Psychonomic Bulletin & Review
https://doi.org/10.3758/s13423-023-02380-5

BRIEF REPORT

Confidence ratings increase response thresholds in decision making

Baike Li[1,2] · Xiao Hu[3] · David R. Shanks[4] · Ningxin Su[5] · Wenbo Zhao[6] · Liu Meng[1,7] · Wei Lei[8] · Liang Luo[2,9] · Chunliang Yang[2,3]

Accepted: 31 August 2023

Abstract

Many mental processes are reactive – they are altered as a result of introspection and monitoring. It has been documented that soliciting trial-by-trial confidence ratings (CRs) reactively improves decision accuracy and lengthens response times (RTs), but the cognitive mechanisms underlying CR reactivity in decision-making remain unknown. The current study conducted two experiments and employed the drift-diffusion model (DDM) to explore why reporting confidence reactively alters the decision-making process. The results showed that CRs led to enhanced decision accuracy, longer RTs, and higher response thresholds. The findings are consistent with an increased conservatism hypothesis which asserts that soliciting CRs provokes feelings of uncertainty and makes individuals more cautious in their decision making.

Keywords Confidence rating · Decision-making · Reactivity effect · Drift diffusion model · Response threshold

Introduction

The past few decades have witnessed a burgeoning of research on metacognition. Research in this field typically requires participants to submit trial-by-trial confidence ratings (CRs) to provide insights into their metacognitive monitoring processes (Busey et al., 2000; Fleming et al., 2010; Hu et al., 2022a, 2022b) and to measure individuals' metacognitive ability by quantifying the relationships between CRs and task performance (Fleming et al., 2012; Fleming et al., 2014; Fleming & Lau, 2014).[1] In previous studies, CR accuracy is typically quantified as signed differences or intra-individual relations between CRs and task performance (e.g., decision accuracy). However, an emerging body of research has found that soliciting CRs can cause *reactivity* in task performance, whereby task performance is altered by the act of being monitored and reported (Bonder & Gopher, 2019; Double & Birney, 2017; Double & Birney, 2018, 2019; Lei et al., 2020).

A concrete example of reactivity is that the requirement to report trial-by-trial response-confidence promotes decision accuracy and lengthens response times (RTs) (Bonder & Gopher, 2019; Lei et al., 2020). Such reactivity effects suggest that CRs may not provide an unbiased measure of metacognition because the requirement to report confidence reactively

Baike Li and Xiao Hu contributed equally to this work.

✉ Liang Luo
luoliang@bnu.edu.cn

✉ Chunliang Yang
chunliang.yang@bnu.edu.cn

1 School of Psychology, Liaoning Normal University, Dalian, China

2 Institute of Developmental Psychology, Faculty of Psychology, Beijing Normal University, Beijing, China

3 Beijing Key Laboratory of Applied Experimental Psychology, National Demonstration Center for Experimental Psychology Education, Faculty of Psychology, Beijing Normal University, Beijing, China

4 Division of Psychology and Language Sciences, University College London, London, UK

5 Collaborative Innovation Center of Assessment for Basic Education Quality, Beijing Normal University, Beijing, China

6 School of Social Development and Public Policy, Beijing Normal University, Beijing, China

7 School of Psychology, South China Normal University, Guangzhou, China

8 Department of Psychiatry, The Affiliated Hospital of Southwest Medical University, Luzhou, China

9 State Laboratory of Cognitive Neuroscience and Learning, Beijing Normal University, Beijing, China

[1] For the sake of conciseness, hereafter confidence rating is referred to as CR.

Published online: 06 October 2023

changes the very entity being monitored. Although several studies have observed reactivity effects of CRs on decision-making (Bonder & Gopher, 2019; Lei et al., 2020), little research has been conducted to investigate its cognitive underpinnings. The current study aims to fill this gap by proposing two theoretical explanations and empirically testing them.

Confidence Rating (CR) reactivity in decision making

A handful of recent studies demonstrated that the requirement to report confidence following each decision can reactively alter the decision-making process per se (e.g., Birney et al., 2017; Bonder & Gopher, 2019; Double & Birney, 2018, 2019; Lei et al., 2020). For instance, in a perceptual decision-making task, Lei et al. (2020) presented participants with a rectangle filled with orange and blue colors and instructed them to judge which color (orange/blue) filled more of the area. After making each perceptual decision, participants reported either confidence about their decision accuracy or not. The results showed that the requirement of reporting confidence significantly increased both decision accuracy and RTs. Similar findings were obtained by Bonder and Gopher (2019). In their study, participants judged the location of a small gap. Participants in a CR group were asked to report response-confidence in each trial while those in a control group did not. Bonder and Gopher (2019) found that participants responded more accurately and slowly in the CR than in the control group. Furthermore, the enhancing effect of CRs on decision accuracy transferred to a subsequent task in which there was no requirement to report response confidence.

It should be noted that previous findings about the reactivity effect of CRs on decision accuracy are somewhat inconsistent. For instance, Baranski and Petrusic (2001) found that, although reporting confidence increased RTs in a line-length comparison task, it had no significant influence on decision accuracy. Similarly, Petrusic and Baranski (2003) found that, although reporting confidence resulted in longer decision RTs in a forced-choice sensory discrimination task, there was no significant reactive effect of CRs on sensory discrimination accuracy.

In sum, prior findings about the effect of CRs on decision accuracy are somewhat inconsistent, with some showing CRs facilitating accuracy and others showing minimal influence. By contrast, the reactivity effect of CRs on decision RTs is quite robust: that is, reporting response-confidence lengthens decision RTs. Hence, below we focus on the effect of CRs on decision RTs.

Putative mechanisms underlying CR reactivity in decision Response Times (RTs)

Although many previous studies observed that decision RTs are reactive to CRs, explanations about why this happens are lacking. Here, we propose two possibilities: (1) *dual-task costs* and (2) *increased conservatism.* The dual-task costs account asserts that the requirement of providing CRs functions as a secondary task that borrows cognitive resources from the primary decision-making task, leading to a delay of decision responses and longer decision RTs (Mitchum et al., 2016). Specifically, participants have to search for diagnostic cues to provide an appropriate CR for each decision response, and the cue-search process may borrow cognitive resources from the primary decision task (Baranski & Petrusic, 1998, 2001; Petrusic & Baranski, 2003). In addition, translating subjective feelings of confidence into a specific numerical judgment (e.g., 0–100) is also resource consuming. Many studies have demonstrated that including a secondary task delays responses to the primary task (e.g., Craik et al., 1996).

The second potential explanation for CR reactivity in decision RTs is increased conservatism, which assumes that repeatedly asking participants to report their confidence provokes feelings of uncertainty about decision accuracy, in turn making them more cautious (i.e., conservative) in their decisions (Banca et al., 2015; Theisen et al., 2021). Specifically, asking participants to report decision confidence induces conscious reflection about decision accuracy (Konstantinidis & Shanks, 2014) and invites them to consider that not all of their decisions are correct (Double & Birney, 2017). Put differently, repeatedly prompting participants to provide response-confidence judgments enhances feelings of uncertainty and induces individuals to gather more information (or evidence) before making a decision, resulting in longer RTs.

Drift-diffusion model

Computational models of decision RTs, such as the drift-diffusion models (DDMs), can be employed to directly test the two theoretical explanations discussed above (Hu et al., 2022b; Ratcliff et al., 2016; Voss et al., 2004; Wiecki et al., 2013). DDMs are widely used to decompose RTs in binary decision-making tasks (Ratcliff et al., 2016; Wiecki et al., 2013). Specifically, DDMs assume that the process of making binary decisions is a process of information accumulation over time toward one or the other decision boundary (i.e., decision threshold). DDMs decompose decision-making into four components: boundary separation or threshold (a), drift rate (v), starting point (z), and non-decision time (t_0) (see Fig. 1).

Boundary separation (a) represents the amount of relative evidence required to reach a decision to select one choice over the other. It can be altered by a speed/accuracy trade-off manipulation (Voss et al., 2004). Drift rate (v) describes the mean rate for approaching a boundary within a trial and reflects the quality of information gathered from the stimulus (Ratcliff et al., 2004). Specifically, drift rate reflects the extent to which evidence accumulation is affected by noise

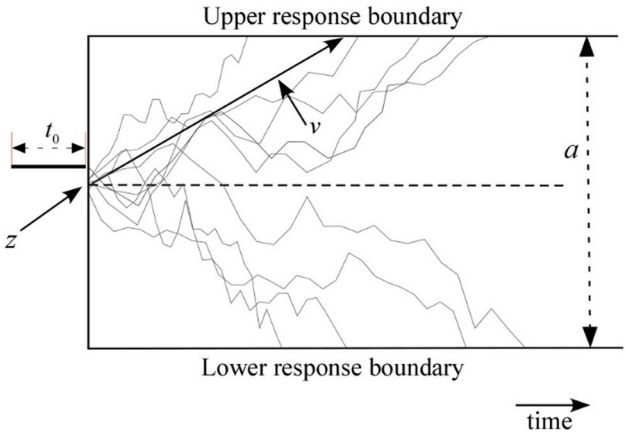

Fig. 1 Visual illustration of the drift-diffusion model (DDM) parameters. v = drift rate; a = boundary separation or threshold; t_0 = non-decision time; z = starting point

(i.e., irrelevant factors). The starting point (z) refers to a bias toward one of the two response options. Researchers often assume that there is no bias between response options which have an equal chance of being correct, and therefore set the starting point at the midpoint between the two boundaries (Stafford et al., 2020), although in other circumstances a bias may be present. Non-decision time (t_0) is the amount of time that is not dedicated to the decision-making process. Typically, it is assumed to be equal to the sum of the time of stimulus preprocessing before evidence accumulation and the motor response time for performing a response after the decision has been reached (for a review, see Ratcliff et al., 2016).

DDMs can be used to decompose decision RTs and hence employed to directly test the dual-task costs and increased conservatism theories of CR reactivity in decision RTs. According to the dual-task costs theory, CR reactivity in decision RTs derives from the fact that reporting confidence, as a secondary task, borrows cognitive resources from the primary decision task. This additional requirement may interfere with information accumulation and introduce more noise into the decision-making process, leading to a smaller drift rate (v). In addition, it may interfere with the non-decision components (such as interfering with stimulus preprocessing and motor response), resulting in longer non-decision time (t_0). In contrast to the dual-task costs theory, the increased conservatism theory proposes that CR reactivity in decision RTs results from the fact that reporting confidence increases feelings of uncertainty about decision accuracy and makes individuals more cautious (Banca et al., 2015; Theisen et al., 2021). Accordingly, it predicts that reporting confidence should increase boundary separation (a).

Overall, the dual-task costs theory assumes that reporting confidence reactively lengthens decision RTs through decreasing information accumulation rate (v) and/or increasing non-decision time (t_0). By contrast, the increased conservatism hypothesis predicts that reporting confidence delays decision responses through increasing response threshold (a). DDMs can be employed to directly test these two theories. Of course, it should be noted beforehand that the dual-task costs and increased conservatism theories are not necessarily mutually exclusive, and the mechanisms proposed by these two theories may jointly contribute to CR reactivity in decision RTs.

Overview of the current study

In the current study, we employed a DDM to test the two theories described above. To achieve this aim, we asked participants to complete perceptual decision-making tasks adapted from Lei et al. (2020) and Fleming et al. (2014). During the decision-making tasks, participants in the CR condition were instructed to report their confidence following each decision, whereas they were asked to press a number key randomly selected by the computer (i.e., not reporting decision confidence) in the control condition. A DDM was employed to decompose decision RTs in both conditions and then we compared the key parameters of drift rate (v), non-decision time (t_0), and boundary separation (a) between the two conditions.

Experiments 1 and 2

Experiment 1 was conducted to test the dual-task costs and increased conservatism theories. The dual-task costs theory predicts lower drift rate (v) and/or longer non-decision time (t_0) in the CR than in the control condition, whereas the increased conservatism theory predicts greater boundary separation (a) in the CR than in the control condition.

Experiment 2 was conducted to conceptually replicate the main findings of Experiment 1. Additionally, it employed a new decision-making task (i.e., a dot number comparison task) to test the generalizability of the findings observed in Experiment 1. Experiments 1 and 2 had the same experimental design and analysis pipeline. Hence, for the sake of conciseness, these two experiments are reported together.

Methods

Participants

For Experiment 1, a pilot study was conducted to roughly determine the effect size of the reactive impact of CRs on decision RTs. According to the pilot results (Cohen's $d = 0.55$), we estimated that approximately 28 participants were required to observe a significant ($\alpha = .05$) effect at .80 power. Accordingly, Experient 1 recruited 28 participants (M age = 20.21 years,

SD = 1.99 years; 26 females) from the Beijing Normal University (BNU) participant pool. Similar to Experiment 1, Experiment 2 recruited 28 new participants (*M* age = 21.68 years, *SD* = 3.03 years; 22 females) from the same participant pool.

In both experiments, participants provided informed consent, were tested individually in a sound-proofed cubicle, and received financial remuneration. The protocol was approved by the Institutional Review Board of the BNU Faculty of Psychology.

Materials

In Experiment 1, the stimuli were pictures depicting a rectangle of 768 × 624 pixels, which were divided into two areas by a random jagged line, with one area filled in orange and the other in blue (see Fig. 2A). The average difference in area filled in blue and orange was 9.31% of the rectangle (*SD* = 3.79%). There were 420 pictures in total, produced via a MATLAB script. Twenty of them were used for practice and the other 400 were used in the main experiment.

Similar to the materials used by Fleming et al. (2014), the stimuli employed in Experiment 2 were pictures that contained two circles (diameter of 11.5°) containing a number of dots (see Fig. 2B). One of the two circles always contained 50 dots and the other contained more than 50 dots (ranging from 51 to 75). The two circles were placed randomly on the left and right sides of the screen. 420 pictures were produced via a MATLAB script. Twenty were used for practice and the other 400 were used in the main experiment.

To prevent any item-selection effects, for each participant in each experiment, the computer randomly divided the 400 pictures into four lists, with two lists randomly assigned to the CR condition and the other two to the control condition. In addition, for each participant, the presentation sequence of the pictures in each list and the list sequence were randomly decided by the computer. All stimuli were presented via the MATLAB *Psychtoolbox* package (Kleiner et al., 2007).

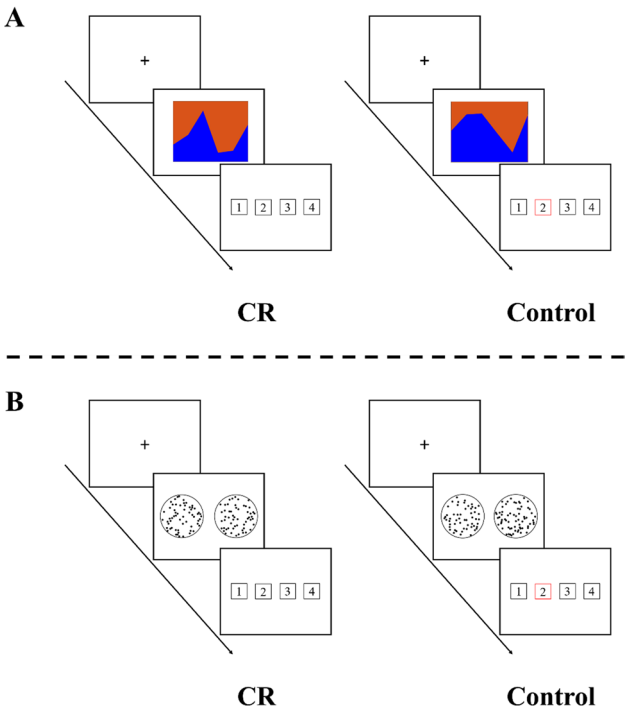

Fig. 2 Flow charts depicting the stimuli and task procedures in Experiments 1 (**A**) and 2 (**B**), respectively. CR = confidence rating

Procedure

Both experiments involved a within-subjects design (condition: CR vs. control). The procedure was adapted from Lei et al. (2020). In Experiment 1, participants were informed that they would perform a four-list area size comparison task in which they needed to make a binary decision on each trial about which color area (orange/blue) was larger. In Experiment 2, participants were told that they would perform a four-list dot number comparison task and would make a binary decision about which (right/left) circle contained more dots.

Participants were further informed that, for two lists of pictures (i.e., CR lists), they would report their decision confidence after making each decision. For the other two lists (i.e., control lists), they would press a number key on the keyboard (1–4) in response to a digit highlighted by a red frame after making each decision. Importantly, they were explicitly told that they should complete the decision task as accurately and quickly as possible regardless of whether they needed to make a confidence rating or not because their compensation was dependent on their decision accuracy in all four lists.

Before the main experiment, participants completed a practice task to familiarize themselves with the experimental procedure. The procedure was the same as that of the main experiment (see below for details). Then the experiment started and participants viewed four lists of pictures, with 100 pictures in each list. Before viewing each list, the computer informed participants whether they needed to make confidence ratings or not for the following list of pictures (see Figs. 2A and B).

For a control list, the 100 pictures were presented one by one in a random order. Before the presentation of each picture, a cross sign appeared at the center of the screen

for 0.5 s to mark the inter-stimulus interval. Then a picture appeared at the center of the screen, and participants were instructed to press a key to judge which color area (orange/blue) was larger (Experiment 1) or which circle contained more dots (Experiment 2). The assignment of keys to decisions was counterbalanced across participants. Once a key was pressed, the picture disappeared and four digits (1–4) were presented on screen, with one of them randomly highlighted by a red frame. Participants were instructed to press a number key (1–4) in response to the highlighted digit. If they did not respond to the highlighted digit within 6 s, a message box appeared to remind them to carefully make a response for the subsequent trials. Participants pressed the space key to remove the message box and trigger the next trial. This cycle repeated until the end of the list.

The procedure in the CR condition was identical to that in the control condition but with one difference. Specifically, after participants made each decision, the four digits were also shown on the screen, but none of them was highlighted by a red frame. Participants were instructed to report their confidence regarding the accuracy of their decision on a scale ranging from 1 (*not confident at all*) to 4 (*very confident*). They reported their confidence by pressing a 1–4 number key on the keyboard.

There was no feedback during the task. At the end of the task, bonuses were awarded according to overall performance on the binary decision-making task. Specifically, the bonus for each correct response was 0.1 RMB. The monetary compensation for a given participant was 20 RMB plus the total bonus he or she earned.

Data analyses

In both experiments, Bayesian and traditional frequentist paired *t*-tests were conducted via JASP Version 0.15.0.0 (https://jasp-stats.org) to compare decision accuracy and median decision RTs between the CR and control conditions. Bayes factors (BF_{10}) can be interpreted as the relative strength of the evidence in favor of the alternative hypothesis over the null (i.e., evidence supporting the existence over the absence of an effect).

The HDDM Toolbox was used to fit a hierarchical Bayesian version of DDM to decision choices and RTs (Wiecki et al., 2013). The HDDM employed a Markov chain Monte Carlo (MCMC) sampling method to estimate posterior parameter distributions. The parameters in the model included drift rate (v), boundary separation (a), starting point (z), non-decision time (t_0), and parameters representing the inter-trial variabilities of v, z, and t_0 (i.e., sv, sz, and st). Following previous studies (Stafford et al., 2020), the starting point was fixed to the mid-point, and sz was set to 0. In addition, the parameters for the inter-trial variabilities (sv and st) were constrained to be the same across participants and experimental conditions, because these parameters often have a very small effect on the likelihood and a large number of data points are needed to estimate them (Wiecki et al., 2013).

Regression models were used to estimate the effect of condition (CR vs. control, in which CR condition was dummy coded as 1 and control condition as 0) on each of the three main parameters in the HDDM (i.e., v, a, and t) at both individual and group levels, with correct responses coded as 1 and incorrect as 0. If condition has a statistically detectable effect on the main parameters, the 95% credible interval (CrI) in the posterior distributions of the group-level regression coefficients should not contain 0.

We fitted the model with four MCMC chains each containing 15,000 samples, with 2,000 samples per chain discarded as burn-in, resulting in 52,000 stored samples in total. Gelman and Rubin's potential scale reduction factor $\hat{R}$ was lower than 1.02 for all parameters in the model, indicating good convergence (Gelman & Rubin, 1992; Wiecki et al., 2013).

To evaluate whether the fitted HDDM model reproduced key patterns of RT data, posterior predictive validity checks were performed. The results are reported in the Online Supplemental Material (OSM). Overall, both experiments showed that the estimated parameters reproduced key patterns of the actual data. In addition, because the accuracy of CRs is not a major focus of interest, the corresponding results are also reported in the OSM. In brief, both experiments found that participants were metacognitively able to discriminate correct decisions from incorrect ones.

Results and discussion

Experiment 1 found that decision accuracy was greater in the CR (M = .82, SD = .08) than in the control condition (M = .79, SD = .12), difference = .03, 95% confidence interval = [.01, .05], $t(27)$ = 2.67, p = .01, Cohen's d = 0.50, BF_{10} = 3.76, implying that the requirement to report confidence reactively enhanced decision accuracy (see Fig. 3A). Additionally, median RTs were longer in the CR (M = 1.74, SD = 0.56) than in the control condition (M = 1.52, SD = 0.46), difference = 0.22 [0.13, 0.32], $t(27)$ = 4.83, $p < .001$, d = 0.91, BF_{10} = 509.80 (see Fig. 3B), reflecting strong evidence that decision speed is reactive to, and specifically, slowed down by the requirement to report decision confidence (Lei et al., 2020).

The reactivity findings of Experiment 1 were successfully replicated in Experiment 2. Specifically, decision accuracy was greater in the CR (M = .87, SD = .04) than in the control condition (M = .85, SD = .04), difference = .014 [.00, .03], $t(27)$ = 2.68, p = .01, d = 0.51, BF_{10} = 3.84 (see Figure 3C).

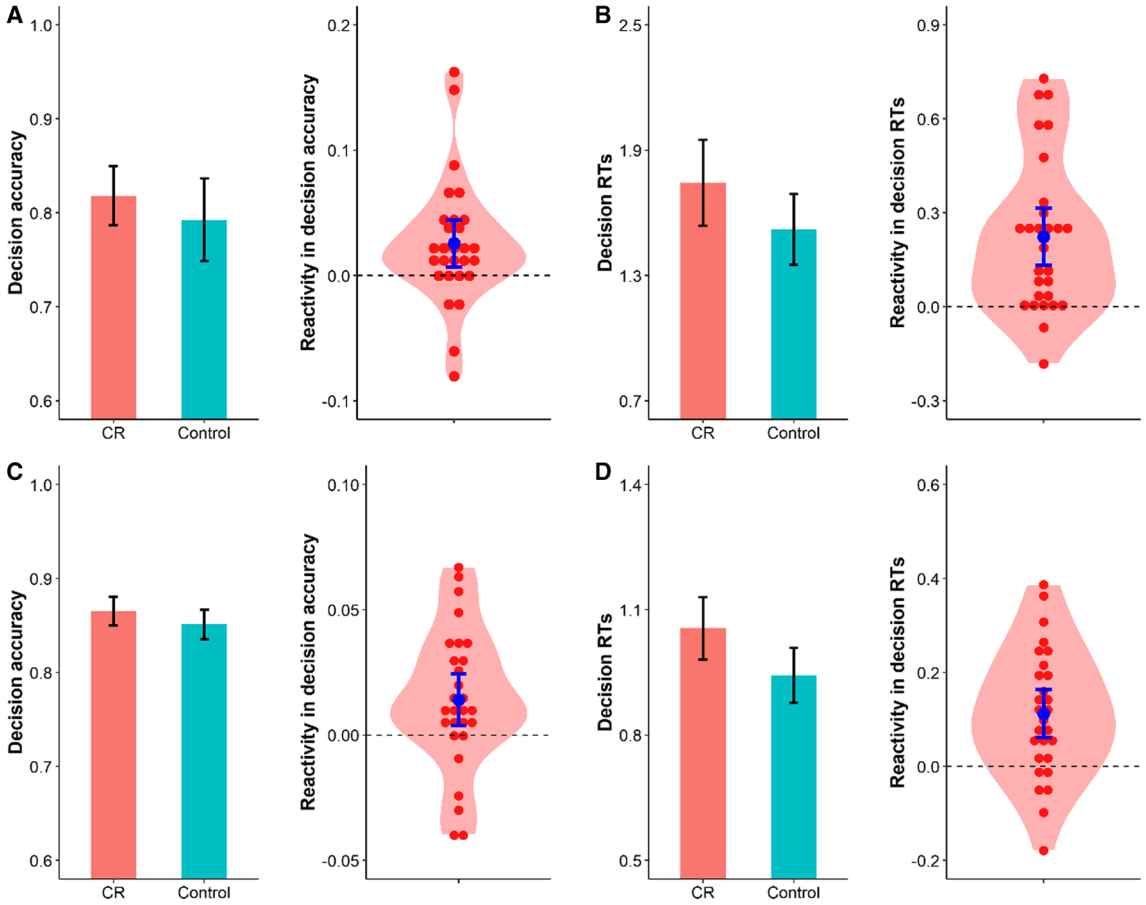

Fig. 3 Decision accuracy (**A**) and response times (RTs) (**B**) in Experiment 1 and decision accuracy (**C**) and RTs (**D**) in Experiment 2 as a function of condition. The violin plots represent the distributions of the reactivity effect of confidence ratings (CRs; i.e., the difference in decision accuracy and RTs between the CR and control conditions). Each red dot represents one participant's reactivity effect and the blue points represent group averages. Error bars represent the 95% confidence interval

Median RTs were longer in the CR ($M = 1.06$, $SD = 0.20$) than in the control condition ($M = 0.94$, $SD = 0.18$), difference = 0.11 [0.06, 0.17], $t(27) = 4.65$, $p < .001$, $d = 0.82$, $BF_{10} = 159.20$ (see Fig. 3D).

Figures 4A and B show posterior densities of the group-level regression coefficients provided by the HDDM in Experiments 1 and 2, respectively. In both experiments, the results showed that boundary separation (a) was greater in the CR than in the control condition, Experiment 1: $b = 0.26$, 95% CrI [0.18, 0.35]; Experiment 2: $b = 0.22$, 95% CrI [0.13, 0.31]. There was no detectable difference in drift rate (v) between the two conditions, Experiment 1: $b = 0.02$, 95% CrI [-0.05, 0.09]; Experiment 2: $b = 0.03$, 95% CrI [-0.08, 0.14]. Furthermore, there was no detectable difference in non-decision time (t_0) between the two conditions, Experiment 1: $b = 0.01$, 95% CrI [-0.04, 0.05]; Experiment 2: $b = 0.03$, 95% CrI [-0.01, 0.06] (see Table 1).

Overall, Experiments 1 and 2 employed different perceptual decision tasks and consistently demonstrated that participants moved their decision boundaries further apart and were hence more conservative in the CR than in the control condition: they required more information before making a choice. In addition, the requirement of reporting confidence had minimal reactive influence on drift rate or non-decision time. These results are in line with the increased conservatism theory, and inconsistent with the dual-task costs theory.

General Discussion

The current study conducted two experiments to test the dual-task costs and increased conservatism theories of CR reactivity in decision RTs. Participants were instructed to either make a confidence rating or not following each decision in an area size (Experiment 1) or dot number comparison task (Experiment 2). The principal findings were that the requirement to report trial-by-trial confidence significantly enhanced decision accuracy (as reflected by more correct decisions in the CR than in the control condition) and slowed down decision speed (as reflected by longer

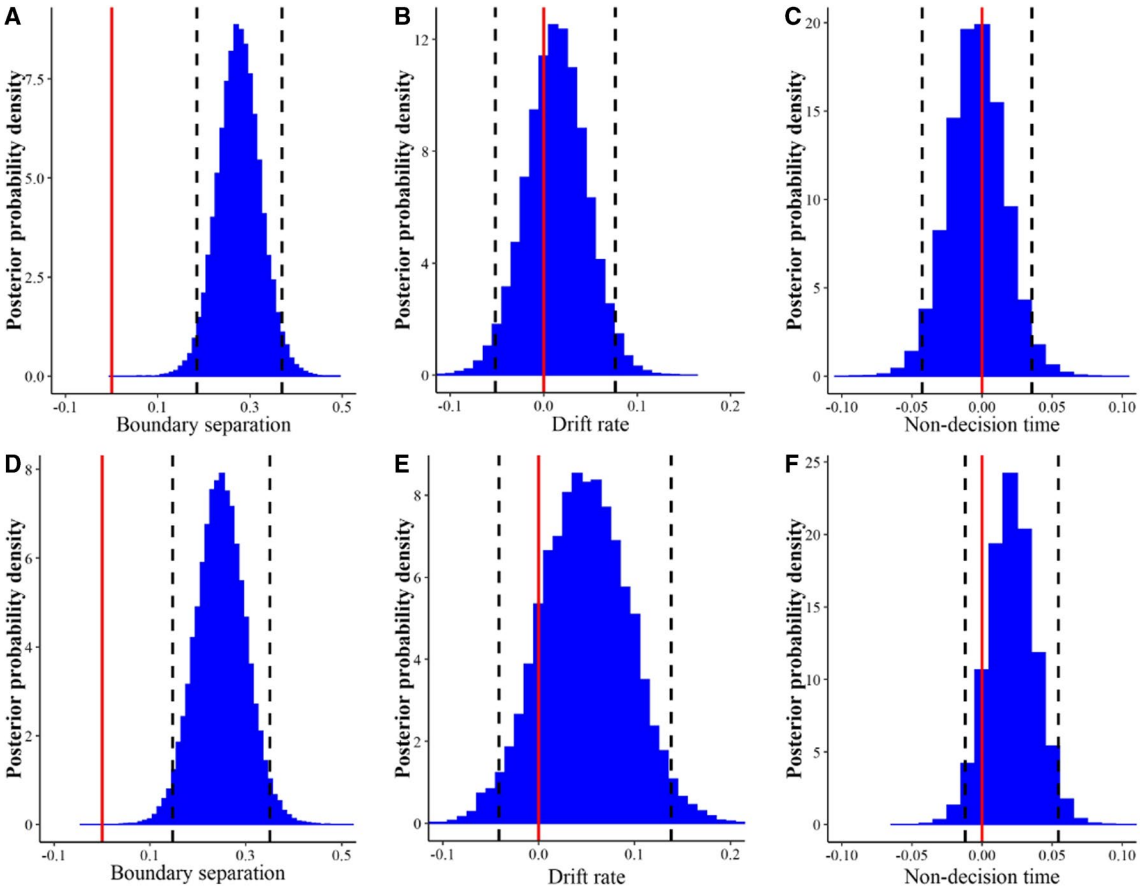

Fig. 4 Posterior densities for the group-level regression coefficients from the HDDM in Experiments 1 (**A–C**) and 2 (**D–F**), respectively. The two black dashed lines represent the 95% credible intervals, with the red solid line marking 0

decision RTs in the CR than in the control condition). More importantly, the HDDM results demonstrated that reporting confidence increased boundary separation but had minimal influence on drift rate or non-decision time.

Although recent studies demonstrated that reporting confidence reactively slows down decision speed (Baranski & Petrusic, 1998, 2001; Lei et al., 2020; Petrusic & Baranski, 2003), little research has been conducted to explore the potential cognitive mechanisms underlying this phenomenon. To fill this gap, the current research proposed two theories and empirically tested them.

According to DDMs, the decision-making process consists of accumulating evidence with noise, with drift rate reflecting the mean rate of evidence accumulation and non-decision time indicating the time taken by processes that are irrelevant for the decision-making process itself (Hu et al., 2022b; Stafford et al., 2020; Wiecki et al., 2013). In the current study, Experiments 1 and 2 consistently found that reporting confidence had no detectable effects on drift rate or non-decision time. These results are inconsistent with the dual-task costs theory, which assumes that the additional requirement of reporting confidence should interfere with evidence accumulation

Table 1 Posterior distribution parameters of the group-level regression coefficients

Parameters	Experiment 1			Experiment 2		
	M (*SD*)	2.5% CrI	97.5% CrI	*M* (*SD*)	2.5% CrI	97.5% CrI
a	0.26 (0.16)	0.18	0.35	0.22 (0.21)	0.13	0.31
v	0.02 (0.10)	-0.05	0.09	0.03 (0.04)	-0.08	0.14
t_0	0.01 (0.09)	-0.04	0.05	0.03 (0.08)	-0.01	0.06

In the regression analyses, confidence rating (CR) was coded as 1 and control as 0. a = boundary separation; v = drift rate; t_0 = non-decision time; CrI = credible interval

(leading to a lower drift rate) and/or disrupt stimulus preprocessing and motor response speed (leading to longer nondecision time) (Ariel et al., 2009; Griffin et al., 2008; Hertzog et al., 2002; Kanfer & Ackerman, 1989; Mitchum et al., 2016).

In contrast to the dual-task cost theory, the increased conservatism theory assumes that asking participants to report their confidence focuses their attention on and increases feelings of uncertainty about their response accuracy, which in turn makes them more cautious. Accordingly, this hypothesis predicts that reporting confidence should lengthen decision RTs through increasing boundary separation (Beste et al., 2018; Stafford et al., 2020; Voss et al., 2004; Wiecki et al., 2013). The observed findings are wholly consistent with this theoretical prediction by showing greater boundary separation in the CR than in the control condition.

In line with some previous studies (Bonder & Gopher, 2019; Lei et al., 2020), the current experiments found that making CRs reactively improved decision accuracy. However, other studies detected no significant enhancing effect of CRs on decision accuracy (Baranski & Petrusic, 2001; Petrusic & Baranski, 2003). Prior research identified some factors (e.g., stimulus quality, task difficulty, fatigue, and attention) that appear to moderate the accuracy of binary decisions by changing boundary separation (Ratcliff et al., 2004; Ratcliff et al., 2016). Additionally, it has been proposed that excessive numbers of trials in a given task may induce fatigue, which in turn affects drift rate during evidence accumulation and reduces decision accuracy (Ratcliff et al., 2004; Ratcliff et al., 2016). Compared to previous studies that did not observe any reactivity effect of CRs on decision accuracy (e.g., Petrusic & Baranski, 2003), the current study employed fewer trials, which might have allowed participants to more effectively focus their attention on gathering information for decision making. Future studies can profitably investigate the boundary conditions of the reactivity effect of CRs on decision accuracy from the perspective of variations in experimental procedure.

The increased conservatism and improved binary-decision accuracy observed in the present experiments suggest that making retrospective metacognitive judgments may enhance engagement in perceptual decision tasks. Previous research in the area of memory has also found that making prospective metacognitive judgments, such as judgments of learning (JOLs), improves memory of word lists and pictures by increasing learning engagement (Li et al., 2021; Li et al., 2023; Shi et al., 2022; Zhao et al., 2022), and that the formation processes of JOLs and CRs share similar underlying mechanisms (Luna & Albuquerque, 2022). These findings suggest that JOLs and CRs may alter memory performance through the same (or similar) mechanism(s). However, it should be noted that, in perceptual tasks, prospective and retrospective metacognitive judgments rely on different cues and the accuracy of prospective and retrospective metacognitive monitoring differs substantially, with higher accuracy for retrospective than for prospective judgments (Fleming et al., 2016; Siedlecka et al., 2016; Siedlecka et al., 2019). Whether prospective and retrospective metacognitive judgments reactively change perceptual task performance through the same or different mechanism(s) needs to be further explored. Additionally, prior research found that the reactivity effects of CRs on decision accuracy and RTs transfer to a subsequent task even when there is no need to report confidence in the subsequent task (Bonder & Gopher, 2019). Future research could usefully explore whether the effect of CRs on boundary separation is transferable. More importantly, whether making CRs can improve decision accuracy in other domains (e.g., economic decisions) needs to be determined, which will provide further indications about the practical uses of the CR reactivity effect.

The observed findings bear some implications for eyewitness identification, one of the major sources of evidence in criminal investigations (Lindsay et al., 2013). Witnesses who have observed a crime describe the suspect or try to make an identification through a live line-up or photograph. However, eyewitness identification accuracy can be biased by a variety of factors such as stress (Sauerland et al., 2016), memory distortion (Douglass & Steblay, 2006), and prejudice (Buckhout et al., 1975). How to improve eyewitness identification accuracy is a matter of considerable importance. Judges and the police often ask witnesses to report their level of confidence after they have given an eyewitness statement to evaluate whether the identification is credible (Brigham et al., 2007; Wixted & Wells, 2017). The current study suggests that reporting CRs is not only a way to evaluate the reliability of testimony but may also improve its accuracy. However, we highlight that all findings documented here are derived from perceptual decision tasks. Future research should directly test whether the requirement of reporting CRs can improve accuracy of eyewitness testimony.

Conclusion

Reporting confidence reactively increases decision accuracy and slows down decision speed. The increased conservatism hypothesis is a viable framework to account for the reactivity effect of CRs on decision RTs.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.3758/s13423-023-02380-5.

Acknowledgements This research was supported by the Natural Science Foundation of China (3237070512; 32000742; 32171045; 32200841), the Research Program Funds of the Collaborative Innovation Center of Assessment toward Basic Education Quality at Beijing Normal University (2021-01-132-BZK01), the UK Economic and Social Research Council (ES/S014616/1), and the Fundamental Research Funds for the Central Universities (2022NTSS36).

Data availability The data contained in this project are publicly available via the Open Science Framework at https://osf.io/x8drf/.

References

Ariel, R., Dunlosky, J., & Bailey, H. (2009). Agenda-based regulation of study-time allocation: When agendas override item-based monitoring. *Journal of Experimental Psychology: General, 138*, 432–447. https://doi.org/10.1037/a0015928

Banca, P., Vestergaard, M. D., Rankov, V., Baek, K., Mitchell, S., Lapa, T., & Voon, V. (2015). Evidence accumulation in obsessive-compulsive disorder: The role of uncertainty and monetary reward on perceptual decision-making thresholds. *Neuropsychopharmacology, 40*, 1192–1202. https://doi.org/10.1038/npp.2014.303

Baranski, J. V., & Petrusic, W. M. (1998). Probing the locus of confidence judgments: Experiments on the time to determine confidence. *Journal of Experimental Psychology: Human Perception and Performance, 24*, 929–945. https://doi.org/10.1037/0096-1523.24.3.929

Baranski, J. V., & Petrusic, W. M. (2001). Testing architectures of the decision–confidence relation. *Canadian Journal of Experimental Psychology, 55*, 195–206. https://doi.org/10.1037/h0087366

Beste, C., Adelhöfer, N., Gohil, K., Passow, S., Roessner, V., & Li, S.-C. (2018). Dopamine modulates the efficiency of sensory evidence accumulation during perceptual decision making. *International Journal of Neuropsychopharmacology, 21*, 649–655. https://doi.org/10.1093/ijnp/pyy019

Birney, D. P., Beckmann, J. F., Beckmann, N., & Double, K. S. (2017). Beyond the intellect: Complexity and learning trajectories in Raven's progressive matrices depend on self-regulatory processes and conative dispositions. *Intelligence, 61*, 63–77. https://doi.org/10.1016/j.intell.2017.01.005

Bonder, T., & Gopher, D. (2019). The effect of confidence rating on a primary visual task. *Frontiers in Psychology, 10*, 2674. https://doi.org/10.3389/fpsyg.2019.02674

Brigham, J. C., Bennett, L. B., Meissner, C. A., & Mitchell, T. L. (2007). The influence of race on eyewitness memory. In *The handbook of eyewitness psychology: Volume II* (pp. 257–281). Psychology Press.

Buckhout, R., Figueroa, D., & Hoff, E. (1975). Eyewitness identification: Effects of suggestion and bias in identification from photographs. *Bulletin of the Psychonomic Society, 6*, 71–74. https://doi.org/10.3758/BF03333151

Busey, T. A., Tunnicliff, J., Loftus, G. R., & Loftus, E. F. (2000). Accounts of the confidence-accuracy relation in recognition memory. *Psychonomic Bulletin & Review, 7*, 26–48. https://doi.org/10.3758/BF03210724

Craik, F. I. M., Govoni, R., Naveh-Benjamin, M., & Anderson, N. D. (1996). The effects of divided attention on encoding and retrieval processes in human memory. *Journal of Experimental Psychology: General, 125*, 159–180. https://doi.org/10.1037/0096-3445.125.2.159

Double, K. S., & Birney, D. P. (2017). Are you sure about that? Eliciting confidence ratings may influence performance on Raven's progressive matrices. *Thinking Reasoning, 23*, 190–206. https://doi.org/10.1080/13546783.2017.1289121

Double, K. S., & Birney, D. P. (2018). Reactivity to confidence ratings in older individuals performing the Latin square task. *Metacognition and Learning, 13*, 309–326. https://doi.org/10.1007/s11409-018-9186-5

Double, K. S., & Birney, D. P. (2019). Do confidence ratings prime confidence? *Psychonomic Bulletin & Review, 26*, 1035–1042. https://doi.org/10.3758/s13423-018-1553-3

Douglass, A. B., & Steblay, N. (2006). Memory distortion in eyewitnesses: A meta-analysis of the post-identification feedback effect. *Applied Cognitive Psychology, 20*, 859–869. https://doi.org/10.1002/acp.1237

Fleming, S. M., Huijgen, J., & Dolan, R. J. (2012). Prefrontal contributions to metacognition in perceptual decision making. *The Journal of Neuroscience, 32*, 6117–6125. 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. https://doi.org/10.3389/fnhum.2014.00443

Fleming, S. M., Massoni, S., Gajdos, T., & Vergnaud, 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*, niw018. https://doi.org/10.1093/nc/niw018

Fleming, S. M., Ryu, J., Golfinos, J. G., & Blackmon, K. E. (2014). Domain-specific impairment in metacognitive accuracy following anterior prefrontal lesions. *Brain, 137*, 2811–2822. https://doi.org/10.1093/brain/awu221

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*, 1541–1543. https://doi.org/10.1126/science.1191883

Gelman, A., & Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science, 7*, 457–472. https://doi.org/10.1214/ss/1177011136

Griffin, T. D., Wiley, J., & Thiede, K. W. (2008). Individual differences, rereading, and self-explanation: Concurrent processing and cue validity as constraints on metacomprehension accuracy. *Memory & Cognition, 36*, 93–103. https://doi.org/10.3758/MC.36.1.93

Hertzog, C., Kidder, D. P., Powell-Moman, A., & Dunlosky, J. (2002). Aging and monitoring associative learning: Is monitoring accuracy spared or impaired? *Psychology and Aging, 17*, 209–225. https://doi.org/10.1037/0882-7974.17.2.209

Hu, X., Yang, C., & Luo, L. (2022a). Are the contributions of processing experience and prior beliefs to confidence ratings domain-general or domain-specific? *Journal of Experimental Psychology: General, 152*, 28–44. https://doi.org/10.1037/xge0001257

Hu, X., Yang, C., & Luo, L. (2022b). Retrospective confidence rating about memory performance is affected by both retrieval fluency and non-decision time. *Metacognition and Learning, 17*, 651–681. https://doi.org/10.1007/s11409-022-09303-0

Kanfer, R., & Ackerman, P. L. (1989). Motivation and cognitive abilities: An integrative/aptitude-treatment interaction approach to skill acquisition. *Journal of Applied Psychology, 74*, 657–690. https://doi.org/10.1037/0021-9010.74.4.657

Kleiner, M., Brainard, D., & Pelli, D. (2007). What's new in Psychtoolbox-3? *Perception 36 ECVP abstract supplement.*

Konstantinidis, E., & Shanks, D. R. (2014). Don't bet on it! Wagering as a measure of awareness in decision making under uncertainty. *Journal Experiment Psychology: General, 143*, 2111–2134. https://doi.org/10.1037/a0037977

Lei, W., Chen, J., Yang, C., Guo, Y., Feng, P., Feng, T., & Li, H. (2020). Metacognition-related regions modulate the reactivity effect of confidence ratings on perceptual decision-making. *Neuropsychologia, 144*, 107502. https://doi.org/10.1016/j.neuropsychologia.2020.107502

Li, B., Zhao, W., Shi, A., Zhong, Y., Hu, X., Liu, M., et al. (2023). Does the reactivity effect of judgments of learning transfer to learning of new information? *Memory, 31*, 918–930. https://doi.org/10.1080/09658211.2023.2208792

Li, B., Zhao, W., Zheng, J., Hu, X., Su, N., Fan, T., et al. (2021). Soliciting judgments of forgetting reactively enhances memory as well as making judgments of learning: Empirical and meta-analytic tests. *Memory & Cognition, 50,* 1061–1077. https://doi.org/10.3758/s13421-021-01258-y

Lindsay, R. C., Ross, D. F., Read, J. D., & Toglia, M. P. (2013). *The handbook of eyewitness psychology: Volume ii: Memory for people* (Vol. 2): Psychology Press.

Luna, K., & Albuquerque, P. B. (2022). Do beliefs about font size affect retrospective metamemory judgments in addition to prospective judgments? *Experimental Psychology, 69*, 172–184. https://doi.org/10.1027/1618-3169/a000549

Mitchum, A. L., Kelley, C. M., & Fox, M. C. (2016). When asking the question changes the ultimate answer: Metamemory judgments change memory. *Journal of Experimental Psychology: General, 145*, 200–219. https://doi.org/10.1037/a0039923

Petrusic, W. M., & Baranski, J. V. (2003). Judging confidence influences decision processing in comparative judgments. *Psychonomic Bulletin & Review, 10*, 177–183. https://doi.org/10.3758/BF03196482

Ratcliff, R., Gomez, P., & McKoon, G. (2004). A diffusion model account of the lexical decision task. *Psychological Review, 111*, 159–182. https://doi.org/10.1037/0033-295X.111.1.159

Ratcliff, R., Smith, P. L., Brown, S. D., & McKoon, G. (2016). Diffusion decision model: Current issues and history. *Trends in Cognitive Sciences, 20*, 260–281. https://doi.org/10.1016/j.tics.2016.01.007

Sauerland, M., Raymaekers, L. H. C., Otgaar, H., Memon, A., Waltjen, T. T., Nivo, M., et al. (2016). Stress, stress-induced cortisol responses, and eyewitness identification performance. *Behavioral Sciences & the Law, 34*, 580–594. https://doi.org/10.1002/bsl.2249

Shi, A., Xu, C., Zhao, W., Shanks, D. R., Hu, X., Luo, L., & Yang, C. (2022). Judgments of learning reactively facilitate visual memory by enhancing learning engagement. *Psychonomic Bulletin & Review, 30, 676–687*. doi: https://doi.org/10.3758/s13423-022-02174-1

Siedlecka, M., Paulewicz, B., & Wierzchoń, M. (2016). But I was so sure! Metacognitive judgments are less accurate given prospectively than retrospectively. *Frontiers in Psychology, 7*, 1–8. https://doi.org/10.3389/fpsyg.2016.00218

Siedlecka, M., Skóra, Z., Paulewicz, B., Fijałkowska, S., Timmermans, B., & Wierzchoń, M. (2019). Responses improve the accuracy of confidence judgements in memory tasks. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 45*, 712–723. https://doi.org/10.1037/xlm0000608

Stafford, T., Pirrone, A., Croucher, M., & Krystalli, A. (2020). Quantifying the benefits of using decision models with response time and accuracy data. *Behavior Research Methods, 52*, 2142–2155. https://doi.org/10.3758/s13428-020-01372-w

Theisen, M., Lerche, V., von Krause, M., & Voss, A. (2021). Age differences in diffusion model parameters: A meta-analysis. *Psychological Research, 85*, 2012–2021. https://doi.org/10.1007/s00426-020-01371-8

Voss, A., Rothermund, K., & Voss, J. (2004). Interpreting the parameters of the diffusion model: An empirical validation. *Memory & Cognition, 32*, 1206–1220. https://doi.org/10.3758/BF03196893

Wiecki, T., Sofer, I., & Frank, M. (2013). HDDM: Hierarchical Bayesian estimation of the drift-diffusion model in python. *Frontiers in Neuroinformatics, 7*, 1–10. https://doi.org/10.3389/fninf.2013.00014

Wixted, J. T., & Wells, G. L. (2017). The relationship between eyewitness confidence and identification accuracy: A new synthesis. *Psychological Science in the Public Interest, 18*, 10–65. https://doi.org/10.1177/1529100616686966

Zhao, W., Li, B., Shanks, D. R., Zhao, W., Zheng, J., Hu, X., ... Yang, C. (2022). When judging what you know changes what you really know: Soliciting metamemory judgments reactively enhances children's learning. *Child Development, 93*, 405–417. https://doi.org/10.1111/cdev.13689