Resilience of perceptual metacognition in a dual-task paradigm

Mahiko Konishi[1,2] · Clémence Compain[1,2] · Bruno Berberian[3] · Jérôme Sackur[1,4] · Vincent de Gardelle[5]

Abstract
When people do multiple tasks at the same time, it is often found that their performance is worse relative to when they do those same tasks in isolation. However, one aspect that has received little empirical attention is whether the ability to monitor and evaluate one's task performance is also affected by multitasking. How does dual-tasking affect metacognition and its relation to performance? We investigated this question through the use of a visual dual-task paradigm with confidence judgments. Participants categorized both the color and the motion direction of moving dots, and then rated their confidence in both responses. Across four experiments, participants ($N = 87$) exhibited a clear dual-task cost at the perceptual level, but no cost at the metacognitive level. We discuss this resilience of metacognition to multitasking costs, and examine how our results fit onto current models of perceptual metacognition.

Keywords Metacognition · Confidence · Multitasking · Dual-task

Imagine an aircraft pilot encountering a storm during a flight, just as the flight closes in to the destination airport. As the pilot scrambles to monitor different gauges, communicate via radio with air traffic control, and keep the plane steady, the pilot makes a crucial mistake. Is the aircraft pilot unaware of the mistake, precisely because of the multitasking situation the pilot is in? In other words, how is the ability to evaluate one's own performance affected, when we have to perform and monitor multiple tasks at the same time? This is the question that we set out to answer in the current investigation.

✉ Mahiko Konishi
konishi@gmail.com

1 Laboratoire de Sciences Cognitives et de Psycholinguistique, Dept d'Etudes Cognitives, ENS, PSL University, EHESS, CNRS, Paris France

2 Centre d'Economie de la Sorbonne, CNRS & Université Paris 1, Paris France

3 ONERA, The French Aerospace Lab, Information Processing and Systems Department, 13661 Salon Cedex Air France

4 École Polytechnique, Palaiseau France

5 Paris School of Economics and CNRS, Paris France

It has long been known that people can report feelings of confidence that correlate with their objective performance in tasks (Henmon, 1911; Metcalfe & Shimamura, 1994; Nelson & Narens, 1990; Peirce & Jastrow, 1884). In other words, people are able to perform a second-order task (Type 2 task, e.g., a confidence rating) on top of a first-order task (Type 1 task, e.g., a perceptual decision). This second-order cognitive monitoring of first-order processes, also known as metacognition, has been shown to play a crucial role in informing future decisions (Folke, Jacobsen, Fleming, & De Martino, 2017) and in guiding learning (Boldt, Blundell, & De Martino, 2019; Guggenmos, Wilbertz, Hebart, & Sterzer, 2016; Hainguerlot, Vergnaud, & de Gardelle, 2018). As these feelings of confidence seem to be computed at an abstract, task-independent level (de Gardelle, Le Corre, & Mamassian, 2016; de Gardelle & Mamassian, 2014), it is now considered that metacognition involves a higher-order system evaluating information coming from lower, first-order systems (Fleming & Daw, 2017; Maniscalco & Lau, 2016) and relying on pre-frontal areas of the brain (Fleming & Dolan, 2012; Fleming, Huijgen, & Dolan, 2012; Fleming, Weil, Nagy, Dolan, & Rees, 2010).

People often have trouble performing even two simple tasks at the same time (Pashler, 1994), in particular when these two tasks engage the same resources, either peripheral or central (Salvucci & Taatgen, 2008; Wickens, 2002). Surprisingly, despite the growing interest in metacognition in decision-making, there has been only a handful of studies assessing

how individuals evaluate their performance in dual-task conditions (Finley, Benjamin, & McCarley, 2014; Horrey, Lesch, & Garabet, 2009; Maniscalco & Lau, 2015; Sanbonmatsu, Strayer, Biondi, Behrends, & Moore, 2016). In these studies, participants had to evaluate their performance in a primary task while also doing a secondary task, such as evaluating how driving is affected by phone usage (Sanbonmatsu et al., 2016), or how visual discrimination is affected by working memory (Maniscalco & Lau, 2015). In particular, Maniscalco and Lau (2015) reasoned that, as both metacognition and working memory solicit overlapping brain networks, engaging in a concurrent working memory task should impair perceptual metacognition, which is indeed what they found.

However, as these studies only used one metacognitive task, the potential cost of performing two concurrent metacognitive tasks remains unknown. One reason to expect that dual-tasking costs at the metacognitive level is that metacognition is thought to rely on central processes, which have shown to be limited from a processing bottleneck (Collette et al., 2005; Dux, Ivanoff, Asplund, & Marois, 2006; Sigman & Dehaene, 2005; Szameitat, Schubert, Müller, & Von Cramon, 2002). In other words, performing and monitoring multiple tasks concurrently might not only degrade performance in Type 1 tasks, it may also deteriorate Type 2 evaluations, over and above the expected impact that a loss in performance has on metacognition. This would lead to a

potentially catastrophic scenario: The pilot not only makes more errors but also becomes less able to detect these errors.

The aforementioned studies suggest that when two tasks are to be performed concurrently, metacognition is degraded. Here, we wanted to analyze the impact of multitasking on perceptual metacognition in the simplest case, where two equally important tasks are to be performed concurrently and based on the same visual stimulus. In a way, this situation maximizes the overlap between the two tasks (Wickens, 2002), increasing the chances to obtain a cost at the metacognitive level, too. We thus designed a visual dual-task paradigm, using a random-dot kinematogram stimulus that supports two orthogonal tasks. In this task, participants perform blocks of either a motion or a color discrimination task in isolation, or they performed both perceptual tasks together, resulting in a 2 (task: color or motion) × 2 (condition: single or dual) within-subjects experimental design. Each perceptual judgment is followed by a confidence judgment. The stimuli and experimental design are shown in Fig. 1.

We report four experiments conducted with these stimuli and tasks (see Table 1 for a summary). Our first two experiments both used an adaptive staircase procedure, but with different stimulus duration (1,500 and 300 ms) so as to vary task demands and create a time pressure that would minimize the chance that participants performed the two tasks serially instead of concurrently. The results from these first two
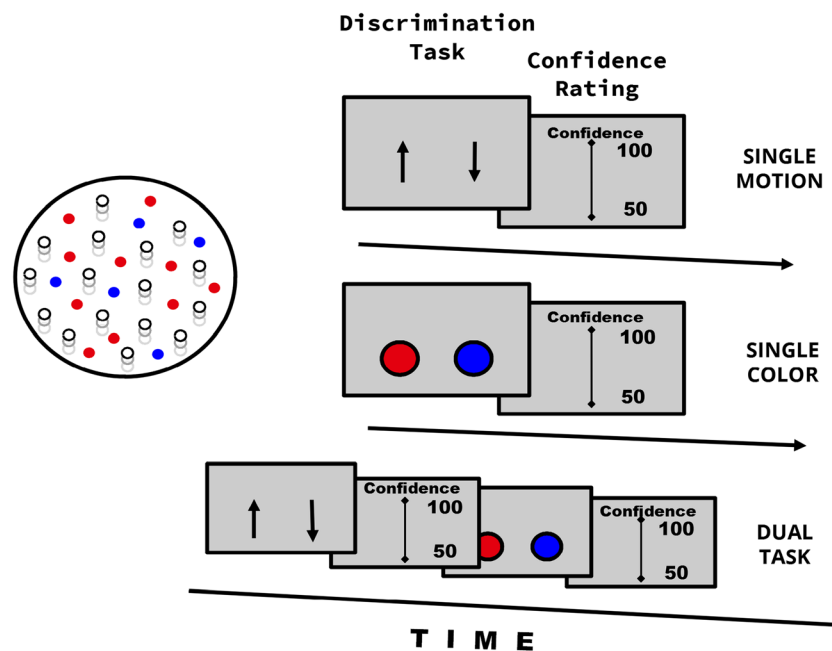
Fig. 1 Dual-task paradigm. Participants were presented with a random-dot kinematogram (shown on the left), and then proceeded to give a perceptual discrimination response, followed by a confidence rating (on a scale from 50 to 100) on that response. On each trial there were either more blue or more red dots (color task), while the white dots moved either upwards or downwards (motion task) with a certain coherence. The design was blocked, so that participants did either 30-trial blocks of motion discrimination, color numerosity discrimination, or both (dual-task condition, shown at the bottom). For the dual-task condition, the order of the responses for the two tasks was counterbalanced across blocks. (Color figure online)

Table 1 Summary of experiments and participants

Experiment	Stimulus duration	Procedure	Initial participants	# excluded for ceiling/floor performance	# excluded for confidence variance	Participants analyzed
1	1,500 ms	Staircase	30	0	3	27
2	300 ms	Staircase	30	1	2	27
3	300 ms	Staircase	33	0	0	33
4	300 ms	Constant stimuli	- 31	7	6	18

Note. Thirty-one participants, the same that completed the short-duration/staircase replication, also completed the short-duration/constant experiment (two dropped out in between sessions)

experiments were surprising: We found that on average, metacognitive efficiency did not degrade in the dual-task condition (it rather *improved* in one task). To evaluate whether this surprising result would replicate, we conducted two additional experiments with short-duration stimuli, one using a staircase method and one using constant stimuli. Using constant stimuli ensures that participants are given the same amount of perceptual information in the single and dual conditions, though their ability to use this information may differ due to dual-tasking costs. Across all four experiments, we found no evidence for a dual-task cost in terms of metacognitive efficiency, despite a clear dual-task cost at the Type 1 level.

Method

Participants

The first two experiments (long-duration/staircase procedure and short-duration/staircase procedure) were conducted at the Laboratory of Experimental Economics in Paris (LEEP), in the Maison des Sciences Économiques of Université Paris 1 Panthéon-Sorbonne, where participants were tested in a large room with 20 individual computers, while the two other experiments (short/staircase replication and short/constant stimuli) were conducted at the Département d'Études Cognitives of the École Normale Supérieure, where participants were tested individually in isolated booths. Participants were recruited from the volunteer databases of the LEEP and of the Relais d'Information en Sciences de la Cognition (RISC) unit, which is largely composed of French students. Each session lasted between 60 and 90 minutes. Ninety-three participants in total (between 18 and 38 years old; 63 women) took part in this study (see Table 1). Participants in the first two experiments were paid a fixed rate plus a bonus depending on a combination of their Type 1 and Type 2 performance: this was calculated using a matching probability rule that was explained to participants beforehand, and which has been shown

to be the best elicitator of people's subjective feelings of confidence (Hollard, Massoni, & Vergnaud, 2016; Karni, 2009). They were paid on average 14€. Participants in the last two experiments were paid a fixed rate of 20€.

Stimuli and tasks

The stimuli and tasks were programmed using Psychtoolbox 3.0.14 (Kleiner et al., 2007) in MATLAB 9.4.0 (R2018a). The main stimulus used in all experiments was a random-dot kinematogram (RDK) consisting of 200 dots (radius of 0.2 deg each approximately with a distance from the screen of ~60 cm, with a finite lifetime of 167 ms) presented in a circular area (radius of approximately 5°) in the center of the screen. One hundred of these dots were white, and these were used for the motion discrimination task. These dots moved with a speed of 1.2 deg/s either upwards or downwards with a coherence that varied on a trial-to-trial basis in the staircase experiments, or was determined after a calibration phase and then fixed throughout the experiment in the constant-stimuli experiment. Participants had to discriminate whether the dots were moving either upwards or downwards on each trial, and gave their response using two keys ("D" for upward, "F" for downwards). The other 100 dots were used for the color task. These always moved without coherence with a speed of 0.3 deg/s. These dots were either red or blue in a certain ratio, which again changed on a trial-to-trial basis in the staircase experiments, or was determined after a calibration phase and then fixed throughout the experiment in the constant-stimuli experiment. Participants had to discriminate if there were more blue or red dots on the screen, on every trial, and responded using two keys ("E" for blue, "R" for red). After each discrimination trial (the Type 1 task), participants rated their confidence on their decision using the mouse cursor on a vertical scale ranging from 100 ("I'm sure I made the correct decision") to 50 ("I responded randomly"). Response times (RTs) were collected for both discrimination and confidence responses, although due to a bug in the program confidence response times were not recorded in two experiments

(Experiments 1 and 2 in Table 1). For the dual-task condition, the order of the responses for the two tasks was counterbalanced across blocks.

Procedure and instructions to participants

Participants signed a consent form and were then instructed on the nature of the Type 1 (the discrimination tasks) and Type 2 (the confidence ratings) tasks. Furthermore, they were encouraged to reflect on their subjective feelings of confidence in their Type 1 decisions and to try to use the whole range of the confidence scale throughout the experiment. They then completed a training session that consisted of three blocks (a single-color, a single-motion, and a dual-task block each) of 30 trials each, for which the staircase procedure was already implemented. In the actual experimental session, the stimulus level started from where it left off in the training session.

Experimental design

All the experiments followed a simple 2 (task) × 2 (condition) within-subjects, blocked design. Participants performed blocks of 30 trials. There were three block types: single-color, single-movement, or dual-task condition. There were 18 blocks, for a total of 540 trials (180 per condition). The 18 blocks were divided into six macroblocks, within which blocks of all three block types were presented to participants once. The order of presentation of conditions in the macroblocks was randomized across participants.

The four experiments presented here all follow the same design and procedure. The only differences were the duration of stimulus presentation, and the procedure used to vary the stimulus difficulty throughout the task. In one study the main stimulus was presented for 1,500 ms (long-duration experiment), and in the other three studies for 300 ms (short-duration experiments). Likewise, three studies used an adaptive staircase procedure, while the last one used constant stimuli. The first two experiments (long-duration and short-duration staircase) had unique subjects participating in them, while the same cohort of participants performed the replication and the constant-stimuli experiments: they completed the staircase experiment in the first session, after which their psychometric curves were fitted to their response data in order to estimate the appropriate stimulus level to elicit 78% correct performance. This was then used in the constant stimulus experiment, for which participants returned for a second session between 1 and 4 days after the first one.

Staircase procedure and constant stimuli

In all staircase experiments, we used a 1-up-1-down staircase with unequal step sizes: step size was of 4% the maximum value of the stimulus, and the ratio of down-step/up-step was 0.2845, which results in the staircase converging to ~78% accuracy (García-Pérez, 1998). In the constant-stimuli experiment, stimulus difficulty for both tasks was estimated using participants' psychometric curves from the first staircase session, in order to achieve ~78% accuracy in the single condition.

Exclusion criteria

We used two exclusion criteria at the participant level. First, we excluded participants who had over 90% accuracy or below 60% accuracy for Type 1 discrimination performance in any of the tasks or conditions. There are two reasons to do so: firstly, this is a sign that the staircase failed to converge for such participants. Secondly, it is problematic to study metacognition in participants that perform close to ceiling or floor level. Because of the relationship between Type 1 and Type 2 information, the former will tend to have perfect metacognition, while the latter will show the opposite pattern. This resulted in the exclusion of 11 participants; these were mostly from the constant stimuli experiment, as using the same stimulus level for both single-task and dual-task conditions raised the chances that participants performed either too close to ceiling or floor level, in one of the two conditions. We also excluded participants that never varied their Type 2 confidence response (for example, by always responding "100"). This is a necessary step because, with no variability in the confidence responses, a measure of metacognitive efficiency cannot be reliably computed. We thus looked at the distribution of the standard deviations of participants' confidence responses in the two tasks and conditions, and decided to exclude participants with a standard deviation lower than five in any task or condition (see Fig. 10 in the Supplementary Plots of the Appendix). This resulted in the exclusion of an additional eight participants.

After applying these exclusion criteria, we analyzed data from 87 participants (13 performed well in the short/staircase replication, but did not pass the exclusion criteria in the short/constant experiment, and their data for the latter was consequently removed and not analyzed), 18 of which took part in both of the last two experiments. A summary of the experiments and participants is available in Table 1. Finally, at the trial level, we excluded every first trial of each new block (for task-switching costs), and every trial for which the log value of the response time was above or below 2.5 times the standard deviation of the log-RT distribution (~2.5% of all trials).

Statistical analyses

Discrimination performance

For each participant, and for each condition, we fitted cumulative Gaussian functions describing the proportion of

responses indicating "upwards" responses in the motion task, and "red" responses in the color task, respectively, as a function of motion coherence, and proportion of red dots relative to blue dots (see Fig. 9 in the Supplementary Plots of the Appendix for an example). This was done with the *quickpsy* package (Linares & López-Moliner, 2016) in the R environment (R Core Team, 2018). The mean and standard deviation were free parameters, while the lapse rate was set to a free parameter with an upper bound of 2% and a lower bound of 1%. It is known that if lapses are not taken into account in the fitting process, it can bias both threshold and slope estimates (Wichmann & Hill, 2001); however, letting the lapse rate vary freely can also induce biases (Prins, 2012). We thus chose these bounds to allow for some degree of variation while keeping the estimates to lapses rates that are usually found in the field. After fitting, we extracted the slopes of the psychometric curves (the inverse of the standard deviation of the cumulative Gaussians) as a measure of discrimination sensitivity. The goodness of fit was never unacceptably poor ($p >$ .05 of the deviance for all the participants, using bootstrapping; Linares & López-Moliner, 2016).

The logic behind our four experiments was to understand whether some specific experimental manipulation (such as stimulus duration, or the staircase procedures) could influence the cost (or the absence of a cost) at the metacognitive level. Given that the stimuli used (the RDK) and the experimental procedures were identical across experiments, we reasoned that the best way to analyze the effects of task and experimental condition (single vs. dual) on discrimination and metacognitive performance, and to increase overall statistical power, was to run linear mixed models across our data set. However, we included the factors (e.g., stimulus duration) that could have influenced performance, as well as having each experiment as a random effect. To analyze the effect of task type (motion vs. color), stimulus duration (300 ms vs. 1,500 ms), and condition (single vs. dual) on participants' discrimination sensitivity, we ran linear mixed models with the R package *lme4* (Bates, Mächler, Bolker, & Walker, 2015). All other data manipulation and plotting was done with the *tidyverse* package (Wickham, 2017). Task type, condition, and duration were included in the model as fixed effects, while participants and experiments were included as random effects. Experimental procedure (staircase/constant stimuli) was initially also added to the models, but had no effect and was never selected as a factor in the best describing models. The p values for fixed effects were calculated using the *lmerTest* (Kuznetsova, Brockhoff, & Christensen, 2017) package, which implements Satterthwaite's method to judge the significance of fixed effects. We then compared all possible models through the *step* function (which performs backward elimination of nonsignificant fixed and random effects) of the *lmerTest* package to find the best fitting model. A list of the best models tested for each analysis is available in Table 3, in

the Appendix. We also conducted post hoc ANOVAs in order to better understand the significant interactions observed with these linear mixed models. Eta square and partial eta square were calculated using the *sjStats* package (Lüdecke, 2020). Bayes factors were calculated with the *BayesFactor* package (Morey & Rouder, 2018).

Metacognitive efficiency

The confidence ratings collected after every perceptual decision were analyzed to measure the efficiency of the metacognitive system. Metacognitive efficiency is measured by the M-ratio (Fleming & Lau, 2014), which quantifies the ability of an observer to discriminate between correct responses and errors, relative to the information they have about the stimulus. Importantly, unlike other measures of metacognition, the M-ratio measure controls for the level of Type 1 performance in the task. This is crucial, as a decrease in Type 2 performance, such as one induced by multitasking, is to be expected following a decrease in Type 1 performance. The M-ratio can thus reveal the true impact of dual-tasking on metacognition. Metacognitive efficiency was computed for each participant, in each task and condition separately, using the *metaSDT* package (Craddock, 2018) in the R environment. To compute M-ratio, the continuous confidence ratings for each participant were median split onto high and low confidence, separately by task and condition (Maniscalco & Lau, 2014). The analysis of the effects of task and condition on metacognitive efficiency was performed in a parallel manner to the one described for discrimination performance in the previous section, using linear mixed models.

Results

Discrimination performance

Figure 2 shows the discrimination sensitivity values for each participant in the two conditions, separately for the two tasks. A dual-task cost can be seen in Figure 2 for both tasks, as most participants showed increased sensitivity in the single-task condition; we tested for the statistical significance of this cost with linear mixed models. The best fitting model here had task, condition, stimulus duration, and their interactions as fixed effects, a random intercept for each participant, and a random intercept by experiment. This model (intercept = 12.7, participant SD = 2.24, experiment SD = 1.29) revealed the following effects of interest: a significant effect of condition, so that the dual-task condition predicted lower discrimination sensitivity (β_{single} = 7.18, 95% CI [4.73, 9.62], t = 5.73, $p <$ .001); a significant effect of stimulus duration (β_{short} = −6.22, 95% CI [−9.53, −2.92], t = 3.29, p = .043), so that the longer duration predicted higher sensitivity; a significant interaction
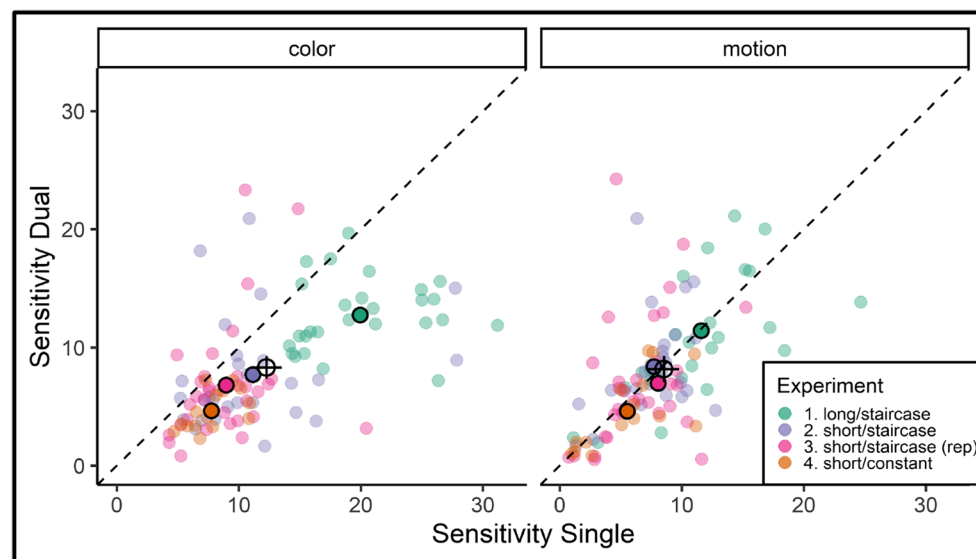
Fig. 2 Discrimination performance. Participants' discrimination sensitivity (as measured by the slope of their psychometric curves) in the single-task and dual-task conditions, for each task. The dashed diagonal lines show equality in sensitivity between single and dual conditions. A point below the dashed line means that the participant had better sensitivity in the single-task than in the dual-task condition(i.e., they experienced a dual-tasking cost). Every point is one subject in one of the two tasks; colors refer to different experiments (see legend). Group averages for each experiment are shown with the black outlined, color-filled shapes. Group averages for each task are shown with the black, empty shapes; error bars within these shapes represent 95% confidence intervals. (Color figure online)

between task and condition, to the effect that the dual-task cost on sensitivity was increased in the color task ($\beta_{motion \times single}$ = −7.01, 95% CI [−10.46, −3.55], t = 3.95, p < .001). To further understand the interaction between task and condition, a post hoc repeated-measures ANOVA, with condition as a within-subjects factor, was ran separately in each task, suggesting that this dual-task cost in discrimination sensitivity was evident in the color task, $F(1, 86) = 45.88$, p < .001, $\eta^2 = .108$, partial $\eta^2 = .765$, but not in the motion task $F(1, 86) = 0.23$, $p = .634$, $BF_{01} = 6.06$.
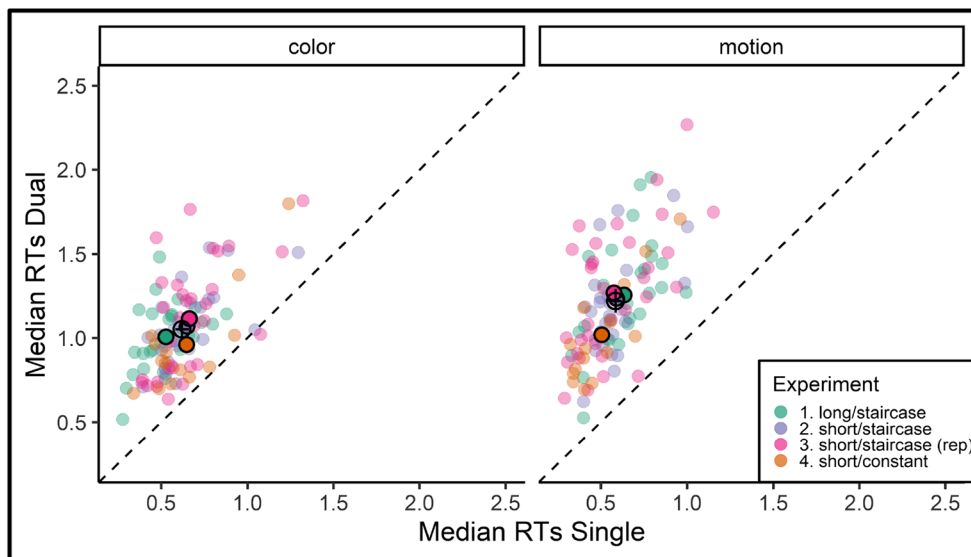
Response times

The median RTs for the Type 1 responses were calculated for each participant and then analyzed with linear mixed models in a parallel manner to the one described for discrimination performance. For the dual-task condition, we only analyzed the RTs for the first of the two tasks, given that responses were done in a sequential manner. Figure 3 shows the median response times for each participant in the two conditions, separately for the two tasks. The best fitting model here had task, condition, and their interactions as fixed effects, a random intercept for each participant and for each experiment. This model intercept = 1.04, participant SD = 0.20, experiment SD = 0.06 showed the following effects: a significant effect of condition, so that the dual-task condition predicted longer RTs (β_{single} = −0.43, 95% CI [−0.48, −0.38], t = 17.59, p < .001); an effect of task (β_{motion} = 0.17, 95% CI [0.12, 0.22], t = 6.92, p < .001), so

that RTs were longer in the motion task; and finally a significant interaction between task and condition, so that this dual-task cost on RTs was more pronounced in the motion task ($\beta_{motion \times single}$ = −0.20, 95% CI [−0.27, −0.14], t = 5.96, p < .001).

Metacognitive efficiency

Figure 4 shows the M-ratio measure of metacognitive efficiency for every participant in the two conditions. Here, the best fitting model had task, condition, and their interaction as fixed effects, and a random intercept for every participant. This model (intercept = 0.86, participant SD = 0.24) had a significant effect of task (β_{motion} = −0.24, 95% CI [−0.32, −0.16], t = 5.63, p < .001), describing participants' better metacognitive efficiency in the color task; a significant effect of condition (β_{single} = −0.17, 95% CI [−0.25, −0.09], t = 4.00, p < .001), showing participants' overall worse metacognition in the single-task condition; and a significant Task × Condition interaction ($\beta_{motion \times single}$ = 0.22, 95% CI [0.10, 0.34], t = 3.62, p < .001). Post hoc repeated-measure ANOVAs, with condition as a within-subjects factor, were run separately in each task to interpret this interaction, confirming what is shown in Fig. 4: Participants showed a metacognitive dual-condition *benefit* in the color task, $F(86) = 17.52$, p < .001, $\eta^2 = .044$, partial $\eta^2 = .226$, while showing no differences in metacognitive efficiency between conditions in the motion task, $F(86) = 1.67$, $p = .20$, $BF_{01} = 3.33$.

Fig. 3 Response times. Participants' median response times in the single-task and dual-task conditions, for each task. The dashed diagonal lines show equality in RTs between single and dual conditions. A point above the dashed line means that the participant had slower RTs in the dual-task relative to the single-task condition (i.e., they experienced a dual-tasking cost). Every point is one subject in one of the two tasks; colors refer to different experiments (see legend). Group averages for each experiment are shown with the black outlined, color-filled shapes. Group averages for each task are shown with the black, empty shapes (in this case heavily overlapping with the experiment averages); error bars within these shapes represent the 95% confidence intervals. For the dual-task condition, we only analyzed the RTs for the first of the two tasks on each trial, given that responses were given in a sequential manner. (Color figure online)

Control analyses

The results from our four experiments were surprising, in part. Not only did we not find evidence for a dual-tasking cost on metacognition, but we also observed a small *benefit* in the task that showed the larger Type 1 cost (the color task). We thus decided to run some control analyses and an additional experiment to rule out several potential simple explanations. We explored the possibility that this benefit could be due to an increase in response times (see Appendix 1.1) or in participants' motivation in the dual-task condition (Appendix 1.4); in a confidence leak between the two tasks (Appendix 1.2); in
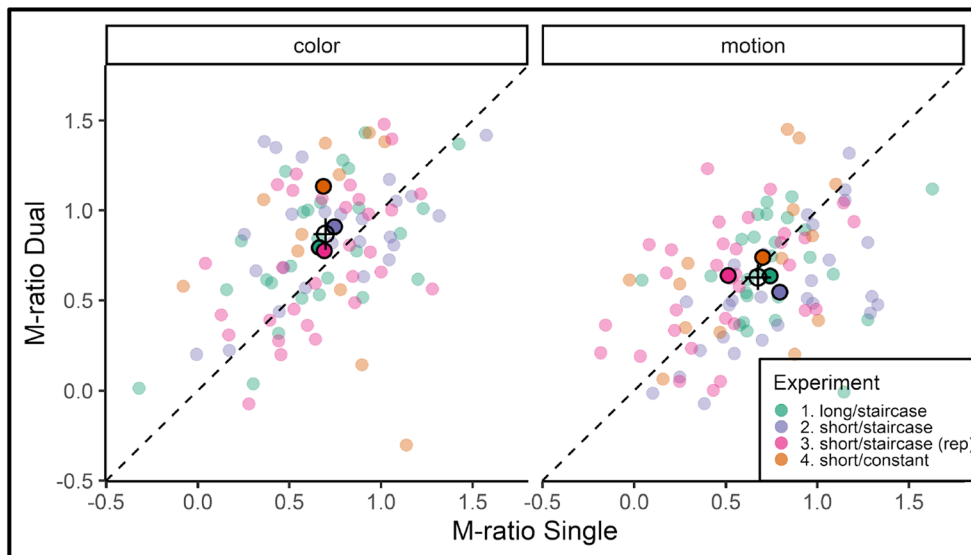
Fig. 4 Metacognitive efficiency. Participants' metacognitive efficiency (as measured by the M-ratio) in the single-task and dual-task conditions, in the two tasks. The dashed diagonal lines show equality in metacognitive efficiency in single-task and dual-task conditions. A point below the dashed line means that the participant had better metacognitive efficiency in the single than in the dual condition. Every point is one subject in one of the two tasks; colors refer to different experiments. Group averages for each experiment are shown with the black outlined, color-filled shapes (the one for the long/staircase experiment is hidden underneath the other ones). Group averages for each task are shown with the black, empty shapes; error bars within the empty circles represent the 95% confidence intervals. (Color figure online)

a between-conditions difference in stimulus variability (Appendix 1.5); or, finally, to a strategy (potentially implemented by some participants) of trading off the two tasks (Appendix 1.3). We also analyzed if the order of response in the dual-task condition (given our design) was affecting discrimination performance, and found evidence against this hypothesis (Appendix 1.9). Ultimately, none of these potential mechanisms were sufficient to explain our results.

Discussion

We adopted a simple visual dual-task paradigm to understand whether the typical dual-task costs in performance associated with multitasking would be accompanied by dual-task costs at the metacognitive level. Across four experiments, we observed a dual-task cost in perceptual discrimination performance and in terms of RTs, but no dual-task cost in metacognitive efficiency. Au contraire, we found evidence for a small, dual-task metacognitive benefit, precisely in the color task. Note that when measuring metacognitive efficiency with the M-diff (a measure closely related to the M-ratio) and metacognitive sensitivity using the area under the Type 2 ROC curve (see Appendix 1.8), we again found no dual-task cost on metacognition, suggesting that our results are independent of which specific measure of metacognition is used. We also found no effect of dual-tasking on metacognitive bias (Appendix 1.7), the general tendency of being overconfident or underconfident, or on the variability of confidence ratings (Appendix 1.10), suggesting that dual-tasking does not affect how individuals anchor or vary their confidence in our setup.

Our study shows that perceptual metacognition can operate efficiently on both tasks in our dual-task condition. This is surprising for two reasons. First, although the information for perceptual decisions is degraded (as can be seen from the dual-task cost at the Type 1 level), it nevertheless remains accessible for a metacognitive judgement in both tasks. In other words, the two metacognitive tasks do not suffer from competing against each other in our paradigm. Second, our results suggest that self-evaluation does not necessarily compete with the central resources that might be needed to coordinate the two concurrent tasks. Further, our results suggest that some information needed for the computation of metacognition (i.e., confidence) is registered early on in the processing stream without degradation caused by the occupation of central processes, and can be read off and recovered effortlessly when a metacognitive report is required.

It has been shown that confidence can be extracted at an abstract level without loss of information (de Gardelle et al., 2016; de Gardelle & Mamassian, 2014). Here, we show that perceptual metacognition is resilient in a different way, in that performing two metacognitive tasks instead of one does not appear to degrade metacognition. This result is consistent with

a recent study in which participants had to identify two targets within a rapid stream of letters, and had to evaluate their confidence in their reports (Recht, Mamassian, & de Gardelle, 2019). In that study, confidence was not perfectly able to capture the accuracy of reports, but, importantly, in a control experiment, metacognitive performance did not depend on whether one or both reports had to be evaluated. Our present result is also in line with the finding of Maniscalco and Lau (2015) that Type 2 efficiency was unimpaired when a perceptual task was performed concurrently with an easy working memory task.

However, it is important to acknowledge that such resilience of metacognition has its limits. For instance, in the same study, Maniscalco and Lau (2015) found that Type 2 efficiency could be degraded if the perceptual task was accompanied by a much more demanding working memory task. In a different paradigm, Corallo, Sackur, Dehaene, and Sigman (2008) had previously shown that participants were introspectively blind to the cost of a secondary concurrent task, when the metacognitive task consisted in the estimation of one's response times. There are thus experimental cases of metacognitive costs induced by multitasking. Nevertheless, these seem to heavily depend on the nature of the interference at the Type 1 level, and on the specificity of the metacognitive task.

One intriguing aspect of our results remains the metacognitive benefit found in the dual task condition for the color task. We explored the possibility that this benefit could be due to an increase in Type 1 response times (see Appendix 1.1) or in participants' motivation in the dual-task condition (Appendix 1.4), in a confidence leak (Rahnev et al. 2015) between the two tasks (Appendix 1.2), in a between-conditions difference in stimulus variability (Rahnev & Fleming, 2019; Appendix 1.5), or to the strategic forfeiting of one task in favor of the other. None of these potential mechanisms could explain our results. Similarly, it would have been interesting to also explore the link between metacognitive efficiency and confidence RTs (Type 2 RTs). Unfortunately, confidence RTs were not available for all of our experiments (see Stimuli and Tasks subsection in the Method section). Further, our paradigm was not optimized to collect such information, as confidence responses were collected on a continuous scale using a mouse, while dynamic models of confidence (Pleskac & Busemeyer, 2010; Yu, Pleskac, & Zeigenfuse, 2015) typically rely on a few confidence categories. Thus, further research would be needed to explore the dynamic construction of confidence from confidence RTs in dual-tasking.

Understanding this increase in metacognitive efficiency thus remains an important open issue for future investigations. Notably, our result is consistent with a recent study (Bang, Shekhar, & Rahnev, 2019) showing that, counterintuitively, increasing sensory noise can result not only in a *decrease* in Type 1 sensitivity but also an *increase* in metacognitive

efficiency. Indeed, our dual-task manipulation lowered Type 1 discrimination sensitivity, while metacognitive efficiency increased (in the color task) or was left untouched (in the motion task). As argued by Bang et al. (2019), this pattern is consistent with hierarchical models of metacognition (Fleming & Daw, 2017; Maniscalco & Lau, 2016), which allow in principle for dissociations between Type 2 and Type 1 performance like the one we observe in the present study.

But let us go back to our endangered and multitasking pilot from the Introduction. Will the pilot be able to recognize the mistake and safely land the plane? Clearly, there is yet much work to be done to understand the impact of multitasking on metacognition in real-life situations. For example, in our study, we asked participants to explicitly rate their confidence in their decisions. Although there is evidence that confidence is spontaneously computed when taking a decision (Aguilar-Lleyda, Konishi, Sackur, & de Gardelle, 2019; Lebreton, Abitbol, Daunizeau, & Pessiglione, 2015), continuously probing someone to evaluate their confidence could help them to maintain awareness on their performance and thus boost their metacognition. One other potentially intriguing factor that could interact in how multitasking affects metacognition, is the level of expertise of the agent: Will an expert pilot be more (because of increased task knowledge) or less (task learning is related to automatization, which is in turn related to decreased conscious awareness) likely to detect the mistake? This is an open question.

The present investigation also used a simple dual-task visual paradigm to measure dual-task costs at the level of perceptual performance and at the metacognitive level. Future experiments could investigate more challenging cases, to further characterize the extent of the resilience of metacognition. For instance, one might use tasks that require exactly the same Type 1 resources (e.g., two independent motion discrimination tasks). Indeed, following current models of multitasking interference (Salvucci & Taatgen, 2008; Wickens, 2002), this situation might give even more leverage to study the factors by which Type 2 multitasking costs relate to Type 1 costs. One other option for further research is to increase the demands put on the cognitive system by engaging the participant in more than two concurrent tasks, so as to reach the multitasking limit of the agent. Further, it might be possible to induce a metacognitive cost by use of concurrent working memory (e.g., Maniscalco & Lau, 2015) and/or sustained attention tasks. Indeed, the paradigm used in the present study faced participants with brief trials, requiring focused attention in a limited time frame. Conversely, multitasking situations in real life often involve a continued effort in time. In our lab, we are currently working to investigate a combination of the situations listed above.

Contributions Design: M.K., C.C., B.B., J.S., V.d.G.; Data collection: M.K., C.C.; Data analysis: M.K., C.C., J.S., V.d.G.; Text preparation: M.K., J.S., V.d.G.

Open practices statement The raw data, the tasks' stimuli, and the scripts used for analyses are available at https://osf.io/e2jpz/.

References

Aguilar-Lleyda, D., Konishi, M., Sackur, J., & de Gardelle, V. (2019). Confidence can be automatically integrated across two visual decisions. *PsyArXiv.* doi:https://doi.org/10.31234/osf.io/3465b

Bang, J. W., Shekhar, M., & Rahnev, D. (2019). Sensory noise increases metacognitive efficiency. *Journal of Experimental Psychology: General, 148*(3), 437–452. doi:https://doi.org/10.1037/xge0000511

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

Boldt, A., Blundell, C., & De Martino, B. (2019). Confidence modulates exploration and exploitation in value-based learning. *Neuroscience of Consciousness, 2019*(1), niz004. doi:https://doi.org/10.1093/nc/niz004

Collette, F., Olivier, L., Van der Linden, M., Laureys, S., Delfiore, G., Luxen, A., & Salmon, E. (2005). Involvement of both prefrontal and inferior parietal cortex in dual-task performance. *Cognitive Brain Research, 24*(2), 237–251.

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. doi:https://doi.org/10.1111/j.1467-9280.2008.02211.x

Craddock, M. (2018). MetaSDT: Calculate Type 1 and Type 2 signal detection measures [Computer software]. Retrieved from https://github.com/craddm/metaSDT

de Gardelle, V., Le Corre, F., & Mamassian, P. (2016). Confidence as a common currency between vision and audition. *PLoS ONE, 11*(1), e0147901. doi:https://doi.org/10.1371/journal.pone.0147901

de Gardelle, V., & Mamassian, P. (2014). Does confidence use a common currency across two visual tasks? *Psychological Science, 25*(6), 1286–1288. doi:https://doi.org/10.1177/0956797614528956

Dux, P. E., Ivanoff, J., Asplund, C. L., & Marois, R. (2006). Isolation of a central bottleneck of information processing with time-resolved fMRI. *Neuron, 52*(6), 1109–1120.

Finley, J. R., Benjamin, A. S., & McCarley, J. S. (2014). Metacognition of multitasking: How well do we predict the costs of divided attention? *Journal of Experimental Psychology: Applied, 20*(2), 158–165. doi:https://doi.org/10.1037/xap0000010

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. doi:https://doi.org/10.1037/rev0000045

Fleming, S. M., & Dolan, R. J. (2012). The neural basis of metacognitive ability. *Philosophical Transactions of the Royal Society, B: Biological Sciences, 367*(1594), 1338–1349. doi:https://doi.org/10.1098/rstb.2011.0417

Fleming, S. M., Huijgen, J., & Dolan, R. J. (2012). Prefrontal contributions to metacognition in perceptual decision making. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *32*(18), 6117–6125. doi:https://doi.org/10.1523/JNEUROSCI.6489-11.2012

Fleming, S. M., & Lau, H. C. (2014). How to measure metacognition. *Frontiers in Human Neuroscience*, *8*, 443. doi:https://doi.org/10.3389/fnhum.2014.00443

Fleming, S. M., Weil, R. S., Nagy, Z., Dolan, R. J., & Rees, G. (2010). Relating introspective accuracy to individual differences in brain structure. *Science*, *329*(5998), 1541–1543.

Folke, T., Jacobsen, C., Fleming, S. M., & De Martino, B. (2017). Explicit representation of confidence informs future value-based decisions. *Nature Human Behaviour*, *1*(1), 0002.

García-Pérez, M. A. (1998). Forced-choice staircases with fixed step sizes: Asymptotic and small-sample properties. *Vision Research*, *38*(12), 1861–1881. doi:https://doi.org/10.1016/S0042-6989(97)00340-4

Guggenmos, M., Wilbertz, G., Hebart, M. N., & Sterzer, P. (2016). Mesolimbic confidence signals guide perceptual learning in the absence of external feedback. *Elife*, *5*, e13388. doi:https://doi.org/10.7554/eLife.13388

Hainguerlot, 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). doi: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. doi:https://doi.org/10.1037/h0074579

Hollard, G., Massoni, S., & Vergnaud, J.-C. (2016). In search of good probability assessors: An experimental comparison of elicitation rules for confidence judgments. *Theory and Decision*, *80*(3), 363–387.

Horrey, W. J., Lesch, M. F., & Garabet, A. (2009). Dissociation between driving performance and drivers' subjective estimates of performance and workload in dual-task conditions. *Journal of Safety Research*, *40*(1), 7–12. doi:https://doi.org/10.1016/j.jsr.2008.10.011

Karni, E. (2009). A mechanism for eliciting probabilities. *Econometrica*, *77*(2), 603–606.

Kleiner, M., Brainard, D., Pelli, D., Ingling, A., Murray, R., & Broussard, C. (2007). What's new in Psychtoolbox-3. *Perception*, *36*(14), 1.

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, *82*(13), 1–26. doi:https://doi.org/10.18637/jss.v082.i13

Lebreton, M., Abitbol, R., Daunizeau, J., & Pessiglione, M. (2015). Automatic integration of confidence in the brain valuation signal. *Nature Neuroscience*, *18*(8), 1159–1167.

Linares, D., & López-Moliner, J. (2016). quickpsy: An R package to fit psychometric functions for multiple groups. *The R Journal*, *8*(1), 122–131.

Lüdecke, D. (2020). sjstats: Statistical functions for regression models (Version 0.17.9) [Computer software]. doi:https://doi.org/10.5281/zenodo.1284472

Maniscalco, B., & Lau, H. (2014). Signal detection theory analysis of Type 1 and Type 2 data: Meta-*d′*, response-specific meta-*d′*, and the unequal variance SDT model. In S. M. Fleming & C. D. Frith (Eds.), *The cognitive neuroscience of metacognition* (pp. 25–66). Berlin, Germany: Springer. doi:https://doi.org/10.1007/978-3-642-45190-4_3

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). doi:https://doi.org/10.1093/nc/niv002

Maniscalco, B., & Lau, H. (2016). The signal processing architecture underlying subjective reports of sensory awareness. *Neuroscience of Consciousness*, *2016*(1). doi:https://doi.org/10.1093/nc/niw002

Metcalfe, J. E., & Shimamura, A. P. (1994). *Metacognition: Knowing about knowing.* The MIT Press.

Morey, R. D., & Rouder, J. N. (2018). BayesFactor: Computation of Bayes factors for common designs [Computer software]. Retrieved from https://CRAN.R-project.org/package=BayesFactor

Nelson, T. O., & Narens, L. (1990). Metamemory: A theoretical framework and new findings. *Psychology of Learning and Motivation*, *26*, 125–173. doi:https://doi.org/10.1016/S0079-7421(08)60053-5

Pashler, H. (1994). Dual-task interference in simple tasks: Data and theory. *Psychological Bulletin*, *116*(2), 220–244. doi:https://doi.org/10.1037/0033-2909.116.2.220

Peirce, C. S., & Jastrow, J. (1884). *On small differences in sensation*. Baltimore, MD: Johns Hopkins University Press.

Pleskac, T. J., & Busemeyer, J. R. (2010). Two-stage dynamic signal detection: A theory of choice, decision time, and confidence. *Psychological Review*, *117*(3), 864–901. doi:https://doi.org/10.1037/a0019737

Prins, N. (2012). The psychometric function: The lapse rate revisited. *Journal of Vision*, *12*(6), 25–25.

R Core Team. (2018). R: A language and environment for statistical computing [Computer software]. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.R-project.org/

Rahnev, D., & Fleming, S. M. (2019). How experimental procedures influence estimates of metacognitive ability. *Neuroscience of Consciousness*, *2019*(1), niz009.

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. doi:https://doi.org/10.1177/0956797615595037

Recht, S., Mamassian, P., & de Gardelle, V. (2019). Temporal attention causes systematic biases in visual confidence. *Scientific Reports*, *9*(1), 1–9.

Salvucci, D. D., & Taatgen, N. A. (2008). Threaded cognition: An integrated theory of concurrent multitasking. *Psychological Review*, *115*(1), 101–130. doi:https://doi.org/10.1037/0033-295X.115.1.101

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. doi:https://doi.org/10.3758/s13423-015-0922-4

Sigman, M., & Dehaene, S. (2005). Parsing a cognitive task: A characterization of the mind's bottleneck. *PLoS Biology*, *3*(2), e37.

Szameitat, A. J., Schubert, T., Müller, K., & Von Cramon, D. Y. (2002). Localization of executive functions in dual-task performance with fMRI. *Journal of Cognitive Neuroscience*, *14*(8), 1184–1199.

Wichmann, F. A., & Hill, N. J. (2001). The psychometric function: I. Fitting, sampling, and goodness of fit. *Perception & Psychophysics*, *63*(8), 1293–1313.

Wickens, C. D. (2002). Multiple resources and performance prediction. *Theoretical Issues in Ergonomics Science*, *3*(2), 159–177. doi: https://doi.org/10.1080/14639220210123806

Wickham, H. (2017). Tidyverse: Easily install and load the "tidyverse" [Computer software]. Retrieved from https://CRAN.R-project.org/package=tidyverse

Yu, S., Pleskac, T. J., & Zeigenfuse, M. D. (2015). Dynamics of postdecisional processing of confidence. *Journal of Experimental Psychology: General*, *144*(2), 489–510. doi:https://doi.org/10.1037/xge0000062