloading proportion independently from actual performance. However, it should be noted that a biased metacognitive evaluation (e.g., consciously evaluating that using a calculator is more efficient than relying on the brain) does not necessarily translate into behavior, as observed by Virgo *et al.* (Jérémy Virgo, Jonathan Pillon, Jordan Navarro, Emanuelle Reynaud, & François Osiurak, 2017).

Experimental[4] evidence for the Metacognitive Theory was provided by Wiese *et al.* (Wiese, Wykowska, & Müller, 2014). In their study, participants had to identify target letters on the left or the right side of a computer screen. In the middle of the screen, a human face was either gazing towards the target or gazing away from the target. When the gaze of that face matched the side of the target letter, it is known that participants are quicker at identifying the target (*gaze cueing effect*, for a review see Frischen, 2007). The

[4] Most studies used correlational designs to investigate the impact of metacognitive evaluations on cognitive offloading as in the study described in the preceding paragraph ("It was found that participants deemed tasks with two letters to be harder to solve internally than was actually the case, possibly leading to a higher offloading rate"; Risko & Dunn, 2015).

rationale behind is that in everyday social interactions, we can *offload* attentional mechanisms onto other people. Instead of hyper-vigilantly screening an environment, we can rely on the attentive mechanisms of the people surrounding us by observing their gaze. Interestingly, Wiese *et al.* (2014) showed that this outsourcing is less pronounced (i.e., higher gaze cueing effect) when people believe that the gazer is unreliable. This decreasing reliance on an agent's gaze thus represents a metacognitive bias that is independent of the performance benefit obtained by following the gaze.

Follow cultural knowledge

External resources are also used when indicated by our cultural embedding. The underlying assumption is that the rules governing external resource use can be cultivated (Hutchins, 2014). For example, language use is culturally transmitted, languages play an important part in a human's cognitive toolkit (for a review, see Clark, 1998), and using written words wisely is a prime example of cognitive offloading. That way, cultural context not only influences who we are but also how we offload. It is likely that the influence of cultural knowledge on cognitive offloading is mediated by metacognitive evaluations (see section *Determinants of cognitive offloading: Follow metacognitive evaluations*).

Augmenting cognition beyond internal ability

External resources enable cognitive operations that are more complex than what humans can achieve internally (e.g., using pen and paper calculations and external representations to build a model of planetary movements based on gravitational forces), have different properties than internal resources (e.g., paper-based memory does not change or is forgotten over time), and can be used for inter-individual cognitive efforts (e.g., using an external representation like a graph to solve a problem in joint effort; Kirsh, 1995, 2010, 2013). The idea that external resources can augment cognitive abilities rather than only shift the locus of cognitive processing has been termed the *complementarity principle* (Sutton, 2010) and considerably increases the difficulty of interpreting cognitive offloading behavior as it introduces additional possible reasons to use an external resource that might result in the same overt behavior as performance minimization or reliance on metacognitive evaluations would.

However, it should be noted that using external resources does not necessarily raise abilities of people with poor internal skills to a level on par with people with strong internal skills. People with poor internal cognitive ability in a respective domain might fail to use an external resource efficiently or simply benefit less than people with high internal ability (Cherkaoui & Gilbert, 2017; Osiurak et al., 2018).

Rationale of the dissertation project

In the current PhD project, the main objective lies in improving the understanding of parameters that influence a human problem solver's decision to use external resources instead of using brain-based internal resources. Several of these determinants are currently debated in the scientific literature (see section *Determinants of cognitive offloading*) but a coherent picture is yet to emerge. There is a broad agreement that the comprehension of how humans navigate cognitive environments is only at the beginning (compare to *Outstanding Questions* in Risko & Gilbert, 2016, p 685; Anderson, 1990; Marewski & Schooler, 2011; Scaife & Rogers, 1996; Kirsh, 2013) and that improving our comprehension will be highly rewarding. The latter is exemplified in Kirsh's (2013) prediction of a "magical future" of human-computer-interaction.

Such a magical future is of course dependent on progress in research areas beyond the determinants of cognitive offloading. For example, long-term effects of cognitive offloading are largely unknown (Risko & Gilbert, 2016), individual differences in cognitive offloading poorly understood (Risko & Gilbert, 2016), it is unclear how closely external resources can be included into action repertoires and cognitive routines and how such incorporation differs from incorporating internal resources (Kirsh, 2013), and the mechanisms by which external representations like graphs can support and scaffold internal thought are hardly understood (Scaife & Rogers, 1996)[5], just to name a few. So why is the focus of the current project set on the determinants of cognitive offloading rather than on one of the other topics mentioned in the preceding paragraph? The answer is twofold. Firstly, I deem the question about the determinants of cognitive offloading to be of relevance for most other outstanding questions. For example, long-term consequences might be quite different for situations in which the human problem solver is actively engaged in monitoring the external resources as to maximize overall performance in comparison to situations in which the human problem solver is mentally unengaged, using the external resource simply to unburden the brain. In other words, long-term consequences might depend on the reasons the human problem solver engaged in cognitive offloading in the first place. Secondly, interventions to in- or decrease an individual's propensity to engage in cognitive offloading are hardly available (Risko &

[5] Though some progress has been made since 1996 (see Kirsh, 2010).

Gilbert, 2016) even though such guidance has the potential to increase the individual's performance (Gilbert et al., 2019). A better understanding of the determinants of cognitive offloading would allow designing interventions that leverage this knowledge. For example, Gilbert *et al.* (2019) presented metacognitive information ("*According to your performance so far, we have calculated that you will probably score more points if you choose to perform with/without reminders*") to alter their participants propensity to offload prospective memory[6]. Similar interventions could be possible after cognitive offloading research has yielded more knowledge about determinants beyond metacognitive information.

So far, it has been argued that cognitive offloading is an activity of societal importance, that it is still rather poorly understood, and that specifically furthering the understanding of the determinants of cognitive offloading has considerable potential for societal welfare. In the remainder of this section, two specifics of how the understanding of the determinants of cognitive offloading could be advanced will be carved out:

(1) Creating new experimental and interactive paradigms. As noted by Risko and Gilbert (2016, p. 686), expanding the focus from brain-based to include environment-based cognitive processing requires novel methodological approaches: *"[...] investigating cognitive offloading often requires allowing research participants to [...] manipulate and interact with their environment.*

[6] Interestingly, participants followed the advice in the overwhelming majority of trials. If the main determinant of cognitive offloading was to unburden the brain, participants should not have switched to an internal processing mode just because metacognitive information suggested improved performance.

Methods in cognitive science, however, have traditionally been designed to restrict this type of natural behavior. […] Thus, understanding cognitive offloading will require an expansion of the cognitive scientist's methodological toolbox.". One such methodological approach is to increase the breadth of experimental paradigms available to research cognitive offloading. This approach comes along with two opportunities. First, it allows causal rather than correlational inferences, which is of importance given that the currently available evidence for some of the determinants is based on correlational evidence (e.g., the evidence for the influence of metacognitive information; see section *Determinants of cognitive offloading: Follow metacognitive evaluations*). Second, the approach allows investigating the external validity of already established determinants, i.e. it allows investigating whether existing cognitive offloading paradigms are representative. For example, it is unclear whether the same determinants hold in settings where the own body is used to offload cognition (e.g., turning the head to read a rotated paragraph rather than mentally rotating the paragraph; Risko et al., 2014) in comparison to settings where a computer is used to offload cognition. Likewise, it is unclear whether determinants are stable across different cognitive tasks, e.g. whether the same determinants hold for offloading arithmetic tasks as for offloading navigational tasks. Establishing new experimental and interactive paradigms thus expands the cognitive scientist's toolbox while allowing for causal inference and investigating external validity of established determinants.

(2) **Gauging how different determinants of cognitive offloading interact.** Although there is substantial evidence for a wide range of determinants (see section *Determinants of cognitive offloading*), it is not clear how to judge the likely contributions of each determinant in a specific setting with specific internal and external resources available. For example, it is known that people can use external resources to improve performance beyond internal capabilities (see section *Determinants of cognitive offloading: Optimizing performance*). However, it is not known how relevant performance optimization is if such optimization simultaneously required considerable cognitive effort and is in conflict with cultural knowledge and other metacognitive information. To understand how humans offload cognition in certain scenarios, it is imperative to know the importance of different determinants and to know if some determinants supplant others.

In the present manuscript, three studies that have been conducted are summarized (section *Conducted research*) and three other possible lines of research are suggested as future directions (section *General discussion: Future directions*). All conducted studies are catering to both aspects *(1)* and *(2)* that have been described in the above. For each study, before going into its specific details, the study's relevance for the dissertation-specific questions is explicated and its main results are embedded into the context of the dissertation as a whole (sections *Rationale* for Study 1, 2, and 3).

CONDUCTED STUDIES

Three studies have been conducted to research the influence of 1) metacognitive information 2) performability of internal resources and 3) performance goals on cognitive offloading. The studies will be described in what follows and be embedded into the dissertation in the *Rationale* subsections.

Study 1

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

Author contributions: EW, PPW conceived and designed research. PPW performed research. PPW analyzed data. EW, PPW wrote the paper.

Weis, P. P., & Wiese, E. (2019). Using Tools to Help Us Think: Actual But also Believed Reliability Modulates Cognitive Offloading. *Human Factors*, 61(2), 243–254. doi.org/10.1177/0018720818797553

© 2018, Human Factors and Ergonomics Society. This paper is not the copy of record and may not exactly replicate the final, authoritative version of the article. Please do not copy or cite without authors' permission. The final article is available via its DOI: 10.1177/0018720818797553

Rationale

In this study, we altered the mental rotation paradigm (Shepard & Metzler, 1971) in a way that allowed participants to offload cognitive processing (see **Figure 1**). In the altered paradigm, participants could 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: On the one hand the knob's actual reliability (AR) and on the other hand an instruction altering participants' beliefs about the knob's reliability which supposedly manipulated the participants' metacognitive evaluations concerning the external resource (believed reliability; BR). Cognitive offloading proportion and perceived knob utility were measured. The main focus of the study was to use the BR manipulation to experimentally validate correlational findings (e.g., Dunn & Risko, 2016; Risko & Dunn, 2015) concerning the influence of metacognitive judgments on cognitive offloading (compare *(1)* in the section *Rationale of the dissertation project*). A second focus was on exploring whether AR contributes to cognitive offloading independently of BR, thereby increasing the understanding of how different determinants synergize to influence cognitive offloading (compare *(2)* in the section *Rationale of the dissertation project*).

In a nutshell, results showed that participants were able to quickly and dynamically adjust their cognitive offloading proportion in response to AR independently of BR, suggesting a high level of cognitive 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. Thus, how frequently human problem solvers offload their cognition is based on both AR-related performance monitoring as well as on possibly erroneous pre-existing beliefs. These results (a) confirm correlational findings regarding the importance of metacognitive evaluations when using external resources with an experimental paradigm and (b) suggest a comparable contribution of both performance monitoring and metacognitive information in determining cognitive offloading behavior.

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.

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 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 offloaded cognitive resources unnecessarily often when internal processing was efficient (i.e., two letters), and did not fully make use of external resources when they were 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 currently available 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* (Dunn & Risko, 2016, 2016; 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. 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 and 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 all trials), resulting in a final sample size of 122 (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 created 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**).

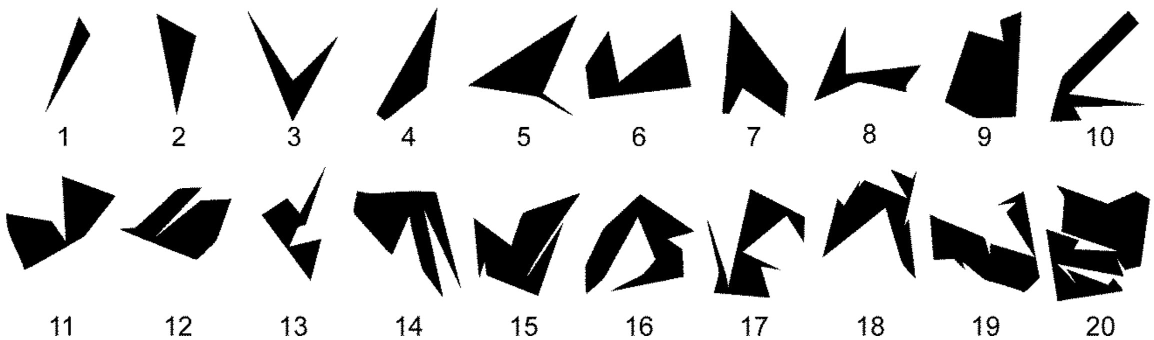


Figure 1 Stimuli used for the extended rotation task. The twenty stimuli were created using the Attneave procedure (see Stimuli).

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 right and a working stimulus on the left 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.

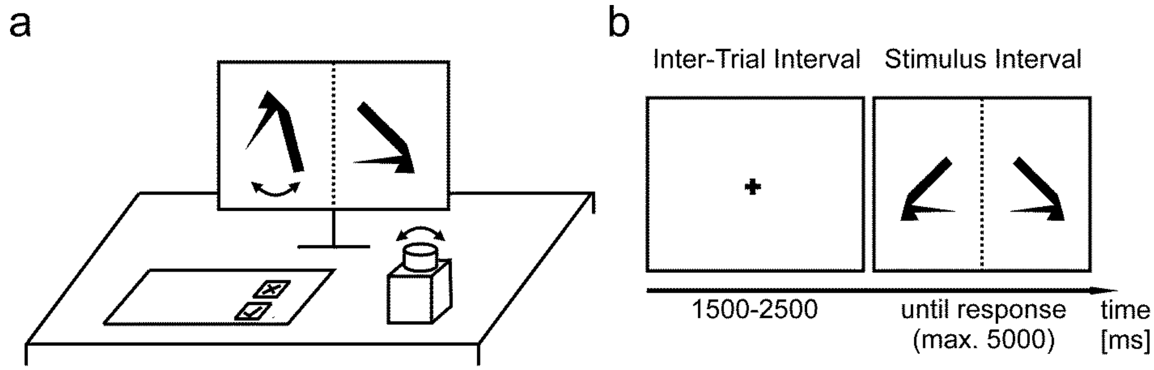


Figure 2 Extended rotation paradigm. (a) The experimental set-up contained a computer screen, a standard keyboard, and a rotation knob. (b) Participant's task was to determine whether the base stimulus has the same handedness as the working stimulus. Participants could solve the task by mentally rotating 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 external resource's utility, 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). P-values are reported Greenhouse-Geisser-corrected where applicable.

Results

Performance

Neither reaction time ($F(2, 119) = 1.49$, $p = .229$, $\eta_G^2 = .016$) nor accuracy ($F(2, 119) = .12$, $p = .883$, $\eta_G^2 = .001$) differed between belief conditions, suggesting comparable overall performance across groups. The ANOVA results are summarized in the Supplemental Material, **Table 3 and 4**.

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* ($F(5, 595) = 23.69$, $p < .001$, $\eta_G^2 = .042$), and *Beliefs* regarding the knob's reliability ($F(2, 119) = 3.49$, $p = .034$, $\eta_G^2 = .035$) had a significant impact on the extent to which participants used the rotation knob (i.e., cognitive offloading proportion). The *Reliability* x *Belief* interaction did not reach the level of significance ($F(10, 595) = 1.64$, $p = .115$, $\eta_G^2 = .005$). As expected, but of minor interest for the purposes of this study, *Angle* ($F(1, 119) = 71.62$, $p < .001$, $\eta_G^2 = .004$, $M(60°) =$

64.3%, *M*(120°) = 68.6%) and *Handedness* ($F(1, 119) = 5.85$, $p = .017$, $\eta_G^2 = .0002$, *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 $\eta_G^2 < .006$, **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*(incongruent) = 55.71) or when given congruent information (i.e. *congruent* vs. *incongruent*, $W = 1051.5$, $p = .036$) about the external resource's reliability. Thus, correct utility beliefs, in contrast to incorrect utility beliefs, had no influence on cognitive offloading proportion. All p-values for the post-hoc tests were corrected for multiple comparisons using the Bonferroni-Hochberg method (BH; Benjamini & Hochberg, 1995).

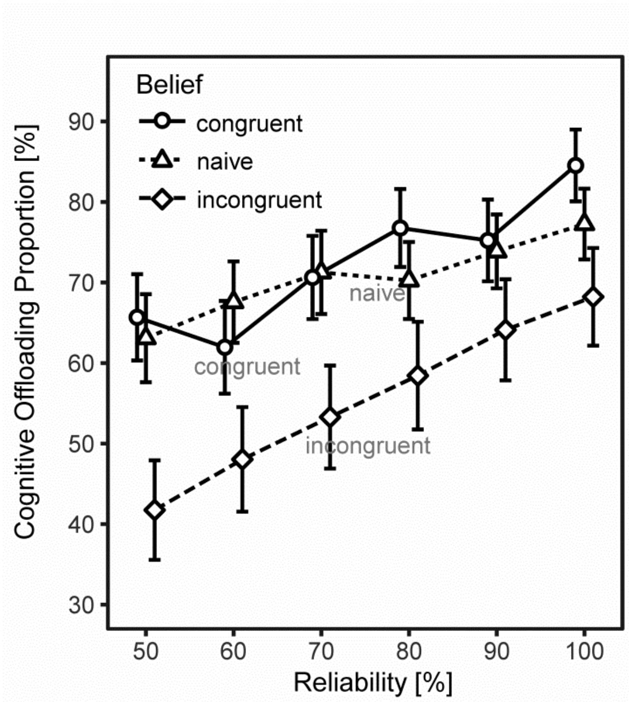


Figure 3 Cognitive offloading proportion as a function of actual and believed reliability. Participant's cognitive offloading behavior depends on both actual (x-axis) and believed (line types) reliabilities. Error bars depict SEM.

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 5**.

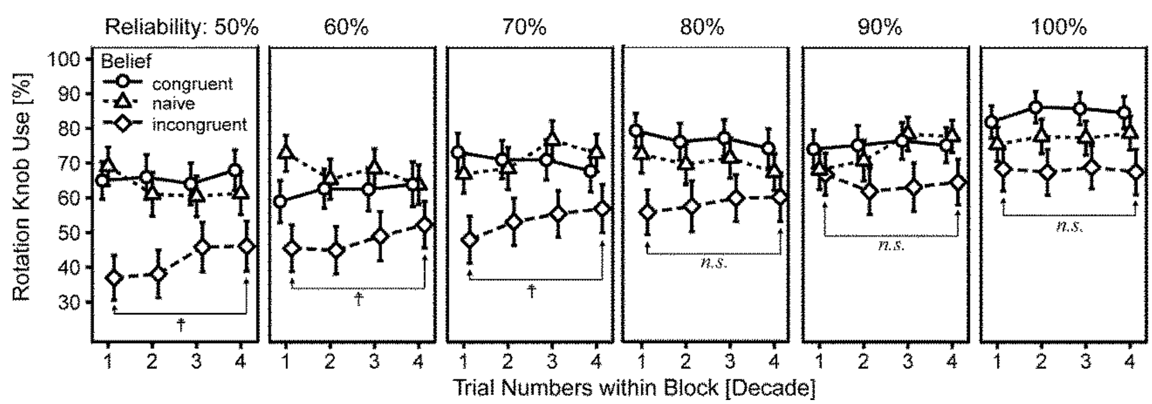


Figure 4 Exploration of the stability of false beliefs**.** As indicated by post-hoc pairwise comparisons (lines with arrows), for low reliabilities (50%, 60%, 70%), participants with incongruent beliefs seem to converge towards naïve behavior over time whereas for higher reliabilities (80%, 90%, 100%), 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 (i.e., after 40 trials). This procedure enabled comparing usefulness ratings of different reliabilities and beliefs simultaneously, statistically 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 (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 7*0%* vs. 8*0%*, 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.

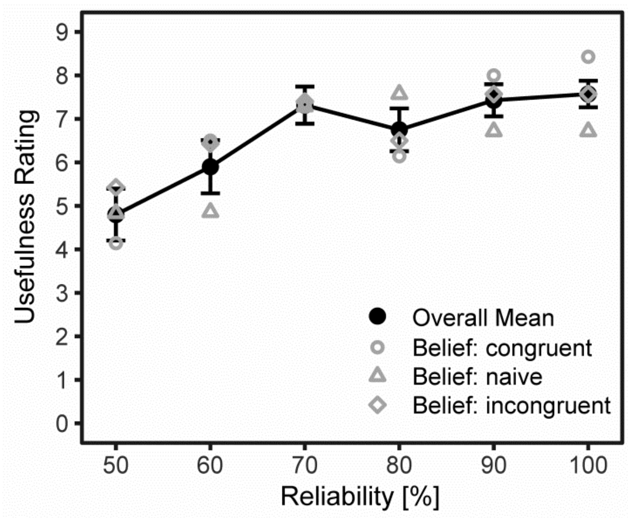


Figure 5 External resource usefulness evaluation. Only Reliablity, not Beliefs about reliability altered usefulness evaluations (see **Figure 3** for offloading behavior; see **Table 2** for ANOVA results). Usefulness was rated on a 10-point scale ranging from 0 to 9. Error bars depict SEM.

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 indicate 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 external 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 would the belief manipulation only affect 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, Figure 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 cases (e.g., when outsourcing memory onto a cockpit; Hutchins, 1995) but not in others (e.g., when overrelying on a vehicle's autopilot; National Transportation Safety Board, 1994; Parasuraman & Riley, 1997). 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 ex perience 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 their 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, Figure 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

Supplementary material

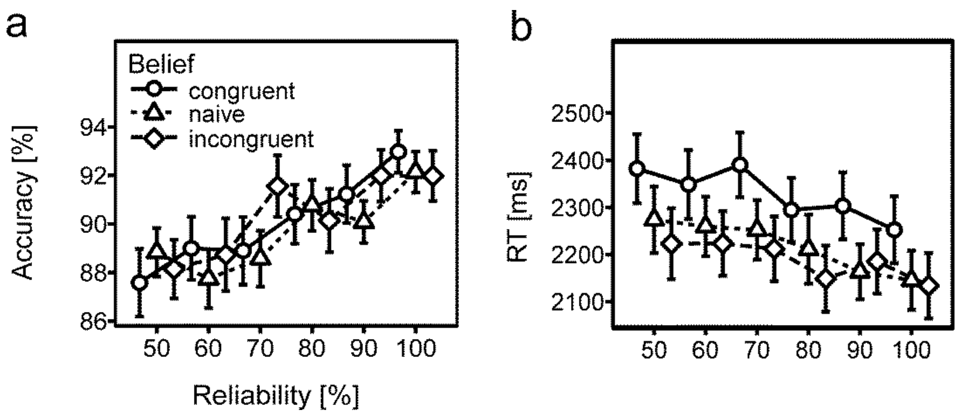


Figure 6 Performance data. Accuracy (a) and reaction time (b) data are displayed for exploratory purposes only. Error bars depict SEM.

Table 3 ANOVA results for accuracy

	DF1	DF2	F	p	η_G^2
Belief	2	119	0.12	0.8830	0.0005
Reliability ***	5	595	9.42	< 0.0001	0.0157
Angle *	1	119	4.69	0.0323	0.0016
Handedness	1	119	0.05	0.8200	0.0000
Belief x Reliability	10	595	1.16	0.3130	0.0039
Belief x Angle	2	119	0.54	0.5840	0.0004
Belief x Handedness	2	119	0.24	0.7830	0.0003
Reliability x Angle	5	595	1.73	0.1260	0.0019
Reliability x Handedness	5	595	2.73	0.0188	0.0033
Angle x Handedness	1	119	22.84	< 0.0001	0.0059
Belief x Reliability x Angle	10	595	1.27	0.2420	0.0028
Belief x Reliability x Handedness	10	595	1.36	0.1960	0.0033
Belief x Angle x Handedness	2	119	2.09	0.1280	0.0011
Reliability x Angle x Handedness	5	595	0.57	0.7220	0.0007
Belief x Reliability x Angle x Hand.	10	595	0.71	0.7110	0.0017

Notes. *** p < 0.001, * p < 0.05

Table 4 ANOVA results for reaction time

	DF1	DF2	F	p	η_G^2
Belief	2	119	1.49	0.2290	0.0156
Reliability ***	5	595	4.61	0.0004	0.0079
Angle ***	1	119	285.24	< 0.0001	0.0260

Handedness ***	1	119	94.77	< 0.0001	0.0223
Belief x Reliability	10	595	0.26	0.9890	0.0009
Belief x Angle	2	119	2.12	0.1240	0.0004
Belief x Handedness	2	119	2.27	0.1080	0.0011
Reliability x Angle	5	595	0.96	0.4390	0.0003
Reliability x Handedness	5	595	0.46	0.8070	0.0002
Angle x Handedness ***	1	119	31.09	< 0.0001	0.0024
Belief x Reliability x Angle	10	595	1.07	0.3860	0.0006
Belief x Reliability x Handedness	10	595	1.44	0.1600	0.0011
Belief x Angle x Handedness	2	119	1.31	0.2740	0.0002
Reliability x Angle x Handedness	5	595	0.54	0.7490	0.0002
Belief x Reliability x Angle x Hand.	10	595	0.93	0.5040	0.0006

Notes. *** p < 0.001, * p < 0.05

Table 5 ANOVA results for external resource use over time

	DF1	DF2	F	p	η_G^2
Belief *	2	119	3.48	0.0339	0.0387
Reliability ***	5	595	23.18	< 0.0001	0.0319
Time *	3	2142	2.73	0.0428	0.0006
Belief x Reliability	10	595	1.64	0.0918	0.0046
Belief x Time	6	2142	1.62	0.1370	0.0007
Reliability x Time	15	2142	0.61	0.8660	0.0006
Belief x Reliability x Time *	30	2142	1.56	0.0267	0.0031

Notes. *** p < 0.001, * p < 0.05

Study 2

Investing in brain-based memory leads to decreased use of technology-based memory

Patrick P. Weis & Eva Wiese
George Mason University, Fairfax, VA, USA

Author contributions: PPW and EW conceived and designed the study. Data collection was performed under supervision of PPW. PPW performed the data analysis and interpreted the data together with EW. PPW and EW wrote the manuscript.

Weis, P. P., & Wiese, E. (in press). Investing in brain-based memory leads to decreased use of technology-based memory. *Journal of Experimental Psychology: Applied.*

© 2019, American Psychological Association. This paper is not the copy of record and may not exactly replicate the final, authoritative version of the article. Please do not copy or cite without authors' permission. The final article will be available, upon publication, via its DOI: 10.1037/xap0000259.

Rationale

In this study, we altered the *alphanumeric equation validation* (also called *alphabet arithmetic*) task (e.g., Compton & Logan, 1991; Logan, 1988; Logan & Klapp, 1991) to incorporate the possibility to offload cognitive processing (see **Figure 7**). In the original task, participants have to count upwards the alphabet starting from a first letter given in an alphanumeric equation for as many letters as given by a number in the equation and then check whether this yields the second letter given in the equation (e.g., B + 3 = E, which is correct). After prolonged exposure to specific equations, participants are also able to recall the correct solution from memory. In our altered paradigm, participants could not only use counting or internal retrieval—as in the original paradigm—, but could also externally retrieve the solution by hovering the mouse cursor over a black box depicted on a computer screen that would then vanish and reveal the correct solution.

The study's main focus was to confirm correlational findings (compare *(1)* in the section *Rationale of the dissertation project*) in that human problem solvers are adjusting cognitive offloading behavior based on internal brain-based task performance even when internal performance is at odds with the task's apparent difficulty. The study also investigates whether internal performability can supplant apparent task difficulty in determining cognitive offloading frequency (compare *(2)* in the section *Rationale of the dissertation project*).

To this end, two parameters were manipulated: The apparent difficulty of the addend (2, 3, or 4) and how frequently equations with the specific addend had been learned before (Learning 2: many trials with Addend 2 and few with Addend 4; Learning 4: few trials with Addend 2 and many with Addend 4). The Learning manipulation allowed us to alter the difficulty of specific alphanumeric problems independently of apparent difficulty. For example, the apparently difficult equation “H + 4 = M“ is easy to validate if one has already solved the equation frequently and is able to recall the correct solution (here: L) from memory.

In a nutshell, results showed that participants were able to adaptively adjust their cognitive offloading proportion in response to the learning frequency even when apparent task difficulty would suggest otherwise. Thus, how frequently human problem solvers offload their cognition is not merely based on interpreting a task’s visual features but also on how well they can solve the task internally. These results (a) confirm correlational findings regarding the importance of internal performability when using external resources with an experimental paradigm and (b) suggest that high internal performability can supplant visual features in determining cognitive offloading behavior.

Abstract

Humans frequently use external (environment-based) strategies to supplement their internal (brain-based) thought. In the memory domain, whether to solve a problem using external or internal retrieval depends on the accessibility of external information, judgment of mnemonic ability, and on the problem’s visual features. It likely also depends on the accessibility of internal information. Here, we asked whether internal accessibility contributes to strategy choice even when visual features bear no information on internal accessibility. Specifically, 114 participants were to validate alphanumerical equations (e.g., A + 2 = C) whose visual appearance (addends 2, 3, or 4) signified different difficulty levels. First, some equations were presented more frequently than others, allowing participants to establish efficient internal access to the correct solution via memory retrieval rather than counting up the alphabet. Second, participants viewed the equations again but could access the correct solution externally using a computer mouse. We hypothesized that external strategy use should selectively decrease for frequently learned equations and irrespectively of the task's visual features. Results mostly confirm our hypothesis. Exploratory analyses further suggest that participants partially used a sequential "try-internal-retrieval-first" mechanism to establish the adaptive behavior. Implications for intervention methods aimed at improving interactive cognition are discussed.

Introduction

Imagine you are in your kitchen and about to prepare your new favorite meal that got a five-star rating on your go-to recipe website. You prepared it once and are known for your good memory. The next time you prepare the same dish, would you try to find the recipe online again or would you rely on your mnemonic abilities? Human problem solvers face similar problems, i.e. whether to retrieve information from internal (brain-based) or external (environment-based; e.g., internet, paper) storage, on a daily basis. The present study is designed to illuminate the underlying decision process. Its focus is on investigating the impact of internal information accessibility (i.e., performance of memory retrieval) on the use frequency of external information storages.

The paradigm: Solving alphanumerical equations with external storage

To facilitate the understanding of the remaining introduction, we now briefly describe the paradigm we designed for the present inquiry before continuing with theoretical considerations. Each trial, participants were faced with alphanumerical problems of the format “Letter + Number = Letter” and asked to indicate whether counting the indicated number up the alphabet from the former letter equals in the latter letter. In a mixed design, we altered the number (factor Addend: 2, 3, or 4) within participants and the frequency with which participants learned solving specific problems involving the different Addends (factor Learning: 2 or 4)[7] between participants. After solving specific alphanumeric problems frequently (here, 128 times), human problem solvers are known to shift away from a slow counting to a fast memory retrieval strategy (Compton & Logan, 1991). This strategy shift prominently alters the information’s internal accessibility. After the accessibility of internal information had been altered, participants gained access to an external storage that could be used to replace internal cognitive strategies. Specifically, participants gained access to a black box that revealed the correct solution whenever the mouse cursor was being moved on top of it (e.g., it would reveal “D” if the verification task was “A + 2 = C”). This design allowed analyzing whether participants accessed the black box less frequently for equations with high internal accessibility. More generally, it afforded insight into how proficiently

---

[7] Learning 2: 128 trials with the “2” Addend, 64 trials with the “3” Addend, and 32 trials with the “4” Addend; Learning 4: 32 trials with the “2” Addend, 64 trials with the “3” Addend, and 128 trials with the “4” Addend

human problem solvers incorporate technology-infused environments into their memory processing.

Interactive cognition: Distributed, embodied, and situated perspectives on Cognitive Science

Researching human cognition using interactive paradigms like the one just described has been a focus in recent cognitive science research. The subfields of distributed (e.g., Hol-lan, Hutchins, & Kirsh, 2000), embodied (e.g., Clark, 1999; Wilson, 2002), and situated (e.g., Robbins & Aydede, 2009) cognition are focused on researching the cross-play between sensory input, brain-based cognition, motor output, and manipulation of the environment unfolding over time. One upshot of such a holistic perspective on cognition is that it enables researching cognition-related behaviors in real-world complex environments like today’s and tomorrow’s technologized homes and workplaces. Relating back to the cooking example, this means that a recipe does not need to be retrieved from brain-based memory (knowledge in the head; Norman, 1988) but could instead be retrieved from environment-based sources like paper notes or the internet (knowledge in the world; Norman, 1988; for a review comparing brain-based and internet-based information retrieval, see Clowes, 2013). Without knowledge in the world, the only option to access information would be to use brain-based memory. With knowledge in the world however, the cost structure of the human inferential landscape (Kirsh, 2010) changes. The world can then provide problem solvers with knowledge similar to the one available in the brain but associated with different retrieval costs. Thus, to understand cognitive operations like information retrieval in situations in which the environment can be exploited, it is imperative to broaden the focus beyond the brain. To illustrate this issue, imagine the following scenario that is focused on understanding Peter's cognitive operations: you observed Peter preparing meals throughout the whole last week. You noticed that he is hardly ever looking up recipes on his smartphone. Why would that be?

Accessibility of external information influences external information retrieval

Proficient problem solvers need to refrain from integrating external information when it is associated with higher costs than relying on internal strategies. Thus, possibly, Peter is experiencing poor network coverage and loading the recipe page would take too long to be beneficial. Current literature suggests that there is some truth to this option: human problem solvers proficiently adjust the frequency of external information retrieval based on the accessibility of the external information.

In previous studies, accessibility of external information has been altered via a delay between the time when externally stored information was requested and when it eventually showed up (Gray, Sims, Fu, & Schoelles, 2006; Morgan, Patrick, Waldron, King, & Patrick, 2009; Walsh & Anderson, 2009), by altering the size of the interface elements needed to access the information (which manipulates time costs in a more natural manner via Fitts' law; Gray et al., 2006), by altering the number of key strokes needed to change what external information is shown (O'Hara & Payne, 1998), or by altering the distance between the problem solver and a computer from which the relevant information could be accessed from (Storm, Stone, & Benjamin, 2017). In all cases, decreased accessibility led to less external information retrieval and, equivalently, to more internal information retrieval. In one study, an ideal performer analysis suggested that problem solvers switched between using internal and external information retrieval in a way that maximizes speed (Gray et al., 2006); in another study, problem solvers were found to proficiently switch between internal and external strategies as to maximize a monetary performance-related reward (though there was a slight bias for internal strategies and, strictly speaking, participants were found to probability match rather than optimize; Walsh & Anderson, 2009).

Taken together, the studies suggest that problem solvers adjust their use of external information based on properties of the external information source (i.e., accessibility). Changing the accessibility of externally stored information can alter the cost structure of the inferential landscape which causes agents equipped with a rational and adaptive cognitive system (which might be close to the human cognitive system; Anderson, 1990) to adjust how they incorporate the environment into their cognitive processing.

Additional factors influencing external information retrieval

If human problem solvers are indeed able to mix internal and external cognitive strategies to maximize performance, they should not only be sensitive to properties of external information sources. Instead, they should pay equal attention to features of the task (e.g., does the task look difficult?) and properties of internal brain-based information sources (e.g., how fast is it to access the solution internally?). For example, after using a written recipe for preparing a meal for more than ten times, a substantial brain-based memory trace of the recipe should have built up that might make it more efficient for a culinary problem solver to omit using the written recipe and rely on internal information instead.

And indeed, humans can be quite proficient in deciding between internal and external strategies even when the cost of external information access is kept constant (e.g.; Risko, Medimorec, Chisholm, & Kingstone, 2014; Siegler & Lemaire, 1997; Walsh & Anderson, 2009). When allowed to use internal (e.g., internal counting) and external (i.e., using a calculator) strategies for arithmetic problem solving, participants were repeatedly shown to mix the strategies in a way that led to better speed performance than when allowed to only use one of the strategies (e.g., Siegler & Lemaire, 1997; Walsh & Anderson, 2009). More specifically, participants quite adaptively preferred using mental arithmetic for equations in which one multiplicand was "10" (e.g., 17 x 10) and the calculator for tasks in which no multiplicand was "10" (e.g., 17 x 13; Siegler & Lemaire, 1997).

But how is such adaptive external strategy use being achieved and what is the driving force behind adaptive cognitive strategy selection? In the following, we are discussing four possible candidates: 1. *feature-specific strategy selection* (e.g., seeing a "10" as a factor in an arithmetic task affords the specific internal strategy of simply adding a 0 to the other factor to get the product), 2. *feature-based apparent difficulty of the problem* (e.g., seeing a "10" as a factor is associated with an easy problem through metacognitive reasoning), 3. *evaluation of internal accessibility of the solution* (e.g., perceiving oneself as being bad at math and thus preferring external over internal strategies), and 4. *actual internal accessibility of the solution* (e.g., the solution to “17 x 10” is quick to calculate internally).

1. Thoroughly analyzing the visual features of a problem is likely a highly relevant process for adaptive cognitive strategy selection. Feature-specific strategy selection has been frequently researched by observing arithmetic problem solvers. For example, simply seeing a “10” had profound consequences for problem solvers’ strategy selection: it accounted for a 14 percentage point increase in explained strategy choice variance on top of the variance explained by reaction time differences between internal and external strategies (Siegler & Lemaire, 1997; Experiment 1). A similar effect of visual features on strategy choice has been triggered by a “5” as a factor in arithmetic-based problems (Lemaire & Reder, 1999; Experiment 3)[8] or by a letter that enabled rule-based instead of retrieval-based processing in a string classification task (e.g., if the first letter of the string is a consonant, classify the string as “code” instead of “noncode”; Bourne, Raymond, & Healy, 2010).
2. A second relevant process requiring analysis of a problem’s visual features constitutes metacognitive judgments related to the problem’s difficulty. For example, it has been suggested that some problem solvers assume that it is manageable to keep an array of ten letters in working memory, which is why they

[8] However, note that participants do not always use that rule but, adaptively, skip the rule in experimental sessions in which the last digit hardly violates the five rule (Lemaire & Reder, 1999; Experiment 3). Human problem solvers thus also exhibit feature-independent strategy adaptation.

skip writing the letters down and end up with attenuated task performance (Risko & Dunn, 2015). Similarly, it has been suggested that whether adopting an external or an internal strategy heavily depends on the familiarity of the respective problem rather than how well the solution can be accessed internally (Schunn, Reder, & Nhouyvanisvong, 1997). Note that familiarity-based judgments might not be made on a conscious level but depend on an implicit frequency tracker (Onyper, Hoyer, & Cerella, 2006).

3. The third process relevant for cognitive strategy choice is the evaluation of own performance. Such evaluation can have components independent of actual skill (Gilbert, 2015) and might be a reason for memory avoidance in older adults (Hines, Hertzog, & Touron, 2012; Touron, 2015).
4. A fourth relevant process might be the actual – in contrast to the estimated – internal accessibility of the solution: how fast can one produce a correct solution with mental processing? Although adaptive selection between an internal and an external strategy has been frequently shown, it is hard to say whether the problem solver's sensitivity to brain-based task performance actually contributed to strategy selection. Instead, as pointed out in the above, problem solvers might rely on what can be directly seen. For example, differences in task difficulty are usually obvious (e.g., 17 x 13 looks harder than 7 x 10) and properties of external information sources are oftentimes equally observable (e.g., accessing information using a slow internet connection makes you stare at a mostly empty screen for a while). In comparison, perceiving the properties of internal information sources

might prove challenging. As mentioned in the preceding paragraph, judgments of task-related internal memory ability can be independent of actual ability (Gilbert, 2015; Experiment 1). Given the difficulty of accurately judging own ability and the frequent reliance on visible task features for deciding between internal and external cognitive processing, it seems questionable how big the influence of internal memory accessibility on proficient external strategy use really is.

Influence of accessibility of internal information on external information retrieval

So far, we have argued the accessibility of external information, judgments of own skill, and visual features of the task influence how problem solvers mix internal and external strategies. We have also argued that, based on the current evidence, it is questionable whether the same holds for the accessibility of internal information. Consequently, with the current study, we aim to investigate whether problem solvers monitor their internal information access to inform the orchestration process between internal and external information retrieval or whether they prefer possibly misleading external cues and higher-level metacognitive evaluations (like evaluations of own skill) for the same end. However, drawing on the previous studies on cognitive strategy selection and cognitive offloading that are discussed in what follows, we hypothesize that human problem solvers can indeed adapt their interactive behavior based on internal information accessibility and independent of solely feature-based reasoning.

In one study (Howes, Duggan, Kalidindi, Tseng, & Lewis, 2016), problem solvers were tested on their ability to copy name lists. Initially, in a no-choice condition, problem solvers had to copy lists consisting of between three and nine names. Then, in a choice condition, people were able to select their preferred list length on their own. Results show that problem solvers tended to choose the list length they performed best with in the no-choice condition. Promisingly, the study shows that problem solvers are able to adapt their interaction behavior depending on how well information can be stored in internal working memory. Importantly however, the study allowed participants only to alter the set size parameter of an internal process (i.e., storing words in working memory). Though while the results sound promising, the study did not investigate how participants chose between internal and external strategies, which is the focus of the current investigation .

In another study (Touron & Hertzog, 2004), problem solvers had to decide whether one noun was “correctly” paired with another noun. To find out whether a noun was paired correctly, problem solvers could rely on two strategies. At the beginning, they only could use an external strategy: they could search through a list of noun pairs provided on screen and compare whether the nouns that were paired in the current problem were also paired in the list on screen. After a while, problem solvers could also use an internal strategy: if they already had had the opportunity to learn parts of the list in earlier trials, they could consult their memory. Results showed that problem solvers relied less on the external search strategy once they had established a good enough memory to use the internal strategy. The only limiting factor of this study is that internal strategies had become available through learning while external strategies were available from the very beginning. Thus, it was to be expected that problem solvers used external information for unlearned problems since no internal options were available. Other paradigms like arithmetic (Walsh & Anderson, 2009) or alphanumeric (Compton &

Logan, 1991; Zbrodoff, 1995) problem solving avoid this lack of an internal option by providing an internal strategy that is computation- instead of retrieval-based. In the current study, we decided to use an alphanumeric task to avoid the unavailability of an internal strategy.

Some correlational evidence also suggests that problem solvers adjust how frequently they accessed external information based on internal information accessibility. For example, the older elderly people get, the more they use external information, which is likely an adjustment made due to declining internal information accessibility (Dixon & de Frias, 2004; Touron, 2015) [9]. Similar adjustments are compatible with the results of a study examining prospective memory: the better participants performed without the opportunity to use external information, the less they made use of external information when gaining the opportunity to do so (Gilbert, 2015). Further support for the importance of the accessibility of internal information comes from a study comparing two different internal computational strategies (which, in the terminology of this paper, would be equal to two different internal information sources with differential accessibilities) to convert currencies: after briefly practicing the two competing strategies, participants were more likely to use the faster one (Lemaire & Lecacheur, 2001). Thus, problem solvers can adapt to the accessibility of internal information when selecting between internal strategies. However, note that selecting between internal and external strategies involves additional mechanisms like metacognitive evaluation of the external resource (reviewed

[9] Please note, however, that elderly adults likely *over*-use external information due to low confidence in their internal memory (reviewed in Touron, 2015).

by Risko & Gilbert, 2016) and evaluation of observable performance cues (e.g., size of a button or delay after pressing a button; Gray et al., 2006).

Current study

Humans are known to be proficient problem solvers who make use of a variety of in-ternal cognitive strategies to meet their goals (Compton & Logan, 1991; Lemaire & Lecacheur, 2001; Lemaire & Reder, 1999). Evidence is accumulating that humans are also proficient in incorporating external strategies when solving problems (e.g., using a calculator instead of internal strategies; Siegler & Lemaire, 1997; Walsh & Anderson, 2009). However, the importance of feature-based considerations for the selection between internal and external strategies makes it hard to gauge how important sensitivity to the performance of internal strategies for this selection process is (see section: What else Influences External Information Retrieval?).

In the present study, we manipulated the performance of internal problem solving strategies and observed whether participants adjust their use of an external problem solving strategy accordingly. Importantly, we were controlling for the visual features of the task. In other words, we used a paradigm in which an identical problem will be easy to solve internally for one group of participants but hard to solve internally for another group of participants. Based on the studies presented in the preceding section (see section: Does Accessibility of Internal Information Influence External Information Retrieval as well?), we assume changes in the efficiency of the internal strategies to drive changes in how frequently the external strategy is being employed. We expect such adaptive changes in external strategy use despite the fact that participants cannot rely on feature-based reasoning to make adaptive strategy choices. Such a finding would confirm the human problem solver's proficiency in using the environment for cognitive processing.

The present paradigm combines three features that are of particular importance for the interpretability of the study's outcome: (1) The cognitive process investigated here is information retrieval. Thus, external validity is likely highest in the memory domain. (2) In many paradigms, internal accessibility can be derived from appearance. Here, internal information accessibility is manipulated independently from apparent task difficulty. (3) The external strategy used in the current paradigm constitutes using the mouse to access task-relevant information. Thus, external validity is likely highest in the human-computer-interaction domain.

Methods

Participants

In total, 114 undergraduate students participated in the experiment. Two participants were excluded because they reported that they did not understand the task, one because of technical problems, and twelve due to poor task performance (i.e. answering incorrectly in more than 15% of the problems), resulting in a final sample size of 99 participants (62 fe-males; mean age: 20.3; range: 18 – 50; 79 right handed). 51 participants (31 females; mean age: 21.2; range: 18 – 50; 37 right handed) were assigned to the Learning 2 and 48 partici-pants (31 females; mean age: 19.3; range: 18 – 27; 42 right handed) to the Learning 4 condi-tion (see Design and Procedure for details on the Learning factor). Our targeted sample size (N of 100) was based on an a-priori power analysis for a within-between interaction at medium effect size conducted in G*Power (Faul, Erdfelder, Lang, & Buchner, 2007). All participants were recruited from the psychology undergraduate student pool at George Mason University and reimbursed via research participation credits. All participants were at least 18 years old and reported normal or corrected to normal vision. The Ethics Committee at George Mason University approved the experiment and participants provided informed consent prior to participation.

Apparatus

The experiment was presented at a distance of about 57 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 Psy-chophysics Toolbox (Brainard, 1997; Pelli, 1997). Responses were recorded using a USB-connected standard keyboard and a USB-connected optical mouse with a resolution of 800 dpi. The mouse cursor speed was set in a way that moving the mouse for 1 cm would move the cursor on screen for 1.4 cm.

Stimuli

Stimuli consisted of equations that started with one of six letters (A to F) and had one of three addends (2 to 4), e.g. "A + 2 = C". Each equation was presented with a correct or an incorrect solution. The incorrect solution was always one letter further up the alphabet than the correct solution. Thus, in total, 36 different equations were used (18 correct, 18 incorrect). For each participant, each starting letter was uniquely associated with one addend. Associations between starting letters and addends were balanced between participants within each condition in a way that each addend was associated with each target letter with equal probability. Equations were presented on the left side of the screen with an eccentricity of 7° visual angle. Equations had a width of 5° and a height of 0.6°. During the second block, a square target box was presented at the right side of the screen with an eccentricity of 7° and a width of 0.8°.

Task

During the main experiment, participants had to examine the correctness of alphanu-merical equations (e.g., A + 2 = C; see Stimuli) as used by Compton and Logan (1991) to study the transition from solving a task algorithmically to solving it through memory retrieval. Participants had to press the downward arrow key labeled with a checkmark to indicate a correct equation and the upward arrow key labeled with a cross to indicate an incorrect equation. During different parts of the experiment, participants had different options to arrive at their answer.

During the first part of the experiment, the learning block, participants had the two options investigated by Compton and Logan (1991). First, participants could count upwards the alphabet starting from the first letter given in the equation (internal counting strategy; e.g., when given the equation "C + 3 = G", counting C + 1 = D, D + 1 = E, and E + 1 = F, would lead to the conclusion that C + 3 != G). Second, with increasing exposure to a specific equation, the counting strategy could be replaced by a more efficient memory-based strategy, i.e. participants could recall the solution from memory (internal retrieval strategy). The likely reason for the strategy switch over time is that with increased exposure, the memory trace linking cue (here, the left side of the equation, e.g. C + 3) and the correct solution (i.e.., F) grows stronger, leading to both increased frequency and increased speed of internal memory recall (as argued by Compton & Logan, 1991). During the second part of the experiment, the choice block, participants additionally were able to access the solution by hovering the mouse cursor over a black box that then would disappear and reveal the correct solution (external strategy). If an incorrect answer was given or an unassigned key pressed, a feedback message was displayed for 500 ms immediately after their response. To keep timing constant, the inter-trial interval was shortened to 1500 ms after incorrect answers. The task is illustrated in **Figure 7**.

For a more detailed treatise on how problem solvers could automate cue-specific in-formation retrieval, the interested reader can consult Logan's (1988) Instance Theory of Automatization. For a more general overview of how humans create internal problem solving routines, interested readers should consult Anderson's (1987) review article about skill acquisition. For the purpose of the current study, it should be sufficient to know that our problem solvers can choose between an automatic internal retrieval strategy (as in Logan, 1988) and two algorithmic strategies (one internal, one external) to solve a problem and that increased exposure with a specific alphanumeric equation increases the likelihood that its solution can be automatically retrieved from internal memory (Logan & Klapp, 1991).

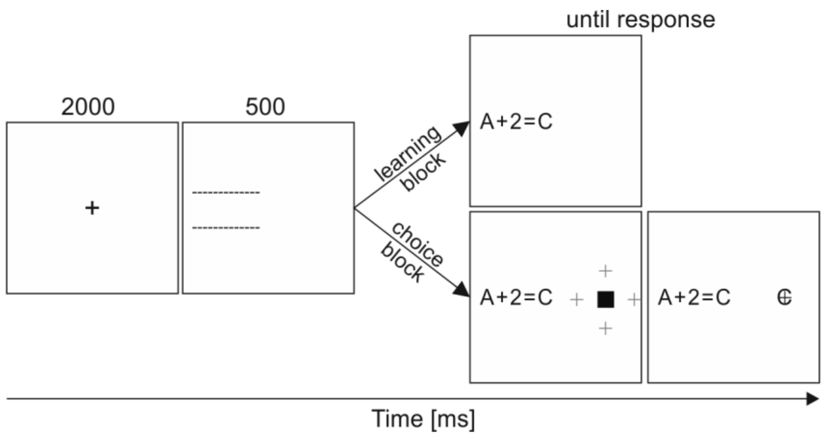


Figure 7 Trial Sequence. Participants have to check alphanumerical equations for correctness via button press (see Task for details). At the beginning of a trial, participants fixate on a fixation cross for 2000 ms (inter-trial interval). Afterward, to guide the participant's spatial attention, the location of the upcoming equation is indicated for 500 ms. In the learning block, participants have to rely on brain-based internal strategies to solve the equation. In the choice block, participants can additionally rely on an external strategy: hovering with the cursor over the black box will reveal the correct answer hidden under the box (here: "C"). At the beginning of each choice trial, the cursor will appear randomly at one of the four locations indicated by the gray cursors.

Design and procedure

The main task followed a 3 x 2 mixed design with the within-participants factor Addend (2, 3, or 4) and the between-participants factor Learning (2 or 4). The Addend manipulation refers to the addend used in the respective equation (e.g., "2" in the equation A + 2 = C). The Learning manipulation refers to the addend-specific learning process that took place during the learning block: participants either solved 128 equations with the "2" Addend, 64 equations with the "3" Addend, and 32 equations with the "4" Addend (Learning 2) or 32 equations with the "2" Addend, 64 equations with the "3" Addend, and 128 equations with the "4" Addend (Learning 4). This differential learning is known to alter internal information accessibility because problem solvers transition from the time-consuming internal counting strategy to the more efficient but learning-dependent internal retrieval strategy (compare Figure 3 in Compton & Logan, 1991). It is important to note that this design affords researching the influence of internal information accessibility on external information retrieval independently of appearance-related effects (e.g., "A + 4" looks inefficient to solve by using internal strategies but is actually very efficient to solve internally for problem solvers in the Learning 4 condition). Also note that we have not altered the internal accessibility of solutions to trials with the "3" Addend. Equations with the "3" Addend were presented at medium frequency (i.e., 64 trials) for all participants to serve as a baseline or control condition and to make results more comparable to Compton and Logan (1991).

In total, participants engaged in 224 learning trials (128 trials with the well-learned Addend, 64 control trials with the Addend "3", and 32 trials with the remaining Addend) in block one and 96 choice trials (32 with each Addend) in block two. Trial order was random-ized. At the beginning of each block, participants additionally engaged in four practice trials (two with addend 1, two with addend 5) with letters different from the ones used in the main task.

Upon entering the lab, participants were welcomed, seated in front of a computer screen, and provided informed consent. During each session, up to three participants were tested simultaneously. Participants then engaged in solving alphanumerical equations during the learning and the choice block, took a brief demographic survey, and finally took a brief metacognitive survey. Overall, the experiment took about 30 minutes to complete. Partici-pants were instructed to respond as quickly and accurately as possible. We refrained from asking participants to exclusively rely on either accuracy or speed because focusing on speed bares the risk of fast but random answers (i.e., participants make use of neither internal nor external strategies) while focusing on accuracy bares the risk of time-intensive double-checking (i.e., participants make use of both internal and external strategies for the same problem). Also, since the Learning manipulation likely increases both speed and accuracy for internally solving well-learned problems (compare Compton & Logan, 1991), asking participants to consider speed and accuracy likely maximizes our effect size.

To keep body posture constant between blocks, participants were asked to respond with their left hand only during the learning block. During the choice block, participants were asked to respond with their left hand and use their right hand to move the mouse when needed.

Analysis

All trials with extreme reaction times above 15s (0.1% of all trials) were excluded from analysis. After removing these trials, all trials that deviated more than three standard deviations from the individual reaction time means of the respective problem size condition in the respective block (1.7 % of all trials) were excluded as well because they likely presented either motor slips (low RTs) or inattentiveness (high RTs). To increase readability, statistical analyses are described directly preceding the respective result. All p-values are reported Greenhouse-Geisser-corrected where indicated.

Results

Does learning alter how efficiently a solution can be accessed internally?

This analysis served as a manipulation check. The main experiment was divided into two blocks, the learning and the choice block. The purpose of the learning block was to ma-nipulate the ease of internal access to the solutions of the alphanumerical problems (Learning manipulation). Participants in the Learning 2 condition are thus expected to get more efficient in solving equations with the Addend 2 whereas participants in the Learning 4 condition are expected to get more efficient in solving equations with the Addend 4 throughout the learning block. Efficiency in solving equations with the Addend 3 should be comparable across both Learning conditions since learning was kept constant. Meeting these expectations is a crucial prerequisite for the validity of our main analysis (see next section).

To check whether the expectations are met, a 2 x 3 x 2 mixed ANOVA with the be-tween-participants factor Learning and the within-participants factors Addend and Block was deployed. Block was included as a factor to explore efficiency patterns between learning and choice block. Our focus was on internal efficiency as dependent variable since, in line with Compton and Logan (1991), we expected learning to increase both speed and accuracy. Inverse efficiency is defined as the reaction time of correct responses divided by the accuracy of all responses (Townsend & Ashby, 1978) and thus captures both measures of interest simultaneously. To enable exploration of possible speed-accuracy-tradeoffs, we also report speed and accuracy data.

Learning, Addend, and Block interacted in their influence on inverse efficiency ($F(2, 194) = 43.0$, $p_{GG} = 8.04 \times 10^{-13}$, $\eta_G^2 = .046$). All other effects were also significant at the .05 significance level and are reported in more detail in **Table 6** in the supplemental materials. The three-way interaction reflects the nontrivial consequences of introducing the external resource in the choice block. More specifically, dependent post-hoc t-tests revealed that participants in the Learning 2 condition were, as expected, more efficient in solving equations with Addend 2 in comparison to the control equations with Addend 3 ($t(50) = 10.50$, $p = 3.04 \times 10^{-14}$, $M_{Delta} = 1120$ ms) and less efficient in solving equations with Addend 4 in comparison to the control equations with Addend 3 ($t(50) = 6.96$, $p = 6.85 \times 10^{-9}$, $M_{Delta} = 1132$ ms); **Figure 8a**, left. The reverse was true for participants in the Learning 4 condition: they were more efficient in solving equations with Addend 4 in comparison to the control equations with Addend 3 ($t(47) = 3.96$, $p = 2.55 \times 10^{-4}$, $M_{Delta} = 493$ ms). However, they were also more efficient in solving equations with Addend 2 in comparison to the control equations with Addend 3 ($t(47) = 4.53$, $p = 4.05 \times 10^{-5}$, $M_{Delta} = 429$ ms), which might be due to the high efficiency of the counting strategy for equations with Addend 2 and despite the low efficiency of the retrieval strategy. There was no evidence for inverse efficiency differences for equations with the control Addend 3 but differential Learning conditions (independent t-test: $t(97) = .94$, $p = .351$, $M_{+2} = 3112$ ms, $M_{+4} = 2944$ ms).

In sum, these results confirm that learning specific equations indeed established efficient internal information access for these equations. The learning-driven rise in efficiency originated from both increased speed and increased accuracy simultaneously; **Figure 8b** and **c**, left. Detailed ANOVA results for inverse efficiency are summarized in the supplemental material, **Figure 14**. ANOVA results for accuracy and speed are also reported in the supplemental materials (**Figure 15** and **16**) so the reader can further investigate the data for a possible speed-accuracy-tradeoff; there are no signs of a speed-accuracy tradeoff. Note that these results strongly suggest an increase in the retrieval strategy even for the difficult Learning 4 condition, thereby excluding the possibility that the difficult Learning 4 condition draws too many processing resources to allow for memory consolidation (as shown for a similar task by Hoyer, Cerella, & Onyper, 2003).

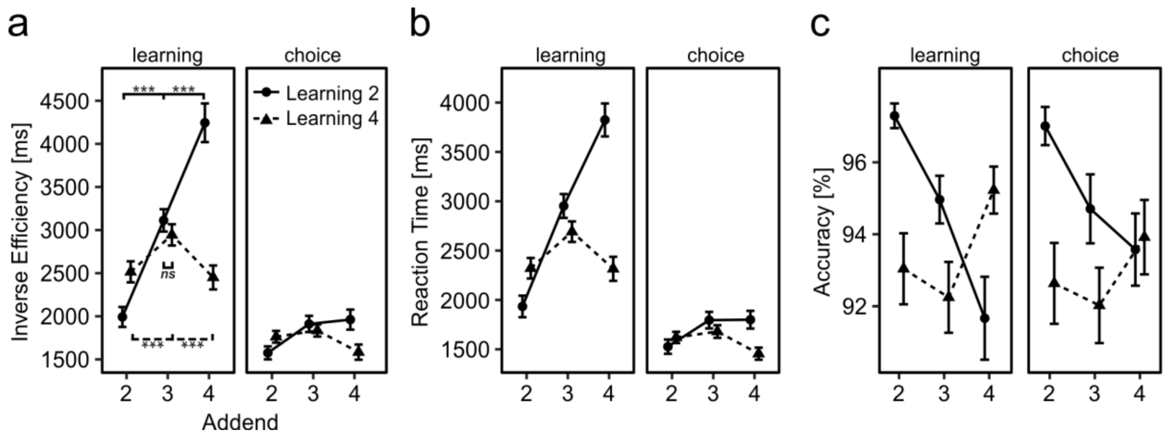


Figure 8 Performance during learning and choice blocks for whole sample. The Learning manipulation successfully altered how efficiently participants could solve a problem internally (a, left). Participants were more effi-cient in solving the equations they have been learning frequently (Learned 2 and Addend 2; Learned 4 and Addend 4) than in solving control equations (Addend 3). To allow the curious reader to inspect the full data, choice performance (a, right) as well as reaction time (b) and accuracy (c) data is presented as well. Error bars depict SEM. *** $p < 0.001$, *ns* $p = .35$

Does external retrieval depend on how efficiently solutions can be accessed internally?

External information retrieval is defined as the proportion in which a participant ac-cessed the black box with the mouse cursor during the choice block (**Figure 7**, right, bottom). As indicated by inverse efficiency during the learning block (**Figure 8a**, left), participants established efficient internal access to the solutions of frequently but not infrequently practiced alphanumerical problems. Here, we expect the learning effect indicated by the inverse efficiency analysis to be reflected by the frequencies of external information retrieval during the choice block.

To test this hypothesis, a 3 x 2 mixed ANOVA with external information retrieval (in % of trials) as DV, Addend as within-participants factor, and Learning as between-participants factor was used. Only data from the choice block was included in the analysis. The ANOVA was followed up with dependent and independent t-tests where appropriate.

Results of an initial analysis with all participants showed the hypothesized interaction between Learning and Addend (($F(2, 194) = 12.7$, $p = 6.85 \times 10^{-6}$, $\eta_G^2 = .005$). However, data exploration revealed that a significant proportion of participants hardly showed any variance in their external information retrieval; **Figure 9a**. Seventeen participants used the external information in less than 5% of trials (internal group) of the choice block and thirty-two participants in more than 95% (external group). Since these two groups exhibit little variance to be explained by the experimental manipulations and introduce normality assumption violations for our analyses, we decided to limit the current analysis to participants that used external information in at least 5% and at most 95% of trials in the choice block (mixed group). For the curious reader, performance data for all groups and both blocks is depicted in **Figure 9b**. Post-hoc t-tests for the mixed group inverse efficiency scores mirror the results obtained from the whole sample. Also note that, at least on a descriptive level, a participant's choice to avoid internal or external strategies was likely at least somewhat adaptive: participants who decided to avoid the internal strategies (i.e., External Group in **Figure 9a**) performed especially poor in the internal learning block (External Group: learning in **Figure 9b**).

Results for the mixed group confirm the initial whole sample analysis in showing that Learning and Addend interacted in their influence on external information retrieval ($F(2, 96) = 12.26$, $p = 1.81 \times 10^{-5}$, $\eta_G^2 = .025$). There also was a main effect of Addend ($F(2, 96) = 8.96$, $p = 4.76 \times 10^{-4}$, $\eta_G^2 = .017$)) but no main effect of Learning ($F(1, 48) = 0.07$, $p = .790$, $\eta_G^2 = .001$)). Cell means are illustrated in **Figure 10a**. A post-hoc dependent t-test confirmed that participants in the Learning 2 condition used external information less when solving equations with the well-learned Addend 2 than with control Addend 3 ($t(27) = 5.07$, $p = 2.54 \times 10^{-5}$, $M_{Delta} = 18.4$ %). Participants in the Learning 4 condition analogously used external information less when solving equations with well-learned Addend 4 than when solving control equations with control Addend 3 ($t(21) = 2.82$, $p = 0.0103$, $M_{Delta} = 10.5$ %). Thus, participants that established highly efficient internal access to the solution of the respective alphanumerical problem relied less on external information. Surprisingly, there was no difference in external information retrieval between problems with medium and problems with low internal accessibility (Learning 4 and Addend 2 vs Learning 4 and Addend 3: $t(21) = 0.80$, $p = .430$, $M_{Delta} = 2.7$ %; Learning 2 and Addend 4 vs Learning 2 and Addend 3: $t(27) = .19$, $p = .849$, $M_{Delta} < 0.1$ %). However, when considering that using the counting strategy is faster for Addend 3 than Addend 2 (as reflected by **Figure 9c**), the missing external retrieval differences for Learning 4 are not that surprising anymore. For Learning 2 however, this explanation does not hold: quite in contrast, the counting strategy is slower for Addend 4 and equations with Addend 4 additionally had been learned less. This will be discussed in more detail in the next section. External information retrieval for the control condition, i.e. trials with the Addend 3, did not differ between Learning conditions ($t(48) = 0.48$, $p = 0.641$, $M_{+2} = 55.4$ %, $M_{+4} = 51.0$ %). Thus, there is no evidence for differences in baseline performance between the Learning groups. To account for aberrations from normality, robust non-parametric Wilcoxon signed rank tests and a Wilcoxon rank sum test were used to confirm the results of the significant t-tests.

In sum, these results confirm our hypothesis in that problem solvers adjusted external information retrieval based on how efficiently they can access solutions using internal strategies. However, our data shows one aberration from this pattern. Participants in the Learning 2 condition were significantly less efficient in solving equations with Addend 4 than equations with Addend 3 and nevertheless exhibited no differential external information retrieval. This aberration will be addressed in the next section.

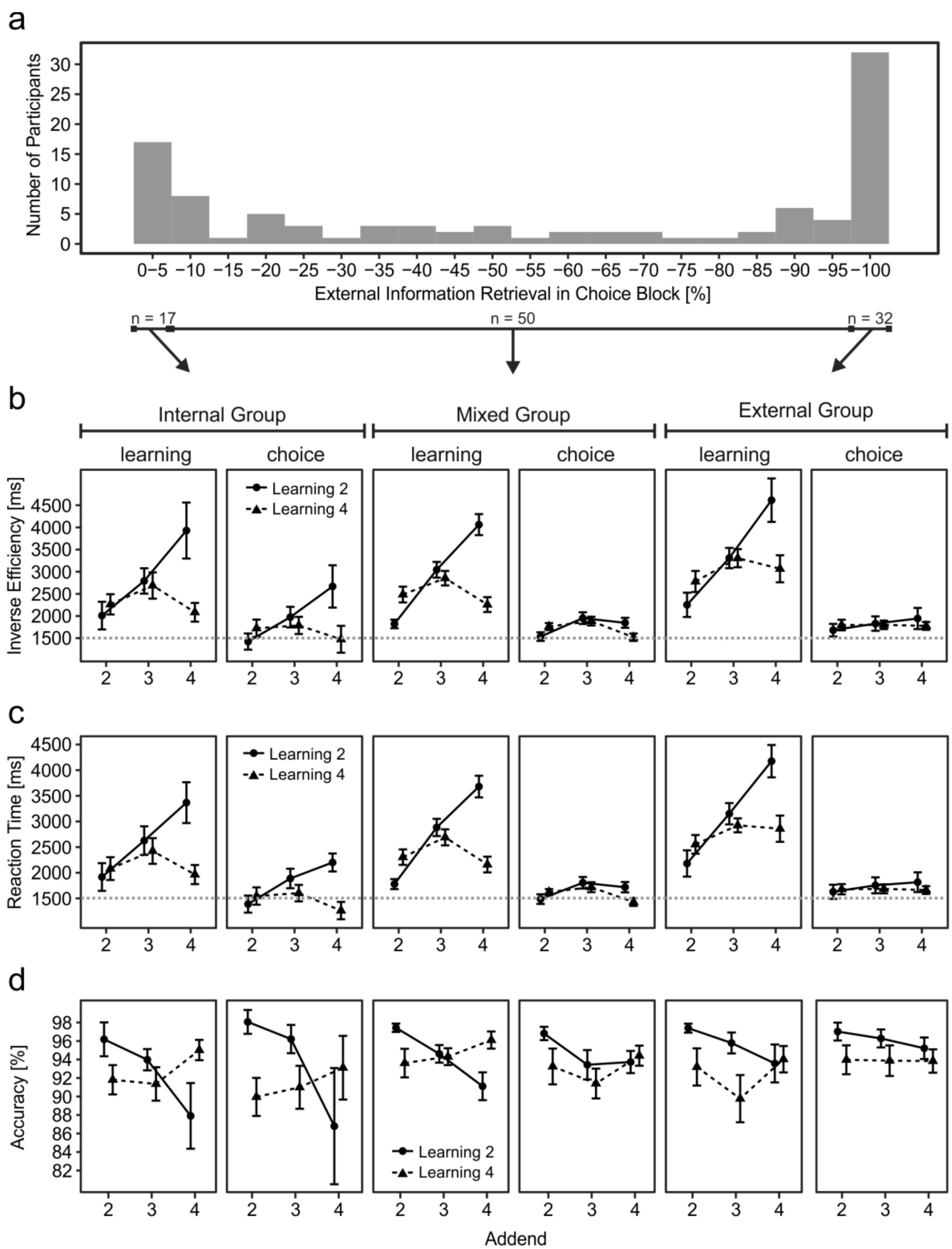


Figure 9 Performance during learning and choice blocks split by groups based on frequency of external infor-mation retrieval. Many participants showed hardly any or permanent external information retrieval (a). Performance data was therefore split into three different groups to allow eyeballing of possible performance differences between groups and to allow for a more powerful analysis of external information retrieval data (b, c, d). The figure depends on the same data as **Figure 8**. Error bars depict SEM.

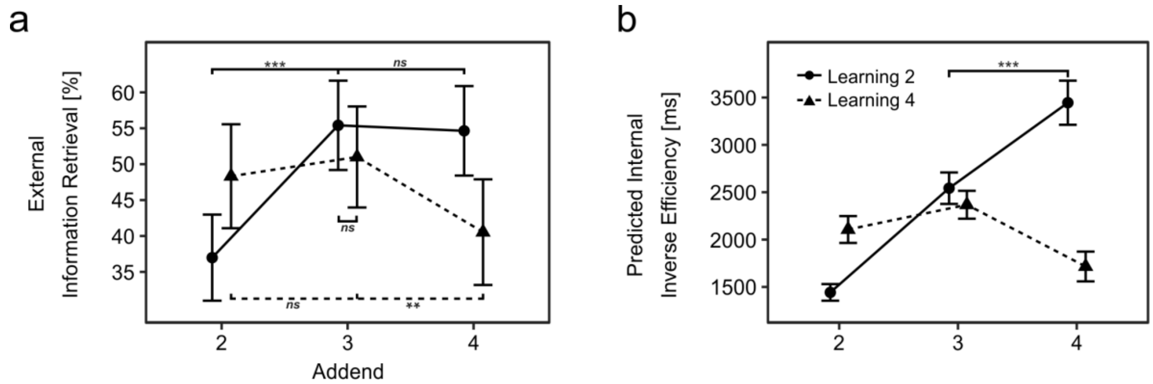


Figure 10 Comparison of external information retrieval and predicted efficiency of internal strategies for the mixed group. Learning and Addend interacted in their influence on external information retrieval (a). The pre-dicted efficiencies for the internal strategies mostly mirror the pattern of external information retrieval (b). It thus appears likely that internal information accessibility informed the decision to retrieve information externally. However, note that participants in the Learning 2 condition exhibit great performance differences between Addend 3 and 4 that are unexpectedly not mirrored in external information retrieval. For the definition of the mixed group, see **Figure 9**. Error bars depict SEM. Note that the error bars only provide information about effects between, not within, participants. *** $p < 0.001$, ** $p <= .01$, *ns* $p > .1$

Why does external retrieval not exclusively depend on internal information accessibility?

The following analysis was conducted to explain an unexpected finding: participants in the Learning 2 condition exhibit comparable external information retrieval for Addend 3 and for Addend 4 (**Figure 9a**) despite prominent differences in inverse efficiencies during the learning block (**Figure 8a**: learning). However, average inverse efficiencies during the learning block as reported in **Figure 8a**: learning do not factor possible practice effects throughout the block in. Thus, the poor average internal performance in the “Learning 2, Addend 4” condition reported in **Figure 8a**: learning might be an artifact stemming from poor initial internal performance that had been compensated for by the end of the learning block. To account for this possibility, we fitted the power law function

Equation 1 Fitting the power law function

$$RT\ or\ Accuracy\ =\ a\ *\ t^{-k}$$

to individual accuracy and reaction time data as it developed over time during the learning block. Based on the two fitted functions for each participant, we computed the participant's predicted inverse efficiency for the first trial in the choice block.

Qualitatively, results of a two-factorial mixed ANOVA (Addend x Learning) were comparable with the results obtained by the ANOVA on the inverse efficiency averages, i.e. Addend and Learning interacted in their influence on predicted inverse efficiency ($F(2, 96) = 51.91$, $p_{GG} = 1.67 \times 10^{-13}$, $\eta_G^2 = .270$; see **Figure 10b**). There also were main effects of Learning ($F(1, 48) = 4.68$, $p = .0356$, $\eta_G^2 = .060$) and Addend ($F(2, 96) = 35.34$, $p_{GG} = 2.10 \times 10^{-10}$, $\eta_G^2 = .201$). Participants in the "Learning 2, Addend 4" condition were still predicted to perform worse than participants in the "Learning 2, Addend 3" condition ($t(27) = 4.55$, $p = 1.01 \times 10^{-4}$, $M_{Delta} = 903$ ms), despite the differences found for external information retrieval. Thus, we currently have no data-backed explanation for this unexpected finding but will refer to possible other theoretical considerations in the *Discussion*.

Exploration: how did participants establish adaptive external information retrieval?

Except for the single unexpected finding reported above, participants reduced their external information retrieval selectively for equations for which high internal information accessibility had been established. To explore possible underlying mechanisms, four exploratory analyses were conducted.

(1) *Do participants evaluate addend-specific learning frequencies and adjusted external information retrieval accordingly?* In a post-experimental survey, we asked participants whether they preferred internal or external information retrieval and to briefly explain why they preferred the one over the other in an open answer format. Interestingly, no single participant mentioned differential learning frequencies for equations with different addends. Participants frequently mentioned that one strategy was "quicker" or "faster" (35 times) or needed less "effort" or was "easier" (26 times). They however did not report any addend-specific strategies. Although we did not explicitly ask participants whether they noticed the Learning manipulation, the fact that no single participant mentioned it during the survey lets it appear unlikely that a strategy based on conscious reflection about the learning process is the prime reason for the adaptive external resource use.

    We aimed to confirm this interpretation in a follow-up study. Specifically, we hypothesized that participants did not consciously notice the "learning" manipulation, which would make strategy selection based on a conscious metacognitive evaluation of stimulus frequencies highly unlikely. In the follow-up study, participants followed the identical experimental procedure as in Experiment 1 with one exception: after the learning block, participants now were to judge how frequently they saw equations with different Addends. Specifically, participants had to answer the question "How often have you been solving problems involving '+ X'" three times, where X would be 2, 3, or 4, respectively. Answers were to be given via a cross on a paper-based visual analogue scale labelled “not at all” on the left-hand and “very often” on the right-hand side.

The sample consisted of eighteen participants that were drawn from the same student population as the main study, were at least 18 years old, reported normal or corrected normal vision, and did not participate in the main experiment. The Ethics Committee at George Mason University approved the experiment and participants provided informed consent prior to participation. Two participants had to be excluded from analysis due to low accuracy in the learning block (< 85% correct), leading to a final sample size of sixteen (nine Learning 2, seven Learning 4; 10 females, mean age: 19.9, age range: 18 – 26, 15 right handed). Only perceived performance will be reported because performance in the learning and choice blocks is not of primary interest for question at hand.

Results of a mixed ANOVA showed that Addend and Learning interacted in their influence on perceived frequency ($F(2, 28) = 12.58$, $p = 1.27 \times 10^{-4}$, $\eta_G^2 = .22$; see **Figure 11**). The main effects did not reach significance at the .05 alpha level (both $p > .59$). These results imply that participants were sensitive to changes in learning frequency, which falsifies our initial hypothesis.

Thus, possibly, our participants relied on perceived frequency to decide for a specific cognitive strategy. We now wanted to explore whether perceived frequency can explain the aberration in Experiment 1, i.e. that participants in in the Learning 2 condition did not access the external information more frequently for Addend 4 than for Addend 3. A post-hoc dependent t-test revealed that the perceived frequency data cannot explain the aberration: while participants in the Learning 2 condition of Experiment 1 showed no difference in external information retrieval for Addend 3 and 4, participants in the follow-up study were sensitive to the different learning frequencies (Learning 2 and Addend 3 vs Learning 2 and Addend 4: $t(8) = 2.72$, $p = .0262$, $M_{\text{Delta}} = 15.4$ %).

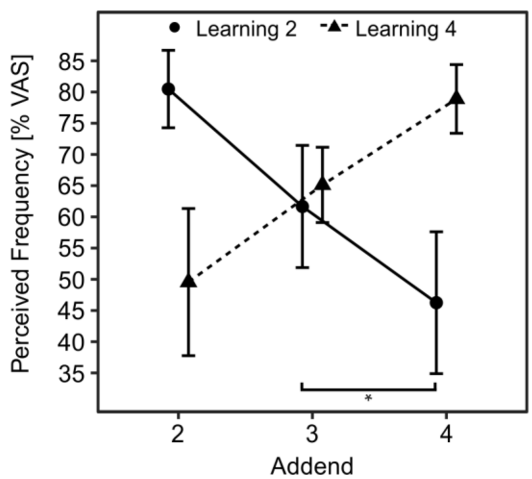


Figure 11 Perceived frequency of alphanumerical equations during the learning block. Note that this graph is based on data from a follow-up study (see text). VAS: visual analogue scale

In sum, this exploration shows that participants did not consciously report an Ad-dend-specific metacognitive strategy to select between internal and external information retrieval. However, data from the follow-up analysis shows that participants are sensitive to the Learning manipulation, which is a prerequisite for any Addend-specific selection strategy. We conclude that the sensitivity to the Learning manipulation might have contributed to strategy selection, though likely not on a conscious level. Instead, strategy selection might have been influenced implicitly by the familiarity of the specific items (Onyper et al., 2006; Schunn et al., 1997).

(2) *Do participants rely on performance monitoring to establish adaptive external information retrieval?* If participants relied on performance monitoring, adaptive use of external information should emerge only after participants got the chance to compare their performance when using internal with their performance when using external information retrieval. To address this exploratory hypothesis, we looked at the time course of external information retrieval in the choice block. First, to reduce noise in the visual representation, we averaged across four adjacent trials for each Addend separately. Second, we ran two ANOVAs with the within-participants factor Addend on the averages of the first four trials separately for both Learning conditions. For this analysis, only participants in the mixed group were used. We did not run one combined ANOVA instead since we were not interested in a possible main effect of Learning or the interaction between Learning and Addend.

Results indicate that Addend significantly influenced external information retrieval during the first four choice trials for both Learning 2 ($F(2, 54) = 10.13$, $p = 1.84 \times 10^{-4}$, $\eta_G^2 = .108$) and Learning 4 ($F(2, 42) = 6.74$, $p = .00290$, $\eta_G^2 = .104$). Post-hoc dependent t-tests confirmed that participants in the Learning 2 condition used the external information less with respect to the control condition (i.e., Addend 3) when solving equations for whose solutions high internal accessibility had been established ($t(27) = 3.34$, $p = 0.00248$, $M_{+3} - M_{+2} = 21.4$ %); **Figure 12a**. The same difference was trending for participants in the Learning 4 condition ($t(21) = 2.02$, $p = 0.0563$, $M_{+3} - M_{+4} = 12.1$ %)) group; **Figure 12b**. Given that decreased external information retrieval for well-learned solutions was exhibited already at the very beginning of the choice block, these results let it appear unlikely that the adaptive behavior was based on performance monitoring during the choice block. [10]

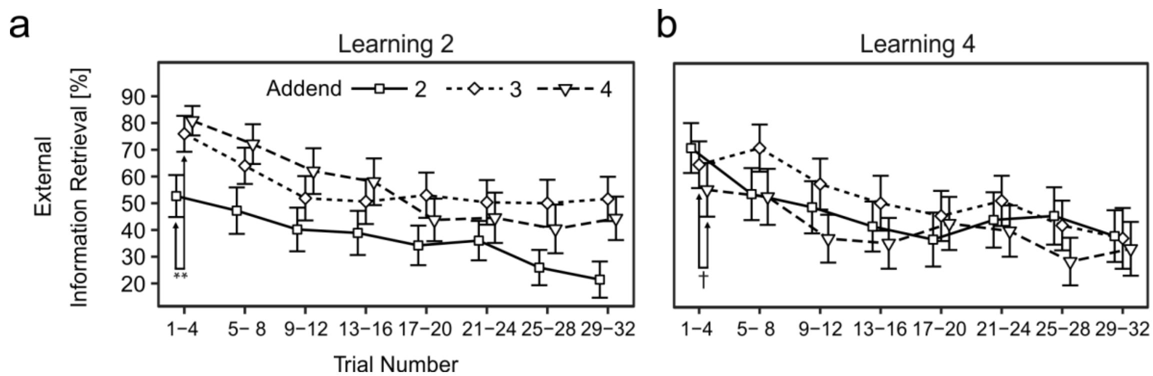


Figure 12 Time course of external information retrieval during choice block for the Learning 2 (a) and the Learning 4 (b) condition. Error bars depict SEM. ** $p < .01$, † $p = .06$

(3) *Do participants rely on parallel strategy use to establish the adaptive external information retrieval?* Participants might have executed internal and external strategies at the same time and have used the solution of whichever strategy was finished first (i.e., a race between strategies Compton & Logan, 1991; Logan, 1988). If so, our

[10] Also note that the negative slope exhibited in **Figure 12** suggests an ongoing increase in efficiency of internal information retrieval. To allow the interested reader to inspect the time course of performance during the choice block, we also provide graphs illustrating inverse efficiency, reaction time, as well as accuracy in the supplemental materials (**Figures 14**, **15**, and **16**, respectively; note the roughly constant internal performance over time despite the decreasing use of external information retrieval).

finding of de-creased external information retrieval for equations with high internal accessibility might be due to the fact that internal solution retrieval was faster than accessing the solution externally (i.e., faster than moving the cursor from the starting position to the black box). To investigate this post-hoc hypothesis, we compared (a) the count of trials in which participants did not move the cursor for even a single pixel to (b) the count of trials in which participants started to move the cursor but did not reach the black box and to (c) the count of trials in which the cursor was moved inside the black box.

Out of all trials across all participants, a substantial proportion acquired the solution internally without even initiating the external strategy (a: 3168 trials or 33.7 %). Thus, in a third of trials, participants did unambiguously not rely on parallel processing. This upfront preference for internal strategies might be due to an initial strategy selection phase that is informed by higher-level metacognitive evaluations (e.g., being convinced of one's own abilities, Gilbert, 2015; or being suspicious about the usefulness of the external strategy, Weis & Wiese, 2019) or lower-level item-specific or strategy-specific learning (as claimed by the CMPL, Rickard, 1997; and the ASCM, Siegler & Lemaire, 1997 models). If we are willing to assume that participants have not started moving the mouse without cognitive intent (e.g., due to muscle jitter), results however also show that participants did sometimes use both strategies in parallel (b: 615 trials or 6.5 %). This result is in line with a study by Walsh and Anderson (2009): arithmetic problem solvers sometimes started moving the mouse towards a screen-based calculator but changed their trajectory towards the answer box before reaching the calculator. Lastly, as already indicated by the main analysis, the external strategy was fully executed for most trials (c: 5619 trials or 59.8%). However, with the current data, it is impossible to tell whether participants followed a purely external or a parallel strategy during these trials. Nevertheless, the existence of parallel strategy execution (b; Walsh & Anderson, 2009) and the fact that some participants mentioned that they used the external strategy specifically to save effort (see Exploration: Do participants evaluate individual learning frequencies metacognitively and adjusted external information retrieval accordingly) makes it likely that participants made use of both purely external and parallel options, though the exact proportions cannot be determined with the current data. The present analysis thus suggests the existence of all possible, i.e. purely internal, purely external, and parallel, processing strategies.

(4) *Did participants mostly rely on a sequential strategy to establish the adaptive external information retrieval?* Participants might have tried to recall the solution first internally and only in a second step consider other options. Such a sequential strategy has been suggested to be the "best of both worlds": it does improve retrieval from internal memory even when the recall fails and still makes use of the external strategy to omit costly internal strategies like counting (Pyke & LeFevre, 2011). Here, we investigate this exploratory hypothesis using mouse movement onset data. The general idea is that if participants showed different mouse movement onsets depending on their internal information accessibility (compare **Figure 10b**) it would speak for some sequential mechanism. Else, there would be no reason for onsets to

be different. We now present one possible underlying mechanism for differential mouse movement onsets:

The higher the internal information accessibility, the faster and more accurate internal retrieval becomes (e.g., Logan, 1988; Ratcliff, 1978). However, analogously, the higher internal information accessibility is, the earlier a problem solver might stop the retrieval process and continue with another (in the present case, internal counting or external mouse movement) strategy. In other words, if one "knows" that internal accessibility should be high for a specific solution, one would predict an earlier retrieval success and might be willing to declare the retrieval as unsuccessful earlier. Such a prediction would also be made by a 2-choice diffusion model with the choices 'retrieval' and 'retrieval error': error RT with a high drift parameter (as for frequently learned alphanumeric equations) would be predicted to be lower than error RT with a low drift parameter (as for infrequently learned alphanumeric equations; Ratcliff & McKoon, 2008). Note that this sort of decision process would imply that participants learned to associate specific features (e.g. "+4") with specific drift rates. Such an assumption is not implausible as problem solvers are known to be sensitive to "featural data" (Siegler & Lemaire, 1997, p. 72) and to use such data to inform their cognitive strategy selection for novel problems (as claimed by the ASCM, e.g. Siegler & Lemaire, 1997). Following that rationale, we hypothesize that if participants relied on a sequential strategy, they should have started moving the mouse earlier for well-trained equations than for less trained equations.

A mixed ANOVA (Addend x Learning) and one-sided post-hoc dependent t-tests were used to test this explorative hypothesis. Trials with mouse movement onsets deviating more than three standard deviations from the individual mean were filtered. Only participants in the mixed group were used for this analysis. Six participants were excluded because they moved the mouse less than three times in at least one of the three Addend conditions. We chose a threshold of three trials as a liberal criterion to avoid noisy individual estimates.

In accordance with the hypothesis, Addend and Learning interacted in their influence on mouse movement start ($F(2, 84) = 3.37$, $p = .0390$, $\eta_G^2 = .018$); **Figure 13a**. Post-hoc tests confirmed that participants in the Learning 2 condition started moving the mouse earlier for equations with Addend 2 than with Addend 4 ($t(25) = 2.37$, $p = 0.0128$, $M_{+4} - M_{+2} = 89$ ms)). The reverse comparison was trending for participants in the Learning 4 condition ($t(17) = 1.47$, $p = 0.0805$, $M_{+2} - M_{+4} = 55$ ms)). These results suggest that our problem solvers, at least in some trials, used a sequential approach in which they first tried to recall the solution from internal memory and only then started considering other options. Given the exploratory nature of this analysis, this result should not be over-interpreted.

To explore possible differences between groups (see **Figure 9**) we also include an analogous ANOVA for the external group. Results showed no interaction effect ($F(2, 60) = .41$, $p_{GG} = .596$, $\eta_G^2 < .001$); **Figure 13b**. The main effect of Learning was trending ($F(1, 30) = 3.69$, $p = .0642$, $\eta_G^2 = .11$) and there was no main effect of Addend ($F(2, 60) = .29$, $p_{GG} = .672$, $\eta_G^2 < .001$). The results for the external group suggest that these participants did not rely on a sequential strategy when deciding which strategy to use. Results rather suggest that participants decided for the external strategy early on and further on did not try to rely on internal retrieval at all.

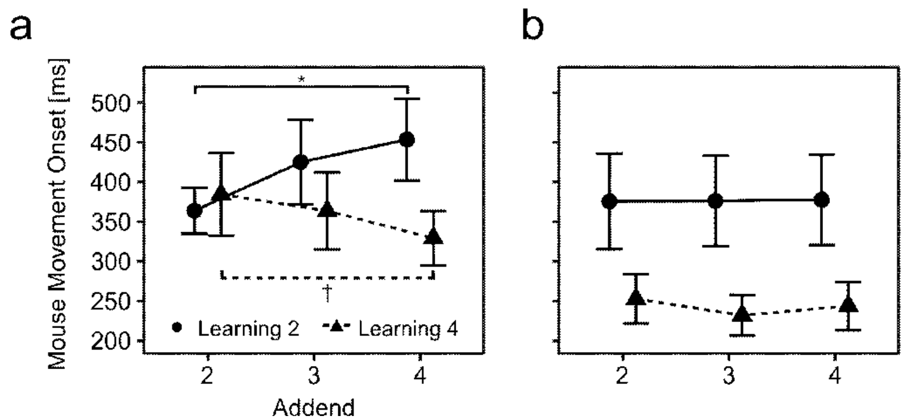


Figure 13 Mouse movement onset in the choice block. Data is shown separately for the mixed (a) and the external (b) group. Error bars depict SEM. * $p = .01$, † $p = .08$

Discussion

In the present study, a novel human-computer-interaction paradigm was used to investigate whether problem solvers choose between internal and external information retrieval based on the accessibility of internal information. By and large, we found this to be the case: increasing the internal accessibility of a problem's solution decreased how frequently participants retrieved information externally. Crucially, this relationship was present even when visual features of a task would suggest the opposite (e.g., "A + 4" looks harder to solve internally than "A + 2" but is nevertheless solved more frequently internally when the solution's internal accessibility is high). Participants thus were sensitive to their internal information access and used that sensitivity to choose between internal and external cognitive strategies rather than using sensitivity-independent strategies based on the task's visual features. Four exploratory analyses were conducted that (1) suggest adaptive choice between internal and external information retrieval to be, at least in some instances, realized using a sequential "try internal retrieval first and then consider other options" heuristic; (2) are inconclusive about whether the decision between internal and external retrieval was additionally dependent on an implicit metacognitive process that evaluates differential learning frequencies or familiarity in the learning block, (3) let it appear unlikely that the choice was dependent on performance monitoring during the choice block, and (4) suggest that internal and external retrieval are not executed in parallel in some but might be in other instances.

Results are consistent with previous studies suggesting humans to be impartial about whether to use internal or external strategies for cognitive processing (Gray & Fu, 2004; Gray et al., 2006; Morgan et al., 2009), ultimately preferring the strategy with the lower costs. Results extend their findings by providing possible underlying mechanisms of strategy choice. Results also support theories that proclaim the cognitive system to be rational and adaptive (e.g., Anderson, 1990) and question theories that proclaim a strong bias against mental effort (e.g. Ballard, Hayhoe, Pook, & Rao, 1997; Kool, McGuire, Rosen, & Botvinick, 2010).

How do problem solvers decide between internal and external cognitive strategies? Previous studies suggest that problem solvers might rely on knowledge about the efficiencies of differential internal strategies that might be partially based on the analysis of the problem's visual features (Bourne et al., 2010; Lemaire & Reder, 1999;

Siegler & Lemaire, 1997), on the familiarity of the specific problem (Schunn et al., 1997), or subjective judgments of own ability (Gilbert, 2015; Touron, 2015). Problem solvers might also rely on performance monitoring using error feedback and/or response time (for a review, see Ullsperger, Fischer, Nigbur, & Endrass, 2014) or follow less monitoring-sensitive strategies like trying to access the solution via internal and external strategies in parallel. Conversely, problem solvers might first try to recall the information from memory and, if that fails, access the solution via the external resource in a second step (for competing internal strategies, the parallel option seems more plausible; Logan, 1988). The current study allows us to compare the strategy choice process for the alphanumeric task at hand with what has been proposed in these previous studies:

(1) Given that no addend-specific strategies were reported in the questionnaire at the end of the study, we find it to be unlikely that our participants made their adaptive decision to use an external strategy based on a conscious metacognitive strategy. However, additional data showed that participants were sensitive to the addend-specific Learning manipulation (i.e., knew which types of equations they learned more and which ones they learned less frequently), which would enable them to base their strategy choice on the familiarity of the problem (i.e., use the external strategy for unfamiliar problems; Schunn et al., 1997). Thus, at this point, it is unclear whether our participants have implicitly used that sensitivity to inform strategy selection.

(2) We deem it unlikely that performance monitoring during the choice block was the prime cause for adaptive external strategy use given that the frequency of

external information retrieval was adjusted to the Learning condition right from the beginning of the choice block.

(3) It appears equally unlikely that participants consistently executed internal and external retrieval in parallel since in about a third of all trials the mouse was not moved at all. This clear preference for the internal strategy in about a third of the trials is consistent with findings of a study that used mouse trajectories to examine problem solvers' uncertainty about whether to use an internal or an external strategy: participants had to solve math equations and could either move their mouse towards a calculator first (i.e., external strategy) or immediately towards the answer box (i.e., internal strategy). Though participants sometimes adjusted their mouse movement throughout the trial, they also frequently and unambiguously preferred not to use the external resource at all, as indicated by a mouse trajectory directly leading towards the answer box without any curvature towards the calculator (Walsh & Anderson, 2009).

(4) We investigated whether our problem solvers might have used a sequential mechanism where participants retrieve the solution externally only if internal retrieval had failed beforehand. This sequential mechanism was supported by our data: in accordance with theoretical predictions based on a diffusion model with the outcomes "retrieval" or "retrieval error" (see Results: Exploration (4) Did participants mostly rely on a sequential strategy to establish the adaptive external information retrieval?), our problem solvers started external retrieval

(i.e., mouse movements) earlier for frequently than for infrequently trained equations. A similar sequential mechanism has been reported for strategy choice in mental arithmetic: people likely first tried to verify equations using the "five rule" before engaging in standard arithmetic operations (Lemaire & Reder, 1999). Interestingly, this effect vanished for participants that almost exclusively relied on external retrieval, thus suggesting a different mechanism for those participants.

Taken together, the exploratory findings suggest a divergence of how internal (e.g., memory retrieval and mental arithmetic) and competing mixed (e.g., memory retrieval and external retrieval) strategies are employed. Internal strategies might be more prone to be employed in parallel (Logan, 1988; but also see Lemaire & Reder, 1999, for sequential pro-cessing) while the existence of even comparably easy external strategies like the one used in the current paradigm might encourage solitarily strategy use (current study). Future studies are needed to consolidate this finding and should also address the underlying reasons, for example the possibility that external strategies are oftentimes too resource-draining to allow simultaneous execution of internal strategies.

Though we deem it likely that a sequential mechanism was frequently used for strategy selection in the current study, we want to stress that we do by no mean deny the existence of other mechanisms, for example:

(1) Monitoring accuracy and time feedback during the learning block. However, note that speed- and accuracy-related performance does not exhaustingly predict strat-egy selection (Gray et al., 2006; Risko et al., 2014; Walsh &

Anderson, 2009; Weis & Wiese, 2019), which was also true in the present study: even though par-ticipants in the Learning 2 condition relied less on the external strategy when in-ternal accessibility of the solution was high (i.e., Addend 2) rather than medium (i.e., Addend 3), this behavior was not mirrored when comparing medium with low (i.e., Addend 4) accessibilities; **Figure 10a**.

(2) Metacognitive misconceptions about their performance. The behavior might not have been mirrored because of people's metacognitive misconceptions about their performance (Dunn & Risko, 2016; Pauszek & Gibson, 2018; Risko & Dunn, 2015; Weis & Wiese, 2019): our participants might have underestimated how slow they are at solving equations which they had little experience with.

Further research is needed to clarify the interplay of different parameters like moni-toring-based efficiency optimization and metacognitive misjudgments on external strategy use. Equally importantly, it is yet to be examined whether findings in the domain of declara-tive long-term memory, like in the current study, transfer to other areas of cognition like working memory or spatial navigation (see also Risko & Gilbert, 2016, p. 685) and if similar efficiency-dependent mechanisms of external strategy use hold when outsourcing memory to humans (i.e., transactive memory; Wegner, 1987) rather than computers. Lastly, we want to direct the reader's attention towards the fact that about half of our participants (i.e., the internal and external group) exhibited hardly any variance in strategy choice, which would be compatible with the view that many participants make a strategy choice once rather than at the beginning of each trial (similar to some participants in Bourne et al., 2010). Understanding the individual differences in cognitive strategy choice will be key to an improved understanding of how humans solve problems in cognitive environments (as discussed by Risko & Gilbert, 2016).

From an applied perspective, the present results inform possible intervention methods aimed at remediating external information retrieval in particular and, possibly, external strategy use in general. Currently, such intervention methods are hardly available (Risko & Gilbert, 2016, p. 685). From the current data, we can extrapolate that improving the efficiency of internal strategies should by itself suffice to remediate externalization behavior. A similar approach but altering the efficiency of external rather than internal strategies has been suggested by O'Hara and Payne (1998): increasing the time costs associated with using (an interface that was needed for) an external strategy encouraged more frequent internal strategy use. This mechanism of decreasing efficiency of an external or increasing efficiency of an internal strategy could be used to guide externalization behavior whenever internal have more favorable attributes than external strategies. For example, bolstering internal strategy use can be important and beneficial when external strategies are regularly unavailable or when insightful knowledge transfer is needed, the latter of which can oftentimes only be achieved internally. A proof of concept for this mechanism, but targeting the efficiency of external rather than internal strategies, was provided by O'Hara and Payne (1998). Analogously, internal strategies should intentionally not be relied upon when they have unfavorable properties. For example, in internal memory, similar stimuli are often grouped together to reduce representational complexity (Nosofsky, 1992), leading to a decreased ability to discriminate attributes associated with these similar stimuli and decreased performance in discrimination-based tasks (i.e., similar stimuli are basically underrepresented since they are conceived as one rather than two separate entities). One possibility to restore performance is to rely on external strategies without biased representations (Fu, 2011). Providing easy access to the external strategy (Fu, 2011) and avoiding learning-induced increases in the efficiency of the internal strategy (current study) would guide the user towards the beneficial reliance on external strategies. We thus argue that the efficiency of both internal and external strategies can be intentionally manipulated as to maximize specific performance outcomes in cognitive tasks.

The present findings also have implications for problem solvers in static cognitive environments where properties like task difficulty or external information accessibility cannot be changed. For example, the speed of accessing information online depends heavily on the quality of the internet connection and cannot be directly controlled by the problem solver. In contrast, changing internal information accessibility oftentimes depends on deploying appropriate mental strategies that can be taught or discovered (Lemaire & Lecacheur, 2001) as well as on learning (e.g., Logan, 1988; or the present study). Both options leave the initiative with the problem solver rather than some extrinsic force like network coverage. The present finding that human problem solvers adjust their use of externally stored information based on the internal accessibility of that information speaks for the human ability to proficiently exploit a technologized environment for their own benefit. Establishing high internal information accessibility is thus a viable option to become less dependent on that environment whenever desired.

Acknowledgments

We thank our undergraduate RA Caitlin Mahoney for collecting the data. All data and the R analysis script are available at https://osf.io/scbj3/.

Conclusion

In an increasingly computerized future, being an efficient problem solver in interac-tive environments will gain importance. Former studies have shown that increasing the time needed to access externally stored information increases reliance on brain-based information. Here, we strengthen the prevalent notion that increasing the efficiency of brain-based information retrieval increases reliance on brain-based information in an analogous manner and supply possible underlying mechanisms. Our study thereby increases the understanding of human behavior in interactive settings that afford external information storage.

References

Anderson, J. R. (1987). Skill acquisition: Compilation of weak-method problem situations. *Psychological Review*, 94(2), 192.

Anderson, J. R. (1990). *The Adaptive Character of Thought*. https://doi.org/10.4324/9780203771730

Ballard, D. H., Hayhoe, M. M., Pook, P. K., & Rao, R. P. (1997). Deictic codes for the em-bodiment of cognition. *Behavioral and Brain Sciences*, 20(4), 723–742.

Bourne, L. E., Raymond, W. D., & Healy, A. F. (2010). Strategy selection and use during classification skill acquisition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 36(2), 500–514. https://doi.org/10.1037/a0018599

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

Clark, A. (1999). An embodied cognitive science? *Trends in Cognitive Sciences*, 3(9), 345–351.

Compton, B. J., & Logan, G. D. (1991). The transition from algorithm to retrieval in memory-based theories of automaticity. *Memory & Cognition*, 19(2), 151–158. https://doi.org/10.3758/BF03197111

Dixon, R. A., & de Frias, C. M. (2004). The Victoria Longitudinal Study: From Characterizing Cognitive Aging to Illustrating Changes in Memory Compensation. *Aging, Neuropsychology, and Cognition*, 11(2–3), 346–376. https://doi.org/10.1080/13825580490511161

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

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G* Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, 39(2), 175–191.

Fu, W.-T. (2011). A Dynamic Context Model of Interactive Behavior: Cognitive Science. *Cognitive Science*, 35(5), 874–904. https://doi.org/10.1111/j.1551-6709.2011.01173.x

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

Hines, J., Hertzog, C., & Touron, D. (2012). A prelearning manipulation falsifies a pure associational deficit account of retrieval shift during skill acquisition. *Aging, Neuropsychology, and Cognition*, 19(4), 449–478. https://doi.org/10.1080/13825585.2011.630718

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.

Howes, A., Duggan, G. B., Kalidindi, K., Tseng, Y.-C., & Lewis, R. L. (2016). Predicting Short-Term Remembering as Boundedly Optimal Strategy Choice. *Cognitive Science*, 40(5), 1192–1223.

Hoyer, W. J., Cerella, J., & Onyper, S. V. (2003). Item learning in cognitive skill training: Effects of item difficulty. *Memory & Cognition*, 31(8), 1260–1270. https://doi.org/10.3758/BF03195809

Kool, W., McGuire, J. T., Rosen, Z. B., & Botvinick, M. M. (2010). Decision making and the avoidance of cognitive demand. *Journal of Experimental Psychology: General*, 139(4), 665–682. https://doi.org/10.1037/a0020198

Lemaire, P., & Lecacheur, M. (2001). Older and younger adults' strategy use and execution in currency conversion tasks: Insights from French franc to euro and euro to French franc conversions. *Journal of Experimental Psychology: Applied*, 7(3), 195–206. https://doi.org/10.1037//1076-898X.7.3.195

Lemaire, P., & Reder, L. (1999). What affects strategy selection in arithmetic? The example of parity and five effects on product verification. *Memory & Cognition*, 27(2), 364–382. https://doi.org/10.3758/BF03211420

Logan, G. D. (1988). Toward an instance theory of automatization. *Psychological Review*, 95(4), 492.

Logan, G. D., & Klapp, S. T. (1991). Automatizing Alphabet Arithmetic: I. Is Extended Practice Necessary to Produce Automaticity*? Journal of Experimental Psychology: Learning, Memory, and Cognition*, 17(2), 179–195.

Morgan, P. L., Patrick, J., Waldron, S. M., King, S. L., & Patrick, T. (2009). Improving memory after interruption: Exploiting soft constraints and manipulating information access cost. *Journal of Experimental Psychology: Applied*, 15(4), 291–306. https://doi.org/10.1037/a0018008

Neth, H., & Payne, S. J. (2001). Addition as interactive problem solving. *Proceedings of the Annual Meeting of the Cognitive Science Society*, 23.

Norman, D. A. (1988). *The psychology of everyday things.(The design of everyday things)*. Basic Books.

Nosofsky, R. M. (1992). Exemplar-based approach to relating categorization, identification, and recognition. In F. G. Ashby (Ed.), *Scientific psychology series. Multidimensional models of perception and cognition* (pp. 363–393). Hillsdale, NJ, US: Lawrence Erlbaum Associates, Inc.

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

Onyper, S. V., Hoyer, W. J., & Cerella, J. (2006). Determinants of retrieval solutions during cognitive skill training: Source confusions. *Memory & Cognition*, 34(3), 538–549.

Pauszek, J. R., & Gibson, B. S. (2018). The Least Costs Hypothesis: A rational analysis ap-proach to the voluntary symbolic control of attention. *Journal of Experimental Psychology: Human Perception and Performance*, 44(8), 1199.

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

Pyke, A. A., & LeFevre, J.-A. (2011). Calculator use need not undermine direct-access ability: The roles of retrieval, calculation, and calculator use in the acquisition of arithmetic facts. *Journal of Educational Psychology*, 103(3), 607–616. https://doi.org/10.1037/a0023291

Ratcliff, R. (1978). A Theory of Memory Retrieval. *Psychological Review*, 85(2), 59–108. https://doi.org/10.1037/0033-295X.85.2.59

Ratcliff, R., & McKoon, G. (2008). The Diffusion Decision Model: Theory and Data for Two-Choice Decision Tasks. *Neural Computation*, 20(4), 873–922. https://doi.org/10.1162/neco.2008.12-06-420

Rickard, T. C. (1997). Bending the power law: A CMPL theory of strategy shifts and the automatization of cognitive skills. *Journal of Experimental Psychology: General*, 126(3), 288.

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.

Robbins, P., & Aydede, M. (2009). A short primer on situated cognition. In *The Cambridge Handbook of Situated Cognition*, 3–10.

Schunn, C. D., Reder, L. M., & Nhouyvanisvong, A. (1997). To Calculate or Not to Calcu-late: A Source Activation Confusion Model of Problem Familiarity's Role in Strategy Selection. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 23(1), 3–29. https://doi.org/10.1037/0278-7393.23.1.3

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.

Storm, B. C., Stone, S. M., & Benjamin, A. S. (2017). Using the Internet to access information inflates future use of the Internet to access other information. *Memory*, 25(6), 717–723.

Touron, D. R. (2015). Memory avoidance by older adults: When "old dogs" won't perform their "new tricks." *Current Directions in Psychological Science*, 24(3), 170–176.

Touron, D. R., & Hertzog, C. (2004). Distinguishing Age Differences in Knowledge, Strategy Use, and Confidence During Strategic Skill Acquisition. *Psychology and Aging*, 19(3), 452–466. https://doi.org/10.1037/0882-7974.19.3.452

Townsend, J. T., & Ashby, F. G. (1978). Methods of modeling capacity in simple processing systems. *Cognitive Theory*, 3, 200–239.

Ullsperger, M., Fischer, A. G., Nigbur, R., & Endrass, T. (2014). Neural mechanisms and temporal dynamics of performance monitoring. *Trends in Cognitive Sciences*, 18(5), 259–267. https://doi.org/10.1016/j.tics.2014.02.009

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

Wegner, D. M. (1987). Transactive memory: A contemporary analysis of the group mind. In *Theories of group behavio*r (pp. 185–208). Springer.

Weis, P. P., & Wiese, E. (2019). Using Tools to Help Us Think: Actual But also Believed Reliability Modulates Cognitive Offloading. *Human Factors*, 61(2), 243–254. https://doi.org/10.1177/0018720818797553

Wilson, M. (2002). Six views of embodied cognition. *Psychonomic Bulletin & Review*, 9(4), 625–636.

Zbrodoff, N. J. (1995). Why is 9+ 7 harder than 2+ 3? Strength and interference as explana-tions of the problem-size effect. *Memory & Cognition*, 23(6), 689–700.

Supplementary material

ANOVA statistics for inverse efficiency

Table 6 ANOVA results for inverse efficiency

Factor	*DFn*	*DFd*	*F*	*p*	η_G^2
Inverse Efficiency					
Learning *	1	97	4.41	0.0221	0.027
Addend ***	2	194	54.59	$2.11 \times 10^{-17}{}_{GG}$	0.091
Block ***	1	97	197.05	4.28×10^{-25}	0.300
Learning x Addend ***	2	194	68.11	$1.48 \times 10^{-20}{}_{GG}$	0.111
Learning x Block *	1	97	6.10	0.0152	0.013
Addend x Block ***	2	194	53.52	$3.78 \times 10^{-15}{}_{GG}$	0.056
Learning x Addend x Block ***	2	194	42.97	$8.04 \times 10^{-13}{}_{GG}$	0.046

Notes. *** $p < 0.001$, * $p < 0.05$, $_{GG}$ Greenhouse-Geisser-corrected value

Influence of Learning on reaction time and accuracy

Learning, Addend, and Block interacted in their influence on reaction time ($F(2, 194) = 55.25$, $p_{GG} = 2.45 \times 10^{-16}$; see **Table 7** in the supplemental materials). The three-way interaction reflects the nontrivial consequences of introducing the external resource in the choice block. More specifically, dependent post-hoc t-tests revealed that participants in the Learning 2 condition were, as expected, quicker in solving equations with Addend 2 in comparison to the control equations with Addend 3 ($t(50) = 11.34$, $p = 1.95 \times 10^{-15}$, $M_{Delta} = 1018$ ms) and slower in solving equations with Addend 4 in comparison to the control equations with Addend 3 ($t(50) = 8.37$, $p = 4.44 \times 10^{-11}$, M_{Delta}

= 872 ms); **Figure 8b**. The reverse was true for participants in the Learning 4 condition: they were quicker in solving equations with Addend 4 in comparison to the control equations with Addend 3 ($t(47) = 3.85$, $p = 3.53 \times 10^{-4}$, $M_{\mathrm{Delta}} = 377$ ms). However, they were also quicker in solving equations with Addend 2 in comparison to the control equations with Addend 3 ($t(47) = 4.92$, $p = 1.12 \times 10^{-5}$, $M_{\mathrm{Delta}} = 371$ ms), which might be due to the high efficiency of the counting strategy for equations with Addend 2 despite the lower efficiency of the retrieval strategy. There was no evidence for reaction time differences in equations with the control Addend 3 between Learning conditions (independent t-test: $t(97) = 1.62$, $p = 0.108$, $M_{+2} = 2952$ ms, $M_{+4} = 2692$ ms).

Learning interacted with Addend in its influence on accuracy ($F(2, 194) = 18.23$, $p_{GG} = 1.82 \times 10^{-7}$; see **Table 8** in the supplemental materials). Participants in the Learning 2 condition answered equations with Addend 2 more accurately than the control equations with Addend 3 ($t(50) = 3.88$, $p = 3.05 \times 10^{-4}$, $M_{\mathrm{Delta}} = 2.3$ %). The reverse was true for participants in the Learning 4 condition ($t(47) = 3.45$, $p = 0.00120$, $M_{\mathrm{Delta}} = 3.0$ %). Unexpectedly, accuracy also differed between Learning conditions for the control condition (i.e. Addend 3: $t(97) = 2.31$, $p = 0.0228$, $M_{+2} = 95.0$ %, $M_{+4} = 92.2$ %). Robust non-parametric Wilcoxon signed rank tests for the paired t-testes and a Wilcoxon rank sum test for the independent t-test were used to confirm the results of the t-tests. All p-values were significant at an alpha of 0.05 and of comparable magnitude.

In sum, these results disprove the existence of a within-participants speed-accuracy-tradeoff since higher speed is associated with higher accuracy rather than the reverse.

Table 7 ANOVA results for reaction time

Factor	*DFn*	*DFd*	*F*	*p*	η_G^2
Reaction Time					
Learning **	1	97	7.25	0.00837	0.041
Addend ***	2	194	73.61	1.63×10^{-24}	0.093
Block ***	1	97	235.94	1.01×10^{-27}	0.347
Learning x Addend ***	2	194	83.12	8.48×10^{-27}	0.104
Learning x Block *	1	97	6.35	0.0134	0.014
Addend x Block ***	2	194	82.68	5.78×10^{-22} $_{GG}$	0.065
Learning x Addend x Block ***	2	194	55.25	2.45×10^{-16} $_{GG}$	0.044

Notes. *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$, $_{GG}$ Greenhouse-Geisser-corrected value

Table 8 ANOVA results for accuracy

Factor	*DFn*	*DFd*	*F*	*p*	η_G^2
Accuracy					
Learning	1	97	3.92	0.0504	0.018
Addend *	2	194	4.87	0.0108 $_{GG}$	0.012
Block	1	97	0.03	0.856	< 0.001
Learning x Addend ***	2	194	18.23	1.82×10^{-7} $_{GG}$	0.0426
Learning x Block	1	97	1.12	0.2926	0.002
Addend x Block	2	194	0.32	0.7281	0.001
Learning x Addend x Block	2	194	2.23	0.1052	.004

Notes. *** $p < 0.001$, * $p < 0.05$, $_{GG}$ Greenhouse-Geisser-corrected value

Time course of inverse efficiency, accuracy, and reaction time in the choice block

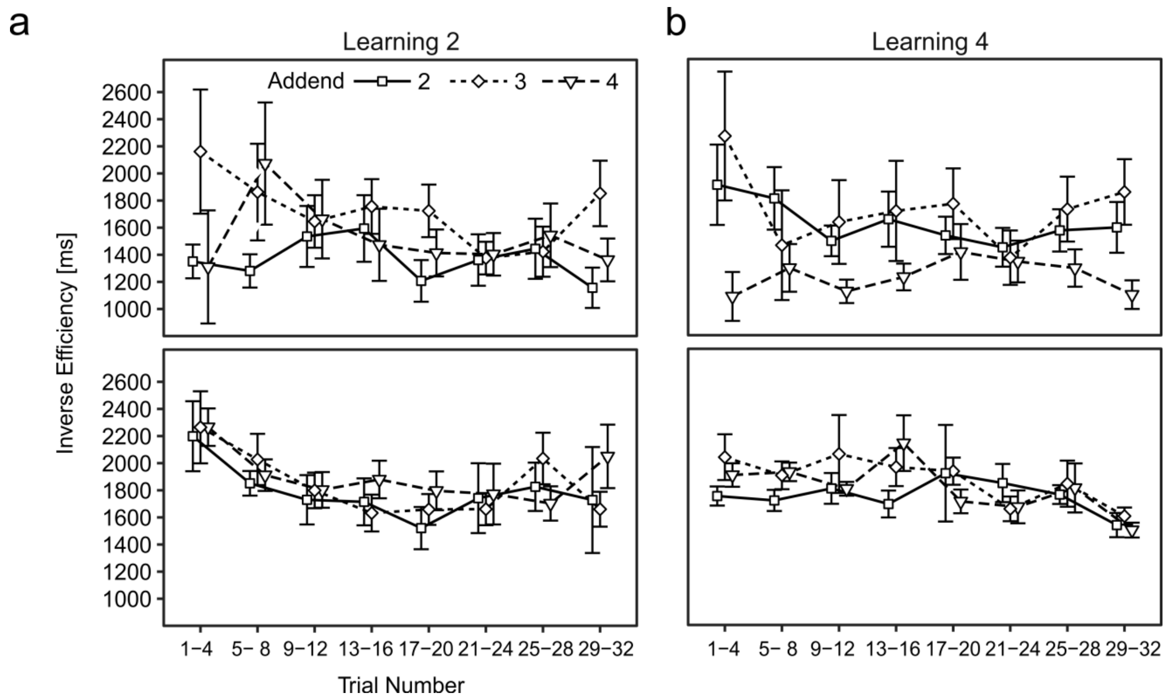


Figure 14 Time course of reaction times of correct trials during choice block for the Learning 2 (a) and the Learning 4 (b) condition. Answers given via internal retrieval are depicted in the top row, answers given via external retrieval are depicted in the bottom row. Error bars depict SEM

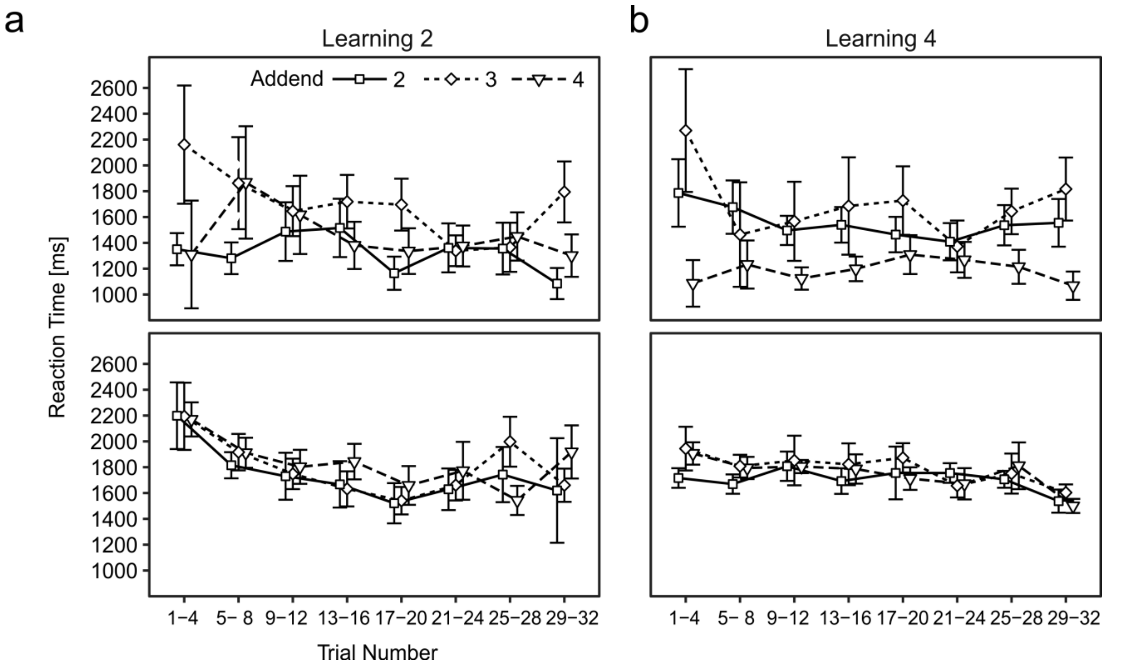


Figure 15 Time course of reaction times during choice block for the Learning 2 (a) and the Learning 4 (b) condition. Answers given via internal retrieval are depicted in the top row, answers given via external retrieval are depicted in the bottom row. Error bars depict SEM.

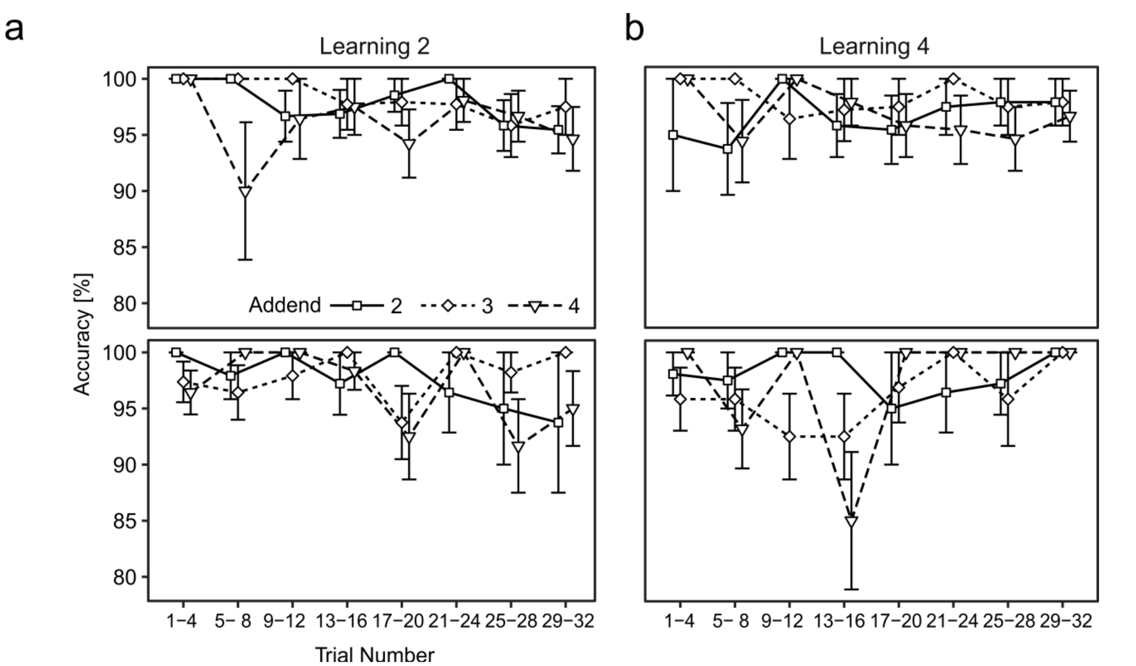


Figure 16 Time course of accuracies during choice block for the Learning 2 (a) and the Learning 4 (b) condition. Answers given via internal retrieval are depicted in the top row, answers given via external retrieval are depicted in the bottom row. Error bars depict SEM.

Study 3

Problem solvers adjust cognitive offloading based on performance goals

Patrick P. Weis & Eva Wiese
George Mason University, Fairfax, VA, USA

Author contributions: EW, PPW conceived and designed research. PPW performed research. PPW analyzed data. EW, PPW wrote the paper.

Weis, P. P., & Wiese, E. (in press). Problem Solvers Adjust Cognitive Offloading based on Performance Goals. *Cognitive Science*.

© 2019, Cognitive Science Society. This paper is not the copy of record and may not exactly replicate the final, authoritative version of the article. Please do not copy or cite without authors' permission. The final article will be available, upon publication, via its DOI: 10.1111/cogs.12802

Rationale

In this study, the same extension of the mental rotation paradigm (Shepard & Metzler, 1971) as used in Study 1 had been employed (see **Figure 17**). It was investigated whether human problem solvers exhibit different (Experiment 1) and adaptively goal-directed (Experiment 2) cognitive offloading behavior when confronted with different performance goals.

To this end, two parameters had been manipulated: The participants' performance goals (maximize speed or maximize accuracy) and the way our problem solvers were allowed to use the external resource (forced internal – they were instructed to solve the task internally; forced external – they were instructed to never rely on mental resources for the rotation process and always use the knob instead; free choice – they were free to choose between internal and external processing). The manipulations afforded investigating whether participants in the free choice condition strategically employed cognitive offloading to specifically boost goal-related behavior while possibly sacrificing goal-unrelated behavior. The forced conditions allowed us to gauge whether offloading could be beneficial for maximizing the respective performance goal in the free choice condition.

In a nutshell, results suggest that our participants did indeed employ cognitive offloading to boost goal-related behavior while sacrificing goal-unrelated behavior. Thus, at least in rather simple environments that provide direct performance feedback—as in the current study—, human problem solvers seem to be able to exploit their environments to pursue their cognitive goals. Complementing theories suggesting the importance of maximizing speed (Gray, Sims, Fu, & Schoelles, 2006) or minimizing mental effort (Kool, McGuire, Rosen, & Botvinick, 2010), the present results show that human problem solvers both sacrifice speed and invest mental resources in situations where it helps achieve their current cognitive performance goals. We thus argue that clear cognitive performance goals can supersede more generic heuristics like maximizing speed or minimizing effort (compare *(2)* in the section *Rationale of the dissertation project*).

Abstract

When incorporating the environment into mental processing (cf., *cognitive offloading*), one creates novel cognitive strategies that have the potential to improve task performance. Improved performance can, for example, mean faster problem solving, more accurate solutions, or even higher grades at university[11]. Although cognitive offloading has frequently been associated with improved performance, it is yet unclear how flexible problem solvers are at matching their offloading habits with their current performance goals (can people improve goal-related instead of generic performance, e.g., when being in a hurry and aiming for a “quick and dirty” solution?). Here, we asked participants to solve a cognitive task, provided them with different goals – maximizing speed (SPD) or accuracy (ACC), respectively – and measured how frequently (Experiment 1) and how proficiently (Experiment 2) they made use of a novel external resource to support their cognitive processing. Experiment 1 showed that offloading behavior varied with goals: participants offloaded less in the SPD than in the ACC

[11] Bocanegra, Poletiek, Ftitache, & Clark, 2019

condition. Experiment 2 showed that this differential offloading behavior was associated with high goal-related performance: fast answers in the SPD, accurate answers in the ACC condition. Simultaneously, goal-unrelated performance was sacrificed: inaccurate answers in the SPD, slow answers in the ACC condition. The findings support the notion of humans as canny offloaders who are able to successfully incorporate their environment in pursuit of their current cognitive goals. Future efforts should be focused on the finding's generalizability, e.g. to settings without feedback or with high mental workload.

Introduction

Saving a door code on the smartphone, outsourcing arithmetic to a calculator, or relying on cloud-based rather than brain-based knowledge: the contemporary ubiquity of computerized equipment has con-siderably increased the availability of external strategies to support human cognizing (e.g., Clark, 2004; Clowes, 2013; Dror & Harnad, 2008). Such incorporation of external resources into the cognitive repertoire can be quite rewarding as it can change a cognitive task's cost structure (Kirsh, 2010) and, if used wisely, improve task-related performance. In other words, internal and external strategies are associated with distinct performance profiles[12] (e.g.; Lemaire & Lecacheur, 2001; Risko, Medimorec, Chisholm, & Kingstone, 2014; Siegler & Lemaire, 1997; Touron & Hertzog, 2014; Walsh & Anderson, 2009), which makes it important to choose the right

[12] Please note that those performance profiles are not static. Performance profiles can change with increasing expertise and in many settings. For example, with increasing expertise, novel strategies that interleave internal and external processing can be discovered and used (Maglio & Kirsh, 1996). For a model that incorporates the effectiveness of different strategies over time in a problem-specific way, see Siegler and Lemaire (1997).

strategy at the right time. For example, outsourcing arithmetic to a calculator can be superior to mental processing because the former might afford increased speed and accuracy with respect to the latter (Siegler & Lemaire, 1997). In general, it was found that humans frequently (e.g., Gray, Sims, Fu, & Schoelles, 2006; Lemaire & Lecacheur, 2001; Walsh & Anderson, 2009) - though not always (Gilbert et al., 2019; e.g., Risko & Dunn, 2015; Touron, 2015; Weis & Wiese, 2019) - show high proficiency in mixing internal and external cognitive strategies. However, there is currently no con-sensus in the literature as to how humans achieve this proficiency (Risko & Gilbert, 2016, p 685; Anderson, 1990; Marewski & Schooler, 2011; Scaife & Rogers, 1996; Kirsh, 2013).

In the current paper, we focus on a hitherto neglected antecedent of a problem solver's decision to use an external strategy: performance goals. Affording the pursuit of a user's goal is a hallmark of humane technology; without it, a device would not empower but rather distract its users from what is important to them (Bosker, 2016). To shed light onto this topic of societal relevance, we used the current study to ask whether human problem solvers possess the skills to pursue their goals in technologically enhanced environments.

Cognitive offloading: Using the environment to (help us) think

The general idea of using cognitive strategies that incorporate a problem solver's environment to decrease brain-based processing costs is subsumed under the term cognitive offloading (Risko & Gilbert, 2016; for a review). Cognitive offloading overlaps with other approaches that also expand cognitive science's classic focus of what's happening inside the brain and include body (Embodied Cognition; e.g. Wilson, 2002) and environment (Situated Cognition; e.g. Kirsh, 2009; Extended Cognition; e.g. Clark & Chalmers, 1998; and Distributed Cognition; e.g. Hollan, Hutchins, & Kirsh, 2000)[13]. A related concept has been termed epistemic action, which is defined as an action undertaken to advance in a cognitive task rather than to alter the physical environment for non-cognitive purposes (Kirsh & Maglio, 1994)[14]. It should be noted that cognitive offloading can constitute very simple operations like replacing brain-based with paper-based retrieval or complex and dynamic operations like the ones that take place when a pilot is interacting with an airplane's cockpit (Hutchins, 1995). To reduce complexity, the current paper focuses on the former rather than the latter.

Are the problem solver's goals considered in the decision to offload cognition?

Goal-efficiency set aside, many studies suggest that human problem solvers are quite proficient in deciding when to offload cognition. For example, human problem solvers were shown to stop using exter-nal resources with high access costs (Gray et al., 2006; Walsh & Anderson, 2009), increase offloading with increased difficulty of the cognitive task (Experiment 5: Risko & Gilbert, 2016; Risko et al., 2014; Walsh & Anderson, 2009), decrease offloading if the external resource is unreliable (Weis &

---

[13] From a more philosophical perspective, it is currently debated whether epistemic actions directly replace internal cognitive processes (see *parity argument* and *extended mind hypothesis* in Clark & Chalmers, 1998; and *first wave extended mind* in Sutton, 2010) or complement and augment internal cognitive processing (*second wave extended mind*; Sutton, 2010).

[14] For example, reordering Scrabble tiles is an epistemic action as it unburdens working memory and thereby supports the cognitive task of finding words, possibly by providing a scaffold to start the word search from (Maglio, Matlock, Raphaely, Chernicky, & Kirsh, 1999).

Wiese, 2019), and are able to adjust a computer program based on their own memory capabilities (Howes, Duggan, Kalidindi, Tseng, & Lewis, 2016).

What is unclear at this point is whether humans are adaptive enough to adjust offloading based on their current goals. In most studies, only task difficulty (e.g., Risko & Dunn, 2015; Risko et al., 2014; Walsh & Anderson, 2009; Weis & Wiese, 2019) or accessibility of the external resource (e..g, Gray et al., 2006; Walsh & Anderson, 2009) was manipulated. Consequentially, they have not been sufficient to silence concern in the public (e.g., Bowles, 2018; Lewis, 2017) and the academic community (e.g., Turkle, 2012; Risko & Dunn, 2015; Weis & Wiese, 2019) about whether people are able to recruit external re-sources 'for their own good'. This concern seems reasonable because it can be hard to gauge whether seemingly proficient behavior is related to the problem solver's current needs and goals. That is, even though the way people use external resources might maximize speed (Gray et al., 2006) or monetary reward (Walsh & Anderson, 2009), it is hard to gauge whether that person's priority was to optimize for the respective metric in a goal-oriented manner (i.e., time or money, respectively) or used a generic cognitive processing approach instead (e.g., maximizing speed irrespective of current goals; Gray et al., 2006). People also do aim for optimizing different metrics in the same task (e.g., effort and accuracy; Risko & Dunn, 2015) and retroactively determining that metric is difficult. Lastly, problem solvers frequently prioritize local over global performance (Fu & Gray, 2006), making it difficult to infer whether poorly performing participants were unable to pursue their performance goals, pursued local rather than global goals, or had performance-independent goals like minimizing effort.

To make informed conclusions about the importance of problem solvers' goals for their decision to offload cognition, it is thus imperative to clearly communicate and manipulate these goals. Such informed conclusions are currently not available but would be highly valuable as they provided insight in how adaptively a human problem solver can navigate the cognitively enhanced environments of today and tomorrow.

Current investigation

In the present study, we controlled for well-established contributors to cognitive strategy selection (i.e., task difficulty and properties of the external resource) to investigate whether problem solvers are adaptive enough to adjust cognitive offloading based on their current goals. For this purpose, a novel human-computer-interaction paradigm has been developed (see *2.1.1: Extended Rotation Task*). Specifically, we provided participants with different performance goals and tracked whether they differed in how fre-quently they recruited an external resource (Experiment 1) and whether they were able to mix internal and external resources in a way compatible with their current goals (Experiment 2). If internal and external strategies differed in their goal-related performance profiles, we would expect participants to employ differential offloading behaviors when confronted with differential performance goals (*H1,* Experiment 1). This differential offloading behavior should be exhibited despite the availability of identical internal and external resources. Furthermore, if differential offloading behavior is exhibited, we expect it to be associ-ated with performance benefits specifically related to the current performance goal (*H2-1*) while possibly being associated with performance detriments related to performance metrics not relevant for the current goal (*H2-2*;

Experiments 2A and B). The hypotheses are described in more detail in the first paragraphs of sections 2 and 3.

Experiment 1: Free choice

Experiment 1 was conducted to investigate whether problem solvers employ differential offloading behaviors when confronted with differential performance goals (*H1*): in the accuracy goal condition, participants were incentivized for answering correctly; in the speed goal condition, participants were incentivized for answering fast.

Methods and materials

In total, 100 participants were recruited and assigned equally to an accuracy performance goal and a speed performance goal group. The final sample that entered data analysis consisted of 88 students (47 accuracy, 41 speed performance goal). More information on participants, apparatus, stimuli, procedure, and data filtering can be accessed in the *Supplemental Material*. Data and R analysis script are available through the Open Science Framework at https://osf.io/sh6qa/.

Extended Rotation Task

During each trial, participants had to engage in an expansion of the mental rotation paradigm (Shepard & Cooper, 1986; see also Shepard & Metzler, 1971), a task we termed *Extended Rotation Task* (see **Figure 17**; see Weis & Wiese, 2018a, 2019). In the original paradigm by Shepard and Metzler, the cognitive processes necessary to solve the task rely on mental resources only. In our expanded paradigm, computer-based external resources can be used to outsource the mental rotation part of these cognitive processes (see **Figure 17A**). Designing the external resource in a way that it affords offloading one specific cog-nitive process minimizes variance in usage behavior and sets the stage for researching physiological correlates in future studies.

Study design

The study follows a three-factorial design with the within-participants factors handedness of the working stimulus with respect to the base stimulus (same, opposite; **Figure 17B**) and angle (0°, 60°, 120°, 180°; **Figure 17B**), and the between-participants factor performance goal (speed, accuracy). In the opposite handedness condition, the working stimulus was first mirrored with respect to a vertical axis before the angle transformation took place (**Figure 17B**). Note that the 0° condition is used as baseline condition since the external resource only affords rotation, a cognitive process not necessary to solve problems in the 0° condition. The performance goal condition indicated whether participants were motivated to focus on speed or accuracy, respectively. In the accuracy goal condition, trial-based feedback was given with respect to accuracy only (correct/incorrect). In the speed goal condition, feedback was given with respect to speed for correct answers and with respect to accuracy for incorrect answers (speedy/slow/incorrect; compare **Figure 17B**). Accuracy feedback for incorrect answers had to be given in the speed condition as well to avoid complete negligence of accuracy and thus omitting performing the cognitive task at hand and instead only responding as quickly as possible. Speed feedback was based on a sliding window consisting of the reaction times of the preceding 32 trials. For responses given faster than the 85th percentile of those 32 trials participants received ‘Speedy’, for responses given slower than the 85th percentile participants received ‘Slow’ as feedback. Participants were also able to collect goal-specific points throughout the experiment. The best scoring participants were eligible for a monetary reward (for details, see *Supplemental Material*).

Each participant had to complete 576 trials: three repetitions for each of the twenty-four stimuli in each of the four angle and two handedness conditions. Trials were presented in three blocks, each consisting of 192 non-identical trials. Within blocks, trials were randomized. Every sixteen trials, participants were allowed to take a self-paced break. At the end of each block, participants were reminded that “it is not the best way to always rely on the mind’s eye or to always rely on the rotation knob. Try to use each way when it works best.”. Participants practiced the task for 32 trials with stimuli that ware not used in the main experiment. To get a crude idea of how tiring the extended rotation task is, participants were to fill out the Stanford Sleepiness Scale (Hoddes, Zarcone, & Dement, 1972) before and after the task. The Extended Rotation Task took between 40 and 60 minutes to complete.

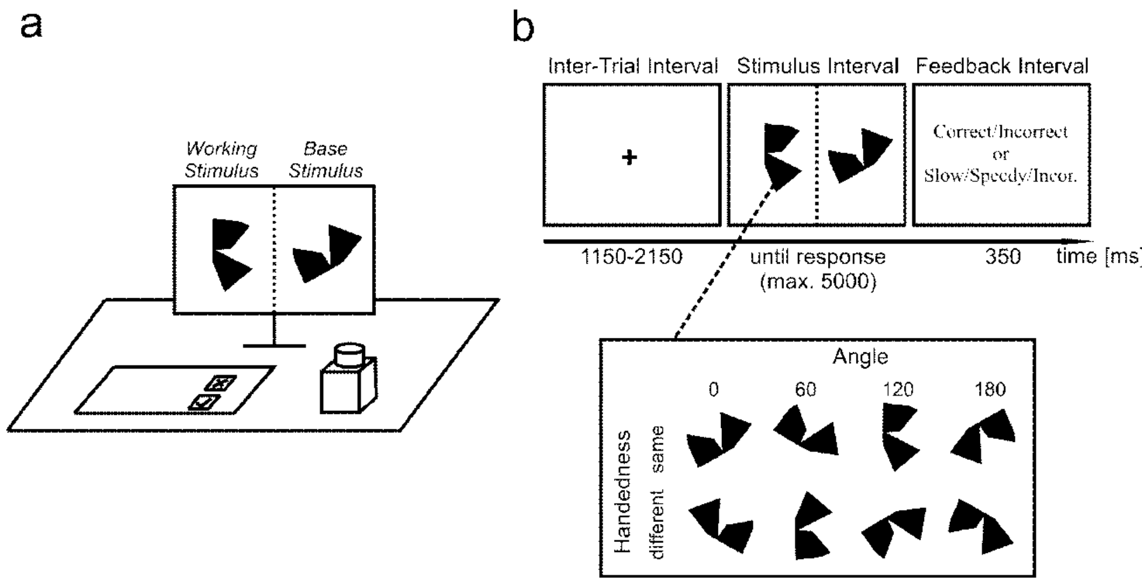


Figure 17 Extended Rotation Task. Participants have to compare the handedness of two stimuli that differ in their angular orientation and decide whether the left is the “same” (only rotated in 2D plane) or a “different” (first mirrored, then rotated in 2D plane) stimulus. For each shape, the base stimulus stays identical whereas the working stimulus is altered using a handedness and angle transformation. To help their decision, participants can offload their mental rotation process onto a physical knob as depicted in (a) that affords rotating the working stimulus on screen. During each trial, the base stimulus is presented in the right half and the working stimulus in the left half of the screen for five seconds or until a response was given (b). The figure is adapted from Weis & Wiese (2019) and depicts one out of twenty-four base stimuli used in this study.

Analysis

To determine whether participants offloaded the mental rotation process onto the knob, a binary variable was created on a trial-by-trial basis that indicated whether the stimulus on the screen was rotated for more than 5° (i.e., offloading) or less than 5° (i.e., no offloading). The threshold of 5° was chosen be-cause it allows simultaneous minimizing of (1) false alarms due to motor jitter and (2) false positives be-cause a rotation of less than 5° is unlikely to help cognitive processing even in the lowest 60° angle condi-tion. To analyze the offloading data, a random coefficient modeling approach that allowed to fit generalized linear models with a logit link function and two random effects, participants and stimuli, was used (for more details on this approach, see Supplemental Material). Models were implemented using R (Team, 2013) and the lme4 package's function glmer (Bates, Mächler, Bolker, & Walker, 2015). Marginal means were computed using the emmeans[15] package.[16]

Results and discussion

Unsurprisingly, both angle and handedness affected offloading ($|Z|$ = 13.1 and $|Z|$ = 12.4, respectively; for estimated marginal means, see **Figure 18**). More interestingly, changing the performance goal from accuracy to speed, when holding all other predictors constant, was associated with a 83% decrease in offloading odds ($odds_{accuracy}$ = 22.6, $odds_{speed}$ = 4.0; $|Z|$ = 4.9) or, equivalently, a drop of 16 percentage points in offloading probability ($p_{accuracy}$ = .96, p_{speed} = .80, see **Figure 18**). Similarly, but of less importance for the current purposes, changing the performance goal also changed the relationship between angle and offloading ($|Z|$ = 6.3) as well as between handedness and offloading ($|Z|$ = 6.1); for details concerning these interactions and other model results, see **Table 10**. To avoid redundancy, accuracy and RT data is reported with Experiment 2A and 2B (see **Figures 19 - 22**; data from Experiment 1 is labelled "free choice" since participants were able to freely choose between internal and external processing in Experiment 1). Increases in reported sleepiness from before to after the rotation task were comparable for

[15] https://CRAN.R-project.org/package=emmeans

[16] Note that the angle 0 condition is omitted in the main analyses as it is not relevant for offloading the mental rotation process. Analyses for the angle 0 condition can be found in the *Supplemental Material*, **Tables 9**, **11**, **13**, **16**, **20** and **Figures 24 - 28**, **30**.

both accuracy and speed goal conditions (independent t-test: $\Delta(\text{after-before})_{\text{speed}} = 1.00$, $\Delta(\text{after-before})_{\text{accuracy}} = 0.91$, $t(84) = 0.31$, $p = 0.76$). Reported difficulty of the extended rotation task was also comparable across goal conditions (independent t-test: $M_{\text{spd}} = 2.78$, $M_{\text{acc}} = 2.55$, $t(84) = 0.31$, $p = 0.76$; scale ranged from 1 to 5).[17]

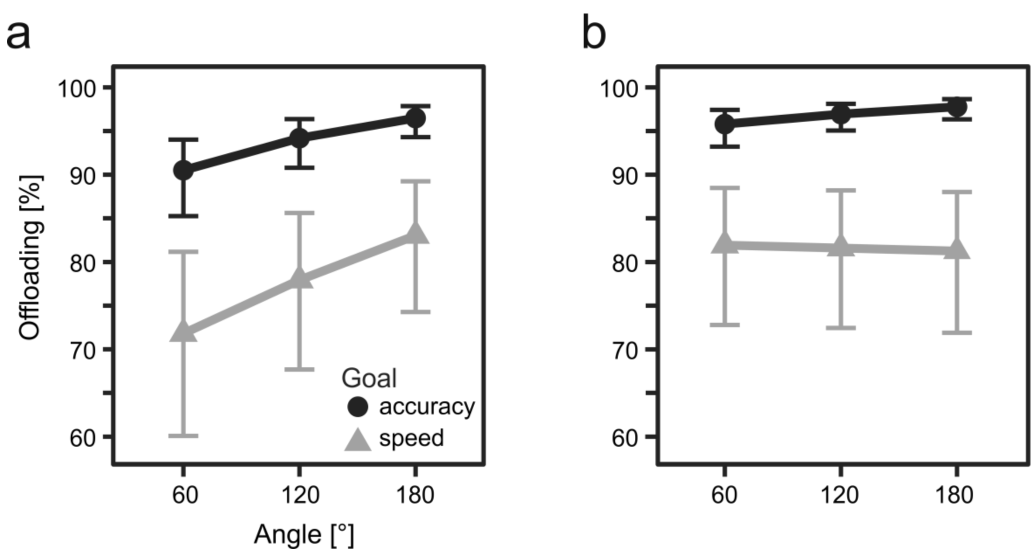


Figure 18 Model-based offloading proportions for different (a) and same (b) handedness. Error bars depict asymmetric 95% CIs that have been back-transformed from the logit scale.

In line with *H1*, problem solvers altered their cognitive offloading behavior based on their performance goals while the available internal and external resources were kept constant. Participants almost exclusively rotated externally when aiming for accuracy and

[17] Two participants had to be excluded from the sleepiness and one participant from the difficulty analysis due to missing data.

relied more on mental rotation when aiming for speed. Note that participants had a pronounced preference for external rotation for both conditions. Possible reasons include minimization of mental effort (Ballard, Hayhoe, Pook, & Rao, 1997), a generally more favorable performance profile of the external strategy, or a large proportion of participants who use internal and external strategies in parallel. Also note that we do not suggest accuracy goals to be always specifically associated with increased offloading. Instead, we conclude performance goals to have substantial impact on the way problem solvers mix internal and external strategies in general.

Experiment 2: Forced choice

The confirmation of *H1* laid the foundation for Experiment 2 in which we investigated whether the differences in offloading behavior exhibited in Experiment 1 were associated with goal-related performance gains. We asked one group of participants to solve the extended rotation task while exclusively relying on their internal resources without availability of an external resource (forced internal *cognition locus* condition) and another group of participants to exclusively rely on the external resource (forced external *cognition locus*)[18]. We then compared performances in these forced conditions to performance in the setting of Experiment 1 (free choice *cognition locus*). This way of comparing forced and free strategy choices has been termed the Choice/No-Choice Method (Siegler & Lemaire, 1997).

[18] Note that the forced external condition might include internal processing as well because participants might not always adhere to the instructions.

Specifically, we expect the offloading behavior exhibited in Experiment 1 to be associated with high goal-related performance. We expect that participants in the free choice condition (Experiment 1) should be at least as accurate as the more accurate of the two forced groups in the accuracy goal condition (Experiment 2A) and at least as fast as the faster of the two forced groups in the speed goal condition (Experiment 2B); *H2-1*. Additionally, we explore the possibility that participants in the free choice condition (Experiment 1) sacrificed performance in the metric not relevant for the current goal (i.e., sacrificed accuracy in the speed goal and speed in the accuracy goal condition); *H2-2*.

Methods and materials

More information on methods and materials, including information about participants, apparatus, stimuli, procedure, and data filtering can be accessed in the *Supplemental Material*. The final sample consisted of 77 students (41 forced external, 36 forced internal) in Experiment 2A and of 75 students (40 forced external, 35 forced internal) in Experiment 2B. Data and R analysis script are available through the Open Science Framework at https://osf.io/sh6qa/.

Design changes

Task and design were identical to Experiment 1 except that participants were not able to freely choose whether or not to recruit the external resource (factor *cognition locus*). Two experiments were conducted: participants were asked to be as accurate (i.e., the accuracy *performance goal* of Experiment 1) in Experiment 2A and to be as fast (i..e, the speed *performance goal* of Experiment 1) as possible in Experiment 2B.

Analysis

Accuracy *performance goal* data from Experiment 1 was added to the analysis of Experiment 2A and speed *performance goal* data from Experiment 1 to the analysis of Experiment 2B and labelled "free choice". The same data-analytic approach as in Experiment 1 has been employed. Note that Experiment 2 was conducted after Experiment 1 and participants were thus not randomly assigned to one of the three performance goal conditions, thereby introducing a possible confound (i.e., time point of data collection).

Experiment 2A: results and discussion (forced choice, accuracy goal)

Accuracy

In comparison to participants in the forced internal *cognitive locus* condition, when holding all other predictors constant, the odds of solving a problem correctly was increased by 116% for participants in the forced external and free choice conditions combined ($|Z| = 5.2$; $\text{odds}_{\text{choice}} = 14.2$, $\text{odds}_{\text{forced external}} = 17.8$, $\text{odds}_{\text{forced internal}} = 7.3$). Equivalently, when transforming the odds back to probability values, participants in the forced internal condition were about five percentage points less accurate than participants in the forced external and choice conditions ($p_{\text{choice}} = .95$, $p_{\text{forced external}} = .96$, $p_{\text{forced internal}} = .91$; p refers to the probability of answering accurately; **Figure 19**). Accuracies between forced external and choice conditions did not differ ($|Z| = 1.39$). The remaining model results are reported in **Table 12**. Increases in reported sleepiness ($\Delta(\text{after-before})_{\text{ext}} = 1.35$, $\Delta(\text{after-before})_{\text{int}} = 1.03$, $t(74) = 1.14$, $p = 0.26$) and reported difficulty of the extended rotation task (M_{ext} = 2.50, M_{int} = 2.53, $t(74) = -0.12$, $p = 0.90$; scale ranged from 1 to 5) were comparable for both forced cognition locus conditions.[19]

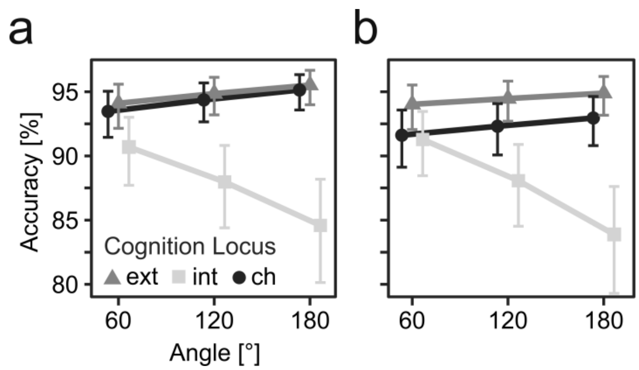


Figure 19 Model-based correct answer probabilities for different (a) and same (b) handedness in Experiment 2A. Error bars depict asymmetric 95% CIs that have been back-transformed from the logit scale. ext: forced external, int: forced internal, ch: free choice

In sum, accuracies in the free choice and forced external conditions were comparable while accuracy in the forced internal condition was considerably lower. Thus, participants in the choice condition employed a combination of internal and external resources that afforded high goal-related performance, suggesting an adaptive use of the external resource (confirming *H2-1*).

Speed

Analyzing RT in addition to accuracy data allows the exploration of whether participants in the choice condition sacrificed speed to achieve high accuracy. Such

[19] One participant had to be excluded from both sleepiness and difficulty analyses due to missing data.

behavior would speak for our participants' ability to choose cognitive strategies in a way that specifically maximizes goal-related rather than generic performance.

Results show that participants did indeed sacrifice speed to maximize accuracy: RTs in the forced external and free choice conditions were similar ($\Delta RT = 51$ ms, $|t| = 0.8$)[20] whereas participants in the forced internal condition answered considerably faster than participants in the external and choice conditions combined ($\Delta RT = 193$ ms, $|t| = 3.4$). Results are also in accordance with the classical finding by Shepard and Metzler (1971) that reaction time increases linearly with angle ($|t| = 27.5$)[21], which can be seen as a manipulation validation. The remaining model results are reported in **Table 14** and illustrated in **Figure 20**. To further illuminate the choice process, we analyzed the onset of external processing: in the free choice condition, participants started using the knob more than 200ms later than in the forced external condition (**Table 15**, **Figure 27**); this suggests either a sequential processing approach or a costly choice process and is discussed in section *7.6* of the *Supplemental Material*. It also suggests that participants in

[20] $|t|$ refers to the absolute value of the Wald statistic as reported by R's lme4 package (Bates, Mächler, Bolker, & Walker, 2015). Here, the *t*-value can be used to gauge whether RTs between conditions are similar or different. Where binary interpretation is necessary, we use a $|t| > 2$ criterion to infer difference rather than similarity.

[21] More precisely, a one standard deviation increase in angle was associated with a 73 ms increase in reaction time. Since one standard deviation equals 49 degrees in our experiment, a one degree increase in angle is associated with a 1.5 ms increase in reaction time. Please note that this value refers to the main effect, holding the interaction effects constant.

the forced external condition followed instructions as they did not exhibit the internal, roughly 200ms-lasting, processing participants in the choice condition engaged in.

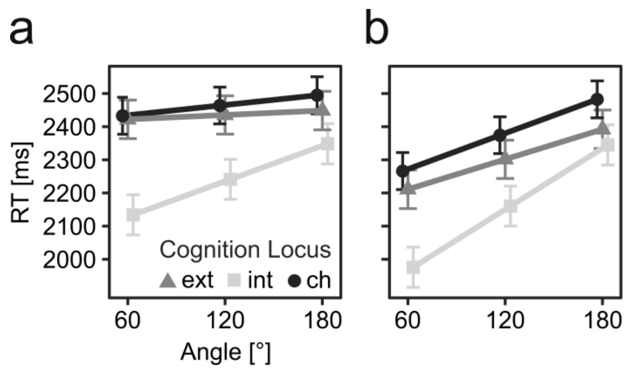


Figure 20 Model-based reaction time estimates for different (a) and same (b) handedness in Experiment 2A. Error bars depict 95% CIs. ext: forced external, int: forced internal, ch: free choice

In sum, RTs in the free choice and forced external conditions were comparable whereas RT in the forced internal condition was considerably lower. Thus, problem solvers in the choice condition have employed a combination of internal and external resources that sacrificed goal-irrelevant performance (confirming *H2-2*). The result also shows that a problem solver's inclination to optimize for speed (Gray et al., 2006; see also Weis & Wiese, 2018) can be superseded by conflicting goals.

Experiment 2B: results and discussion (forced choice, speed goal)

One might argue that the extensive offloading of 96% in the free choice accuracy goal condition (**Figure 18**) was only accidentally related to benefits in goal-related performance while the true underlying motivation was different (e.g., minimizing mental effort; Ballard et al., 1997; Kool, McGuire, Rosen, & Botvinick, 2010; or because incremental feedback on the display when offloading is preferred over no feed-back when not offloading; Fu & Gray, 2004). The purpose of Experiment 2B is to confirm the results of Experiment 2A by investigating whether participants that could freely choose in the speed performance goal condition of Experiment 1 exhibited high goal-related performance despite considerably less offloading (i.e., 80% instead of 96%).

Speed

Participants were equally fast in the forced internal and the free choice conditions (ΔRT = 19 ms, $|t| = 0.3$) whereas participants in the forced external condition were responding considerably slower than participants in the other two conditions combined (ΔRT = 146 ms, $|t| = 2.4$); **Figure 21**. The remaining model results are reported in **Table 17**. Participants in forced internal condition reported higher increases in sleepiness ($\Delta(\text{after-before})_{int} = 1.35$, $\Delta(\text{after-before})_{ext} = 0.78$, $t(72) = -2.14$, $p = 0.04$)[22] and higher difficulty of the extended rotation task ($M_{ext} = 2.37$, $M_{int} = 3.18$, $t(73) = 3.00$, $p = 0.004$; scale ranged from 1 to 5) than participants in the forced external cognition locus condition. As in Experiment 2A, we also analyzed the onset of external processing: in contrast to Experiment 2A, participants in the free choice condition started using the knob equally fast as in the forced external condition (**Table 22**, **Figure 31**); this suggests a non-sequential processing approach and is discussed in in section *8.6* of the *Supplemental Material*.

[22] One participant had to be excluded from the sleepiness analysis due to missing data.

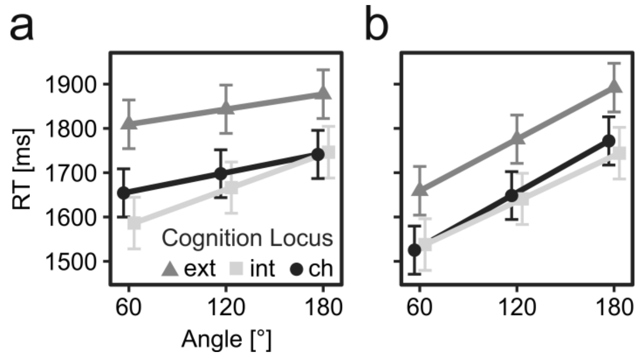


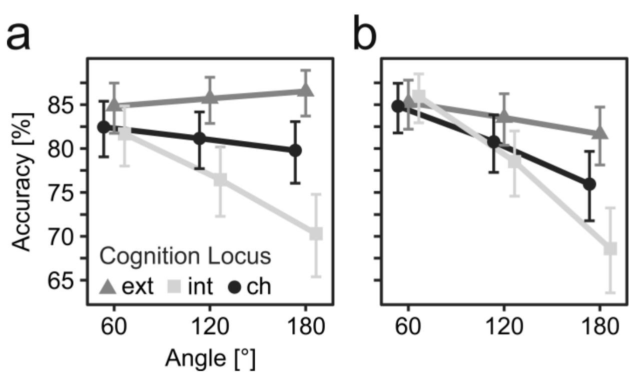
Figure 21 Model-based reaction time estimates for different (a) and same (b) handedness in Experiment 2B. Error bars depict 95% CIs. ext: forced external, int: forced internal, ch: free choice

In sum, RTs in the free choice and forced internal conditions clustered together and were considerably lower than RT in the forced external condition (which further confirms *H2-1*). Interestingly, the exploratory analyses of sleepiness and difficulty of the extended rotation task both suggest internal resource use to be more taxing than external resource use. Note that such a difference could only be shown for the speed goal, not for the accuracy goal condition and that participants were, when given the choice, less likely to offload when speed instead of accuracy was incentivized. This suggests that participants did not offload cognition merely to minimize effort but instead offloaded cognition to meet their performance goals. Lastly, also note that participants were nearly 150ms slower when solving the task externally 100% of the time (forced external) in comparison to solving it externally 0% of the time (forced internal) but also in comparison to solving it externally 80% of the time (free choice). This pattern suggests that adaptively switching strategies in the minority of only 20% of trials made up for the majority of the RT difference, which could have possibly been realized by monitoring strategy performance in a stimulus- (i.e., feature-specific) way (as proposed in the

ASCM; e.g. Siegler & Lemaire, 1997). This possibility is backed by a supplemental analysis that shows that participants were about 265ms faster when solving problems internally in the free choice in comparison to the forced internal condition (**Tables 18**, **19**; **Figure 29**).

Accuracy

In comparison to participants in the forced external condition, when holding all other predictors constant, the odds of solving a problem correctly was decreased by 31% for participants in the forced internal and free choice conditions combined ($|Z| = 3.8$; $odds_{choice} = 4.3$, $odds_{forced\ external} = 5.5$, $odds_{forced\ internal} = 7.3$;). Equivalently[23], participants in the forced external condition were about five percentage points more accurate than participants in the forced external and free choice conditions combined ($p_{choice} = .81$, $p_{forced\ external} = .85$, $p_{forced\ internal} = .78$). Accuracy for the forced internal and the free choice condition did not differ ($|Z| = 1.9$). Remaining model results are reported in **Table 21** in the *Supplemental Material* and illustrated in **Figure 22**.



[23] when transforming the odds back to probability values

Figure 22 Model-based accuracy estimates for different (a) and same (b) handedness in Experiment 2B.Error bars depict asymmetric 95% CIs that have been back-transformed from the logit scale. ext: forced external, int: forced internal, ch: free choice

These results again suggest that problem solvers in the free choice condition employed a combination of internal and external resources that sacrificed goal-irrelevant performance (further conforming *H2-2*). Cognition locus main effects of all experiments combined are summarized in **Figure 23**.

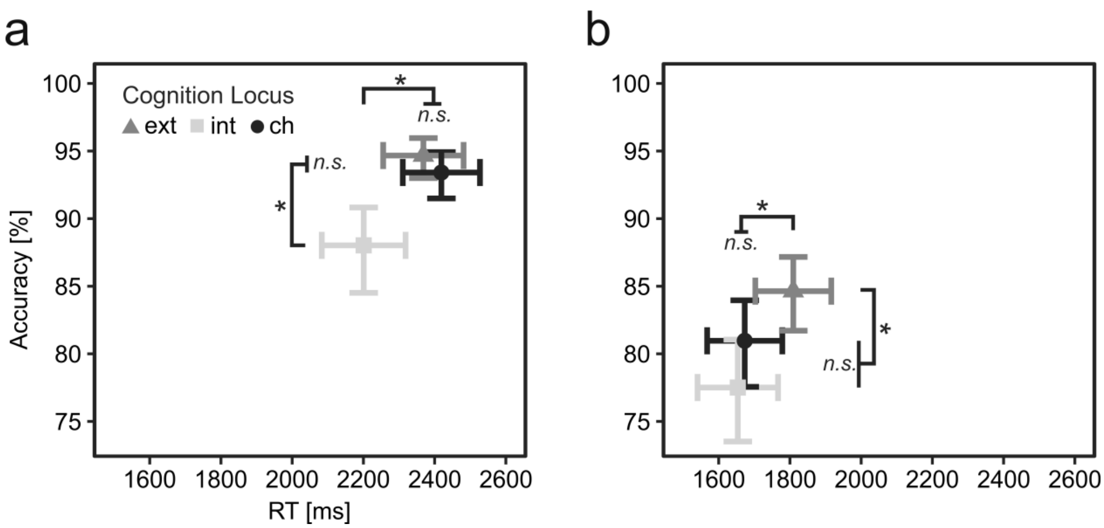


Figure 23 Performance summary of Experiments 1, 2A, and 2B. Data represents estimated marginal means for the accuracy (a) and speed (b) goal conditions. ext: forced external, int: forced internal, ch: free choice, *: |*t*| or |*Z*| >= 2; *n.s.*: |*t*| or |*Z*| < 2

General discussion

We asked participants to solve a cognitive task, provided them with different performance goals – maximizing speed or accuracy, respectively – and measured how frequently (Experiment 1) and how proficiently (Experiment 2) they made use of a novel external resource to support their cognitive processing. Results showed that participants with different performance goals indeed exhibited differential offloading frequencies (*H1*), which reflected the participants' proficiency in distributing cognition between internal and external resources in a goal-directed manner. In particular, the way participants mixed internal and external resources led to high goal-related performance (*H2-1*) whereas goal-unrelated performance was sacrificed *(H2-2)*. In other words, participants were specifically concerned with goal-related rather than generic performance.

How much guidance do problem solvers need to choose between internal and external resources to meet their cognitive goals?

The study's main purpose was to find out whether humans possess the ability to autonomously exploit their technologized environments in pursuit of their cognitive goals. The promising main takeaway is that participants were performing very well without any external guidance in the current extended rotation paradigm. This suggests the human to be capable of proficiently navigating a world with a steadily in-creasing number of possibilities to offload cognition. Humans can thus not only reach high levels of ge-neric performance when mixing internal and external strategies (Gilbert, 2015; Gray et al., 2006; Risko & Dunn, 2015; Walsh & Anderson, 2009) but they can also reach high levels of goal-related performance (current study). Clear performance goals and a steady feedback might be all that is needed for deciding on how to mix internal and external resources. The present results are in line with the notion of the human as an independent and rational problem solver (*cf*. Anderson, 1990).

However, one should be aware that the present finding of good goal-related performance without guidance might not generalize to all possible external resources and environments. For some situations, it is already known that guidance is beneficial, for instance when a problem solver once read faulty information about an external resource's performance. In such a situation, the problem solver would likely include that faulty information in his or her decision whether to use that resource (Weis & Wiese, 2019), leading to poor performance. To alleviate the consequences of false beliefs, verbal advice concerning the preferable strategy, given immediately before solving a problem, was shown to improve offloading performance (Gilbert et al., 2019). In addition to faulty beliefs, one should also consider the impact of the complexity of an environment on the problem solver's ability to proficiently recruit external resources. In complex environments, it can be hard to gauge the utility of a strategy because the associated reward might not be immediate or obvious, or because a wide variety of strategies could be employed and too much effort would be needed to obtain solid estimates of each strategy's utility (*cf*. Lieder & Griffiths, 2017). In a similar vein, it should be noted the cognitive environment available in the present study afforded only one obvious external strategy (i.e., knob-based rotation) and that challenges in other cognitive environments might include discovery of unknown or creation of novel external strategies (as, for example, in the cognitive environment of the computer game TETRIS®; Kirsh & Maglio, 1994). Lastly, it is uncertain whether performance feedback played an important role for establishing the adaptive offloading behavior. Thus, so far, one can conclude that in the absence of faulty beliefs and complex environments, a condition that was likely met in the current study, and the presence of performance feedback, humans are able to employ a well-performing and goal-directed mix of internal and external cognitive strategies without further guidance.

How do problem solvers establish a goal-driven recruitment of external resources?

To the authors, it is intriguing how the participants realized the goal-driven incorporation of the external strategy into their cognitive processing. Two possible underlying mechanisms exist. First, participants might have focused on the goal-related feedback, i.e. used the feedback as an error signal to improve subsequent behavior (i.e. *performance monitoring*; for a review, see Ullsperger, Fischer, Nigbur, & En-drass, 2014). For example, other research suggests that older adults can use accuracy feedback to overcome a bias against using their internal memory (Touron & Hertzog, 2014). Similarly, participants might have monitored their errors and timing independently from the displayed feedback. Second, participants might have made correct metacognitive judgments about the capabilities of the available cognitive resources (for a review, see Risko & Gilbert, 2016). In other words, participants might have metacognitively evaluated the different strategies a priori and opted for the more promising one. Such metacognitive judgments are likely employed (Dunn & Risko, 2016; Weis & Wiese, 2019) but not without fault (Gilbert et al., 2019; Risko & Dunn, 2015)[24]. A third possibility would be that participants chose the path of least effort (Kool et al., 2010) and

[24] Note that such evaluations can be made independently from the actual performance of the respective resource (Gilbert, 2015) but that combined strategies in which participants factor performance feedback into their metacognitive evaluations are also plausible.

ended up with good choice performance more or less by chance, which is however a highly unlikely possibility in the current study (for more details, see 3.3.1).

From the present data, we cannot distinguish the contributions of performance monitoring and met-acognitive evaluations. However, we deem it likely that both mechanisms contributed simultaneously, which has already been proposed for situations in which problem solvers can select between internal and external strategies (Gilbert, 2015; Risko & Gilbert, 2016) and in which they can select between different internal strategies (for a review, see Lieder & Griffiths, 2017). Further studies that capture participants' a priori metacognitive evaluations of different strategies and that track strategy selection and associated performance over time could illuminate the importance of both performance monitoring and metacognitive evaluation for goal-oriented strategy selection. Lastly, it is important to realize that a proficient problem solver is not only able to adaptively choose between given external strategies but is also able to create and use novel strategies in a highly adaptive way (Kirsh, 2013), a topic that is out of the scope of the current article.

Conclusion and outlook

The current findings support the notion of humans as canny offloaders who are able to employ environment-based strategies to pursue their cognitive goals. Such proficiency seems important in an increasingly computerized world that affords an abundance of environment-based strategies. Future efforts should be focused on the mechanisms that underlie the choice to offload and on further illuminating the circumstances under which problem solvers need guidance to fulfill their goals.

Acknowledgments

We thank Chanelle Roegler and Sahar Miller for their help in data collection.

References

Anderson, J. R. (1990). The Adaptive Character of Thought. https://doi.org/10.4324/9780203771730

Ballard, D. H., Hayhoe, M. M., Pook, P. K., & Rao, R. P. (1997). Deictic codes for the embodiment of cognition. *Behavioral and Brain Sciences*, 20(4), 723–742.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Bocanegra, B. R., Poletiek, F. H., Ftitache, B., & Clark, A. (2019). Intelligent problem-solvers externalize cognitive operations. *Nature Human Behaviour*, 3(2), 136–142. https://doi.org/10.1038/s41562-018-0509-y

Bosker, B. (2016, November). The Binge Breaker. *The Atlantic*. Retrieved from https://www.theatlantic.com/magazine/archive/2016/11/the-binge-breaker/501122/

Bowles, N. (2018, August 7). Is the Answer to Phone Addiction a Worse Phone? *The New York Times*. Retrieved from https://www.nytimes.com/2018/01/12/technology/grayscale-phone.html

Clark, A. (2004). *Natural-born cyborgs: Minds, technologies, and the future of human intelligence*. Oxford University Press.

Clark, A., & Chalmers, D. (1998). The extended mind. *Analysis*, 7–19.

Clowes, R. W. (2013). The Cognitive Integration of E-Memory. *Review of Philosophy and Psychology*, 4(1), 107–133. https://doi.org/10.1007/s13164-013-0130-y

Dror, I. E., & Harnad, S. (2008). Offloading cognition onto cognitive technology. In *Cognition Distributed: How cognitive technology extends our minds* (pp. 1–23). John Benjamins Publishing.

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

Fu, W.-T., & Gray, W. (2004). Resolving the paradox of the active user: Stable suboptimal performance in interactive tasks. *Cognitive Science*, 28(6), 901–935. https://doi.org/10.1016/j.cogsci.2004.03.005

Fu, W.-T., & Gray, W. (2006). Suboptimal tradeoffs in information seeking. *Cognitive Psychology*, 52(3), 195–242. https://doi.org/10.1016/j.cogpsych.2005.08.002

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

Gilbert, S. J., Bird, A., Carpenter, J. M., Fleming, S. M., Sachdeva, C., & Tsai, P.-C. (2019). Optimal use of reminders: Metacognition, effort, and cognitive offloading. *Journal of Experimental Psychology: General*.

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.

Howes, A., Duggan, G. B., Kalidindi, K., Tseng, Y.-C., & Lewis, R. L. (2016). Predicting Short-Term Remembering as Boundedly Optimal Strategy Choice. *Cognitive Science*, 40(5), 1192–1223.

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

Kirsh, D. (2009). Problem Solving and Situated Cognition. In *The Cambridge Handbook of Situated Cognition* (pp. 264–306).

Kirsh, D. (2010). Thinking with external representations. *AI & SOCIETY*, 25(4), 441–454. https://doi.org/10.1007/s00146-010-0272-8

Kirsh, D. (2013). Embodied cognition and the magical future of interaction design. *ACM Transactions on Computer-Human Interaction*, 20(1), 1–30. https://doi.org/10.1145/2442106.2442109

Kirsh, D., & Maglio, P. (1994). On distinguishing epistemic from pragmatic action. *Cognitive Science*, 18(4), 513–549.

Kool, W., McGuire, J. T., Rosen, Z. B., & Botvinick, M. M. (2010). Decision making and the avoidance of cognitive demand. *Journal of Experimental Psychology: General*, 139(4), 665–682. https://doi.org/10.1037/a0020198

Lemaire, P., & Lecacheur, M. (2001). Older and younger adults' strategy use and execution in currency conversion tasks: Insights from French franc to euro and euro to French franc conversions. *Journal of Experimental Psychology: Applied*, 7(3), 195–206. https://doi.org/10.1037//1076-898X.7.3.195

Lewis, P. (2017, October 6). "Our minds can be hijacked": The tech insiders who fear a smartphone dystopia. *The Guardian*. Retrieved from https://www.theguardian.com/technology/2017/oct/05/smartphone-addiction-silicon-valley-dystopia

Lieder, F., & Griffiths, T. L. (2017). Strategy selection as rational metareasoning. *Psychological Review*, 124(6), 762.

Maglio, P. P., & Kirsh, D. (1996). Epistemic action increases with skill. *Proceedings of the Eighteenth Annual Conference of the Cognitive Science Society*, 16, 391–396. Erlbaum.

Maglio, P. P., Matlock, T., Raphaely, D., Chernicky, B., & Kirsh, D. (1999). Interactive skill in Scrabble. *Proceedings of the Twenty-First Annual Conference of the Cognitive Science Society*, 326–330. Routledge.

Marewski, J. N., & Schooler, L. J. (2011). Cognitive niches: An ecological model of strategy selection. *Psychological Review*, 118(3), 393.

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.

Scaife, M., & Rogers, Y. (1996). External cognition: How do graphical representations work? *International Journal of Human-Computer Studies*, 45(2), 185–213.

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.

Sutton, J. (2010). Exograms and Interdisciplinarity: History, the Extended Mind, and the Civilizing Process. In R. Menary (Ed.), *The Extended Mind* (pp. 189–225). https://doi.org/10.7551/mitpress/9780262014038.003.0009

Team, R. C. (2013). *R: A language and environment for statistical computing*.

Touron, D. R. (2015). Memory avoidance by older adults: When "old dogs" won't perform their "new tricks." *Current Directions in Psychological Science*, 24(3), 170–176.

Touron, D. R., & Hertzog, C. (2014). Accuracy and Speed Feedback: Global and Local Effects on Strategy Use. *Experimental Aging Research*, 40(3), 332–356. https://doi.org/10.1080/0361073X.2014.897150

Turkle, S. (2012). *Alone together: Why we expect more from technology and less from each other*. Basic books.

Ullsperger, M., Fischer, A. G., Nigbur, R., & Endrass, T. (2014). Neural mechanisms and temporal dynamics of performance monitoring. *Trends in Cognitive Sciences*, 18(5), 259–267. https://doi.org/10.1016/j.tics.2014.02.009

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

Weis, P. P., & Wiese, E. (2018). Speed Considerations Can Be of Little Concern When Outsourcing Thought to External Devices. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 62, 14–18. SAGE Publications Sage CA: Los Angeles, CA.

Weis, P. P., & Wiese, E. (2019). Using Tools to Help Us Think: Actual But also Believed Reliability Modulates Cognitive Offloading. *Human Factors*, 61(2), 243–254. https://doi.org/10.1177/0018720818797553

Wilson, M. (2002). Six views of embodied cognition. *Psychonomic Bulletin & Review*, 9(4), 625–636.

Supplementary material

Apparatus

Participants were seated at a distance of about 80 cm in front of an ASUS VB198T-P 19-inch monitor set to a resolution of 1,280 × 1,024 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). Key presses 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 rota-tion of the rotation knob corresponded to one full rotation of the working stimulus on the screen. Based on a frame-wise video analysis, we could estimate the lag between knob rotation and stimulus rotation to be less than 20ms.

Stimuli

Creation of the twenty-four stimuli used for the Extended Rotation Task was based on a procedure described by Attneave and Amoult (1956) that was implemented in a Matlab script provided by Collin & McMullen (2002). We used Matlab version R2015b (The Mathworks, Inc., Natick, MA) to create the stimuli. All stimuli used for the current experiment differed only in the edge parameter which ranged from three to twenty-one (see **Figure 17** in the main manuscript for an example stimulus with eight edges).

Procedure

After arrival, participants were welcomed and seated in front of computer screens. Participants then pro-vided informed consent before engaging in computer-based tasks. Up to two participants were tested simultaneously. First, each participant completed the shortened versions of three complex working memory span tasks (reading span, operation span, and symmetry span; Oswald, McAbee, Redick, & Hambrick, 2015) and a brief task testing visuo-motor coordination (computer-ported version of the rotary pursuit task; Mueller & Piper, 2014). Participants then engaged in the extended rotation task and were instructed to answer both as accurately and as quickly as possible. To get a crude idea of how tiring the extended rotation task is, participants were to fill out the Stanford Sleepiness Scale (Hoddes, Zarcone, & Dement, 1972) before and after the task.

Additionally, after completion of the Extended Rotation Task, participants were to judge whether they found that "The task involving object rotation was easy" on a five-point Likert scale (strongly agree to strongly disagree). The experimental session ended with a brief demographic sur-vey. The Extended Rotation Task took between 40 and 60 minutes to complete. In total, the study took around 90 minutes to complete. Working memory span and visuo-motor-coordination data are of minor relevance for the current manuscript and will be explored elsewhere.

Data filtering and analysis

Trials in which participants answered slower or faster than their individual mean for all stimuli in the same angle condition plus/minus two standard deviations were excluded from all analyses (4.7% of trials in Experiment 1 and 2B, 4.5% in Experiment 2A). No other trials were filtered. For RT analyses, including offloading onset analyses, only correct trials[25] were used.

We used a random coefficient modeling approach that allowed us to fit generalized linear models with a logit link function and two random effects, participants and stimuli, to the offloading and accuracy data. The same approach was used to fit general linear models (without the logit link function) to reaction time data. We preferred a random coefficient modelling approach over the more standard ANOVA approach since the former affords increased power for the current experiment with two random variables (Judd, Westfall, & Kenny, 2017) and since one easily runs into assumption

[25] All RT-dependent analyses were also conducted without excluding incorrectly answered trials. Results were highly similar and would have led to the same conclusions as with the present analyses.

violations with the latter when using binary outcome data (Dixon, 2008). Models were implemented using R (R Core Team, 2013) and the lme4 package's function glmer (Bates, Mächler, Bolker, & Walker, 2015). Marginal means were computed using the emmeans[26] package. All models include two random effects incorporating random intercepts for participants and stimuli, respectively.

More specifically, three independent variables were entered in each analysis: handedness, angle, and performance goal. For each analysis, we were fitting three models, one with no, one with the two-way, and one with the two-way and the three-way interaction terms. Out of three models, the one with the lowest Bayesian Information Criterion (BIC) is reported. Note that the angle 0 condition is omitted in this analysis as it is not relevant for offloading the mental rotation process (see below for analyses of the angle 0 conditions). The factors handedness and performance goal have been contrast-coded. The factor angle was entered as continuous and z-standardized.

Points

Participants were able to gather goal-specific points throughout the experiment. Participants were instructed about how they could gather points in the respective performance goal condition before the start of the experimental trials. Points for the last block were shown during each of the self-paced breaks between blocks. The shown point value reflected the participant's performance during the preceding sixteen trials (accuracy goal: 100% of trials correct: 5 points; < 100 and ≥ 90% of trials correct: 2 points; < 90 and ≥ 70% of trials correct: 1 point; else: 0 points; speed goal: ≥ 95% of trials speedy: 5

[26] https://CRAN.R-project.org/package=emmeans

points; < 95 and ≥ 80% of trials speedy: 2 points; < 80 and ≥ 60% of trials speedy: 1 point; else: 0 points). Trials were classified as speedy even if answered incorrectly. However, if accuracy was below 67% for the sixteen trials, participants received no points and were prompted to increase their accuracy. This was necessary to avoid that participants skip the rotation process altogether (i.e., neither do it internally nor externally) in favor of random answers. The three participants with the overall highest scores were awarded Amazon vouchers (of values USD 15, 10, and 5 respectively).

Experiment 1

Participants

In total, 100 participants were recruited and assigned equally to an accuracy *performance goal* and a speed *performance goal* group. Nine participants (two accuracy goal, seven speed goal) were excluded due to extremely poor task performance (less than 80% correct answers when working and base stimulus, see **Figure 17A** in the main manuscript, were identical). Additionally, three participants were excluded because their mean RT deviated more than 2.5 SDs from the grand mean of participants with the same performance goal, resulting in a final sample of 47 (M_{age} = 19.8, $\text{range}_{\text{age =}}$ 18 to 27, 30 females, 2 left-handed) and 41 participants (M_{age} = 19.4, $\text{range}_{\text{age =}}$ 18 to 25, 22 females, 5 left-handed), respectively. Participants were recruited from the undergraduate psychology student pool at a large American University and received course credit for their participation. Participants were at least 18 years old and had normal or corrected to normal vision. The experiment was approved by the local ethics committee and participants provided informed consent prior to participation.

Offloading at angle 0

To assess possible baseline offloading differences, two generalized linear models with a logit link function, one with and one without the interaction term, were used to model binary offloading choices based on handedness and performance goal at angle 0. IVs have been contrast coded. Based on the Bayesian Information Criterion, the model with the interaction term is reported.

Same-handed stimuli were, given that the stimuli looked identical unsurprisingly, solved less frequently with help of the external resource than stimuli with different handedness ($|Z|$ = 25.5). Although there was no main effect of performance goal ($|Z|$ = 1.27), handedness and performance goal interacted with a difference between speed and accuracy goal only emerging for same handedness ($|Z|$ = 6.7). For model details, see **Table 9**. The model's estimates are illustrated in **Figure 24**.

Table 9 Generalized linear model results for cognitive offloading in Experiment 1 at angle 0

	Offloading (log odds)			
Random Effects	Variance	SD		
Participants	1.65	1.29		
Items	0.18	0.43		
Fixed Effects	Estimate[log odds]	Exp(Estimate)	SE	$|Z|$
Intercept	-1.73	0.18	0.17	10.38
Handedness	-1.34	0.26	0.05	25.52
Goal	0.36	1.43	0.28	1.27
Handedness x Goal	0.70	2.02	0.10	6.71

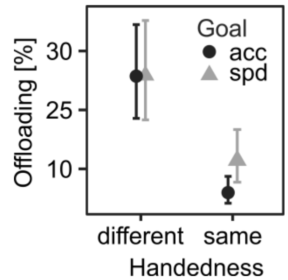


Figure 24 Model-based offloading estimates as predicted by handedness and performance goal at angle 0. Error bars depict asymmetric 95% Cis that are back-transformed from the logit scale. Goal: Performance Goal Factor; acc: accuracy, spd: speed

Offloading at angles 60 to 180

Table 10 Generalized linear model results for cognitive offloading in Experiment 1

Proportion Offloaded (log odds)				
Random Effects	Variance	SD		
Participants	2.73	1.65		
Items	0.07	0.26		
Fixed Effects	Estimate[log odds]	Exp(Estimate)	SE	|*Z*|
Intercept	2.25	9.47	0.19	12.06
Angle	0.24	1.27	0.02	13.11
Handedness	0.45	1.57	0.04	12.40
Goal	-1.74	0.18	0.36	4.86
Angle x Handedness	-0.22	0.80	0.04	6.23
Angle x Goal	-0.23	0.80	0.04	6.25
Handedness x Goal	-0.44	0.64	0.07	6.13
Angle x Handedness x Goal	-0.12	0.89	0.07	1.69

Notes. To interpret significance of a fixed effect, a |*Z*| > 2 criterion has been used.

Experiment 2A

Participants

The same sample size (50 forced internal, 50 forced external), exclusion criteria, recruitment, and ethical guidelines as in Experiment 1 were used. Two participants were excluded due to technical problems, twenty participants due to poor task performance, and one because RT mean deviated more than 2.5 SDs from the grand mean, resulting in a final sample of 41 participants in the forced external (M_{age} = 20.9, $range_{age\,=}$ 18 to 35, 25 females, 4 left-handed) and 36 in the forced internal (M_{age} = 20.9, $range_{age\,=}$ 18 to 50, 20 females, 3 left-handed) condition.

Accuracy at angle 0

To assess possible baseline accuracy differences, two generalized linear models with a logit link function, one with and one without the interaction term, were used to model binary accuracies based on handedness and cognition locus at angle 0. IVs have been contrast coded. Based on the Bayesian Information Criterion, the model with main effects only is reported.

Problems with same-handed stimuli were, given that the stimuli looked identical unsurprisingly, solved more accurately than problems with stimuli of opposite handedness ($|Z|$ = 8.0). Cognition locus did not alter accuracy ($|Z|$ of both contrasts < 2). For model details, see **Table 11**. The model's estimates are illustrated in **Figure 25**.

Table 11 Generalized linear mixed model results for accuracy in Experiment 2A at angle 0

	Accuracy (log odds)

Random Effects	Variance	SD		
Participants	0.64	0.80		
Items	0.13	0.36		
Fixed Effects	Estimate	Exp(Estimate)	SE	|*Z*|
Intercept	3.81	45.27	0.12	31.73
Handedness	0.72	2.05	0.09	7.96
Cognition Locus (ext + ch > int)	0.28	1.32	0.19	1.48
Cognition Locus (ch > ext)	0.30	1.35	0.21	1.42

Notes. ext: forced external, int: forced internal, ch: free choice.

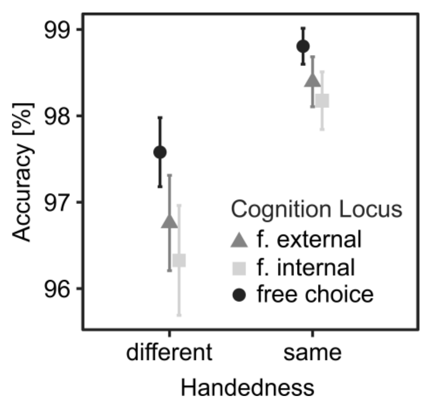


Figure 25 Model-based accuracy estimates as predicted by handedness and cognition locus at angle 0. Error bars depict asymmetric 95% CIs that are back-transformed from the logit scale. f. external: forced external, f. internal: forced internal

Accuracy at angles 60 to 180

Table 12 Generalized linear mixed model results for accuracy in Experiment 2A

	Accuracy (log odds)	
Random Effects	Variance	SD
Participants	0.53	0.72

Items	0.19	0.43		
Fixed Effects	Estimate	Exp(Estimate)	SE	|*Z*|
Intercept	2.51	12.28	0.11	22.50
Angle	-0.02	0.98	0.02	1.37
Handedness	-0.13	0.88	0.03	4.24
Cognition Locus (ext + ch > int)	0.77	2.16	0.15	5.20
Cognition Locus (ch > ext)	-0.22	0.80	0.16	1.39
Angle x Handedness	-0.05	0.95	0.03	1.63
Handedness x Cognition Locus (ext + ch > int)	-0.21	0.81	0.06	3.36
Handedness x Cognition Locus (ch > ext)	-0.26	0.77	0.08	3.18
Angle x Cognition Locus (ext + ch > int)	0.36	1.43	0.03	11.25
Angle x Cognition Locus (ch > ext)	0.01	1.01	0.04	0.23

Notes. ext: forced external, int: forced internal, ch: free choice. Free choice refers to the participants' free choice between using internal or external processing as available in Experiment 1.

RT at angle 0

To assess possible baseline RT differences, two general linear models, one with and one without the interaction term, were used to model binary accuracies based on handedness and cognition locus at angle 0. IVs have been contrast coded. Based on the Bayesian Information Criterion, the model with main effects only is reported.

Participants that had to use the external resource answered considerable slower in the angle 0 condition than participants who could freely choose between internal and external processing (169 ms, $|t| = 3.0$). We interpret that as evidence suggesting that participants in the forced external condition did closely follow the instructions, i.e. relied on the external resource to come to a conclusion. The time cost incurred by recruiting the external resource at angle 0 might be what is reflected by the RT difference. For the remaining effects and model details, see **Table 13**. The model's estimates are illustrated in **Figure 26**.

Table 13 General linear mixed model results for reaction time in Experiment 2A at angle 0

	RT [s]		
Random Effects	Variance	SD	
Participants	0.07	0.26	
Items	0.02	0.12	
Residual	0.21	0.45	
Fixed Effects	Estimate	SE	|*t*|
Intercept	1.54	0.03	44.65
Handedness	-0.33	0.01	46.67
Cognition Locus (ext + ch > int)	0.14	0.05	2.69
Cognition Locus (ch > ext)	-0.17	0.06	3.03

Notes. ext: forced external, int: forced internal, ch: free choice.

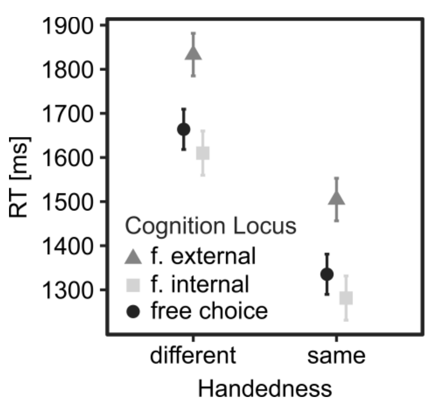


Figure 26 Model-based RT estimates as predicted by handedness and cognition locus at angle 0. Error bars depict asymmetric 95% CIs. f. external: forced external, f. internal: forced internal

RT at angles 60 to 180

Table 14 General linear mixed model results for reaction time in Experiment 2A

	RT [s]		
Random Effects	Variance	SD	
Participants	0.08	0.29	
Items	0.03	0.18	
Residual	0.32	0.56	
Fixed Effects	Estimate	SE	|*t*|
Intercept	2.32	0.04	52.55
Angle	0.07	< 0.01	27.46
Handedness	-0.10	0.01	19.14
Cognition Locus (ext + ch > int)	0.19	0.06	3.37
Cognition Locus (ch > ext)	0.05	0.06	0.82
Angle x Handedness	0.06	0.01	12.00
Handedness x Cognition Locus (ext + ch > int)	-0.03	0.01	2.63
Handedness x Cognition Locus (ch > ext)	0.04	0.01	3.59
Angle x Cognition Locus (ext + ch > int)	-0.07	0.01	11.81
Angle x Cognition Locus (ch > ext)	0.01	0.01	2.34

Notes. ext: forced external, int: forced internal, ch: free choice. Free choice refers to the participants' free choice between using internal or external processing as available in Experiment 1. To interpret significance of a fixed effect, a $|t| > 2$ criterion has been used.

Offloading onset at angles 60 to 180

To explore the nature of the choice process, we also analyzed the delay until participants started using the external strategy. To do so, we employed the same type of general linear model as in the previous analysis but instead of RT, we used the time between stimulus onset and when the knob was turned for more than 5 degrees as dependent variable. Results show that in the free choice condition, participants started using the knob 216ms later than in the forced external condition (**Table 15**, **Figure 27**), which suggests a sequential process: participants might have tried some sort of internal strategy first and if that failed switched over to the external strategy or initiated parallel processing. Alternatively, the 216ms time cost might be attributable to a costly choice process that preceded the employment of the external strategy.

Table 15 General linear mixed model results for offloading onset in Experiment 2A

Offloading Onset [s]			
Random Effects	Variance	SD	
Participants	0.03	0.18	
Items	0.01	0.08	
Residual	0.09	0.30	
Fixed Effects	Estimate	SE	|*t*|
Intercept	0.78	0.03	29.15
Angle	-0.01	< 0.01	3.71
Handedness	-0.03	< 0.01	9.47
Cognition Locus (ext > ch)	-0.22	0.04	5.31

Notes. ext: forced external, ch: free choice. Free choice refers to the participants' free choice between using internal or external processing as available in Experiment 1. To interpret significance of a fixed effect, a $|t| > 2$ criterion has been used.

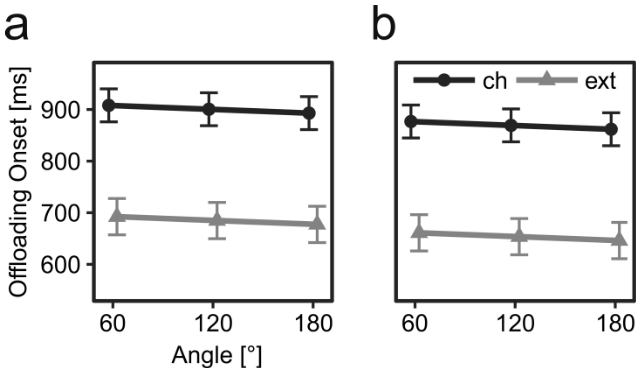


Figure 27 Model-based offloading onset estimates for different (a) and same (b) handedness in Experiment 2A. Error bars depict 95% CIs. ext: forced external, ch: free choice

Experiment 2B

Participants

The same sample size, exclusion criteria, recruitment, and ethical guidelines as in Experiment 2A were used. Two participants were excluded due to technical problems, twenty due to poor task performance and three because their RT means deviated more than 2.5 SD from the grand mean, resulting in a final sample of 40 participants in the forced external (M_{age} = 19.3, $range_{age\,=}$ 18 to 22, 23 females, 5 left-handed) and 35 participants in the forced internal (M_{age} = 19.5, $range_{age\,=}$ 18 to 25, 25 females, none left-handed) condition.

RT at angle 0

To assess possible baseline RT differences, two general linear models, one with and one without the interaction term, were used to model binary accuracies based on handedness and cognition locus at angle 0. IVs have been contrast coded. Based on the Bayesian Information Criterion, the model with the main effects only is reported.

Participants that had to use the external resource answered considerable slower in the angle 0 condition than participants who could freely choose between internal and external processing or had no external processing option available (70 ms, $|t| = 2.0$). The t-statistic falls right at the border of our binary decision criterion but, given the analogous data in *Experiment 2A*, we interpret it as evidence suggesting that participants did closely follow the instructions. In other words we suggest that the 70 ms difference reflects the time cost incurred by recruiting the external resource even at angle 0. For the remaining effects and model details, see **Table 16**. The model's estimates are illustrated in **Figure 28**.

Table 16 General Linear mixed model results for RT in Experiment 2B at angle 0

	RT [s]		
Random Effects	Variance	SD	
Participants	0.03	0.18	
Items	0.01	0.07	
Residual	0.08	0.29	
Fixed Effects	Estimate	SE	$\|t\|$
Intercept	1.17	0.02	52.44
Handedness	-0.21	< 0.01	44.00
Cognition Locus (int + ch > ext)	-0.07	0.04	1.96
Cognition Locus (ch > int)	-0.05	0.04	1.28

Notes. ext: forced external, int: forced internal, ch: free choice.

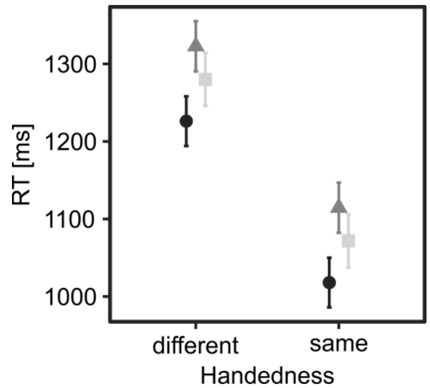


Figure 28 Model-based RT estimates as predicted by handedness and cognition locus at angle 0. Error bars depict 95% CIs. f. external: forced external, f. internal: forced internal

RT at angles 60 to 180

Table 17 General linear mixed model results for reaction time in Experiment 2B

	RT [s]		
Random Effects	Variance	SD	
Participants	0.10	0.32	
Items	0.01	0.10	
Residual	0.19	0.44	
Fixed Effects	Estimate	SE	|*t*|
Intercept	1.71	0.04	48.4
Angle	0.07	< 0.01	30.18
Handedness	-0.05	< 0.01	10.50
Cognition Locus (int + ch > ext)	-0.15	0.06	2.35
Cognition Locus (ch > int)	-0.019	0.07	0.27
Angle x Handedness	0.05	< 0.01	11.18
Handedness x Cognition Locus (int + ch > ext)	0.03	0.01	3.24
Handedness x Cognition Locus	-0.02	0.01	2.11

(ch > int)			
Angle x Cognition Locus (int + ch > ext)	0.01	0.01	2.14
Angle x Cognition Locus (ch > int)	-0.01	0.01	1.19
Angle x Cognition Locus (int + ch > ext) x Handedness	-0.03	0.01	2.70
Angle x Cognition Locus (ch > int) x Handedness	0.05	0.01	4.09

Notes. ext: forced external, int: forced internal, ch: free choice. Free choice refers to the participants' free choice between using internal or external processing as available in Experiment 1. To interpret significance of a fixed effect, a $|t| > 2$ criterion has been used.

Problem solvers achieved the fast RTs in the free choice condition despite the still rather high offloading rate of 80%. This was surprising as it was shown that RT is rather high in the forced external condition (compare **Figure 21** in the main manuscript). We ran two general linear mixed models, analogous to the one used above, to elucidate how participants achieved this high speed performance in offloading trials. Specifically, we compared the free choice and the forced external condition for offloaded trials (**Table 18**) and the free choice and the forced internal condition for trials that had not been offloaded (**Table 19**). Results (compare **Figure 29**) show that participants in the choice condition needed less time when employing the internal strategy in comparison to participants in the forced internal condition. This suggests the ability to offload costly trials was a major contributor to adaptive strategy use.

Table 18 General linear mixed model results for reaction time in Experiment 2B: offloading trials only

	RT [s]

Random Effects	Variance	SD	
Participants	0.11	0.33	
Items	0.02	0.13	
Residual	0.18	0.42	
Fixed Effects	Estimate	SE	|*t*|
Intercept	1.80	0.05	39.99
Angle	0.06	< 0.01	20.74
Handedness	-0.09	0.01	15.81
Cognition Locus (ch > ext)	0.08	0.07	1.05
Angle x Handedness	0.08	0.01	14.84
Handedness x Cognition Locus (ch > ext)	-0.01	0.01	0.44
Angle x Cognition Locus (ch > ext)	-0.01	0.01	1.83

Notes. ext: forced external, int, ch: free choice. Free choice refers to the participants' free choice between using internal or external processing as available in Experiment 1. To interpret significance of a fixed effect, a |*t*| > 2 criterion has been used.

Table 19 General linear mixed model results for reaction time in Experiment 2B: internal trials only

	RT [s]		
Random Effects	Variance	SD	
Participants	0.09	0.29	
Items	0.01	0.07	
Residual	0.19	0.43	
Fixed Effects	Estimate	SE	|*t*|
Intercept	1.52	0.04	39.86
Angle	0.07	< 0.01	18.11
Handedness	-0.03	0.01	3.88
Cognition Locus (ch > int)	0.26	0.07	3.80

Notes. int: forced internal, ch: free choice. Free choice refers to the participants' free choice between using internal or external processing as available in Experiment 1. To interpret significance of a fixed effect, a |t| > 2 criterion has been used.

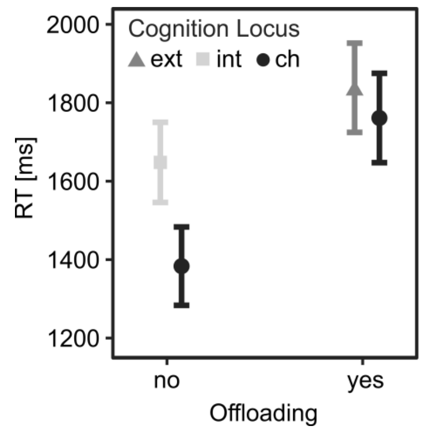


Figure 29 Model-based reaction time estimates split by offloading. Data represents estimated marginal means for the speed goal conditions split by whether participants did or did not offload during the respective trials. Error bars depict 95% CIs. Note that CIs are asymmetric for accuracy estimates. ext: forced external, int: forced internal, ch: free choice. Error bars depict 95% CIs. ext: forced external, int: forced internal, ch: free choice

Accuracy at angle 0

To assess possible baseline accuracy differences, two generalized linear models with a logit link function, one with and one without the interaction term, were used to model binary accuracies based on handedness and cognition locus at angle 0. IVs have been contrast coded. Based on the Bayesian Information Criterion, the model with main effects only is reported.

Problems with same-handed stimuli were, analogously to Experiment 2A, solved more accurately than problems with stimuli of opposite handedness ($|Z| = 8.8$). Cognition locus did not alter accuracy ($|Z|$ of both contrasts < 1.8). For model details, see **Table 20**. The model's estimates are illustrated in **Figure 30**.

Table 20 Generalized linear mixed model results for accuracy in Experiment 2B at angle 0

	Accuracy (log odds)			
Random Effects	Variance	SD	SD	
Participants	0.25	0.50	0.50	
Items	0.13	0.36	0.36	
Fixed Effects	Estimate	Exp(Estimate)	SE	|Z|
Intercept	2.87	17.72	0.09	30.35
Handedness	0.58	1.79	0.07	8.83
Cognition Locus (ext + ch > int)	0.21	1.23	0.12	1.73
Cognition Locus (ch > ext)	-0.17	0.84	0.14	1.20

Notes. ext: forced external, int: forced internal, ch: free choice.

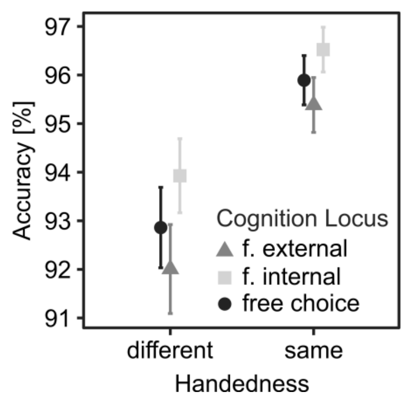


Figure 30 Model-based accuracy estimates as predicted by handedness and cognition locus at angle 0. Error bars depict asymmetric 95% CIs that are back-transformed from the logit scale. f. external: forced external, f. internal: forced internal

Accuracy at angle 60 to 180

Table 21 Generalized linear mixed model results for accuracy in Experiment 2B

	Accuracy (log odds)	
Random Effects	Variance	SD

Participants	0.23	0.48		
Items	0.13	0.36		
Fixed Effects	Estimate	Exp(Estimate)	SE	|*Z*|
Intercept	1.46	4.32	0.09	17.05
Angle	-0.17	0.84	0.01	14.52
Handedness	-0.02	0.98	0.02	1.04
Cognition Locus (int + ch > ext)	-0.36	0.69	0.10	3.76
Cognition Locus (ch > int)	0.21	1.23	0.11	1.86
Angle x Handedness	-0.16	0.85	0.02	6.85
Handedness x Cognition Locus (int + ch > ext)	0.21	1.24	0.05	4.18
Handedness x Cognition Locus (ch > int)	-0.14	0.87	0.06	2.54
Angle x Cognition Locus (int + ch > ext)	-0.22	0.80	0.036	8.62
Angle x Cognition Locus (ch > int)	0.19	1.21	0.03	6.59

Notes. ext: forced external, int: forced internal, ch: free choice. Free choice refers to the participants' free choice between using internal or external processing as available in Experiment 1.

Offloading onset at angles 60 to 180

The same analysis as describe above in section *Offloading onset at angles 60 to 180* of *Experiment 2A* was conducted for Experiment 2B. Results show that in the free choice condition, participants showed no pronounced differences in offloading onset (44ms, not significant) between the free choice and the forced external condition (**Table 22**, **Figure 31**), which speaks for a non-sequential process. When under time pressure, participants might have shifted from sequential processing to either an external processing or a parallel processing strategy.

Table 22 General linear mixed model results for offloading onset in Experiment 2B

Offloading Onset [s]			
Random Effects	Variance	SD	
Participants	0.03	0.17	
Items	< 0.01	0.05	
Residual	0.07	0.26	
Fixed Effects	Estimate	SE	|*t*|
Intercept	0.64	0.02	30.01
Angle	< 0.01	< 0.01	1.24
Handedness	< 0.01	< 0.01	0.29
Cognition Locus (ext > ch)	-0.04	0.04	1.18

Notes. ext: forced external, ch: free choice. Free choice refers to the participants' free choice between using internal or external processing as available in Experiment 1. To interpret significance of a fixed effect, a $|t| > 2$ criterion has been used.

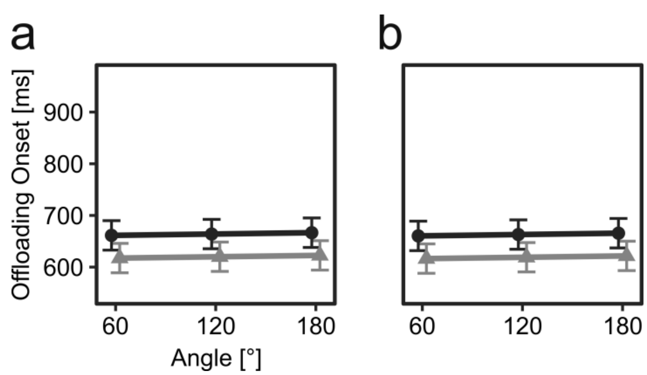


Figure 31 Model-based offloading onset estimates for different (a) and same (b) handedness in Experiment 2B. Error bars depict 95% CIs. ext: forced external, ch: free choice

References

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

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1), 1–48. https://doi.org/10.18637/jss.v067.i01

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

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.

Dixon, P. (2008). Models of accuracy in repeated-measures designs. *Journal of Memory and Language*, 59(4), 447–456. https://doi.org/10.1016/j.jml.2007.11.004

Hoddes, E., Zarcone, V., & Dement, W. (1972). Development and use of Stanford Sleepiness scale (SSS). *Psychophysiology*, 9, 150.

Judd, C. M., Westfall, J., & Kenny, D. A. (2017). Experiments with More Than One Random Factor: Designs, Analytic Models, and Statistical Power. *Annual Review of Psychology*, 68(1), 601–625. https://doi.org/10.1146/annurev-psych-122414-033702

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

Oswald, F. L., McAbee, S. T., Redick, T. S., & Hambrick, D. Z. (2015). The development of a short do-main-general measure of working memory capacity.

Behavior Research Methods, 47(4), 1343–1355. https://doi.org/10.3758/s13428-014-0543-2

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

R Core Team. (2013). *R: A language and environment for statistical computing*.

GENERAL DISCUSSION

Synopsis of results

The common main objectives across *Study 1* to *3* lie in improving the understanding of parameters that influence a human problem solver's decision to use environment-based external instead of brain-based internal resources. In *Study 1*, it was shown that a human problem solver's inclination to offload his or her cognitive processing depends to similar degrees on monitoring the external resource's actual performance as well as on pre-existing beliefs about the external resource's performability. In *Study 2*, it was shown that a human problem solver's inclination to offload his or her cognition also depends on the performability of task-relevant *internal* resources. Good internal performability decreased cognitive offloading frequency even when the task at hand looked comparably difficult, thereby suggesting that performability can supplant task appearance in determining cognitive offloading propensity. Both *Study 1* and *Study 2* confirmed former studies in that human problem solvers are sensitive to performance parameters when deciding whether to offload cognitive processing. In *Study 3*, this sensitivity was investigated further. Specifically, it was investigated whether human problem solvers are able to mix internal and external processing in a way that matches their current performance goals or whether they are biased toward speed or accuracy maximization, or mental effort minimization, without considering whether those heuristics meet their current goals. Results of *Study 3* showed that participants were indeed mixing internal and external processing in a way congruent with their cognitive performance goals, indicating that human problem solvers obtain a high proficiency in incorporating external resources into their cognitive processing.

The present findings picture the human problem solver as proficiently navigating cognitive environments, offloading cognition mostly when internal processing is inefficient (Study 2) or when it helps to achieve current cognitive performance goals (Study 3). The findings however also show that this proficiency can be profoundly altered once false beliefs about the external resource are introduced (Study 1). Taken together, the findings emphasize importance of balancing first-hand *experience* with an external resource, e.g. in the shape of performance feedback, with more abstract *information*, e.g. in the shape of communicated beliefs about an external resource.

Future directions

How do people keep track of an external resource's usefulness?

It has been shown that human problem solvers are more inclined to incorporate an external resource into their cognitive processing if that external resource improves performance (see introductory section *Determinants of cognitive offloading: Optimizing performance*). However, the mechanisms that keep track of the external resource's—or , similarly, of an internal resource's—performance remain largely in the dark as Dunn and Risko (2016, p. 1083) noted:

> *„The idea that performance/effort is driving the decision to try to offload cognition via adopting an external strategy implies the existence of some mechanism dedicated to performing online monitoring of an individual's performance. For example, Gray et*

al. (2006) suggest that individuals possess implicit knowledge of performance costs (in terms of time) associated with the retrieval of items in memory, and strategy selection is sensitive to such knowledge. In most cases, however, the nature of this mechanism is left open. Nevertheless, there exist a number of candidate mechanisms that could perform the requisite work. For example, research investigating conflict monitoring, error monitoring (e.g., Blais, Robidoux, Risko, & Besner, 2007; Botvinick, Braver, Barch, Carter, & Cohen, 2001), and introspective response time (Marti, Sackur, Sigman, & Dehaene, 2010)"

Analogously, it is not clear how experience-based information like performance feedback is incorporated with metacognitive information like verbally communicated information about an external resource's performability. In *Study 1*, it has been shown that experience-based information on the one hand and faulty prior information about an external resource's performability on the other hand can alter people's inclination to recruit an external resource simultaneously and independently (also compare to Dunn & Risko, 2016). In other words, both performance monitoring as well as communicated information contribute to the decision to offload cognition, though it is not yet clear how both factors are integrated.

A first hint however was given in a study in which participants had to decide for one of two stimuli at a time that were each associated with a different reward probability (M. M. Walsh & Anderson, 2011). In one condition, participants received prior information about the reward rates associated with each stimulus. In another condition, participants received no prior information. Results showed that behavior, i.e. which stimulus is selected, is heavily depending on whether prior information was given. That influence on behavior was present from the very beginning of the experiment. However, the FRN—an ERP that is thought to signal feedback-related prediction error (reviewed in M. M. Walsh & Anderson, 2011)—does not seem to incorporate the prior information and thus varies in a feedback- or experience-dependent manner just as in the condition in which no prior information was given. In sum, these results suggest that metacognitive evaluation based on prior information and feedback-related performance monitoring are realized by different neural mechanisms and are likely integrated at a hierarchically higher level than the FRN is based at.

As a next step, it would be (1) promising to validate the importance of the FRN in monitoring the performance benefit associated with an external cognitive resource rather than with a stimulus-related reward as done by Walsh & Anderson (2011). Subsequently, it would be promising to (2) find a neural correlate that signifies the prior information about an external resource's usefulness and to (3) use these correlates to advance the understanding of how[27] both factors affect behavior.

Impact of the appearance of an external resource

When solving a social task, people are more inclined to follow advice from a fellow human than from a computer while the reverse is true for analytic tasks (Hertz &

[27] Walsh and Anderson (2011) showed that the FRN can signify reward monitoring independently from exhibited behavior. In *Study 1* of the current manuscript, it was shown that both first-hand experience as well as prior information can simultaneously contribute to behavior. Thus, there is some knowledge about *how* both factors affect behavior. What is meant here with "how" is a more thorough understanding of the underlying mechanism incorporating both factors both neutrally and behaviorally.

Wiese, 2019; Smith, Allaham, & Wiese, 2016). In other words, the appearance of an agent and the type of the task interact in influencing the degree to which external task-related information is used for solving the task. Reliance on information given by other agents to solve cognitive tasks follows the same fundamental mechanism as cognitive offloading does: the incorporation of external information into cognitive processing. In the study by Smith, Allaham, and Wiese (2016), participants were always confronted with the advice of different agents and could choose whether to follow that advice or not. As a next step, one could investigate whether people (1) prefer actively accessing an external resource rather than more passively following its advice based on its appearance as done in the study reported above. If so, one could (2) explore the mechanism behind, potentially encompassing agent-related trust (compare Parasuraman & Manzey, 2010; Smith et al., 2016) and/or mind ascription (Epley & Waytz, 2010; Waytz, Gray, Epley, & Wegner, 2010). Potential studies could draw on the *transactive memory* literature that captures how people access information that is distributed in their the social surrounding (e.g.; Sutton, Harris, Keil, & Barnier, 2010; Wegner, 1987). Lastly, it could be worthwhile to (3) explore the consequences of appearance for external resources that do not resemble an agent. For example, it was shown that the size of a button determines how frequently it is used to access information (Gray et al., 2006). However, following Fitts' law, it is harder to navigate a finger or mouse cursor towards a small than towards a large button. Thus, in this setting, appearance (i.e., size) is confounded with performance. Future studies could investigate appearance parameters in settings in which appearance is not confounded with other parameters.

Simultaneously assessing the impact of multiple determinants of cognitive offloading across different scenarios

In the current project, it was shown that prior performance information about an external resource influences cognitive offloading in addition to performance information gathered via direct interaction experience (*Study 1*). It was also shown that the performability of internal resources influences cognitive offloading in addition to the apparent difficulty of a cognitive task (*Study 2*). What is missing, however, is the big picture: given a specific scenario, which parameters are most likely to influence offloading behavior? Which determinants are able to uniquely predict and which determinants share variance in explaining cognitive offloading? Are different determinants equally important across different scenarios? To find the answers, a multiple regression approach incorporating the key determinants as independent variables (see *Introduction: Determinants of cognitive offloading*) across multiple scenarios and comparing them for stability between scenarios could be employed. The challenge of such an approach is twofold. First, the simultaneous modeling of at least five determinants requires a massive sample size as well as detailed attention to the model's peculiarities like assumption violations. Second, it would be necessary to determine representative scenarios that would—ideally—afford generalizing the findings to everyday problem solving.

REFERENCES (INTRODUCTION AND GENERAL DISCUSSION)

Anderson, J. R. (1990). *The Adaptive Character of Thought* (1st ed.). https://doi.org/10.4324/9780203771730

Ballard, D. H., Hayhoe, M. M., Pook, P. K., & Rao, R. P. (1997). Deictic codes for the embodiment of cognition. *Behavioral and Brain Sciences*, *20*(4), 723–742.

Barr, N., Pennycook, G., Stolz, J. A., & Fugelsang, J. A. (2015). The brain in your pocket: Evidence that Smartphones are used to supplant thinking. *Computers in Human Behavior*, *48*, 473–480. https://doi.org/10.1016/j.chb.2015.02.029

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Blais, C., Robidoux, S., Risko, E. F., & Besner, D. (2007). *Item-specific adaptation and the conflict-monitoring hypothesis: A computational model.*

Bocanegra, B. R., Poletiek, F. H., Ftitache, B., & Clark, A. (2019). Intelligent problem-solvers externalize cognitive operations. *Nature Human Behaviour*, *3*(2), 136–142. https://doi.org/10.1038/s41562-018-0509-y

Botvinick, M. M., Braver, T. S., Barch, D. M., Carter, C. S., & Cohen, J. D. (2001). Conflict monitoring and cognitive control. *Psychological Review*, *108*(3), 624–652.

Carr, N. (2008). Is Google Making Us Stupid? *Yearbook of the National Society for the Study of Education, 107*(2), 89–94. https://doi.org/10.1111/j.1744-7984.2008.00172.x

Cherkaoui, M., & Gilbert, S. J. (2017). Strategic use of reminders in an 'intention offloading' task: Do individuals with autism spectrum conditions compensate for memory difficulties? *Neuropsychologia*, *97*, 140–151. https://doi.org/10.1016/j.neuropsychologia.2017.02.008

Clark, A. (1998). Magic words: How language augments human computation. In P. Carruthers & J. Boucher (Eds.), *Language and thought* (pp. 162–183). https://doi.org/10.1017/CBO9780511597909.011

Clark, A. (1999). An embodied cognitive science? *Trends in Cognitive Sciences*, *3*(9), 345–351.

Clark, A. (2004). *Natural-born cyborgs: Minds, technologies, and the future of human intelligence*. Oxford University Press.

Clark, A., & Chalmers, D. (1998). The extended mind. *Analysis*, 7–19.

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

Epley, N., & Waytz, A. (2010). Mind perception. *Handbook of Social Psychology*.

Fenech, E. P., Drews, F. A., & Bakdash, J. Z. (2010). The effects of acoustic turn-by-turn navigation on wayfinding. *Human Factors and Ergonomics Society Annual Meeting Proceedings*, 1926–1930.

Fu, W.-T., & Gray, W. (2006). Suboptimal tradeoffs in information seeking. *Cognitive Psychology*, *52*(3), 195–242. https://doi.org/10.1016/j.cogpsych.2005.08.002

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

Gilbert, S. J., Bird, A., Carpenter, J. M., Fleming, S. M., Sachdeva, C., & Tsai, P.-C. (2019). Optimal use of reminders: Metacognition, effort, and cognitive offloading. *Journal of Experimental Psychology: General*.

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

Hertz, N., & Wiese, E. (2019). Good advice is beyond all price, but what if it comes from a machine? *Journal of Experimental Psychology. Applied*.

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.

Hutchins, E. (2014). The cultural ecosystem of human cognition. *Philosophical Psychology*, *27*(1), 34–49.

Jérémy Virgo, Jonathan Pillon, Jordan Navarro, Emanuelle Reynaud, & François Osiurak. (2017). Are You Sure You're Faster When Using a Cognitive Tool? *The*

American Journal of Psychology, *130*(4), 493. https://doi.org/10.5406/amerjpsyc.130.4.0493

Kirsh, D. (1995). The intelligent use of space. *Artificial Intelligence, 73*(1), 31–68.

Kirsh, D. (2009). Problem Solving and Situated Cognition. In *The Cambridge Handbook of Situated Cognition* (pp. 264–306).

Kirsh, D. (2010). Thinking with external representations. *AI & SOCIETY*, *25*(4), 441–454. https://doi.org/10.1007/s00146-010-0272-8

Kirsh, D. (2013). Embodied cognition and the magical future of interaction design. *ACM Transactions on Computer-Human Interaction*, *20*(1), 1–30. https://doi.org/10.1145/2442106.2442109

Kirsh, D., & Maglio, P. (1994). On distinguishing epistemic from pragmatic action. *Cognitive Science*, *18*(4), 513–549.

Kool, W., McGuire, J. T., Rosen, Z. B., & Botvinick, M. M. (2010). Decision making and the avoidance of cognitive demand. *Journal of Experimental Psychology: General*, *139*(4), 665–682. https://doi.org/10.1037/a0020198

Lemaire, P., & Reder, L. (1999). What affects strategy selection in arithmetic? The example of parity and five effects on product verification. *Memory & Cognition*, *27*(2), 364–382. https://doi.org/10.3758/BF03211420

Leontjev, A. N. (1981). *Problems of the development of the mind*.

Maglio, P. P., & Kirsh, D. (1996). Epistemic action increases with skill. *Proceedings of the Eighteenth Annual Conference of the Cognitive Science Society*, *16*, 391–396. Erlbaum.

Maglio, P. P., Matlock, T., Raphaely, D., Chernicky, B., & Kirsh, D. (1999). Interactive skill in Scrabble. *Proceedings of the Twenty-First Annual Conference of the Cognitive Science Society*, 326–330. Routledge.

Marewski, J. N., & Schooler, L. J. (2011). Cognitive niches: An ecological model of strategy selection. *Psychological Review*, *118*(3), 393.

Marti, S., Sackur, J., Sigman, M., & Dehaene, S. (2010). Mapping introspection's blind spot: Reconstruction of dual-task phenomenology using quantified introspection. *Cognition*, *115*(2), 303–313.

Nardi, B. A. (1996). *Context and consciousness: Activity theory and human-computer interaction*. mit Press.

Osiurak, F., Navarro, J., Reynaud, E., & Thomas, G. (2018). Tools don't—and won't—make the man: A cognitive look at the future. *Journal of Experimental Psychology: General*, *147*(5), 782–788. https://doi.org/10.1037/xge0000432

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

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.

Scaife, M., & Rogers, Y. (1996). External cognition: How do graphical representations work? *International Journal of Human-Computer Studies*, *45*(2), 185–213.

Shapiro, L. (2010). *Embodied Cognition*. Routledge.

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.

Simon, H. A. (1956). Rational choice and the structure of the environment. *Psychological Review*, *63*(2), 129.

Smith, M. A., Allaham, M. M., & Wiese, E. (2016). Trust in Automated Agents is Modulated by the Combined Influence of Agent and Task Type. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *60*(1), 206–210. https://doi.org/10.1177/1541931213601046

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

Sutton, J. (2010). Exograms and Interdisciplinarity: History, the Extended Mind, and the Civilizing Process. In R. Menary (Ed.), *The Extended Mind* (pp. 189–225). https://doi.org/10.7551/mitpress/9780262014038.003.0009

Sutton, J., Harris, C. B., Keil, P. G., & Barnier, A. J. (2010). The psychology of memory, extended cognition, and socially distributed remembering. *Phenomenology and the Cognitive Sciences*, *9*(4), 521–560. https://doi.org/10.1007/s11097-010-9182-y

Touron, D. R. (2015). Memory avoidance by older adults: When "old dogs" won't perform their "new tricks." *Current Directions in Psychological Science*, *24*(3), 170–176.

Walsh, M. M., & Anderson, J. R. (2011). Modulation of the feedback-related negativity by instruction and experience. *Proceedings of the National Academy of Sciences*, *108*(47), 19048–19053. https://doi.org/10.1073/pnas.1117189108

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

Waytz, A., Gray, K., Epley, N., & Wegner, D. M. (2010). Causes and consequences of mind perception. *Trends in Cognitive Sciences*, *14*(8), 383–388.

Wegner, D. M. (1987). Transactive memory: A contemporary analysis of the group mind. In B. Mullen & G. R. Goethals (Eds.), *Theories of group behavior* (pp. 185–208). Springer.

Wiese, E., Wykowska, A., & Müller, H. J. (2014). What We Observe Is Biased by What Other People Tell Us: Beliefs about the Reliability of Gaze Behavior Modulate

Attentional Orienting to Gaze Cues. *PLoS ONE*, *9*(4), e94529. https://doi.org/10.1371/journal.pone.0094529

Wilson, M. (2002). Six views of embodied cognition. *Psychonomic Bulletin & Review*, *9*(4), 625–636.

BIOGRAPHY

Patrick P. Weis received his Bachelor of Science from Otto-von-Guericke University, Magdeburg, Germany, in 2012 and his Master of Science from Eberhard-Karls University and the International Max Planck Research School, Tuebingen, Germany, in 2014.