Confidence as a Priority Signal

David Aguilar-Lleyda1, Maxime Lemarchand1, and Vincent de Gardelle1,2

1Centre d’Économie de la Sorbonne (Centre National de la Recherche Scientifique [CNRS] and Université Paris 1 Panthéon-Sorbonne), and 2Paris School of Economics

Abstract
When dealing with multiple tasks, we must establish the order in which to tackle them. In multiple experiments, including a preregistered replication (Ns = 16–105), we found that confidence, or the subjective accuracy of decisions, acts as a priority signal, both when ordering responses about tasks already completed or ordering tasks yet to be completed. Specifically, when participants categorized perceptual stimuli along two dimensions, they tended to first give the decision associated with higher confidence. When participants selected which of two tasks they wanted to perform first, they were slightly biased toward the task associated with higher confidence. This finding extends to nonperceptual decisions (mental calculation) and cannot be reduced to effects of task difficulty, response accuracy, response availability, or implicit demands. Our results thus support the role of confidence as a priority signal, thereby suggesting a new way in which it may regulate human behavior.

Keywords
decision-making, confidence, planning, prioritization, metacognition, open data, open materials, preregistered

Received 9/16/19; Revision accepted 4/17/20

It is common to start a day at work by compiling a to-do list of all the tasks we hope to complete that day. But once the list is written, where do we start? Which task shall we tackle first, and which tasks can be postponed? This problem is sometimes solved by taking into account external constraints: scheduled meetings, imminent deadlines, the limited availability of specific tools or collaborators, and so on. In other circumstances, however, we can freely decide in which order to complete our tasks. In such unconstrained situations, we may perform the tasks in a random order, or we may instead exhibit some systematic preference for doing one task before another. Here, we suggest that confidence plays a role in this prioritization problem. When two tasks have to be done, individuals may first address the task they feel more confident about. When two completed tasks have to be reported, individuals may start with the task in which they felt more confident. In both cases, confidence would be a priority signal.

Confidence corresponds to the feeling of success (predicted or achieved) in a task. This feeling is generally valid, as confidence typically correlates with performance (e.g., Dallenbach, 1913; Henmon, 1911; Peirce & Jastrow, 1884; for reviews, see Lichtenstein, Fischhoff, & Phillips, 1982; Schwartz, 1994), although some dissociations exist (Rahnev, Maniscalco, Luber, Lau, & Lisanby, 2012; Wilimzig, Tsuchiya, Fahle, Einhäuser, & Koch, 2008; Zylberberg, Barttfeld, & Sigman, 2012). Recently, several studies have documented how confidence is not only a commentary on past decisions but also a variable that regulates behavior. For instance, our confidence can determine whether we look for additional information before committing to a choice (Desender, Boldt, & Yeung, 2018) and how much time we invest in a decision (van den Berg, Zylberberg, Barttfeld, & Sigman, 2012).

Here, we suggest that in addition to regulating information accumulation for the task at hand, confidence also affects behavior when we have to determine which task to tackle first. This view naturally follows from...
basic principles of rationality: If two tasks are equal in all dimensions except for confidence, agents seeking to maximize their subjective rewards as quickly as possible would first complete the task associated with higher confidence. Indeed, performing a task successfully is rewarding both intrinsically (e.g., when crossing this item off our to-do list) and extrinsically (e.g., when it provides money or food). Because rational agents first seek the highest expected reward (e.g., to minimize temporal discounting effects), they would first complete the task associated with higher confidence, everything else being equal. To do so, they could compare their confidence across the different tasks at hand (de Gardelle, Le Corre, & Mamassian, 2016; de Gardelle & Mamassian, 2014) and prioritize them accordingly.

We evaluated whether confidence serves as a priority signal in multiple experiments involving elementary decisions. In Experiments 1, 2, and 3, we found evidence that confidence drives priority when participants report on two tasks already completed. In Experiments 4 and 5, we found evidence that confidence drives priority when participants select which of two tasks to tackle first. Our experiments involved both perceptual decisions (Experiments 1–4) and nonperceptual decisions, such as basic mental calculus (Experiment 5).

**Experiment 1**

In our first experiment, we investigated whether confidence relates to the prioritization of responses between two perceptual decisions already made. On each trial (see Fig. 1a), participants were presented with an array of colored letters, and they had to report both the dominant color and the dominant letter in their preferred order. After giving both responses, they gave confidence ratings for both choices, and we evaluated whether confidence predicted the order in which participants gave their responses.

**Method**

**Participants.** The sample for Experiment 1 consisted of 28 participants, who were healthy adults recruited from the Laboratoire d’Économie Experimentale de Paris. They reported normal or corrected-to-normal vision and provided informed consent. Participation was compensated on the basis of performance and duration at an average rate of €12 per hour. The Paris School of Economics ethics committee approved the experimental procedures after determining that they complied with all the ethical regulations. The experiment was conducted in a lab that could accommodate up to 20 participants at a time, and we scheduled as many sessions as we needed to recruit at least 20 participants.

**Sample-size calculation.** Because our measure of prioritization was novel, our sample size for Experiment 1 was not based on prior studies measuring this effect. A sensitivity analysis indicated that the effect size of our main result was above the minimum that we should be able to detect (see the Results section).

**Apparatus.** The experiment was run using MATLAB (The MathWorks, Natick, MA) and the Psychophysics Toolbox (Brainard, 1997) on a 17-in. computer screen (1,024 × 768-pixel resolution, 60-Hz refresh rate) viewed from approximately 60 cm. Stimuli appeared on a gray background.

**Stimuli and procedure.** The experiment was divided into four parts of 4 blocks each, with 25 trials per block. Between each block, participants could rest for 15 s, and at the end of each part, they could take self-timed breaks.

On each trial, the task consisted of reporting the dominant letter and the dominant color of an array of 80 letters (Xs and Os), colored blue or orange (see Fig. 1a). Each trial started with a 500-ms fixation cross, followed by a 300-ms blank period, and then by the stimulus array, which was presented for 1 s. Each letter, in Arial font, occupied approximately 0.5° of visual angle. Elements were presented within a 10°-wide imaginary square. These stimuli were based on the code available at https://github.com/DobyRahnev/Confidence-leak and used in a previous publication (Rahnev, Koizumi, McCurdy, D’Esposito, & Lau, 2015).

After the stimuli disappeared, participants saw a response screen that consisted of four square boxes containing, respectively, a blue square, an orange square, a white O, and a white X (see Fig. 1a). The four boxes were randomly arranged in a 2 × 2 grid on each trial. Participants gave their response by clicking on a color box and a letter box in their preferred order. A first response on one dimension (e.g., clicking on the X box) made the corresponding boxes disappear, leaving only the boxes for the other dimension (e.g., the orange box and the blue box), which required a second response. Participants did not receive feedback on the accuracy of their responses.

After participants made their two choices, they had to rate their confidence on each of them. Two confidence scales were simultaneously presented, one horizontal for the color task and one vertical for the letter task (see Fig. 1a). Each scale featured the two response categories at its ends (each of which corresponded to 100% confidence) and an empty region at its center (near which confidence was at 50%). The two choices participants had made on that trial were surrounded by a yellow circle. Participants gave their confidence levels on both tasks by clicking on the corresponding scales.
Fig. 1. (continued on next page)
Fig. 1. Experiment 1: design and results. An example trial is shown in (a). An array of 80 elements appeared for 1 s. Each element was either an O or an X, colored either blue or orange. After the stimulus array disappeared, a response screen appeared. Four boxes were presented, each containing one of the four elements related to participants’ choices. Participants indicated the dominant color in the color task by clicking on the box containing either the blue or the orange square, and they indicated the dominant letter in the letter task by clicking on the box containing either the O or the X. Then two scales, one for the color task and another for the letter task, appeared on the screen in the shape of a cross. Participants then rated their confidence on each task’s decision, from 50 (total uncertainty) to 100 (total certainty). Observed performance (b) is shown as a function of expected performance, separately for hard and easy versions of both the color and letter tasks. Bars represent averages across participants, and error bars denote 95% confidence intervals. Dots represent individual participants. The upper and lower dashed lines indicate where 90% and 60% observed performance would lie, respectively. The average confidence rating given for the second choice within a trial (c) is shown as a function of the average confidence rating given for the first choice. Dots represent individual participants, and the solid line represents the best linear fit of the data. Participants below the dashed identity line expressed more confidence in their first choices, and participants above the dashed identity line expressed more confidence in their second choices. The density of the proportion of trials in which the first choice had a higher confidence rating than the second choice is shown in (d). The black line corresponds to the whole data set; the green line corresponds to easy color, easy letter (EE) and hard color, hard letter (HH) trials in which both choices were correct; and the red line corresponds to EE and HH trials in which both choices were errors. Density lines were drawn using each participant’s data, and each participant’s average is shown as an individual dot at the bottom. Across-participant averages are displayed as large dots with error bars showing 95% confidence intervals. The dashed line corresponds to equal confidence ratings for first and second choices.

and then validated the ratings by clicking at the center of the screen. Participants were told that maximum uncertainty (no preference between their response and the other option within the same task) corresponded to 50% confidence and absolute certainty on the choice corresponded to 100% confidence.

For each perceptual task, difficulty was manipulated by changing the proportion of the dominant over the nondominant feature. After a short training, we used psychophysical staircases to estimate, separately for each task, the proportion of items that should be used for responses to be 90% correct (easy condition) or 60% correct (hard condition). Details of the training and staircase procedure are presented in Section 1.1 of the Supplemental Material available online. We then used these parameters in the main experimental part to manipulate difficulty on the letter (easy vs. hard) and color (easy vs. hard) tasks, orthogonally within each participant, in a 2 x 2 factorial design. Thus, trials of all conditions were intermixed within blocks.

Performance on the main part of the experiment determined the final amount of money participants won. For each of the 16 blocks, a random trial was chosen. Within that trial, the choice for either color or letter was also randomly picked. Then a number was selected from a uniform distribution between 0 and 100. When the number was lower than the confidence given for that choice, the payoff depended on performance: €1 was awarded if the picked choice was correct, and €0 was awarded otherwise. When the drawn number was higher, another random number was sampled from a uniform distribution between 0 and 100. When this number was smaller than the random number previously drawn, €1 was awarded, and €0 was awarded otherwise. This system has previously been adopted (Massoni, Gajdos, & Vergnaud, 2014) to ensure that participants try to accurately rate their confidence, because their payoff depends on it. Further, the random selection of one trial per block ensured that participants maintained consistent performance and did not relax their attention or automate their responses during some trials.

On top of the variable payoff, participants were awarded a fixed €3 for participation. Before starting the experiment, participants were carefully informed about the procedure. The payoff structure and its objective were made clear, and the fact that reward did not depend on their response times was also pointed out. At the end of each part, participants received feedback on performance, expressed as the percentage of correct trials for that part, collapsed across tasks. Payoff was announced only at the end of the experiment.

Data handling and statistics. In all analyses, data from all participants and trials were included, unless specified otherwise. All reported t tests and binomial exact tests are two-tailed. For t tests, effect size is reported by using Cohen’s $d_z$, written here as $d$, which was calculated using Rosenthal’s (1991) formula: The $t$ statistic was simply divided by the square root of the sample size. Logistic regressions were performed with the glmer function in R, using maximum likelihood and a logit link function.

Results

As Figure 1b shows, observed performance in the experiment closely matched expected performance. Critically, although participants were not asked to report their judgments in a specific order, we expected that they would report first the judgment associated with higher confidence. This was indeed the case on average across trials: Confidence in the first choice was higher than confidence in the second choice (95% confidence
interval, or CI, for the mean difference = [4.17, 8.08]),
κ(27) = 6.43, p < .001, d = 1.21 (see Fig. 1c). This was also
the case within each trial: as seen in Figure 1d, the
proportion of trials in which the confidence associated
with the first choice was greater than that associated
with the second was systematically bigger than .5 (95% CI
for the proportion = [.55, .63]), κ(27) = 4.34, p < .001,
d = 0.82. This effect size is bigger than the minimum
(d = 0.55) we should have been able to detect according
to a sensitivity analysis, given our sample size, an α
(error probability) of .05, and a β (aimed power) of .8.
Furthermore, this pattern held even when the two perceptual
dimensions had the same difficulty level and
when responses were both correct (95% CI for the proportion
= [.51, .60]), κ(27) = 2.40, p = .024, d = 0.45 (see
the green line in Fig. 1d), or both incorrect (95% CI for the proportion
= [.50, .64]), κ(27) = 2.15, p = .041, d = 0.41
(see the red line in Fig. 1d). Note that, by construction,
this later analysis confirmed that participants’ priority was
driven by confidence per se and not simply by task
difficulty or response accuracy (and in any case, response
accuracy was not known by participants, as no trial-by-trial feedback was given to them).

To further demonstrate that confidence affects priority
above and beyond task difficulty and response accuracy, we conducted four more analyses on these data.
First, in a logistic regression model, we found that
responding to the color task first was significantly predicted from the difference in reported confidence
between the two tasks (β = 0.05, SE = 0.002, p < .001),
even when the differences in accuracy and in difficulty
between the two tasks were included as predictors in the
same regression (note that these predictors were
significant, too, both for accuracy (β = 0.002, SE < 0.001,
p < .001) and for difficulty (β = 0.003, SE = 0.001, p = .004)). Second, we compared this regression model with
a simpler one that included only the difference in dificulty and accuracy between tasks, and we found that
including confidence significantly improved the model
(full model: log likelihood = −6,083.62, no-confidence model: log likelihood = −6,644.37, χ²(1) = 1,121.49, p
< .001. Third, we could predict priority from confidence
(p < .001) in a regression in which we included as an
offset the prediction of priority by dificulty and accuracy.
These analyses are presented in more detail in
Section 3 of the Supplemental Material.

In our last analysis, our goal was to pit confidence against accuracy by evaluating whether participants
would prioritize an error made with greater confidence
over a less confident but correct response. To do this,
we focused on trials in which participants made an
error for one dimension and a correct response for the
other but expressed a greater confidence in their error
than in their correct response. Pooling all participants
together (because of the limited number of trials per
participant), we found that the error was indeed prioritized more often than the correct response (proportion
of trials = .57, n = 1,784 trials, binomial exact test: p < .001). Results were similar even when we considered
only trials in which the two dimensions had the same
difficulty level (proportion = .59, n = 658 trials, p < .001). Thus, the role of confidence cannot be reduced
to that of accuracy or difficulty: When confidence was
isolated from both accuracy and difficulty, confidence
still drove prioritization of responses.

Some of these results were associated with small effect sizes: This was the case for those analyses in
which we used a reduced subset of trials. A power analysis revealed that to detect such effect sizes with a
high power, a larger participant sample would be
required. We thus conducted a preregistered replication
of Experiment 1, with a sample size of 105 participants,
which replicated all of the results described above. This
supports the credibility of our original effects, making
it very unlikely that they were caused by random sampling error. See Section 4 in the Supplemental Material
for details on this replication study.

Experiment 2
In Experiment 1, participants may have associated their
confidence ratings with task priority simply because they had to explicitly report their confidence. To eliminate this possibility in Experiment 2, we did not ask for
confidence ratings. We instead relied on task dificulty
as a proxy for confidence, capitalizing on a typical finding in the confidence literature (Dallenbach, 1913;
Kepecs, Uchida, Zariwala, & Mainen, 2008; Peirce &
Jastrow, 1884) that was also confirmed in our first experiment (see Section 2.2 in the Supplemental Material). In Experiment 2, we therefore simply asked participants
to perform the same dual task without confidence rat-
ings, expecting that they would prioritize the easy task.

Method
Participants. A total of 20 individuals participated in
the experiment. They were selected from the same participant pool as in Experiment 1 but did not participate in
that experiment. Selection criteria for participation were
the same. We aimed for a minimum of 20 participants,
slightly above the 19 participants estimated to achieve a power of .9, given the effect size of 0.8 observed in
Experiment 1.

Procedure. The group-session format in which the
experiment was run was the same as in Experiment 1.
Trials in Experiment 2 also closely resembled those in
Experiment 1, except that no confidence had to be rated.
Thus, the trial finished after participants made a choice
for each task on the response screen. Since no confidence ratings were given, final payoff depended only on performance on the main part of the experiment. For each block, we randomly chose a trial, as well as one of the two choices within that trial. When that choice was correct, the participant received €1; otherwise, the participant received €0. The present experiment was shorter than Experiment 1 because of the absence of confidence ratings. For this reason, the previous fixed €3 participation payment was not given here.

Results

Statistical procedures were the same as in Experiment 1. Figure 2a depicts the proportion of times the first response was related to the color task in each of the
four possible kind of trials: easy color, easy letter (EE); easy color, hard letter (EH); hard color, easy letter (HE); and hard color, hard letter (HH). If confidence truly guided the priority of the responses, participants should exhibit a bias toward responding to the easy task before the difficult task. We found that, indeed, the color judgment was reported first more in EH trials than in HE trials (42% of trials vs. 37% of trials). This difference was highly consistent across participants (95% CI for the mean difference = [2%, 10%]), \( t(19) = 3.06, p = .006, \ d = 0.68 \), as illustrated by Figure 2b. In other words, as we predicted, participants were affected by task difficulty in their prioritization and tended to report the easy task first. For completeness, we also confirmed that the same results were obtained in Experiment 1 (see Section 2.3 in the Supplemental Material).

**Experiment 3**

Because the response screen in our previous experiments was presented immediately after the stimuli disappeared, one could argue that only the decision for the easy task had been reached by then. If that were true, the easy-first priority effect could be due to response availability rather than confidence. To rule out this possibility, we conducted Experiment 3, in which we introduced a delay (0 s, 2 s, or 4 s) between stimulus offset and response-screen onset. We reasoned that, if the easy-first priority effect was driven only by response availability at response-screen onset, then it should disappear for long delays.

**Method**

**Participants.** A total of 30 individuals took part in the experiment. Selection procedures were the same as in the previous experiments. We aimed for a minimum of 20 participants given the effect size of 0.8 observed in Experiment 1.

**Procedure.** The stimuli and response procedures in Experiment 3 closely resembled those in Experiment 2. However, in this experiment, we manipulated the time between stimulus offset and the presentation of the response screen, which was set to 0 s, 2 s, or 4 s (counterbalanced across trials). Considering the response times for our previous experiments (see Section 2.4 in the Supplemental Material), 4 s ensured that participants had reached both decisions by the time the response screen was presented. Payoff was determined as in Experiment 2.

**Results**

Statistical procedures for \( t \) tests were the same as in Experiment 1. All analyses of variance (ANOVA) were repeated measures. For ANOVA, effect sizes are reported by using \( \eta_p^2 \). For each factor, this was calculated by taking the sums of squares of the condition and dividing them by the sums of squares of the condition plus the sums of squares of the error (Lakens, 2013).

Taking the current sample, we calculated, as in previous experiments, the proportion of trials in which color was responded to first in EH trials and in HE trials, separately for each delay condition. We found that the main effect of trial type was significant overall, \( F(1, 29) = 6.10, p = .020, \ \eta_p^2 = .17 \), with an easy-first bias replicating our findings in the previous studies. Although the main effect of delay was also significant, \( F(2, 29) = 12.76, p < .001, \ \eta_p^2 = .31 \), it did not interact with trial type (\( p = .679 \); see also Fig. 2c). These data thus ruled out the concern that participants responded to the easy task first because it was the only task for which a decision was available by the time the response screen was presented.

In analyses from Experiments 2 and 3, we used task difficulty as a proxy for confidence, because of the typical finding that confidence is greater for easier tasks. However, it is also known that when only errors are considered, confidence decreases when the task becomes easier (Kepecs et al., 2008; see also Fig. S3a in the Supplemental Material). We thus reasoned that prioritization of the easy task might actually reverse when both responses were errors. Pooling across the data from our three experiments so far (to gain statistical power), we confirmed this prediction. Indeed, when both responses were correct, the proportion of trials in which color was responded to first was higher in EH trials than in HE trials (95% CI for the difference in proportions = [.06, .11]), \( t(77) = 6.25, p < .001, \ d = 0.71 \), indicating that participants prioritized the easy task, whereas when both responses were errors, they prioritized the harder task (95% CI for the difference in proportions = [−.28, −.07]), \( t(56) = −3.21, p = .002, \ d = −0.43 \). When we compared these two conditions, the inversion of the effect was clear (95% CI for the inversion = [.14, .37]), \( t(56) = 4.38, p < .001, \ d = 0.58 \). This last analysis unequivocally demonstrates that neither accuracy nor task difficulty alone can fully explain prioritization between responses and that confidence has a key role in driving priority. We obtained the same results when performing these analyses with the preregistered replication of Experiment 1 (see Section 4 in the Supplemental Material).

**Interim Discussion**

The experiments described so far showed that confidence acts as a priority signal at the level of response execution: When engaged in a dual task, participants...
tend to first communicate the response associated with greater confidence. This pattern could not be explained solely by task difficulty and response accuracy, by implicit demand effects (Experiment 2), or by response-availability effects (Experiment 3).

In the final experiments, we investigated prioritization between two tasks instead of between responses. In other words, when two tasks have to be completed, would participants engage first in the task they feel more confident about? In such situations, participants may rely on prospective confidence (their anticipation about the probability of success in a task) based on prior experience with the same task (Fleming, Massoni, Gajdos, & Vergnaud, 2016; Wright, 1982).

**Experiment 4**

In Experiment 4, participants were familiarized with two sets of trials, one easy and one difficult. The two sets were not explicitly labeled as easy and difficult to participants but were instead associated with random animal names. Participants did not receive any feedback on performance. In the subsequent test phase, participants had to complete more trials from both sets, but they could choose whether to start with the easy or the difficult set. Under our hypothesis, we expected participants to prioritize the easy set, which should be related to higher confidence.

**Method**

**Participants.** A total of 31 individuals took part in the experiment. Selection procedures were the same as in the previous experiments. We aimed for a minimum of 20 participants given the effect size of 0.8 observed in Experiment 1.

**Procedure.** In Experiment 4, participants viewed the same stimuli as in preceding experiments, but they only responded to one task on each trial. Therefore, the response screen contained only the two relevant boxes. The main part was organized in 32 blocks. All blocks included a familiarization phase and a test phase (Fig. 3a). During the familiarization phase, participants completed two sets of six trials each. Each set was associated with a single task (color or letter), a difficulty level (easy or hard), and an animal name (randomly chosen without replacement from a predefined list of common names). Using staircases run before the main phase (see Section 1.1 in the Supplemental Material), we experimentally controlled the proportion for the task-relevant dimension such that for each block, one set contained difficult trials and the other easy trials (for the nonrelevant dimension, the proportion was always set at 50%). No feedback was given about accuracy. This manipulation was not revealed to participants, but they could still learn the difficulty of each set by performing the task during the familiarization phase.

After familiarization, the test phase started. Participants saw a screen with two horizontally aligned boxes presenting (in a random position) the animal names of the two sets just completed. There were four trials of each set. Participants knew that performance in these test trials would determine their final payoff. Importantly, participants were given no feedback about their accuracy, and payoffs were revealed only at the end of the experiment. Participants decided the order in which to complete the test trials of each set by clicking on the name of the set they desired to complete first. We used animal names to avoid any obvious order bias (set A vs. set B, or Set 1 vs. Set 2) for this decision (for the complete list of set names, see Section 9.1 in the Supplemental Material). Across blocks, we factorially manipulated the task (color vs. letter) used for the first set in the block, the task (color vs. letter) used for the second set, and the order of the difficulty levels used for the two sets (easy first vs. hard first). This design resulted in eight conditions, presented in a randomized order as a sequence 4 times per participant.

The payoff in Experiment 4 was similar to the payoff in Experiments 2 and 3, but given the more numerous and shorter blocks, rewards for each block were limited to €0.5. The rest between blocks was also shorter than in preceding experiments (only 10 s).

**Results**

Statistical procedures for t tests are the same as in Experiment 1. The Bayes factor was calculated by using the BayesFactor package in R, with priors set to its default value $\sqrt{2}/2$, in accordance with the recommendations of Morey and Rouder (2011).

Results were consistent with what we found in previous experiments with the prioritization of easy tasks: Participants in the present experiment chose more often to face the easy set first (95% CI for the proportion = [0.50, 0.56]), $t(30) = 2.28$, $p = .030$, $d = 0.41$ (Fig. 3c). These results demonstrate that prospective confidence can establish the priority of one task over another. Interestingly, whether the two sets involved different tasks or the same task did not affect this priority effect ($p = .26$, Bayes factor $= 0.35$). This suggests that the comparison of the two sets involved relatively abstract representations, as was found for confidence comparison (de Gardelle et al., 2016; de Gardelle & Mamassian, 2014).

**Experiment 5**

Our final experiment evaluated the generalizability of our findings beyond perceptual decisions. Here, we tested whether individuals would prioritize easier
mental-calculation problems over difficult ones. On each trial, two schematics of mental-calculation problems were presented (Fig. 4a). One was a priori more difficult than the other, in the sense that it included a multiplication or a division operation (e.g., \(x + \frac{y}{z}\)), instead of only addition or subtraction (e.g., \(x + y - z\)). Participants had to click on one of the schematics to reveal the actual problem, which they then had to solve. Immediately after, participants had to solve the non-chosen problem. We expected that participants would select the easy problem first.

**Method**

**Participants.** A total of 101 individuals participated in the experiment. Unlike the samples used in the previous experiments, this sample was not independent. Participants from Experiments 1, 3, and 4 and from the experiment reported in Section 7.2 in the Supplemental Material completed this experiment after their respective initial experiments. We aimed for a minimum of 20 participants given the effect size of 0.8 observed in Experiment 1.

**Procedure.** The stimuli used in Experiment 5 were different from those used in the previous experiments. On each trial, two horizontally aligned boxes appeared (see Fig. 4a), each containing a simple formula (e.g., \(a + 2b\)). Our formulas involved letters connected by basic operators (+, −, ×, ÷) and were sometimes nested within parentheses. On each trial, the two presented problems were similar (they had the same number of elements and operations) except for the operators connecting their respective letters.
Confidence as a Priority Signal

One of the problems (easy) contained addition and subtraction, whereas the other (hard) contained multiplication and division. The easy and hard boxes were randomly positioned on the left or the right of the screen on every trial. Participants were allowed to solve both problems in their preferred order. When they clicked on a box, the letters of the formula in that box were replaced by single-digit integers. Participants then typed their response, which appeared inside a box below the problem. They could change the response as they wished before pressing the Enter key to validate it. Then the problem they had not selected appeared, and participants entered their responses by the same method. Each participant completed the same 20 pairs of questions (listed in Section 9.2 in the Supplemental Material) but in a random order. The initial presentations of formulas ensured that participants could not try to solve the problems before making a choice.

Experiment 5 had a payoff system that mimicked those of the previous experiments. At the end of the experiment, two trials were chosen randomly. For each of these selected trials, the answer to one of the problems was picked. When that answer was correct, participants received €1; otherwise, they received nothing.

Results

We could not find a difference in terms of response accuracy between the easy and hard problems (89% correct vs. 89% correct, p = .883, Bayes factor = 0.11), probably because of the absence of time limits for response. This made us focus on another performance-related variable: response times. These were significantly longer for the hard problems than for the easy problems (95% CI for the mean difference = [3.01, 4.50], t(100) = 10.00, p < .001, d = 0.995 (Fig. 4b), proving that these problems were more demanding. Participants mostly chose to solve the easy problem first (95% CI for the proportion = [0.59, 0.68]), t(100) = 6.15, p < .001, d = 0.61 (Fig. 4c). This experiment thus replicated our previous findings of an easy-first bias in task prioritization and extended those findings beyond perceptual tasks.

General Discussion

We found that when having to prioritize responses or tasks, participants tended to engage first in the response or the task in which they had greater confidence. We replicated this finding in five experiments using confidence ratings or proxies for confidence (e.g., task difficulty). This effect was present in situations involving retrospective confidence (Experiments 1–3) and prospective confidence (Experiments 4 and 5) and involving both perceptual and nonperceptual tasks, such as mental calculation. We also found that our priority effect based on confidence was reliable across time in most cases (see Section 8 in the Supplemental Material).

Situations in which we face multiple tasks are highly diverse, and task prioritization is likely determined by a varied set of factors. For instance, past studies have shown that prioritization can depend on personal interest and familiarity (Spink, Park, & Koshman, 2006), task duration (Ibanez, Clark, Huckman, & Staats, 2018), or the timing of task-relevant information (Sigman & Dehaene, 2006). Prioritization also seems to depend on deadlines in a nontrivial way: When both deadlines are
far off, people dedicate more resources to the task with the most-distant deadline and to tasks with goals intended to achieve positive outcomes (i.e., approach goals), but these patterns reverse when deadlines are imminent (Ballard, Yeo, Loft, Vancouver, & Neal, 2016; Schmidt & DeShon, 2007; Vancouver, Weinhardt, & Schmidt, 2010). In our paradigm, we did not manipulate any of these factors so we could better isolate the role of confidence. We therefore claim only that confidence is one among many forces driving task prioritization, and we acknowledge that its role in real-life situations also depends on the importance of these other forces.

In our data, we found a large interindivdual variability in the general tendency to prioritize the color or the letter task (see Fig. 2a; see also Section 76.1 in the Supplemental Material); this variability remains to be explained. The locations of the different choice options on the screen could also bias task priority (see Section 76.2 in the Supplemental Material). Another factor that deserves more scrutiny is reward. Our experiments purposefully avoided any difference in reward between the two tasks within a trial, and nothing in the instructions or the reward structure pushed participants to prioritize one task over the other or to rely on confidence for prioritization. Some of the aforementioned features could be manipulated in future studies.

Another important issue for future research is whether this strategy of approaching the easy task first is adaptive. A priori, reducing the mental workload before tackling more difficult tasks could be beneficial. In our data, however, we found no clear evidence for this. Participants were neither better nor worse when responding first to the easy task than when responding first to the hard task (see Section 7.1 in the Supplemental Material). To further investigate this issue, we conducted an additional experiment (based on Experiment 2) in which response order was either free or imposed (see Section 7.2 in the Supplemental Material). Participants' performances were similar between these two conditions and did not depend on whether the easy task was responded to first or second, suggesting that, in our paradigm, the confidence-driven prioritization does not lead to an advantage in task performance. In other words, our study found no advantage for confidence-driven prioritization.

Our study highlights a new way in which confidence actively shapes our behavior, adding to a growing literature showing how confidence actually affects individual or collective decisions. Previous research showed how when agents interact, the confidence with which advice is expressed affects the use of this information by others (Paese & Kinnaly, 1993; Zalesny, 1990; Zarnoth & Sniezek, 1997) and how well the group will perform (Massoni & Roux, 2017). At the individual level, confidence may serve as a teaching signal when feedback is unavailable (Daniel & Pollmann, 2012; Guggenmos, Wilbertz, Hebzer, & Sterzer, 2016; Hainguerlot, Vergnaud, & de Gardelle, 2018; Zylberberg, Wolpert, & Shadlen, 2018). It also influences the amount of resources people engage in a task (Desender et al., 2018; van den Berg et al., 2016). The present study suggests that confidence also contributes to shaping not only how people undertake a task but also what task they decide to do first. This might also explain why confidence is such a salient part of our subjective experience.

Understanding how people decide to perform one task before another may have practical consequences for the management of individuals and organizations. This motivated several past studies showing that individuals address easier tasks first when they can decide their task schedule. For instance, students appear to prioritize easier course assignments (Puffer, 1989) or Web-search tasks (Spink et al., 2006) over difficult ones. Physicians at an emergency department prioritize easy patient cases, especially when under heavy workloads (KC, Staats, Kouchaki, & Gino, 2017). Our study makes several unique contributions to this research topic: We extended this result to situations involving immediate decisions, we linked the easy-first bias to confidence, and we offered strict control over performance in the task via our psychophysical procedures.

Finally, although we did not find a clear consequence from the easy-first bias on performance, we note that one previous study on physicians has associated this strategy with both short-term benefits (treating more patients per unit of time) and long-term costs (across individuals, it predicted lower productivity and thus less revenue for the hospital; KC et al., 2017). Understanding whether similar short-term benefits and long-term costs may arise in other contexts (e.g., those involving multiple perceptuomotor tasks) is an exciting topic for future research.

**Transparency**

*Action Editor:* D. Stephen Lindsay  
*Editor:* D. Stephen Lindsay  
*Author Contributions*  
D. Aguilar-Lleyda and V. de Gardelle conceived the study and designed the experiments. D. Aguilar-Lleyda programmed the stimuli. D. Aguilar-Lleyda, M. Lemarchand, and V. de Gardelle collected and analyzed the data. D. Aguilar-Lleyda and V. de Gardelle wrote the manuscript. All authors approved the final version of the manuscript for submission.

*Declaration of Conflicting Interests*  
The author(s) declared that there were no conflicts of interest with respect to the authorship or the publication of this article.
Funding
This work was supported by the Agence Nationale de la Recherche Grant No. ANR-16-CE28-0002 to V. de Gardelle.

Open Practices
Codes to run the stimuli, data sets, and the scripts to analyze the data for all of the experiments have been made publicly available via OSF and can be accessed at https://osf.io/kcvgy/. The design and analysis plans for the replication of Experiment 1 were preregistered at https://osf.io/etzxq. The other experiments were not preregistered. The complete Open Practices Disclosure for this article can be found at http://journals.sagepub.com/doi/suppl/10.1177/0956797620925039. This article has received badges for Open Data, Open Materials, and Preregistration. More information about the Open Practices badges can be found at http://www.psychologicalscience.org/publications/badges.

ORCID iDs
David Aguilar-Lleyda https://orcid.org/0000-0001-6963-4069
Vincent de Gardelle https://orcid.org/0000-0001-5977-5597

Acknowledgments
We thank Maxim Frolov, Audrey Azerot, and Quentin Cavalan for help with data collection. We thank Mahiko Komishi, Jean-Christophe Vergnau, Jérôme Sackur, Bruno Berberian, and Pascal Mamassian for discussions. We thank Dobromir Rahnev for making available his experiments’ code.

Supplemental Material
Additional supporting information can be found at http://journals.sagepub.com/doi/suppl/10.1177/0956797620925039

References
Ballard, T., Yeo, G., Loft, S., Vancouver, J. B., & Neal, A. (2016). An integrative formal model of motivation and decision making: The MGPM*. Journal of Applied Psychology, 101, 1240–1265. doi:10.1037/apl0000121
Brainard, D. H. (1997). The psychophysics toolbox. Spatial Vision, 10, 433–446.
Dallenbach, K. M. (1913). The relation of memory error to time interval. Psychological Review, 20, 323–337. doi:10.1037/h0076105
Daniel, R., & Pollmann, S. (2012). Striatal activations signal prediction errors on confidence in the absence of external feedback. NeuroImage, 59, 3457–3467. doi:10.1016/j.neuroimage.2011.11.058
de Gardelle, V., Le Corre, F., & Mamassian, P. (2016). Confidence as a common currency between vision and audition. PLOS ONE, 11(1), Article e0147901. doi:10.1371/journal.pone.0147901
de Gardelle, V., & Mamassian, P. (2014). Does confidence use a common currency across two visual tasks? Psychological Science, 25, 1286–1288. doi:10.1177/0956797614528956
Desender, K., Boldt, A., & Yeung, N. (2018). Subjective confidence predicts information seeking in decision making. Psychological Science, 29, 761–778. doi:10.1177/0956797617744771
Fleming, S. M., Massoni, S., Gajdos, T., & Vergnau, J.-C. (2016). Metacognition about the past and future: Quantifying common and distinct influences on prospective and retrospective judgments of self-performance. Neuroscience of Consciousness, 2016(1), Article niw018. doi:10.1093/neuropsychologia/niw018
Guggenmos, M., Wilbertz, G., Hebart, M. N., & Sterzer, P. (2016). Mesolimbic confidence signals guide perceptual learning in the absence of external feedback. eLife, 5, Article e13388. doi:10.7554/eLife.13388
Hainguerlot, M., Vergnau, J.-C., & de Gardelle, V. (2018). Metacognitive ability predicts learning cue-stimulus associations in the absence of external feedback. Scientific Reports, 8, Article 5602. doi: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.
Ibanez, M. R., Clark, J. R., Huckman, R. S., & Staats, B. R. (2018). Discretionary task ordering: Queue management in radiological services. Management Science, 64, 4589–4607. doi:10.1287/mnsc.2017.2810
KC, D. S., Staats, B. R., Kouchaki, M., & Gino, F. (2017). Task selection and workload: A focus on completing easy tasks hurts long-term performance (Working Paper 17-112). Harvard Business School. Retrieved from https://www.hbs.edu/faculty/Publication%20Files/17-112_54df950-a08d-4ba8-a718-1150dc8916cb.pdf
Kepecs, A., Ucha, N., Zariwala, H. A., & Mainen, Z. F. (2008). Neural correlations, computation and behavioural impact of decision confidence. Nature, 455, 227–231. doi:10.1038/nature07200
Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: A practical primer for t-tests and ANOVAs. Frontiers in Psychology, 4, Article 863. doi:10.3389/fpsyg.2013.00863
Lichtenstein, S., Fischhoff, B., & Phillips, L. D. (1982). Calibration of probabilities: The state of the art before 1980. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), Judgment under uncertainty: Heuristics and biases (pp. 306–334). Cambridge, England: Cambridge University Press.
Massoni, S., Gajdos, T., & Vergnau, J.-C. (2014). Confidence measurement in the light of signal detection theory. Frontiers in Psychology, 5, Article 1455. doi:10.3389/fpsyg.2014.01455
Massoni, S., & Roux, N. (2017). Optimal group decision: A matter of confidence calibration. Journal of Mathematical Psychology, 79, 121–130. doi:10.1016/j.jmp.2017.04.001
Morey, R. D., & Rouder, J. N. (2011). Bayes factor approaches for testing interval null hypotheses. Psychological Methods, 16, 406–419. doi:10.1037/a0024377
Paese, P. W., & Kinnaly, M. (1993). Peer input and revised judgment: Exploring the effects of (un) biased confidence. Journal of Applied Social Psychology, 23, 1989–2011. doi:10.1111/j.1559-1816.1993.tb01076.x
Peirce, C. S., & Jastrow, J. (1884). On small differences in sensation. Memoirs of the National Academy of Sciences, 3, 75–83.
Puffer, S. M. (1989). Task-completion schedules: Determinants and consequences for performance. *Human Relations, 42*, 937–955. doi:10.1177/001872678904201005

Rahnev, D., Koizumi, A., McCurdy, L. Y., D’Esposito, M., & Lau, H. (2015). Confidence leak in perceptual decision making. *Psychological Science, 26*, 1664–1680. doi:10.1177/0956797615595037

Rahnev, D. A., Maniscalco, B., Luber, B., Lau, H., & Lisanby, S. H. (2012). Direct injection of noise to the visual cortex decreases accuracy but increases decision confidence. *Journal of Neurophysiology, 107*, 1556–1563. doi:10.1152/jn.00985.2011

Rosenthal, R. (1991). *Meta-analytic procedures for social research*. Thousand Oaks, CA: SAGE. doi:10.4135/9781412984997

Schmidt, A. M., & DeShon, R. P. (2007). What to do? The effects of discrepancies, incentives, and time on dynamic goal prioritization. *Journal of Applied Psychology, 92*, 928–941. doi:10.1037/0021-9010.92.4.928

Schwartz, B. L. (1994). Sources of information in metacognition: Judgments of learning and feelings of knowing. *Psychonomic Bulletin & Review, 1*, 357–375. doi:10.3758/BF03213977

Sigman, M., & Dehaene, S. (2006). Dynamics of the central bottleneck: Dual-task and task uncertainty. *PLOS Biology, 4*(7), Article e220. doi:10.1371/journal.pbio.0040220

Spink, A., Park, M., & Koshman, S. (2006). Factors affecting assigned information problem ordering during Web search: An exploratory study. *Information Processing & Management, 42*, 1366–1378. doi:10.1016/j.ipm.2006.01.007

Vancouver, J. B., Weinhardt, J. M., & Schmidt, A. M. (2010). A formal, computational theory of multiple-goal pursuit: Integrating goal-choice and goal-striving processes. *Journal of Applied Psychology, 95*, 985–1008. doi:10.1037/a0020628

van den Berg, R., Zylberberg, A., Kiani, R., Shadlen, M. N., & Wolpert, D. M. (2016). Confidence is the bridge between multi-stage decisions. *Current Biology, 26*, 3157–3168. doi:10.1016/j.cub.2016.10.021

Wilimzig, C., Tsuchiya, N., Fahle, M., Einhäuser, W., & Koch, C. (2008). Spatial attention increases performance but not subjective confidence in a discrimination task. *Journal of Vision, 8*(5), Article 7. doi:10.1167/8.5.7

Wright, G. (1982). Changes in the realism and distribution of probability assessments as a function of question type. *Acta Psychologica, 52*, 165–174. doi:10.1016/0001-6918(82)90033-6

Zalesny, M. D. (1990). Rater confidence and social influence in performance appraisals. *Journal of Applied Psychology, 75*, 274–289. doi:10.1037/0021-9010.75.3.274

Zarnoth, P., & Sniezek, J. A. (1997). The social influence of confidence in group decision making. *Journal of Experimental Social Psychology, 33*, 345–366. doi:10.1006/jesp.1997.1326

Zylberberg, A., Barttfeld, P., & Sigman, M. (2012). The construction of confidence in a perceptual decision. *Frontiers in Integrative Neuroscience, 6*, Article 79. doi:10.3389/fnint.2012.00079

Zylberberg, A., Wolpert, D. M., & Shadlen, M. N. (2018). Counterfactual reasoning underlies the learning of priors in decision making. *Neuron, 99*, 1083–1097. doi:10.1016/j.neuron.2018.07.035