



# Multitasking costs on metacognition in a triple-task paradigm

Mahiko Konishi<sup>1,2</sup> · Bruno Berberian<sup>2</sup> · Vincent de Gardelle<sup>3</sup> · Jérôme Sackur<sup>1,4</sup>

Accepted: 1 June 2021  
© The Psychonomic Society, Inc. 2021

## Abstract

Multitasking situations, such as using one's phone while driving, are increasingly common in everyday life. Experimental psychology has long documented the costs of multitasking on task performance; however, little is known of the effects it has on the metacognitive processes that monitor such performance. The present study is a step toward filling this void by combining psychophysical procedures with complex multitasking. We devised a multimodal paradigm in which participants performed a sensorimotor tracking task, a visual discrimination task, and an auditory 2-back working memory task, either separately or concurrently, while also evaluating their task performance every ~15 s. Our main finding is that multitasking decreased participants' awareness of their performance (metacognitive sensitivity) for all three tasks. Importantly, this result was independent of the multitasking cost on task performance, and could not be attributed to confidence leak, psychological refractory period, or recency effects on self-evaluations. We discuss the implications of this finding for both metacognition and multitasking research.

**Keywords** Multitasking · Metacognition · Self-evaluation · Task-performance · Triple-task

## Introduction

Calling a friend on the phone while driving home; checking the email inbox in-between writing a research article; scrolling through social media during an online class. With the development of new communication technologies, multitasking is increasing in the Western world and in the younger generations (Carrier et al., 2009; Wallis, 2006). As such, the study of multitasking has recently gathered attention from different fields: for example, disruptions due to multitasking have been documented while driving (Levy & Pashler, 2008; Nijboer et al., 2016), in education (Junco, 2012; Sana et al., 2013), in the workplace (Buser & Peter, 2012), and in mental health (Becker et al., 2012; Reinecke et al., 2017). While the

potential costs of multitasking on task performance have long been known (Pashler, 1994; Salvucci & Taatgen, 2010), whether multitasking also hinders people's awareness of their performance is largely unknown: speaking on the phone while driving decreases our ability to drive, but if it also decreases our ability to realize this very fact, it may have catastrophic consequences.

Research on metacognition investigates the ability of individuals to monitor and exert control over their own cognitive processes (Dunlosky & Metcalfe, 2008; Nelson & Narens, 1990). One classically studied example of metacognition is the feeling of confidence we experience after making a simple decision, an internal estimate of the correctness of our decision (Henmon, 1911; Mamassian, 2016). Another, less studied, example is the ability to explicitly evaluate one's own level of performance in a task that is performed over a more extended duration (but see, e.g., Lee et al., 2021). This is often crucial in real life, and is intimately intertwined with the task itself: for example, when driving in our car, we are not just *driving* (i.e., performing the motor tasks of steering the wheel, shifting gears, etc.), but also constantly monitoring and adjusting our performance.

Given the importance of metacognition in regulating behavior (Aguilar-Lleyda, Lemarchand, & de Gardelle, 2020; Balsdon et al., 2020), learning (Hainguerlot et al., 2018), and coordination between agents (Massoni & Roux, 2017), disruptions of metacognition are expected to have deleterious

---

Vincent de Gardelle and Jérôme Sackur contributed equally to this work.

✉ Mahiko Konishi  
konishi@gmail.com

<sup>1</sup> Laboratoire de Sciences Cognitives et de Psycholinguistique, Dept d'Études Cognitives, ENS, PSL University, EHESS, CNRS, Paris, France

<sup>2</sup> ONERA, The French Aerospace Lab, Information Processing and Systems Department, 13661 Salon Cedex Air, France

<sup>3</sup> Paris School of Economics and CNRS, Paris, France

<sup>4</sup> École Polytechnique, Palaiseau, France

consequences on human behavior. Metacognitive failures are associated with a range of psychiatric disorders (David et al., 2012; Sun et al., 2017), including addictive behaviors such as gambling (Spada et al., 2015). Thus, understanding the factors and contexts that can disrupt these metacognitive processes is important. Prior research has shown that, for example, metacognition becomes less efficient under stress (Reyes et al., 2015), with aging (Palmer et al., 2014), or under working memory load (Maniscalco & Lau, 2015).

Theoretically, it is expected that metacognition should be impaired in multitasking as a consequence of overlapping demands on limited cognitive resources (Salvucci & Taatgen, 2008; Wickens, 2002). Metacognition and multitasking are likely to compete for access to central cognitive resources as they each place demands on executive functions. Moreover, if some domain-general metacognitive evaluation is automatically triggered whenever a task is performed (Aguilar-Lleyda, Konishi, et al., 2020; de Gardelle & Mamassian, 2014; Mazancieux et al., 2020), multitasking might result in overlapping and conflicting calls to the domain-general metacognitive component, a situation in which multitasking costs are predicted (Salvucci & Taatgen, 2008; Wickens, 2002).

However, very few studies, based on hard-to-compare paradigms, have addressed the impact of multitasking on metacognition with mixed results. In one ecological study, Sanbonmatsu et al. (2016) found that having a phone conversation while driving impairs participants' abilities to evaluate their own driving performance. Another study (Maniscalco & Lau, 2015) found that manipulating items in working memory (but not simply holding items in memory) deteriorates metacognition in a visual task. Yet, we have found recently that perceptual metacognition was remarkably resilient to dual-task conditions: when participants performed and evaluated themselves on two visual tasks in parallel, they could maintain the same metacognitive efficiency as in single-task conditions (Konishi et al., 2020). Thus, it remains largely unclear whether multitasking affects people's monitoring and awareness of their own performance.

With the present study we set to advance our understanding of the effects of multitasking on metacognition. What is more, we aimed to take a step towards bridging the results of controlled laboratory experiments with real-life scenarios. We took inspiration from the computer-based Multi-Attribute Task Battery (MATB; Comstock Jr & Arnegard, 1992), which simulates a series of concurrent aircraft pilots' tasks, and is developed by the National Aeronautics and Space Administration (NASA) to evaluate operators' performance in overload situations. Our experimental paradigm involved three tasks: a sensorimotor tracking task, a visual discrimination task, and an auditory 2-back task. Compared with extant studies, our study had the following characteristics: we engaged participants in multiple different modalities (visual,

auditory, sensorimotor); in addition, we increased the multi-tasking load as participants performed three tasks. Finally, we studied participants' metacognition over longer trials, in the order of tens of seconds.

Participants performed either single-task or triple-task blocks, and at the end of each trial rated what they thought their performance had been in the tasks. We then quantified how closely their self-evaluations matched their objective performance, and hypothesized that this relationship would be disrupted in the multitasking condition. Results provided support to this hypothesis: participants were less precise in evaluating their own performance in the triple-task condition, for all three tasks, irrespective of the toll multitasking had on first-order performance.

## Methods

### Participants

Eighteen naive participants (mean age: 28.8 years, range: 20–35 years; four males) took part in the study. All had normal or corrected-to-normal vision and reported no psychiatric or neurological disorders. Participants were naive to the specific purpose of the experiment and gave informed consent prior to the experiment. Participants took part in three sessions on three different days and were compensated 60 € for their time; each session lasted approximately 90 mins. No statistical methods were used to pre-determine sample size, which was chosen based on the resources available. Participants were tested in accordance with the ethics requirements of the Ecole Normale Supérieure and the Declaration of Helsinki. One (male) participant and one session for another participant were excluded from all analyses due to their performance being at chance level on one of the three tasks. Power calculations were implemented with the *simr* package in R (Green & MacLeod, 2016). We ran simulations for the effect of multitasking on task performance (the fixed effect of condition in our linear mixed models), and for that of multitasking on metacognitive sensitivity (the condition × self-evaluation interaction). Specifically, we simulated a dataset corresponding to our final sample size (17 participants), number of trials, and linear mixed-model structures; we then modified the beta estimates for our effects of interest, aiming for a small to moderate effect size, and simulated model fitting 5,000 times. Power was estimated to be 100% with 95% confidence interval (CI) (96.38, 100.0) for the effect of multitasking on task performance, and 81% with 95% CI (71.93, 88.16) for the effect of multitasking on metacognitive sensitivity.

## Experimental tasks

All stimuli were displayed on an LCD monitor ( $1,280 \times 800$  pixels, 60 hz), with participants sitting approximately 50 cm from the monitor. The stimuli and task code (in Python 3) are available at <https://osf.io/9ru8j/>. The tasks ran on PsychoPy (Peirce et al., 2019). Trials lasted for 12–21 s (uniformly distributed), during which time participants performed a continuous visual tracking task, a continuous auditory feedback task, and multiple instances of a visual discrimination task.

**Sensorimotor tracking task** The tracking task was chosen as motor tasks pervade day-to-day life and because their continuous nature is likely to ensure a high level of engagement for participants. A black target dot (15-pixel radius) moved randomly inside a bounded circular area (300-pixel radius) centered on the computer screen. Participants were instructed to track the target by keeping the center of the mouse cursor (a white cross,  $30 \times 30$  pixel height  $\times$  width) inside the dot as much as they could during the whole trial duration. Figure 1 illustrates the task. The speed (i.e., the difficulty of the task; speed range:  $\sim 50$ – $70$  pixels/s) of the moving target was staircased in the first experimental session: the staircase aimed at 50% tracking performance (correctly tracking the target for 50% of frames in a trial), with a 1-up/1-down staircase, with equal step sizes. The target started at the center of the screen on every block. Then, a new point in space was chosen every 12 frames, with a distance of 150 pixels and a random angle calculated with a von Mises distribution with a kappa parameter of 2, and centered in the current motion direction. If the new point fell outside of the boundary area of the task, it was resampled. The target smoothly moved towards that new point with a velocity given by a vector of the distance between target and new movement point, multiplied by a maximum movement speed (which is the variable we staircased in our design): this velocity was also affected by a “gravity” vector that “pushed” the target towards the center the more it drifted outside the circular boundary area. After a trial ended, the target froze in space while participants responded to the self-evaluation questions: after all the responses were submitted, a new trial began once the participants moved the cursor on top of the frozen target. The code is freely available at <https://osf.io/9ru8j/> (the functions *newGoalPol*, *movToGoal*, and *trackingTask*, code for the movement of the target).

**Visual discrimination task** We designed the visual discrimination task to resemble the discrimination tasks used in perceptual metacognition studies. A white stimulus was flashed for 50 ms every  $\sim 3$  s (with a uniform random jitter of  $\pm 0.5$  s). The stimulus was randomly located at a fixed distance of 200 pixels from the center (so that it was inside the main circular area). This stimulus ( $55 \times 40$  pixels) represented a space invader that could be “angry” or “friendly” (same number of

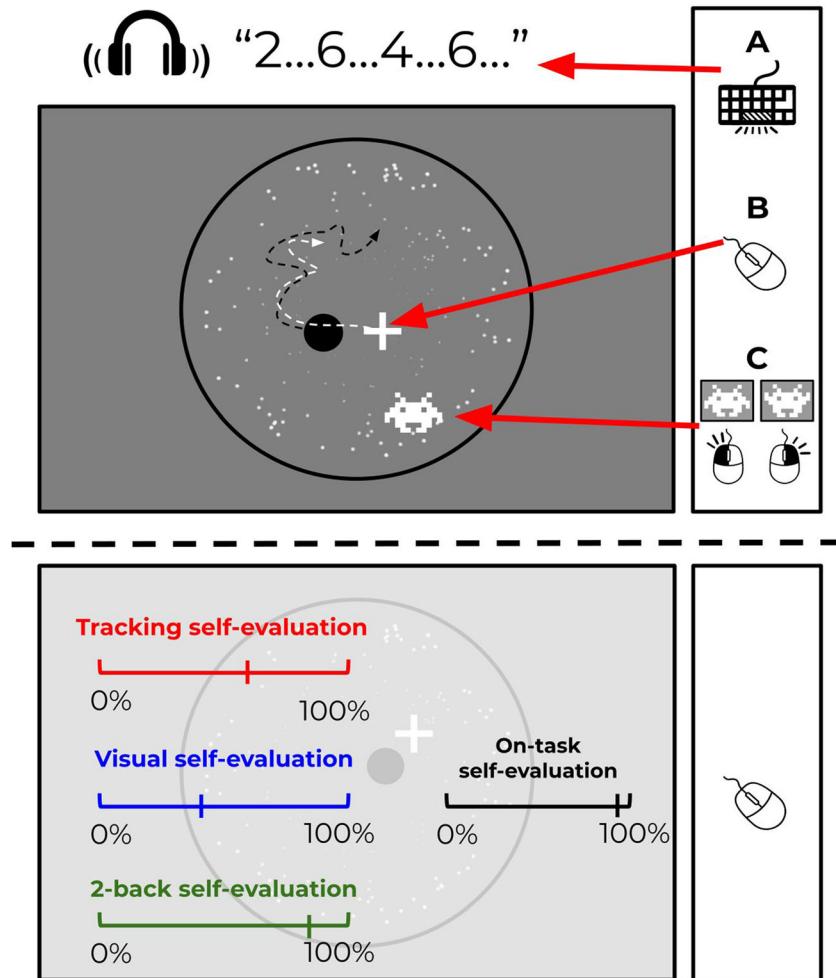
pixels; see Fig. 1 for details): participants were instructed to respond as quickly and accurately as possible, by pressing either the left (for the “angry” space invader) or the right (“friendly” space invader) mouse button. They had a maximum of 2 s after stimulus appearance to respond, after which the response was considered as missing (across participants, missing responses were 3% in the single-task, and 9% in the triple-task condition). Depending on the trial duration, each trial featured between three and six visual targets. The contrast of the visual stimulus was staircased in the first experimental session, with a 1-up/1-down (up-step was 7% of the maximum value; ratio of down-step/up-step = 0.2845) method, targeting 78% performance (García-Pérez, 1998). A cloud of looming white dots ( $n = 200$ ; *max\_size* = 8 pixels; *speed* = 2 pix/frame) was also constantly generated to increase the difficulty of the task.

**2-back task** The auditory 2-back task, standardly used in psychology to measure working memory (WM), was used to simulate the processing load of real-life multitasking situations (e.g., holding content in one’s mind during a conversation while driving). Participants wore headphones through which they listened to a stream of single digits (from 0 to 9; one every 1.5 s). Participants were instructed to press the spacebar as quickly as possible every time they heard a number that was spoken two numbers before. Participants had 1.5 s to respond to the target (after which the next one was presented). To ensure a sufficient number of targets on every trial, every number had a 30% probability of being a target number. Depending on the trial duration, each trial featured between six and 12 spoken numbers (and between zero and six 2-back targets). The stream of numbers was played on top of an auditory noise track generated to increase the difficulty of the task.

**Self-evaluations** At the end of each trial, participants were asked to self-evaluate their performance on that trial, using continuous rating scales (see Fig. 1). These self-evaluation ratings were on the same scale as their corresponding objective performance: in the tracking task, participants subjectively evaluated their performance on a scale from 0% to 100% representing how much time they spent on-target. Similarly, rating scales ranged from 0% to 100% for the visual task (% of correctly responded targets) and for the 2-back task (% of correctly responded numbers). Participants also rated their task focus (from 0%, “completely distracted” to 100%, “completely focused”) on each trial.

## Design and procedure

Our paradigm featured three tasks and two conditions (single-task and triple-task), and followed a within-subject, blocked design. The order of the single- and triple-task blocks was



**Fig. 1 Top panel:** Experimental tasks. The larger panel on the left illustrates the tasks, while the smaller panel on the right shows the response modalities. The stimuli of all three tasks were always presented in all conditions, but participants were selectively instructed to perform only one of the three tasks (single-task condition), or all three together (triple-task condition), depending on block. **(A)** Auditory 2-back task; participants listened to a stream of digits with headphones, and responded to targets with the spacebar key. **(B)** Tracking task; the black target circle moved randomly around the screen (black dashed line), while participants tracked it with the mouse cursor (white cross;

movement visualized with white dashed line). **(C)** Visual discrimination task; participants responded with the left mouse button when the “angry” space invader (hands down) was presented, or the right mouse button for the “friendly” space invader (hands up). The cloud of small white dots represents visual noise (more information in the main text). **Bottom panel:** Self-evaluation screen. At the end of each trial participants were asked to rate their performance on the task(s) they performed (left rating scales), and on their task focus (on the right). The figure shows the case of the triple-task condition, in which participants performed all three tasks

counterbalanced across sessions and participants. The stimuli from all three tasks were also always presented in all conditions, so that trials were equated at the perceptual level; only the instructions before each block changed, ensuring that participants focused on selectively performing only one (single-task blocks) or all three tasks concurrently (triple-task blocks).

Participants performed three experimental sessions over 3 days. On the first day, participants were instructed and practiced the three tasks, both in the single condition (for ten trials in each task) and in the triple-task condition (also ten trials). This initial practice was also present in the second and third sessions. After training in the first session, participants performed blocks of 60 trials for each task in the single-task condition, throughout which the difficulty was staircased for

the tracking and the visual discrimination tasks (as outlined in the previous section). In both the second and third sessions, participants performed blocks of 50 trials for each task in the single-task condition, and 50 trials of the triple-task condition; the difficulty of the tracking and visual tasks’ stimuli was held constant, using the stimulus level found with the staircase procedures. The order of the blocks was counterbalanced across sessions and participants. To disentangle a potential metacognitive cost from a performance cost, we aimed to equate performance between the single-task and triple-task conditions with a manipulation across the second and third sessions. This manipulation is explained in the Session Effect section of the [Online Supplementary Material](#). All statistical analyses and main results in this study used the data

from the second and third sessions (the constant stimuli sessions).

## Statistical analyses

**Task performance** Task performance for the tracking task was indexed by the percentage of time (% of frames) of a trial in which participants' cursor was within the target. For the visual and the 2-back tasks, in addition to the percentage of correct responses, task performance was also measured with  $d'$  as a measure of the observers' discrimination sensitivity. To analyze the effect of triple-tasking on task performance we ran linear mixed models (LMMs) with the R package *lme4* (Bates et al., 2015): task performance on every trial was described using task (tracking, visual, 2-back), condition (single- and triple-task), and their interaction, as fixed effects; participants and sessions (second and third sessions) were added as random intercepts, as well as random slopes for the effect of task condition on each participant. In order to analyze all three tasks concurrently, removing trivial differences due to the different scales in the performance measures used, we standardized trial-by-trial performance within-task, across participants. As performance scores, we used the percentage of correct tracking and  $d'$  for the other two tasks on each trial. Because of the low number of visual and 2-back targets in each trial, and in order to calculate  $d'$ , we applied a standard correction in which hits or false alarms proportions of 0 and 1 were replaced with  $1/(2n)$  and  $1 - (1/(2n))$ , where  $n$  is the number of targets for a task, in a trial. When looking specifically at the effects of session number on performance, similar models were defined for each task separately, with the only difference being the session factor as a fixed instead of a random effect. P-values for fixed effects were calculated using the *lmerTest* (Kuznetsova et al., 2017) package, which implements Satterthwaite's method (backward elimination of non-significant fixed and random effects) to judge the significance of fixed effects. For all LMMs, the most complex model possible in terms of random slopes and intercepts (given the factors of interest) was first tested, and then gradually simplified in case of non-convergence, until the most complex model that converged was found (Barr et al., 2013); this was the model to which the *step* function was applied.

**Response times** Response times (RTs) were available for the visual task (mouse clicks) and for the 2-back task (key press). RTs were log-transformed and then analyzed with linear mixed models, in a fashion analogous to task performance, including the same fixed and random effects (minus the tracking task, having no measurable RTs).

**Metacognitive sensitivity** We measured participants' metacognitive sensitivity (the ability to discern good or bad performance with one's evaluations), separately by condition,

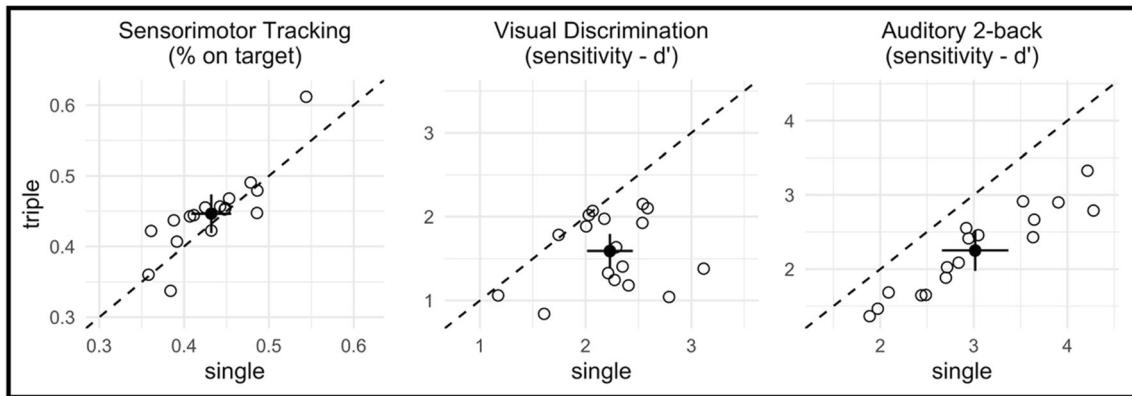
as the slope of the across-trials regression between objective task performance and self-evaluation. A larger slope indicates increased metacognitive sensitivity, as participants are able to track their fluctuations in performance with their self-evaluations. Importantly, we hypothesized that metacognitive sensitivity (and thus the slope of the regression) would decrease in the triple-task condition due to increased cognitive demands. In order to test the effect of multitasking on metacognitive sensitivity, we again ran linear mixed models, this time describing task performance on every trial with the corresponding self-evaluation, condition, and task, and the triple interaction, as fixed effects; participant number and session number were added as random effects. The effect of interest here was the interaction effect between the self-evaluation and the condition factors: this signals a potential difference by condition in the slope of the performance/self-evaluation relationship.

## Results

### Task performance

Figure 2 shows participants' performance in the three tasks and two experimental conditions. On average, participants tracked the target 43% ( $SD = 4.9$ ) of the time in the single-task condition and 45% ( $SD = 5.8$ ) in the triple-task condition. In the visual discrimination task, participants' average accuracy was 82% ( $SD = 5.9$ ) and  $d'$  was 2.23 ( $SD = 0.46$ ) in the single-task condition; mean accuracy was 71% ( $SD = 7.5$ ) and  $d'$  was 1.59 ( $SD = 0.43$ ) in the triple-task condition. In the auditory 2-back task, for which there was no staircase procedure, average accuracy (responding to targets and withholding a response to non-targets) was 93% ( $SD = 4.2$ ) and  $d'$  was 3.01 ( $SD = 0.75$ ) in the single-task; mean accuracy was 88% ( $SD = 5.1$ ) and  $d'$  was 2.25 ( $SD = 0.58$ ) in the triple-task condition.

An LMM was defined, describing (z-scored) performance with condition, task, and their interaction as fixed effects, as well as a random slope and intercept for the effect of condition on each participant. A random intercept for session was initially included but subsequently eliminated from the best fitting model by backwards elimination. For this model ( $\text{intercept} = -0.01$ ,  $\text{participant}_{\text{intercept}} SD = 0.39$ ,  $\text{participant}_{\text{condition}} SD = 0.21$ ), performance was higher in the single-task condition ( $\beta_{\text{single}} = 0.31$ , 95% CI (0.21, 0.41),  $t = 6.18$ ,  $p < .001$ ), indicating a multitasking cost, as expected. A task  $\times$  condition interaction ( $\beta_{\text{trackingXsingle}} = 0.39$ , 95% CI (0.37, 0.41),  $t = 34.65$ ,  $p < .001$ ), however, indicated that this multitasking cost was absent for the tracking task, as seen on Fig. 2. In summary, a multitasking cost was found in the visual and 2-back tasks, while participants appeared unaffected by triple-tasking in the tracking task.



**Fig. 2** Task performance in the three tasks and two experimental conditions. Each empty dot represents an individual participant. The black dot and error bars represent the average across participants and 95% confidence intervals

### Response times

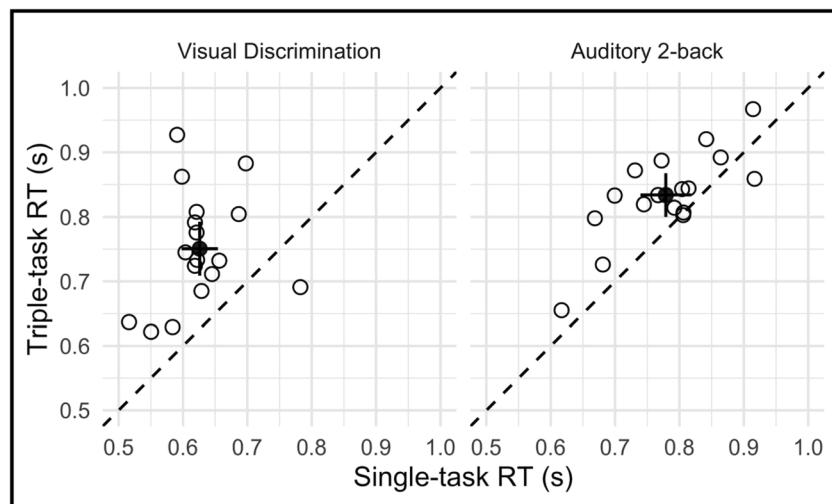
As expected, responses were also slower in the triple-task condition, for both the visual discrimination task and the auditory 2-back task (see Fig. 3). On average, participants' median RTs in the visual discrimination task were 626 ms ( $SD = 60$ ) in the single-task and 751 ms ( $SD = 88$ ) in the triple-task condition. In the auditory 2-back, they were 779 ms ( $SD = 83$ ) for the single-task condition and 834 ms ( $SD = 71$ ) in the triple-task condition.

The effect of triple-tasking on response times was formally analyzed for the visual and the 2-back tasks using an LMM approach, as above. The best fitting LMM (intercept = -0.31, participant<sub>intercept</sub>SD = 0.06, participant<sub>condition</sub>SD = 0.03) described participants' (log-transformed) RTs using condition, task, and their interaction as fixed effects, and a random slope and intercept for the effect of condition for each participant. Participants were slower overall when responding in the triple- relative to the single-task condition ( $\beta_{\text{triple}} = 0.05$ ,

95% CI (0.04, 0.07),  $t = 6.80$ ,  $p < .001$ ), were slower overall in the 2-back relative to the visual discrimination task ( $\beta_{\text{2-back}} = 0.06$ , 95% CI (0.6, 0.7),  $t = 22.49$ ,  $p < .001$ ; although the responses were made via different mediums, key press vs. mouse click), and finally the interaction effect confirmed that the triple-task cost was relatively more pronounced for the visual task compared to the 2-back task ( $\beta_{\text{triple} \times \text{visual}} = 0.04$ , 95% CI (0.03, 0.04),  $t = 12.92$ ,  $p < .001$ ). In summary, for both the visual discrimination and the auditory 2-back task, participants were both less accurate (Task Performance section) and slower to respond in the triple-task, relative to the single-task condition.

### Metacognitive sensitivity

We tested if multitasking induced costs in participants' metacognitive sensitivity, measured as how much self-evaluations predicted performance across trials. Over all participants and tasks, we defined a LMM that described



**Fig. 3** Median response times (RT) for the visual task (left) and auditory 2-back task (right), in the single-task and triple-task conditions. Each empty dot represents an individual participant. The black dot and error bars represent the average across participants and 95% confidence intervals

participants' trial-by-trial performance by their self-evaluation, condition, task, and all possible interactions as fixed effects, and a random slope and intercept for the self-evaluation factor for each participant. Critically, the metacognitive multitasking cost here corresponds to the interaction between the self-evaluation and the condition factors. The best-fitting model (intercept = 0.54,  $\text{participant}_{\text{intercept}}\text{SD} = 0.05$ ,  $\text{participant}_{\text{self-eval}}\text{SD} = 0.08$ ) had a random slope and intercept for the self-evaluation factor for each participant, and a significant main effect of task, condition, and self-evaluation, as well as the three possible double interactions between the three factors: crucially, the self-evaluation  $\times$  condition interaction was significant ( $\beta_{\text{self-eval} \times \text{single}} = 0.04$ , 95% CI (0.02, 0.05),  $t = 4.68$ ,  $p < .001$ ) in the direction hypothesized, with a steeper slope for the self-evaluation factor in the single-task relative to the triple-task condition. We also verified that the decreased metacognitive sensitivity in the triple-task condition was unrelated to the decreased task performance (Online Supplementary Material, section 1.3 – Metacognitive Sensitivity and Task Performance).

As a robustness check, we conducted another analysis based on Spearman's rank correlation ( $\rho$ ) between performance and evaluations as a measure of metacognitive sensitivity (see Fig. 4). An ANOVA on this measure indicated a significant main effect of task ( $F(2,80) = 7.93$ ,  $p < .001$ ,  $\eta^2 = .105$ ,  $\eta^2_p = .165$ ), and critically a significant main effect of condition ( $F(1,80) = 16.66$ ,  $p < .001$ ,  $\eta^2 = .110$ ,  $\eta^2_p = .172$ ), indicating that participants evaluated their performance less accurately in the triple-task condition.

The multitasking cost on metacognitive sensitivity was found in 14 out of 17 participants (or 82%) for the tracking task, and in 12 participants (71%) for the visual and 2-back tasks. In addition, seven participants had a cost for all three tasks and eight had a cost for two out of three tasks.

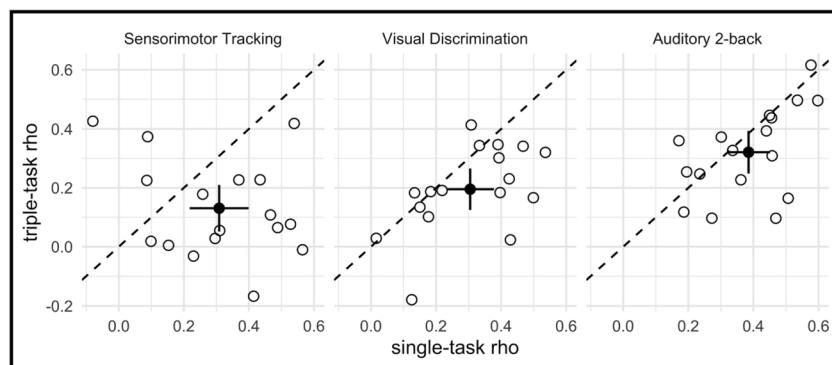
Last, we quantified metacognitive bias as the trial-by-trial differences between objective performance and self-evaluation. We found that 15 out of 17 participants were

underconfident in the triple-task condition (see Online Supplementary Material, section 1.1).

## Supplementary analyses

Our results indicate that multitasking induced a cost in terms of metacognitive sensitivity, in addition to the well-known costs in terms of task performance. We next tested whether the metacognitive decline could be attributed to interference between tasks at the metacognitive level, to inferences at the level of task performance, or to recency effects in self-evaluations (Online Supplementary Material, sections 1.4–1.7).

First, we examined the possible role of a self-evaluation leak (Rahnev et al., 2015) across the three metacognitive evaluations, which may have deteriorated metacognition specifically in the triple task. Although we did find some evidence for such leakage, it did not predict the metacognitive deficit in the triple task across participants. Second, we investigated potential interferences between targets close in time (the psychological refractory period; PRP; Welford, 1952), which may affect performance while remaining inaccessible to self-evaluations (Corallo et al., 2008; Marti et al., 2010). In our data, such PRP effects could indeed be observed on response times. However, they did not impact accuracy or self-evaluations, on a trial-by-trial basis. This suggests that such interference effects were not the cause of the metacognitive deficit in the triple task. Third, we explored the possibility that participants' self-evaluations would suffer from a recency effect when examining performance in the current trial (Locke et al., 2020), and that this limit could be more pronounced in the triple task due to the increased demands. However, although we did find evidence for recency effects in self-evaluations in the tracking task, they were not more pronounced in the triple-task condition.



**Fig. 4** Metacognitive sensitivity (correlation across trials between performance and self-evaluation) for the tracking task (**left**), visual task (**middle**), and auditory 2-back task (**right**), in the single-task (x-axis) and

triple-task (y-axis) conditions. Each empty dot represents an individual participant. The black dot represents the average across participants, with error bars representing the 95% confidence intervals in the two conditions

## Discussion

The goal of this study was to further our understanding of the effects of multitasking on performance metacognition, i.e., the ability to monitor and report one's own performance in a task. We developed a paradigm bridging laboratory experiments to real life scenarios: participants performed three tasks that engaged different modalities, and did so for long (relative to most metacognition studies) periods of time. Participants were then asked to rate how they thought they had performed, on a scale that matched the one of their objective performance: this allowed us to directly compare performance and self-evaluations, and measure participants' metacognitive sensitivity and bias. Nearly all participants showed reduced metacognitive sensitivity in the multitasking condition, in one, two, or all three tasks. In summary, participants incurred a multitasking cost on their ability to assess their own performance. We then ran a series of exploratory analyses to rule out some potential causes of this metacognitive drop: none yielded a single factor that could explain away the metacognitive cost in the triple-task. Thus, the decrement in self-evaluation performance is probably not due to changes in self-evaluation heuristics, but rather demonstrates that multitasking demands globally interfere with metacognitive mechanisms.

This result links together theories in two separate fields. In the multitasking domain, the multiple resources model (Wickens, 2002, 2008) predicts a cost in performance whenever different tasks access the same cognitive resource under time constraints. In the metacognition field, certain aspects of metacognition are thought to be domain-general (de Gardelle & Mamassian, 2014; Mazancieux et al., 2020; Morales et al., 2018); in our paradigm, multiple tasks might concurrently require a central metacognitive resource in order to monitor and report task performance, which in turn will cause a decrease in one's awareness of one's performance. This is precisely what we observed in this study, with participants having worse metacognitive sensitivity in the triple- relative to the single-task condition, even when their actual performance in the task was equated across conditions (i.e., in the tracking task).

In a previous study (Konishi et al., 2020), which used a perceptual dual-task paradigm, we had found no such multitasking cost on metacognition. This resiliency might have different origins: for example, metacognitive processes might use additional sources of information (e.g., global attentional state of the observer) other than the evidence used for the decision (Fleming & Daw, 2017). Moreover, the multiple resource model of multitasking performance cited above predicts that multitasking costs arise only in *overload* situations, that is, when the tasks' demands exceed the cognitive resources available: it is possible that the demands of the

perceptual tasks in our previous study exceeded the resources needed to perform the tasks concurrently (thus causing the first-order performance cost), but not those required to compute the metacognitive decisions. On the contrary, the present study's set-up appeared to overload participants' metacognitive resources.

Importantly, the metacognitive processes studied in this and our previous study are very different: in our previous endeavor we focused on decision confidence, a feeling reflecting an internal estimate of the correctness of a decision. Confidence, like the feeling-of-knowing in the memory domain, can be seen as a fundamental and basic instance of metacognition. On the other hand, the self-evaluations of performance of the present study are a more complex kind of metacognition, requiring a conscious focus of attention over many seconds, and the integration of such performance monitoring in memory. In fact, the processes needed to self-evaluate one's own performance differed across the three tasks, yet all were affected by multitasking. In the tracking task, participants had to continuously monitor the overlap between the cursor and the target. In the visual and 2-back tasks, participants faced a limited number of target events, and had to form a global sense of confidence over a series of discrete decisions (Rouault et al., 2019). Although the continuous evaluation of performance is typically less used in perceptual metacognition studies, it might be equally or more important in real-life contexts such as driving.

Some important questions remain to be answered. Is one task particularly responsible for this metacognitive cost? Future studies might test various combinations of two concurrent tasks to answer this. Similarly, is one specific cognitive process suffering from being shared across tasks? For example, both attention and working memory are fundamental processes needed for these metacognitive evaluations: a different paradigm will have to be devised in order to disentangle their contribution, and their sensitivity, to multitasking. Work by Maniscalco and Lau (2015) provides one clue to this puzzle, showing that working memory manipulation can decrease metacognitive sensitivity in a concurrent task.

We believe the present study to be the first to show that metacognitive sensitivity is impacted by multitasking. Through differing tasks, in differing modalities, the great majority of participants (only one did not have a cost in any task) were less accurate in evaluating their own performance in the multitasking condition, and this was independent of the multitasking cost on task performance. This has far-reaching implications: students media-multitasking while listening to an online course might not just learn less, but also be less aware of what they're missing; a driver on their phone might start to slowly drift off their lane and not realize it; and a researcher too busy switching back and forth from document to email inbox might fail to correct a typo on their manuscript.

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.3758/s13423-021-01967-0>.

**Acknowledgements** The authors acknowledge Isabelle Brunet for help in testing participants; David Aguilar-Lleyda and Tarryn Balsdon for piloting the experiment; and Graeme Godfrey and James Cubiss for manuscript suggestions.

**Author Contributions** Design: MK, BB, VdG, JS. Data collection: MK. Data analysis: MK, VdG, JS. Text preparation: MK, BB, VdG, JS.

## References

- Aguilar-Lleyda, D., Konishi, M., Sackur, J., & de Gardelle, V. (2021). Confidence can be automatically integrated across two visual decisions. *Journal of Experimental Psychology: Human Perception and Performance*, 47(2), 161–171.
- Aguilar-Lleyda, D., Lemarchand, M., & de Gardelle, V. (2020). Confidence as a Priority Signal. *Psychological Science*, 31(9), 1084–1096. <https://doi.org/10.1177/0956797620925039>
- Balsdon, T., Wyart, V., & Mamassian, P. (2020). Confidence controls perceptual evidence accumulation. *Nature Communications*, 11(1), 1753. <https://doi.org/10.1038/s41467-020-15561-w>
- Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, 68(3), 255–278.
- 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>
- Becker, M. W., Alzahabi, R., & Hopwood, C. J. (2012). Media Multitasking Is Associated with Symptoms of Depression and Social Anxiety. *Cyberpsychology, Behavior, and Social Networking*, 16(2), 132–135. <https://doi.org/10.1089/cyber.2012.0291>
- Buser, T., & Peter, N. (2012). Multitasking. *Experimental Economics*, 15(4), 641–655. <https://doi.org/10.1007/s10683-012-9318-8>
- Carrier, L. M., Cheever, N. A., Rosen, L. D., Benitez, S., & Chang, J. (2009). Multitasking across generations: Multitasking choices and difficulty ratings in three generations of Americans. *Computers in Human Behavior*, 25(2), 483–489. <https://doi.org/10.1016/j.chb.2008.10.012>
- Comstock Jr, J. R., & Arnegard, R. J. (1992). *The multi-attribute task battery for human operator workload and strategic behavior research*. National Aeronautics and Space Administration (NASA).
- Corallo, G., Sackur, J., Dehaene, S., & Sigman, M. (2008). Limits on Introspection: Distorted Subjective Time During the Dual-Task Bottleneck. *Psychological Science*, 19(11), 1110–1117. <https://doi.org/10.1111/j.1467-9280.2008.02211.x>
- David, A. S., Bedford, N., Wiffen, B., & Gilleen, J. (2012). Failures of metacognition and lack of insight in neuropsychiatric disorders. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 367(1594), 1379–1390.
- de Gardelle, V., & Mamassian, P. (2014). Does Confidence Use a Common Currency Across Two Visual Tasks? *Psychological Science*, 25(6), 1286–1288. <https://doi.org/10.1177/0956797614528956>
- Dunlosky, J., & Metcalfe, J. (2008). *Metacognition*. Sage Publications.
- Fleming, S. M., & Daw, N. D. (2017). Self-evaluation of decision-making: A general Bayesian framework for metacognitive computation. *Psychological Review*, 124(1), 91–114. <https://doi.org/10.1037/rev0000045>
- García-Pérez, M. A. (1998). Forced-choice staircases with fixed step sizes: Asymptotic and small-sample properties. *Vision Research*, 38(12), 1861–1881. [https://doi.org/10.1016/S0042-6989\(97\)00340-4](https://doi.org/10.1016/S0042-6989(97)00340-4)
- Green, P., & MacLeod, C. J. (2016). SIMR: an R package for power analysis of generalized linear mixed models by simulation. *Methods in Ecology and Evolution*, 7(4), 493–498.
- Hainguierlot, M., Vergnaud, J.-C., & de Gardelle, V. (2018). Metacognitive ability predicts learning cue-stimulus associations in the absence of external feedback. *Scientific Reports*, 8(1), 5602. <https://doi.org/10.1038/s41598-018-23936-9>
- Henmon, V. A. C. (1911). The relation of the time of a judgment to its accuracy. *Psychological Review*, 18(3), 186–201. <https://doi.org/10.1037/h0074579>
- Junco, R. (2012). In-class multitasking and academic performance. *Computers in Human Behavior*, 28(6), 2236–2243. <https://doi.org/10.1016/j.chb.2012.06.031>
- Konishi, M., Compain, C., Berberian, B., Sackur, J., & de Gardelle, V. (2020). Resilience of perceptual metacognition in a dual-task paradigm. *Psychonomic Bulletin & Review*, 1–10.
- Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. (2017). lmerTest package: tests in linear mixed effects models. *Journal of statistical software*, 82(13), 1–26.
- Lee, A. L., de Gardelle, V., & Mamassian, P. (2021). Global visual confidence. *Psychonomic Bulletin & Review*, 1–10.
- Levy, J., & Pashler, H. (2008). Task prioritisation in multitasking during driving: Opportunity to abort a concurrent task does not insulate braking responses from dual-task slowing. *Applied Cognitive Psychology*, 22(4), 507–525. <https://doi.org/10.1002/acp.1378>
- Locke, S. M., Mamassian, P., & Landy, M. S. (2020). Performance monitoring for sensorimotor confidence: A visuomotor tracking study. *Cognition*, 104396.
- Mamassian, P. (2016). Visual Confidence. *Annual Review of Vision Science*, 2(1), 459–481. <https://doi.org/10.1146/annurev-vision-111815-114630>
- Maniscalco, B., & Lau, H. (2015). Manipulation of working memory contents selectively impairs metacognitive sensitivity in a concurrent visual discrimination task. *Neuroscience of Consciousness*, 2015(1). <https://doi.org/10.1093/nc/niv002>
- 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. <https://doi.org/10.1016/j.cognition.2010.01.003>
- Massoni, S., & Roux, N. (2017). Optimal group decision: A matter of confidence calibration. *Journal of Mathematical Psychology*, 79, 121–130. <https://doi.org/10.1016/j.jmp.2017.04.001>
- Mazancieux, A., Fleming, S. M., Souchay, C., & Moulin, C. J. (2020). Is there a G factor for metacognition? Correlations in retrospective metacognitive sensitivity across tasks. : General.
- Morales, J., Lau, H., & Fleming, S. M. (2018). Domain-General and Domain-Specific Patterns of Activity Supporting Metacognition in Human Prefrontal Cortex. *The Journal of Neuroscience*, 38(14), 3534–3546. <https://doi.org/10.1523/JNEUROSCI.2360-17.2018>
- Nelson, T. O., & Narens, L. (1990). Metamemory: A Theoretical Framework and New Findings. *Psychology of Learning and Motivation*, 26, 125.
- Nijboer, M., Borst, J. P., van Rijn, H., & Taatgen, N. A. (2016). Driving and Multitasking: The Good, the Bad, and the Dangerous. *Frontiers in Psychology*, 7. <https://doi.org/10.3389/fpsyg.2016.01718>
- Palmer, E. C., David, A. S., & Fleming, S. M. (2014). Effects of age on metacognitive efficiency. *Consciousness and Cognition*, 28, 151–160.
- Pashler, H. (1994). Dual-Task Interference in Simple Tasks: Data and Theory. *Psychological Bulletin*, 116(2), 25.
- Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., Kastman, E., & Lindeløv, J. K. (2019). PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods*, 51(1), 195–203.

- Rahnev, D., Koizumi, A., McCurdy, L. Y., D'Esposito, M., & Lau, H. (2015). Confidence Leak in Perceptual Decision Making. *Psychological Science*, 26(11), 1664–1680. <https://doi.org/10.1177/0956797615595037>
- Reinecke, L., Aufenanger, S., Beutel, M. E., Dreier, M., Quiring, O., Stark, B., Wölfling, K., & Müller, K. W. (2017). Digital Stress over the Life Span: The Effects of Communication Load and Internet Multitasking on Perceived Stress and Psychological Health Impairments in a German Probability Sample. *Media Psychology*, 20(1), 90–115. <https://doi.org/10.1080/15213269.2015.1121832>
- Reyes, G., Silva, J. R., Jaramillo, K., Rehbein, L., & Sackur, J. (2015). Self-Knowledge Dim-Out: Stress Impairs Metacognitive Accuracy. *PLOS ONE*, 10(8), e0132320. <https://doi.org/10.1371/journal.pone.0132320>
- Rouault, M., Dayan, P., & Fleming, S. M. (2019). Forming global estimates of self-performance from local confidence. *Nature Communications*, 10(1), 1141. <https://doi.org/10.1038/s41467-019-09075-3>
- Salvucci, D. D., & Taatgen, N. A. (2008). Threaded cognition: An integrated theory of concurrent multitasking. *Psychological Review*, 115(1), 101–130. <https://doi.org/10.1037/0033-295X.115.1.101>
- Salvucci, D. D., & Taatgen, N. A. (2010). *The multitasking mind*. Oxford University Press.
- Sana, F., Weston, T., & Cepeda, N. J. (2013). Laptop multitasking hinders classroom learning for both users and nearby peers. *Computers & Education*, 62, 24–31. <https://doi.org/10.1016/j.compedu.2012.10.003>
- Sanbonmatsu, D. M., Strayer, D. L., Biondi, F., Behrends, A. A., & Moore, S. M. (2016). Cell-phone use diminishes self-awareness of impaired driving. *Psychonomic Bulletin & Review*, 23(2), 617–623. <https://doi.org/10.3758/s13423-015-0922-4>
- Spada, M. M., Giustina, L., Rolandi, S., Femie, B. A., & Caselli, G. (2015). Profiling Metacognition in Gambling Disorder. *Behavioural and Cognitive Psychotherapy*, 43(5), 614–622. <https://doi.org/10.1017/S1352465814000101>
- Sun, X., Zhu, C., & So, S. (2017). Dysfunctional metacognition across psychopathologies: A meta-analytic review. *European Psychiatry*, 45, 139–153.
- Wallis, C. (2006). The multitasking generation. *Time Magazine*, 167(13), 48–55.
- Welford, A. T. (1952). The Psychological refractory period and the timing of high-speed performance—A review and a theory. *British Journal of Psychology. General Section; London, Etc.*, 43(1), 2–19.
- Wickens, C. D. (2002). Multiple resources and performance prediction. *Theoretical issues in ergonomics science*, 3(2), 159–177.
- Wickens, C. D. (2008). Multiple resources and mental workload. *Human factors*, 50(3), 449–455.

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.