Metacognition and Learning (2022) 17:651–681
https://doi.org/10.1007/s11409-022-09303-0

Retrospective confidence rating about memory performance is affected by both retrieval fluency and non-decision time

Xiao Hu[2] · Chunliang Yang[1,2] · Liang Luo[1,3]

Received: 14 December 2021 / Accepted: 25 April 2022 / Published online: 21 May 2022

Abstract
Many previous studies observed that higher retrospective confidence ratings about memory performance were associated with shorter response times in memory test. Researchers often interpret response time as a measure of retrieval fluency which is an important cue utilized in confidence formation process. However, the drift diffusion model (DDM) indicates that response time in recognition memory test includes both a decision component representing memory retrieval, and a non-decision component unrelated to retrieval process. Few previous studies have investigated whether retrospective confidence in recognition test is related to the speed of both retrieval and non-decision processes. To address this question, the current study first analyzed data from six published experiments, and found that higher retrospective confidence ratings were associated with both higher drift rate (indicating retrieval fluency) and shorter non-decision time in DDM. Then we manipulated the ease of perception in two new experiments, and the results consistently indicated that difficulty in stimulus perception increased non-decision time in recognition test, which affected retrospective confidence. Furthermore, the documented results also suggest drift rate could partly account for the positive relationship between confidence and memory performance, while the reliance of confidence on non-decision time negatively affected confidence accuracy.

Keywords Metamemory · Retrospective confidence rating · Drift diffusion model · Drift rate · Non-decision time

✉ Liang Luo
luoliang@bnu.edu.cn

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

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

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

Metamemory refers to the processes of monitoring and controlling memory activities (Nelson & Narens, 1990). People rely on the results of metamemory monitoring to guide their learning activities, such as how to allocate subsequent study time and select items for restudy (Metcalfe & Finn, 2008; Metcalfe & Kornell, 2005; Robey et al., 2017), and thus metamemory monitoring plays a critical role in memory process. Metamemory judgments can be made either prospectively or retrospectively: people can prospectively predict their performance in a future memory test, or retrospectively rate confidence about memory performance in a test they have just finished (Nelson & Narens, 1990). Retrospective confidence rating is an important type of metamemory judgment, and can reliably affect people's learning strategy in future study (Robey et al., 2017, 2021).

According to the cue-utilization theory for metamemory, metamemory judgments are inferential in nature and depend on various types of information (or cues) obtained from memory process (Koriat, 1997; Koriat et al., 2008). Researchers believe that one important cue that retrospective confidence ratings rely on is the fluency of memory retrieval, and previous studies often interpreted response time in memory test as an index of retrieval fluency (e.g., Koriat et al., 2008; Koriat & Ackerman, 2010). But in fact, response time in memory test is complicated and may not solely reflect retrieval fluency. For example, the drift diffusion model (DDM), which is a widely accepted computational model about decision making, suggests that response process in recognition memory test includes both a decision component based on memory retrieval, and a non-decision component which is unrelated to retrieval process. The response time in recognition test is the sum of the time for decision and non-decision components (Ratcliff & McKoon, 2008). However, few studies have investigated whether retrospective confidence ratings in recognition test are affected by the speed of both retrieval and non-decision processes.

Below, we briefly summarize empirical findings about the relationship between retrospective confidence ratings and response times in memory test, and then discuss the decomposition of response time based on DDM. Finally, we introduce the rationale of the current study.

Relationship between retrospective confidence ratings and response times

The cue-utilization theory suggests that one important cue that retrospective confidence ratings depend on is retrieval fluency, which refers to the subjective ease or difficulty with which an answer is retrieved from memory (Koriat, 1997; Koriat et al., 2008). It is assumed that retrospective confidence should rely on internal feedback that participants gain from attempting to select or retrieve an answer in memory test, and retrieval fluency is such an important feedback from memory process (Koriat & Ackerman, 2010).

Researchers often use response time in memory test as an index of retrieval fluency, and believe that people need longer time to respond when it is more difficult to retrieve an answer from memory (Koriat et al., 2008). Previous studies consistently documented a negative correlation between retrospective confidence ratings and response times in various types of memory tests: participants gave higher confidence ratings about memory performance when their response time was shorter in the test (Dougherty et al., 2005; Kelley & Lindsay, 1993; Koriat & Ackerman, 2010; Nelson & Narens, 1990; Robinson et al., 1997; Siedlecka et al., 2019; Weidemann & Kahana, 2016). For example, Siedlecka et al., (2019) asked participants to first learn a list of words and then performed an Old/New or 2-alternative-forced-choice (2AFC) recognition test. Their results revealed that shorter response time in recognition test was associated with higher retrospective confidence ratings. Nelson & Narens (1990) also demonstrated a negative correlation between retrospective confidence and response time in a recall test for both correctly answered trials and trials with commission errors (i.e., providing incorrect answers). Furthermore, Kelley & Lindsay (1993) indicated that prior exposure to either correct or incorrect answers before recall test changed both response times and confidence ratings, but did not alter the negative relationship between these two variables, suggesting a stable effect of response time on retrospective confidence.

On the other hand, researchers found that response time in memory test could also negatively predict actual memory performance (Anderson, 1981; Koriat & Ackerman, 2010; Nelson & Narens, 1990; Robinson et al., 1997; Siedlecka et al., 2019). Thus, the reliance of retrospective confidence ratings on response time should strengthen the relationship between confidence and actual memory performance (i.e., increase confidence accuracy or resolution; Vuorre & Metcalfe 2021). To test this possibility, Koriat & Ackerman (2010) examined the correlation between retrospective confidence and memory performance both before and after partialling out the effect of response time. Results indicated that the correlation was reduced significantly when the effect of response time was partialled out, suggesting that the accuracy of retrospective confidence ratings was partly accounted for by the use of response time as a cue for making confidence ratings. In addition, Robinson et al., (1997) found higher accuracy of retrospective confidence in recall than in recognition test, and the difference in the correlation between confidence and response time across recall and recognition tests could mediate the effect of test format on confidence accuracy.

Decomposition of response time

Although response time in memory test is often interpreted as a measure of retrieval fluency in previous studies, using the total response time to represent the fluency of memory retrieval may be a simplified assumption. For example, in order to choose an answer in recognition memory test, participants need to first encode sensory input to perceive the content of the stimulus, then try to retrieve the current stimulus from memory, and finally press a key to indicate their answer. The response time reflects the total amount of time required for pre-retrieval (perception), retrieval and post-retrieval (response execution) processes, rather than only the retrieval process itself (Ratcliff & McKoon, 2008). Thus, response time in recognition test is not a pure measure of retrieval fluency.

In fact, theoretical models about response time, such as the drift diffusion model (DDM), suggest that response time in memory test is complicated and can be broken down into different parts. DDM is a computational model that can explain how people make decisions in two-choice tasks, and has been widely applied in various types of tasks including recognition memory test (for a review, see Ratcliff et al., 2016). It assumes response time in recognition test contains (1) a decision component during which memory evidence accumulates in retrieval process, and (2) a non-decision component which is unrelated to memory retrieval (Ratcliff & McKoon, 2008).

According to DDM, decisions in a recognition test are made by a noisy process accumulating memory evidence over time from a starting point toward one of two boundaries (Ratcliff & McKoon, 2008). A choice is made when accumulated evidence exceeds one threshold. Researchers often assume the upper boundary represents correct response and the lower one represents incorrect response (e.g., Nunez et al., 2017; Ratcliff & McKoon, 2008; Yoon et al., 2021). The evidence accumulation process does not always terminate at the same time and the same boundary due to within-trial noise (see Fig. 1). The mean speed of evidence accumulation, termed as drift rate, is higher than zero when memory evidence generally accumulates toward correct response. Furthermore, DDM suggests that besides evidence accumulation during memory retrieval (i.e., the decision component), there is also a non-decision component in the choice-making process during recognition test, which includes stimulus perception process, response execution process (e.g., pressing a key in recognition test), and so on. Response time is the sum of the time for decision and non-decision components. Thus, instead of using the total response time as a measure of retrieval fluency, it should be better to dissociate the retrieval and non-decision process based on DDM. In the current study, we aimed to examine whether retrospective confidence ratings are related to the speed of both retrieval and non-decision process during recognition test.

There are four main parameters in DDM: drift rate $\boldsymbol{v}$, boundary separation $\boldsymbol{a}$, starting point z, and non-decision time $\boldsymbol{T_{er}}$. The parameters $\boldsymbol{v}$, $\boldsymbol{a}$ and z affect the time needed for memory evidence accumulation during retrieval, and $\boldsymbol{T_{er}}$ indicates the time for cognitive process other than retrieval. In DDM, three of the four parameters ($\boldsymbol{v}$, z and $\boldsymbol{T_{er}}$) are allowed to vary across trials (Ratcliff & McKoon, 2008). However, when modeling correct vs. incorrect responses, researchers often fix the value of starting point to $\boldsymbol{a}/2$ (i.e., the midpoint), assuming there is no bias toward either correct or incorrect choice (Nunez et al., 2017; Voss

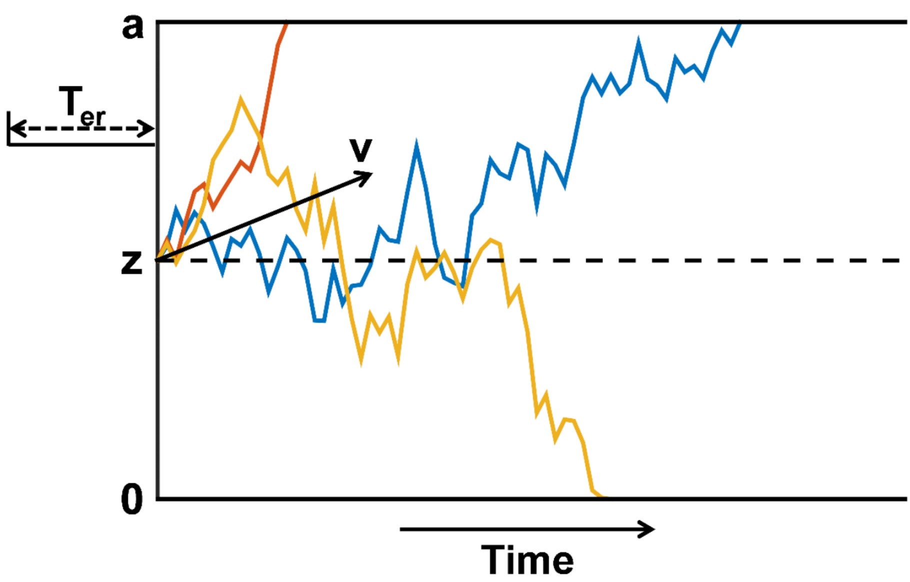

Fig. 1 Illustration of drift diffusion model. The parameter $\boldsymbol{v}$ represents drift rate, $\boldsymbol{a}$ represents boundary separation, z represents starting point, and $\boldsymbol{T_{er}}$ represents non-decision time

et al., 2015). Thus, the two main parameters that may affect the trial-by-trial variation of the speed of retrieval and non-decision process are $\boldsymbol{v}$ and $\boldsymbol{T_{er}}$, respectively.

In DDM, drift rate $\boldsymbol{v}$ directly reflects retrieval fluency: when the retrieval of information from memory storage is faster, the speed at which memory evidence accumulates and reaches decision boundary should also be higher (Ratcliff, 1978). Compared with total response time in recognition test, drift rate should be a better index of retrieval fluency. However, few studies have examined whether drift rate across trials could predict retrospective confidence ratings in recognition test. If retrospective confidence relies on retrieval fluency, as previous studies suggested (Koriat et al., 2008; Koriat & Ackerman, 2010), there should be a positive relationship between confidence ratings and drift rate across trials.

On the other hand, to our knowledge, no study has investigated the relationship between retrospective confidence ratings in recognition test and the time for non-decision process $\boldsymbol{T_{er}}$. One possibility is that people can accurately detect the fluency of retrieval process from the total response time, and base their confidence ratings solely on retrieval fluency. If this is true, then confidence should not be affected by non-decision time. However, it is also possible that the speed of cognitive process other than retrieval may affect confidence. For example, the speed of stimulus perception, which is part of the non-decision process during recognition test (Ratcliff & McKoon, 2008), may influence people's evaluation about memory performance (Wang et al., 2020; Yang et al., 2018). The ease of perceiving a stimulus may lead to a subjective feeling of knowing the answer, which then affects retrospective confidence ratings (Yang et al., 2021).

In addition, higher drift rate results in both higher response accuracy and shorter response time in recognition test, because memory evidence tends to accumulate more accurately and quickly toward the boundary representing correct response. On the other hand, non-decision time only affects the total amount of response time without influencing memory performance, because performance in recognition test is only determined by the threshold (correct or incorrect) that the accumulated memory evidence finally exceeds during retrieval process, and is unrelated to non-decision process (Ratcliff & McKoon, 2008; Voss et al., 2004). Thus, basing retrospective confidence on drift rate (i.e., retrieval fluency) should lead to reliable relationship between confidence and memory performance, and increase the accuracy of confidence ratings. In contrast, it is possible that utilizing non-decision time as a cue during confidence rating process might cause metamemory illusion, leading to the fact that retrospective confidence does not appropriately reflect actual performance in recognition test.

The current study

The main aim of the current study was to first dissociate retrieval and non-decision process in recognition test based on DDM, and then investigate whether retrospective confidence ratings in recognition test are related to drift rate (retrieval fluency), non-decision time, or both. Based on previous studies (Koriat et al., 2008; Koriat & Ackerman, 2010; Ratcliff, 1978), we expected to observe a positive relationship between retrospective confidence and drift rate $\boldsymbol{v}$. We also explored the relationship between confidence and non-decision time $\boldsymbol{T_{er}}$.

In the current study, we first fit DDM to data from recognition tests in six published experiments, and analyzed the relationship between retrospective confidence ratings and parameters T_{er} and v across trials. To foreshadow, our results revealed that both drift rate and non-decision time could significantly predict trial-by-trial confidence ratings. We then conducted two experiments (1a and 1b) to further investigate the relationship between non-decision time and confidence. In Experiments 1a and 1b, the ease of perceiving the stimuli was manipulated to directly affect non-decision time in recognition test. We planned to replicate the effect of both drift rate and non-decision time on retrospective confidence ratings, and further examine whether the change of non-decision time mediated the effect of the ease of perception on confidence.

Analyzing data from previous studies

Method

Experiments

In order to examine the relationship between retrospective confidence ratings and the parameters T_{er} and v in DDM, we analyzed data from six experiments in five published studies (Hu et al., 2017; Kantner & Lindsay, 2014; Sadeghi et al., 2017; Schmidt et al., 2019; Siedlecka et al., 2019). The data were extracted from the Confidence Database, which is a large database containing data of task performance and confidence ratings from published and unpublished studies (for details, see Rahnev et al., 2020). In each experiment, participants were asked to learn one or several word lists, and then perform either an Old/New or 2AFC recogntion test.[1] After making a choice in each trial during the test, participants rated confidence regarding the correctness of their answer. In addition, there are two experimental conditions (before and after meditative training) in Schmidt et al., (2019). Here we only analyzed data from the baseline condition (pre-training condition), which is consistent with the other five experiments.

Data analysis

Data preparation. We excluded all of the trials with missing data (e.g., trials for which participants failed to make a recognition choice or confidence rating during the required time window). Trials with response time that was shorter than 100 ms or differed by more than 3 standard deviations from the mean were also excluded (Rahnev et al., 2020; Whelan, 2008). Furthermore, confidence ratings were given on a 6-point scale in three of the six experiments (Hu et al., 2017; Sadeghi et al., 2017; Schmidt et al., 2019), and on a 3-point or 4-point scale in the other experiments (Kantner & Lindsay, 2014; Siedlecka et al., 2019). To directly compare the results from different experiments, all confidence ratings were transformed into values on a 6-point scale using the following equation:

[1] In the Confidence Database, there are also experiments with study materials other than single words (e.g., word pairs or pictures). However, fitting the Bayesian model to data in each experiment was computationally expensive (see below). Here we only analyzed the experiments asking participants to learn single words, which are one of the most commonly used materials in recognition test and also used in the two experiments of the current study.

$$conf_{trans} = 1 + \frac{(conf - min_{conf})\,(6-1)}{max_{conf} - min_{conf}}$$

In the equation above, *conf* is the original confidence rating, and $conf_{trans}$ is the transformed confidence rating on a 6-point scale. The max_{conf} and min_{conf} are the maximum and minimum value of the original confidence scale, respectively. This transformation was implemented to ensure that the range of confidence ratings (1–6) was the same for all experiments.[2]

Relationships among response time, performance and confidence. Next, we built multilevel logistic regression models (level 1: trial; level 2: participant) to use response time in each trial to predict performance in recognition test (1 = correct, 0 = incorrect) separately for each experiment. We also built multilevel linear models separately for each experiment to (1) regress confidence on response time, (2) regress confidence on performance, and (3) regress confidence on both response time and performance.

In multilevel regression models, the effect of predictor variables on outcome variable can be divided into within-participant and between-participant effect: the within-participant effect refers to whether predictor variables can predict outcome variable across trials within each participant, and the between-participant effect refers to whether the mean of predictor variables for each participant can predict the mean of outcome variable at group level (Enders & Tofighi, 2007; Zhang et al., 2009). In the current study, we only concerned the relationship between variables across trials within each participant. Here we used the *isolate* function from the *bmlm* package in R to extract the between-participant component centered at the grand mean, and the within-participant component centered at the participant mean, for each predictor variable (Vuorre & Bolger, 2018). We added both between-participant and within-participant components into the multilevel regression models, and only focused on the effect of within-participant components.

All multilevel logistic and linear regression analyses were performed in SPSS. We added random intercept and random slope for the within-participant component of each predictor variable in the models. In addition, we set the covariance structure for the random effects as diagonal to remove the correlation between random effects, which often made the model fail to converge (Singmann & Kellen, 2020). After performing the analyses for each experiment, we built 3-level logistic and linear models (level 1: trial; level 2: participant; level 3: experiment) to examine the overall effects across six experiments. We computed partial eta squared (η_p^2) to estimate the effect size in multilevel regression models (see also Hu et al., 2019).

Fitting DDM to data. We fit DDM to data of response time and performance in recognition test with the Wiener module (Wabersich & Vandekerckhove, 2014) for the JAGS modeling software (Plummer, 2003), using *R2jags* package in R as interface (Su & Yajima, 2021). In general, there are two methods to examine the relationship between confidence ratings and the parameters $\boldsymbol{T_{er}}$ and $\boldsymbol{v}$ in DDM across trials. The first method is to build a hierarchical Bayesian DDM model which combines DDM and multilevel linear regression in one model (e.g., Nunez et al., 2017). Hierarchical Bayesian model can consider the uncertainty about estimation of DDM parameters when analyzing the relationship between

[2] The transformation of confidence scale implemented here was based on a simplified assumption that participants considered the intervals between neighboring scale points were equal across the whole scale. Some studies provided evidence against this assumption (Hanczakowski, Zawadzka, et al., 2013; Zawadzka & Higham, 2015), which might be considered in future studies.

DDM parameters and confidence ratings using multilevel regression. However, during model development, we found that in order to make the hierarchical Bayesian DDM model converge, DDM parameters must be regressed on confidence ratings rather than vice versa (as in Nunez et al., 2017). This limitation led to two problems: (1) using DDM parameters as outcome variable and confidence rating as predictor variable (rather than the other way round) deviated from the main aim of the current study (investigating whether $\boldsymbol{T_{er}}$ and $\boldsymbol{v}$ could predict retrospective confidence), and (2) only one parameter ($\boldsymbol{T_{er}}$ or $\boldsymbol{v}$) could be used as outcome variable at a time, preventing us from simultaneously estimating the effect of $\boldsymbol{T_{er}}$ and $\boldsymbol{v}$ on confidence.

The second method is to fit DDM to data separately for each participant, and estimate the posterior distribution of $\boldsymbol{T_{er}}$ and $\boldsymbol{v}$ in each trial. Then we could use the posterior mean of $\boldsymbol{T_{er}}$ and $\boldsymbol{v}$ in each trial as predictor variables to predict confidence ratings in another multilevel linear model. Although simply extracting the posterior mean of DDM parameters ignores the uncertainty in parameter estimation, it allows us to build a multilevel linear model in which confidence rating is the outcome variable and both $\boldsymbol{T_{er}}$ and $\boldsymbol{v}$ are predictor variables. In the main text, we report the results based on the second method. Results from hierarchical Bayesian DDM model (the first method) are reported in Section S1 of the Supplemental Materials.

According to Wabersich & Vandekerckhove (2014), response time for incorrect trials in recognition test was flipped to be negative. The response time for the ith trial within a participant is then distributed as:

$$RT_i \sim WFPT(a, T_{er(i)}, \beta, \nu_i)$$

WFPT refers to Wiener First Passage Time distribution, which characterizes the distribution of response time predicted by DDM (Ratcliff, 1978). The parameter $\boldsymbol{a}$ is boundary separation, $\boldsymbol{T_{er}}$ is non-decision time, and $\boldsymbol{v}$ is drift rate. The parameter $\boldsymbol{\beta}=\boldsymbol{z}/\boldsymbol{a}$ is the relative starting point, with $\boldsymbol{\beta}=0.5$ indicating no bias. Here we set $\boldsymbol{\beta}$ as 0.5 (Nunez et al., 2017; Voss et al., 2015).

In the current model, drift rate $\boldsymbol{v}$ and non-decision time $\boldsymbol{T_{er}}$ are allowed to vary across trials within each participant. As in previous studies, the drift rate in each trial is distributed as a normal distribution (Ratcliff et al., 2004; Ratcliff & Tuerlinckx, 2002). To weaken the dependence between samples from posterior distribution and make the model converge better, we used parameter expansion technique which augments the original model by a redundant multiplicative parameter, introducing an additional random component in the sampling process (Lee & Wagenmakers, 2013):

$$v_i = \mu_v + \xi_v \bullet \delta_{v(i)}$$

$$\delta_{v(i)} \sim Normal(0, {\sigma_{\delta\nu}}^2)$$

$$\sigma_v = |\xi_v| \bullet \sigma_{\delta v}$$

The parameter $\boldsymbol{\mu_v}$ is the mean of drift rate across trials for a participant, and $\boldsymbol{\sigma_v}$ is the standard deviation of drift rate. $\boldsymbol{\delta_v}$ represents the variation of drift rate across trials characterized by a standard deviation parameter $\boldsymbol{\sigma_{\delta v}}$. $\boldsymbol{\xi_v}$ is the redundant multiplicative parameter.

In previous studies, non-decision time across trials is often assumed to be distributed as a uniform distribution (Ratcliff et al., 2004; Ratcliff & Tuerlinckx, 2002; Wiecki et al., 2013). However, during model development, we found that assuming non-decision time comes from a normal distribution could make the model converge better. According to Ratcliff (2013; 2004), using a uniform or normal distribution for non-decision time in DDM should have little effect on model predictions. Similar to drift rate, we also implemented parameter expansion technique for non-decision time:

$$T_{er(i)} = \mu_T + \xi_T \bullet \delta_{T(i)}$$

$$\delta_{T(i)} \sim Normal(0, {\sigma_{\delta T}}^2)$$

$$\sigma_T = |\xi_T| \bullet \sigma_{\delta T}$$

$\boldsymbol{\mu_T}$ is the mean of non-decision time across trials for a participant, and $\boldsymbol{\sigma_T}$ is the standard deviation of non-decision time. $\boldsymbol{\delta_T}$ represents the variation of non-decision time across trials characterized by a standard deviation parameter $\boldsymbol{\sigma_{\delta T}}$. $\boldsymbol{\xi_T}$ is the redundant multiplicative parameter.

We used Markov chain Monte Carlo (MCMC) method implemented in JAGS to sample from posterior distributions of parameters (see Section S2 in the Supplemental Materials for the prior distribution of the parameters). We fit the model separately for each participant with 4 chains and each chain contained 50,000 samples. The first 25,000 samples in each chain were discarded as burn-in and the thin interval was set as 10, resulting in 10,000 stored samples in total. Gelman and Rubin's potential scale reduction factor $\hat{R}$ was calculated for all parameters, and $\hat{R}$ value lower than 1.1 indicates good convergence (Gelman & Rubin, 1992). For the participants with $\hat{R}$ value higher than 1.1, we refit the model with longer chains and added additional redundant parameters to make sure the estimation of all parameters was converged (see Section S2 in Supplemental Materials for details).

Using drift rate and non-decision time to predict confidence. After fitting DDM to data from each participant, we extracted the posterior mean of drift rate $\boldsymbol{v}$ and non-decision time $\boldsymbol{T_{er}}$ for each trial. Then we built a 2-level linear model separately for each experiment, in which retrospective confidence ratings were regressed on both $\boldsymbol{v}$ and $\boldsymbol{T_{er}}$. Similar to the multilevel regression analyses described above, here we only focused on the effect of within-participant component for each predictor variable. We added random intercept and random slope for the within-participant components, and set the covariance structure for the random effects as diagonal. After performing the analyses for each experiment, we built a 3-level linear model to examine the overall effects across six experiments.

Effects of drift rate and non-decision time on confidence accuracy. In DDM, drift rate $\boldsymbol{v}$ directly affects memory performance in recognition test, while non-decision time $\boldsymbol{T_{er}}$ is only related to total response time but not performance (Ratcliff & McKoon, 2008). Thus, basing retrospective confidence on drift rate should make confidence a good predictor of actual memory performance, and increase confidence accuracy. In contrast, the reliance of confidence on non-decision time might lead to inaccurate confidence ratings. To test this possibility, we performed an analysis similar to that in Koriat & Ackerman (2010). For each participant, we first computed Kendall's τ_b correlation between confidence ratings and memory performance as a measure of confidence accuracy.[3] Then we separately computed semi-partial (part) Kendall's τ_b correlation between confidence and performance when the effect of drift rate $\boldsymbol{v}$ on confidence was partialled out, or when the effect of non-decision time $\boldsymbol{T_{er}}$ on confidence was partialled out. The semi-partial Kendall's τ_b correlation was computed using the *ppcor* package in R (Kim, 2015). For each experiment, we performed two paired-sample *t* tests to compare the zero-order Kendall's τ_b correlation and each of the two semi-partial correlation coefficients. If the zero-order correlation between confidence and performance was significantly higher than the semi-partial correlation when the effect of $\boldsymbol{v}$ or $\boldsymbol{T_{er}}$ on confidence was partialled out, we could conclude that the reliance of confidence on $\boldsymbol{v}$ or $\boldsymbol{T_{er}}$ increased confidence accuracy. In contrast, if the semi-partial correlation was higher, then basing confidence on $\boldsymbol{v}$ or $\boldsymbol{T_{er}}$ should negatively affect confidence accuracy.

Results

Relationships among response time, performance and confidence

In all of the six experiments, higher recognition memory performance was associated with shorter response time in recognition test across trials, bs < -0.22, ts > 6.37, ps < 0.001, $\eta_p^2 > 0.63$ (see Table 1). Furthermore, retrospective confidence ratings were higher when response time was shorter, bs < -0.38, ts > 8.23, ps < 0.001, $\eta_p^2 > 0.67$, or when memory performance was higher, bs > 0.40, ts > 7.98, ps < 0.001, $\eta_p^2 > 0.68$ (see Table 2). When regressing confidence ratings on response time and memory performance at the same time, we found that both response time and performance could significantly predict confidence,

Table 1 Within-participant fixed effects for the regression of performance on response time in the six published experiments

Experiments	Performance on response time		
	b	95% CI	η_p^2
Hu et al., 2017	-0.27	[-0.35, -0.19]	0.65
Kantner & Lindsay 2014, Experiment 2	-0.59	[-0.68, -0.49]	0.72
Sadeghi et al., 2017	-0.22	[-0.27, -0.17]	0.64
Schmidt et al., 2019	-0.32	[-0.40, -0.24]	0.72
Siedlecka et al., 2019, Experiment 1	-1.02	[-1.25, -0.80]	0.90
Siedlecka et al., 2019, Experiment 2	-0.93	[-1.23, -0.63]	0.64
Overall effect	-0.53	[-0.87, -0.19]	0.76

Note. CI = confidence interval

[3] Although many previous studies on metamemory use Gamma correlation to measure confidence accuracy (for a review, see Masson & Rotello 2009), there are some studies using Kendall's τ_b (Dougherty et al., 2018; Robey et al., 2017). Here we use Kendall's τ_b as the measure of confidence accuracy because it is easy to compute semi-partial Kendall's τ_b correlation.

Table 2 Within-participant fixed effects for the regression of confidence separately on response time and performance in the six published experiments

Experiments	Confidence on response time			Confidence on performance		
	b	95% CI	η_p^2	*b*	95% CI	η_p^2
Hu et al., 2017	-0.45	[-0.53, -0.37]	0.79	1.29	[1.07, 1.51]	0.80
Kantner & Lindsay 2014, Experiment 2	-0.74	[-0.87, -0.61]	0.68	1.10	[0.97, 1.24]	0.82
Sadeghi et al., 2017	-0.38	[-0.44, -0.32]	0.79	0.41	[0.32, 0.49]	0.69
Schmidt et al., 2019	-0.49	[-0.62, -0.37]	0.74	1.44	[1.20, 1.67]	0.86
Siedlecka et al., 2019, Experiment 1	-1.01	[-1.19, -0.83]	0.89	1.63	[1.20, 2.06]	0.79
Siedlecka et al., 2019, Experiment 2	-1.30	[-1.61, -0.99]	0.78	0.98	[0.78, 1.18]	0.81
Overall effect	-0.71	[-1.07, -0.36]	0.85	1.12	[0.69, 1.55]	0.90

Note. CI = confidence interval

$b_{response\ time}$ < -0.36, *t*s > 7.74, *p*s < 0.001, η_p^2 > 0.63; $b_{performance}$ > 0.29, *t*s > 7.02, *p*s < 0.001, η_p^2 > 0.54 (see Table 3). Results regarding the overall effects across six experiments were consistent with those in each experiment, *t*s > 4.02, *p*s < 0.05, η_p^2 > 0.75 (see Tables 1, 2 and 3).

Effects of drift rate and non-decision time on confidence

When regressing retrospective confidence ratings on both drift rate $\boldsymbol{v}$ and non-decision time $\boldsymbol{T_{er}}$ across trials, we found that higher confidence was associated with higher drift rate in all of the six experiments, *b*s > 0.46, *t*s > 5.23, *p*s < 0.001, η_p^2 > 0.54. Confidence ratings also tended to be higher when non-decision time was shorter. However, the effect of non-decision time on confidence was only statistically detectable in four experiments, *b*s < -2.34, *t*s > 3.03,

Table 3 Within-participant fixed effects for the regression of confidence on both response time and performance in the six published experiments

Experiments	Response time			Performance		
	b	95% CI	η_p^2	*b*	95% CI	η_p^2
Hu et al., 2017	-0.40	[-0.48, -0.33]	0.79	1.00	[0.81, 1.19]	0.77
Kantner & Lindsay 2014, Experiment 2	-0.64	[-0.76, -0.52]	0.64	0.83	[0.72, 0.94]	0.78
Sadeghi et al., 2017	-0.37	[-0.43, -0.30]	0.78	0.29	[0.22, 0.37]	0.55
Schmidt et al., 2019	-0.41	[-0.52, -0.30]	0.71	1.15	[0.95, 1.36]	0.83
Siedlecka et al., 2019, Experiment 1	-0.87	[-1.03, -0.72]	0.90	1.08	[0.75, 1.40]	0.76
Siedlecka et al., 2019, Experiment 2	-1.16	[-1.46, -0.86]	0.76	0.78	[0.60, 0.97]	0.78
Overall effect	-0.63	[-0.94, -0.32]	0.86	0.84	[0.52, 1.16]	0.90

Note. CI = confidence interval

*p*s < 0.02, $\eta_p^2 > 0.44$, but did not reach statistical significance in the other two experiments, *b*s > -2.41, *t*s < 1.98, *p*s > 0.07, $\eta_p^2 < 0.29$ (see Table 4). To further investigate the overall effect of drift rate and non-decision time on confidence in the six experiments, we then conducted 3-level linear regression analysis on data across experiments. Results revealed that drift rate could positively predict confidence, $b = 1.91$, $t = 4.73$, $p = .005$, $\eta_p^2 = 0.82$. More importantly, there was an overall negative effect of non-decision time on confidence, $b = -3.71$, $t = 3.70$, $p = .014$, $\eta_p^2 = 0.74$ (see Table 4).

Effects of drift rate and non-decision time on confidence accuracy

Finally, we investigated whether the reliance of retrospective confidence ratings on drift rate and non-decision time influenced confidence accuracy. In all of the six experiments,

Table 4 Within-participant fixed effects for the regression of confidence on both drift rate and non-decision time in the six published experiments

Experiments	Drift rate v			Non-decision time T_{er}		
	b	95% CI	η_p^2	b	95% CI	η_p^2
Hu et al., 2017	3.70	[2.40, 5.00]	0.55	-2.34	[-4.18, -0.51]	0.57
Kantner & Lindsay 2014, Experiment 2	1.23	[0.97, 1.49]	0.65	-6.72	[-10.21, -3.23]	0.45
Sadeghi et al., 2017	0.46	[0.35, 0.57]	0.75	-15.07	[-21.21, -8.92]	0.77
Schmidt et al., 2019	3.48	[2.09, 4.88]	0.60	-2.41	[-5.13, 0.32]	0.28
Siedlecka et al., 2019, Experiment 1	1.95	[1.18, 2.72]	0.85	-0.20	[-2.49, 2.09]	0.01
Siedlecka et al., 2019, Experiment 2	1.28	[0.88, 1.68]	0.86	-2.48	[-4.08, -0.87]	0.77
Overall effect	1.91	[0.88, 2.94]	0.82	-3.71	[-6.30, -1.12]	0.74

Note. CI = confidence interval

Table 5 Kendall's τ_b correlation between confidence and performance in the six published experiments

Experiments	Zero-order		Semi-partial (partialling out v)		Semi-partial (partialling out T_{er})	
	$M\tau_b$	95% CI	$M\tau_b$	95% CI	$M\tau_b$	95% CI
Hu et al., 2017	0.27	[0.24, 0.30]	0.09	[0.07, 0.12]	0.29	[0.26, 0.32]
Kantner & Lindsay 2014, Experiment 2	0.25	[0.22, 0.27]	0.10	[0.08, 0.12]	0.26	[0.23, 0.28]
Sadeghi et al., 2017	0.14	[0.11, 0.17]	0.02	[0.00, 0.04]	0.15	[0.13, 0.18]
Schmidt et al., 2019	0.30	[0.27, 0.33]	0.11	[0.08, 0.14]	0.31	[0.28, 0.34]
Siedlecka et al., 2019, Experiment 1	0.33	[0.28, 0.38]	0.16	[0.12, 0.20]	0.35	[0.30, 0.40]
Siedlecka et al., 2019, Experiment 2	0.26	[0.22, 0.30]	0.09	[0.06, 0.13]	0.27	[0.23, 0.31]

Note. CI = confidence interval

the zero-order Kendall's τ_b correlation between confidence and memory performance was higher than the semi-partial correlation when the effect of drift rate on confidence was partialled out, $ts > 13.62$, $ps < 0.001$, Cohen's $d > 2.20$, indicating that basing confidence on drift rate could partly account for the positive relationship between confidence and performance. However, the zero-order correlation was lower than the semi-partial correlation in all experiments when the effect of non-decision time on confidence was partialled out, $ts > 2.44$, $ps < 0.05$, Cohen's $d > 0.47$ (see Table 5). Thus, non-decision time acted as a suppressor variable in the relationship between confidence and performance: non-decision time explained part of the variance of confidence that was unrelated to actual memory performance, and removing the effect of non-decision time strengthened the correlation between confidence and performance (Smith et al., 1992).

Discussion

Results from six published experiments showed that retrospective confidence ratings were higher when response time in recognition test was shorter, replicating the effect of response time on confidence ratings in previous studies (Robinson et al., 1997; Siedlecka et al., 2019). Furthermore, response time also negatively predicted memory performance in recognition test, indicating that response time was a reliable cue predicting actual performance (Anderson, 1981; Robinson et al., 1997; Siedlecka et al., 2019). However, there was still a significant effect of response time on confidence ratings when the effect of memory performance was controlled, suggesting that relying on response time might make retrospective confidence unable to fully reflect the variation of actual performance across trials.

Consistent with our hypothesis, higher confidence ratings were associated with higher drift rate across trials in each of the six experiments. Drift rate indexes the speed of memory evidence accumulation during recognition test, and directly reflects the fluency of retrieving information from long-term memory (Ratcliff, 1978; Ratcliff & McKoon, 2008). Our findings were in accordance with previous theoretical account about the utilization of retrieval fluency during retrospective confidence (Koriat et al., 2008; Koriat & Ackerman, 2010).

More importantly, there was also an overall negative effect of non-decision time on confidence ratings across the six experiments, suggesting that retrospective confidence ratings depend on not only retrieval fluency, but also the speed of cognitive process unrelated to memory retrieval. However, in contrary to drift rate, non-decision time did not reliably influence confidence ratings in all of the six experiments. Furthermore, the effect size (η_p^2) was slightly higher for the overall effect of drift rate on confidence than the overall effect of non-decision time across the six experiments. It is possible that although the speed of both retrieval and non-decision process was reliably utilized during retrospective confidence, the extent of the reliance of confidence on non-decision process might be slightly smaller compared with the reliance on retrieval fluency.

When examining the effect of drift rate and non-decision time on confidence accuracy, we found that zero-order Kendall's τ_b correlation between confidence and memory performance was higher than the semi-partial correlation when the effect of drift rate on confidence was partialled out, but lower than the semi-partial correlation when the effect of non-decision time on confidence was partialled out. These results supported our hypothesis that giving retrospective confidence ratings based on drift rate increases confidence accuracy, while the reliance of confidence on non-decision time leads to inaccurate confidence ratings. According to the assumption of DDM, drift rate determines the tendency that memory evidence accumulates toward the correct response, and affects memory performance in recognition test (Ratcliff & McKoon, 2008). Thus, the contribution of drift rate to confidence should strengthen the relationship between confidence ratings and actual performance. On the other hand, non-decision time in DDM only affects the total response time but not memory performance (Ratcliff & McKoon, 2008), and relying on non-decision time during confidence process makes confidence ratings deviate from actual performance in the test.

In brief, the current analysis on data from previous experiments indicated that retrospective confidence ratings are related to the speed of both retrieval and non-decision process during recognition test. However, our analyses above were not able to indicate which non-decision process could affect retrospective confidence ratings. In the next two experiments (1a and 1b), we specifically investigated whether the speed of stimulus perception, which is an important part of non-decision process during recognition (Ratcliff & McKoon, 2008), could predict retrospective confidence ratings. One way to alter the speed of stimulus perception is to manipulate the ease of perceiving the stimuli in recognition test (Yang et al., 2018). Many previous studies indicate that changing the perceptual characteristics of stimuli, such as font size, font type and blurriness, can reliably affect the ease of perception (Diemand-Yauman et al., 2011; Geller & Peterson, 2021; Rosner et al., 2015; Yang et al., 2018; Yue et al., 2013). Furthermore, ease of perception is closely related to people's evaluation of memory performance, and people often predict they will have higher performance in memory test when it is easier to perceive study items (Busey et al., 2000; Undorf et al., 2017; Wang et al., 2020; Yang et al., 2018; Yue et al., 2013). In Experiments 1a and 1b, we manipulated the ease of perceiving the words presented in recognition test based on either blurriness or irregular font type (Diemand-Yauman et al., 2011; Rosner et al., 2015). We aimed to (1) replicate the effects of both drift rate and non-decision time on retrospective confidence ratings, and (2) investigate whether non-decision time mediated the effect of the ease of perception on confidence.

Experiments 1a and 1b

Method

Participants

The sample size in Experiments 1a and 1b was determined based on the summary-statistics-based power analysis for multilevel model developed by Murayama et al., (2020). We aimed to replicate both the positive effect of drift rate and negative effect of non-decision time on retrospective confidence ratings. Because Murayama et al.'s method only receives parameter inputs from a 2-level rather than 3-level model, we combined data from all of the six published experiments, and built a 2-level (level 1: trial; level 2: participant) linear model in which confidence ratings were regressed on both drift rate and non-decision time. The estimated parameters were then input into the web app (https://koumurayama.shinyapps.io/summary_statistics_based_power/) developed by Murayama et al., (2020). Furthermore, the cluster size used in the power analysis was represented by the averaged trial number for each participant across all of the six experiments (141 trials).

Results from the power analysis revealed that when there were 200 trials for each participant (as in our experiments; see below), 26 participants were needed with a power of 0.80. We decided to use a slightly larger sample size because (1) the model for power analysis (2-level model) was different from that for data analysis across the six experiments (3-level model), and (2) the actual trial number for each participant in Experiments 1a and 1b might be lower than 200 after we removed trials with extreme response time (see the data analysis below).

We finally recruited 31 participants in Experiment 1a (19 female; age: $M=28.48$ years, $SD=6.67$), and 31 participants in Experiment 1b (13 female; age: $M=33.97$ years, $SD=11.02$). All participants were recruited online via the Prolific website (https://prolific.ac/). Data from another one participant in Experiment 1b were excluded due to memory performance lower than the chance level (0.5).

Participants signed informed consent prior to the experiment, and received monetary compensation (£4.5) and a bonus (up to £0.5, dependent on memory performance) after the experiment. All participants spoke English as the first language and reported normal or corrected-to-normal vision. All procedures were approved by the Ethics Committee at Beijing Normal University.

Materials

The experimental materials for Experiments 1a and 1b were 400 English words selected from the MRC Psycholinguistic Database (Coltheart, 1981). All words contained 4–8 letters and 1–3 syllables. The Kucera and Francis word frequency for all words was between 8 and 90 ($M=32.38$; $SD=20.83$), and the ratings of familiarity ($M=537.66$; $SD=39.72$), concreteness ($M=561.10$; $SD=47.46$) and imagability ($M=565.74$; $SD=43.10$) were between 450 and 650. The words were randomly divided into two halves for each participant, in which 200 words were used as study materials in the learning phase, with the other 200 words served as lures in the 2AFC recognition memory test.

Procedure

The procedure for the recognition memory task in Experiments 1a and 1b was similar to those in previous studies (Fleming et al., 2014; McCurdy et al., 2013; Palmer et al., 2014). The task contained four blocks, and in each block participants were first presented with 50 words simultaneously on the screen (arranged in 10 rows and 5 columns) and asked to memorize as many words as possible (see Fig. 2a). All words were presented in a clear font. The exposure duration of the words was 30, 50, 70 or 90 s in the four blocks, and the order of the four durations was randomized across participants. When there were 10 s left of the learning phase, participants were notified by a countdown clock presented below the 50 words. Following the learning phase in each block, participants completed 50 trials in the 2AFC recognition memory test (see Fig. 2b). After presenting the fixation for 500 ms, a learned word and a new word were simultaneously shown on the screen, and the positions of the two words were randomly determined in each trial. Participants needed to decide which word had been learned by pressing the number key 1 or 2. The chosen answer would

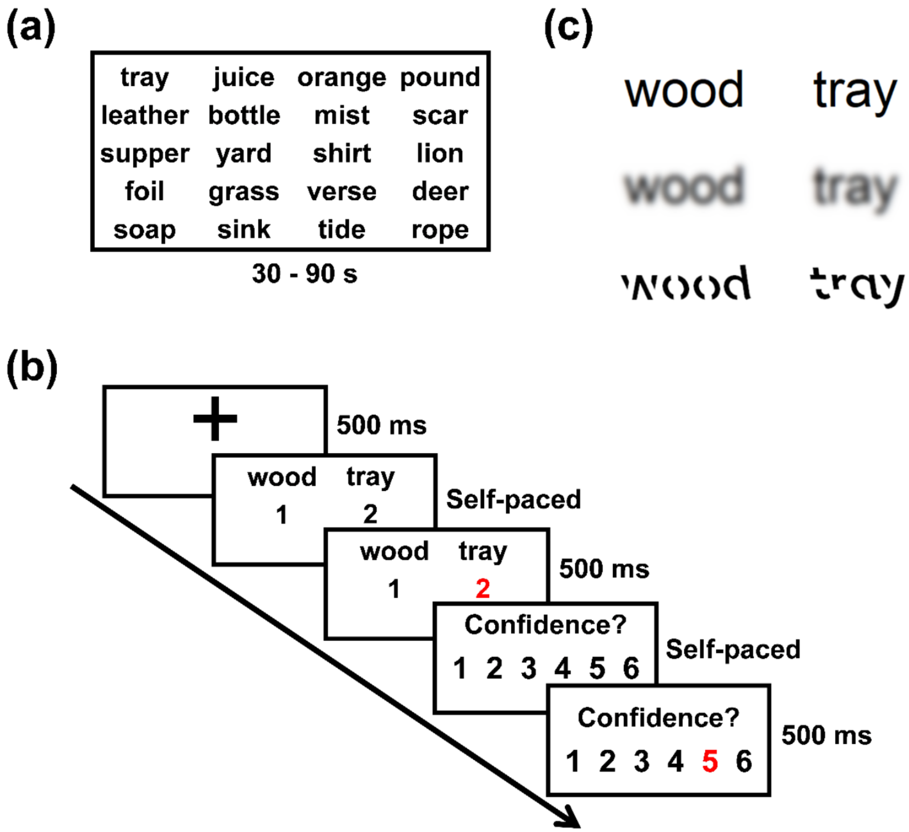

Fig. 2 (a) The learning phase in Experiments 1a and 1b. (b) The recognition test phase in Experiments 1a and 1b. (c) Examples of words for easy-to-perceive (clear font) and hard-to-perceive condition (blurred font in Experiment 1a, and Sans Forgetica font in Experiment 1b)

then be highlighted for 500 ms. Finally, participants were required to press a number key to indicate their confidence regarding answer correctness. The confidence scale ranged from 1 (not confident at all) to 6 (extremely confident). The chosen number in the confidence rating scale would also be highlighted for 500 ms. There was no time pressure on recognition decision or confidence rating.

The 50 trials in the recognition test for each block of Experiments 1a and 1b were randomly divided into easy-to-perceive or hard-to-perceive condition. In each trial of easy-to-perceive condition for both experiments, the two words shown in the recognition test were presented in a clear 32-point Arial font. In each trial of hard-to-perceive condition for Experiment 1a, the two words were presented in a blurred 32-point Arial font. Blurred words were created by applying a Gaussian blur of 10 to each word using GNU Image Manipulation Program (GIMP) (Rosner et al., 2015). In each trial of hard-to-perceive condition for Experiment 1b, the two words were presented in a clear 32-point Sans Forgetica font, which was designed to induce difficulty in the perception of words during reading (see Fig. 2c) (Geller & Peterson, 2021).

Data analysis

Data preparation. Before performing any data analysis, we excluded trials with response time that was shorter than 100 ms or differed by more than 3 standard deviations from the mean, as in our analysis on the six published experiments.

Effect of perceptual manipulation on performance, response time, mean confidence, and confidence accuracy. We first used paired-sample *t* tests to compare memory performance (proportion of correctly answered trials), response time in recognition test, and mean confidence between easy- and hard-to-perceive condition separately in Experiments 1a and 1b. Then we computed Kendall's τ_b correlation as the measure of confidence accuracy separately for each condition within each participant, and investigated the effect of perceptual manipulation on confidence accuracy.[4]

Effect of perceptual manipulation on DDM parameters. We fit DDM to data separately in each condition for each participant (as in the data analysis for previous published experiments; see above), and extracted the posterior mean of boundary separation $\boldsymbol{a}$, overall drift rate across trials $\boldsymbol{\mu_v}$, and overall non-decision time across trials $\boldsymbol{\mu_T}$. We then performed paired-sample *t* tests to examine the difference in $\boldsymbol{a}$, $\boldsymbol{\mu_v}$ and $\boldsymbol{\mu_T}$ between easy- and hard-to-perceive condition in Experiments 1a and 1b (we also built hierarchical Bayesian models to investigate the influence of perceptual manipulation on DDM parameters at trial level, which replicated the results from *t* tests; see Section S3 in the Supplemental Materials for details).

Relationship between performance, response time and confidence. Next, we combined data from two experimental conditions, and built 2-level logistic and linear regression models (level 1: trial; level 2: participant) separately for each experiment to (1) regress confidence on response time, (2) regress confidence on performance, and (3) regress confidence on both response time and performance. As in the data analysis for previous published experiments, here we only focused on the within-participant effects of the predictor variables. We then added the main effect of experimental condition (0 = easy-to-perceive, 1 = hard-to-perceive) and the interaction between condition and each predictor variable into the models, and explored whether the ease of perception could moderate the effects of the predictor variables.

Using drift rate and non-decision time to predict confidence. To replicate the effects of drift rate $\boldsymbol{v}$ and non-decision time $\boldsymbol{T_{er}}$ on retrospective confidence ratings in previous published experiments, we combined data from two experimental conditions and built a 2-level linear model (level 1: trial; level 2: participant) separately for Experiments 1a and 1b, in which trial-by-trial confidence rating was regressed on both $\boldsymbol{v}$ and $\boldsymbol{T_{er}}$. Then we added the main effect of experimental condition and the interaction between condition and each of the two parameters into the model, and examined whether the ease of perception could moderate the effect of $\boldsymbol{v}$ and $\boldsymbol{T_{er}}$ on confidence.

Effects of drift rate and non-decision time on confidence accuracy. To replicate the effects of drift rate $\boldsymbol{v}$ and non-decision time $\boldsymbol{T_{er}}$ on confidence accuracy, we computed zero-order Kendall's τ_b correlation between confidence and performance, semi-partial Kendall's τ_b correlation when the effect of $\boldsymbol{v}$ on confidence was partialled out, and semi-partial correlation when the effect of $\boldsymbol{T_{er}}$ on confidence was partialled out separately for easy- and hard-

[4] The Kendall's τ_b correlation could not be computed in hard-to-perceive condition for one participant in Experiment 1b, because all of the trials were correctly answered in recognition test.

to-perceive condition within each participant. Then for each experiment, we performed a 2 (Correlation type: zero-order vs. semi-partial) × 2 (Condition: easy vs. hard-to-perceive) analysis of variance (ANOVA) on Kendall's τ_b correlation to investigate whether the correlation coefficient significantly changed when the effect of $\boldsymbol{v}$ was partialled out, and whether experimental condition interacted with the type of correlation. Similarly, we conducted a 2 (Correlation type: zero-order vs. semi-partial) × 2 (Condition: easy vs. hard-to-perceive) ANOVA to compare the zero-order correlation and semi-partial correlation when the effect of $\boldsymbol{T_{er}}$ was partialled out.

Mediation effect of non-decision time on the relationship between the ease of perception and confidence. To foreshadow, our results revealed that the ease of perception could reliably affect non-decision time $\boldsymbol{T_{er}}$ in recognition test. Thus, we finally examined whether non-decision time could mediate the effect of the ease of perception on retrospective confidence ratings at trial level. For each experiment, we constructed a multilevel mediation model using the MLmed package in SPSS (Hayes & Rockwood, 2020). The drift rate $\boldsymbol{v}$ for each trial was added into the model as covariate, and we also added all of the random intercepts and random slopes. MLmed could divide the overall mediation effect into within-participant and between-participant effects, and we only focused on the within-participant mediation effect.

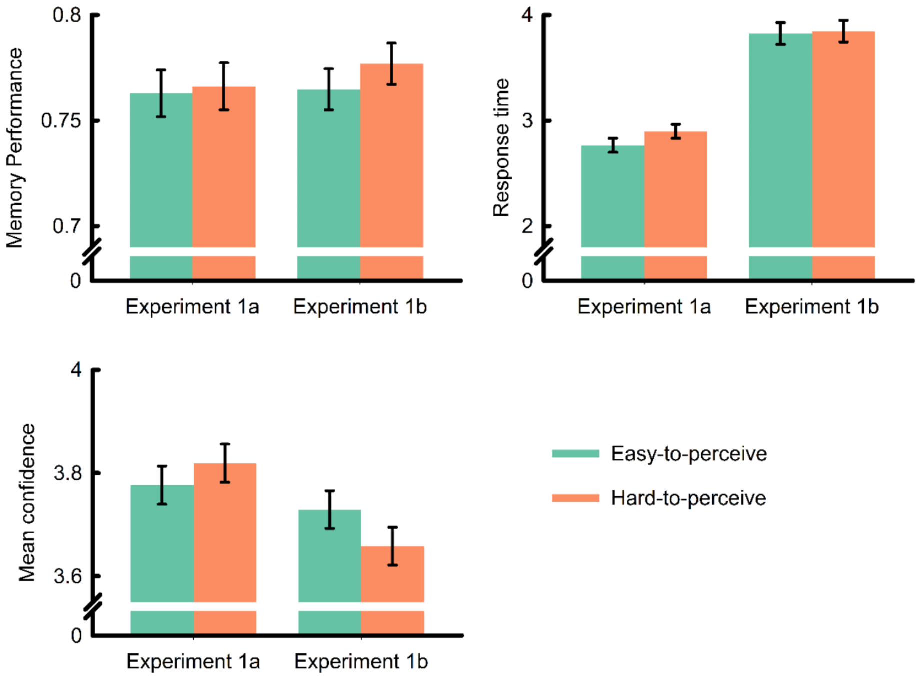

Fig. 3 Memory performance, response time (in seconds) and mean confidence for easy- and hard-to-perceive condition in Experiments 1a and 1b. Error bars represent within-participant 95% confidence interval (Cousineau, 2005)

Results

Effects of perceptual manipulation on performance, response time, mean confidence, and confidence accuracy

Memory performance in recognition test did not differ between easy- and hard-to-perceive condition in both Experiments 1a and 1b, Exp 1a: $t(30)=0.30$, $p=.766$, Cohen's $d=0.05$; Exp 1b: $t(30)=1.27$, $p=.215$, Cohen's $d=0.23$. While response time in recognition test was longer for hard- than easy-to-perceive condition, this difference was not statistically detectable in both experiments (although the effect was marginally significant in Experiment 1a), Exp 1a: $t(30)=2.04$, $p=.051$, Cohen's $d=0.37$; Exp 1b: $t(30)=0.23$, $p=.824$, Cohen's $d=0.04$. Furthermore, mean confidence ratings did not reliably differ between easy- and hard-to-perceive condition in both experiments, although confidence in easy-to-perceive condition was marginally higher in Experiment 1b, Exp 1a: $t(30)=1.18$, $p=.249$, Cohen's $d=0.21$; Exp 1b: $t(30)=1.98$, $p=.057$, Cohen's $d=0.36$ (see Fig. 3).

We then computed Kendall's τ_b correlation between confidence and performance as the measure of confidence accuracy, and examined the effect of perceptual manipulation on confidence accuracy. Our results revealed that there was no reliable difference in Kendall's τ_b correlation between easy- and hard-to-perceive condition, Exp 1a: $t(30)=0.78$, $p=.444$, Cohen's $d=0.14$; Exp 1b: $t(29)=1.18$, $p=.247$, Cohen's $d=0.25$. (see Table 6).

Effect of perceptual manipulation on DDM parameters

We fit DDM separately to each experimental condition within each participant, and examined the effect of perceptual manipulation on DDM parameters at participant level. In both experiments, boundary separation $\boldsymbol{a}$ did not differ between easy- and hard-to-perceive condition, Exp 1a: $t(30)=1.49$, $p=.147$, Cohen's $d=0.27$; Exp 1b: $t(30)=1.05$, $p=.304$, Cohen's $d=0.19$. Similarly, the overall drift rate across trials $\boldsymbol{\mu_v}$ did not differ between the two conditions, Exp 1a: $t(30)=0.43$, $p=.668$, Cohen's $d=0.08$; Exp 1b: $t(30)=0.68$, $p=.501$, Cohen's $d=0.12$. However, the overall non-decision time across trials $\boldsymbol{\mu_T}$ was higher for hard- than easy-to-perceive condition, and this difference was statistically detectable, Exp 1a: $t(30)=3.52$, $p=.001$, Cohen's $d=0.63$; Exp 1b: $t(30)=2.80$, $p=.009$, Cohen's $d=0.50$ (see Fig. 4). Overall, these findings reflect that the manipulation of the ease of perception

Table 6 Kendall's τ_b correlation between confidence and performance in Experiments 1a and 1b

Conditions	Zero-order		Semi-partial (partialling out v)		Semi-partial (partialling out T_{er})	
	$M\tau_b$	95% CI	$M\tau_b$	95% CI	$M\tau_b$	95% CI
Experiment 1a						
Easy-to-perceive	0.32	[0.27, 0.37]	0.15	[0.11, 0.19]	0.33	[0.29, 0.38]
Hard-to-perceive	0.30	[0.26,0.35]	0.14	[0.11, 0.18]	0.31	[0.27, 0.35]
Experiment 1b						
Easy-to-perceive	0.24	[0.18, 0.30]	0.08	[0.04, 0.12]	0.25	[0.19, 0.32]
Hard-to-perceive	0.22	[0.16, 0.28]	0.07	[0.03, 0.12]	0.23	[0.17, 0.29]

Note. CI = confidence interval

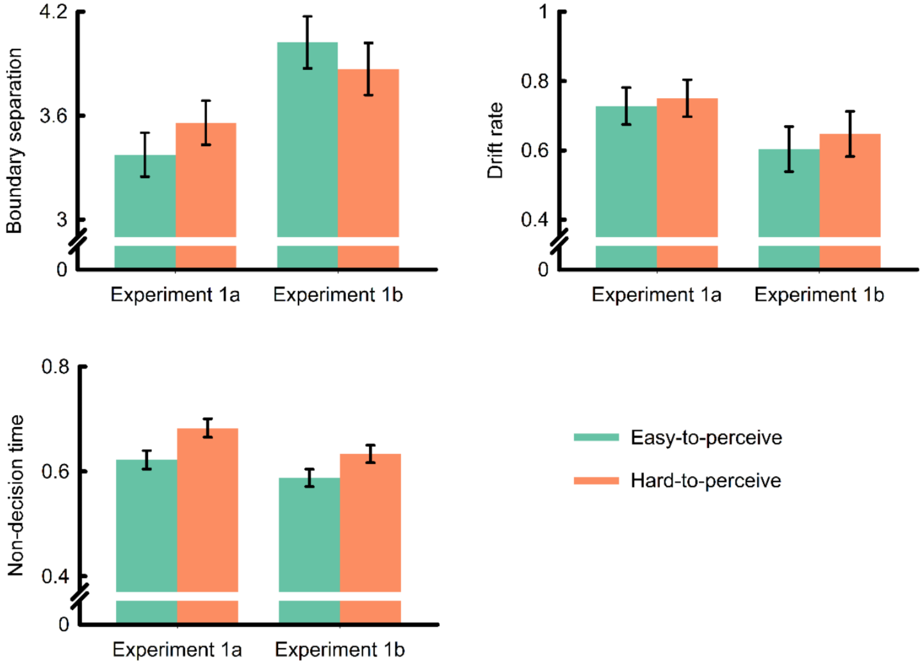

Fig. 4 Boundary separation (***a***), overall drift rate ($\boldsymbol{\mu}_v$) and overall non-decision time ($\boldsymbol{\mu}_T$) for easy- and hard-to-perceive condition in Experiments 1a and 1b. Error bars represent within-participant 95% confidence interval (Cousineau, 2005)

was successful. That is, difficulty in perceiving the stimuli affected non-decision time but not drift rate.

Relationships among response time, performance and confidence

When data from two experimental conditions were combined together, we found that higher recognition memory performance was associated with shorter response time in recognition test across trials in both Experiments 1a and 1b, Exp 1a: b = -0.35, t=5.19, p<.001, η_p^2=0.50; Exp 1b: b = -0.28, t=4.33, p<.001, η_p^2=0.44. Furthermore, retrospective confidence ratings were higher when response time was shorter, Exp 1a: b = -0.55, t=6.12, p<.001, η_p^2=0.58; Exp 1b: b = -0.56, t=5.33, p<.001, η_p^2=0.50, or when memory performance was higher, Exp 1a: b=1.38, t=10.60, p<.001, η_p^2=0.79; Exp 1b: b=1.16, t=7.32, p<.001, η_p^2=0.67. When regressing confidence ratings on response time and memory performance at the same time, we found that both response time and performance could significantly predict confidence, Exp 1a: $b_{response\ time}$ = -0.45, t=6.06, p<.001, η_p^2=0.58; $b_{performance}$ = 1.20, t=9.70, p<.001, η_p^2=0.76; Exp 1b: $b_{response\ time}$ = -0.50, t=5.36, p<.001, η_p^2=0.50; $b_{performance}$ = 0.91, t=6.38, p<.001, η_p^2=0.62.

We then investigated whether experimental condition (easy- vs. hard-to-perceive) could moderate the effect of each predictor variable above. None of the moderation effects reached statistical significance, $|b|$ < 0.11, ts<1.85, ps>0.08, η_p^2<0.40.

Effects of drift rate and non-decision time on confidence

When data from two experimental conditions were combined together, our results revealed that higher confidence was associated with higher drift rate $\boldsymbol{v}$ in both Experiments 1a and 1b, Exp 1a: $b=1.65$, $t=4.91$, $p<.001$, $\eta_p^2=0.54$; Exp 1b: $b=2.01$, $t=5.25$, $p<.001$, $\eta_p^2=0.61$. Furthermore, confidence ratings were also higher when non-decision time $\boldsymbol{T_{er}}$ was shorter, Exp 1a: $b=-1.39$, $t=2.08$, $p=.049$, $\eta_p^2=0.16$; Exp 1b: $b=-2.84$, $t=3.55$, $p=.003$, $\eta_p^2=0.46$.

We then investigated whether experimental condition could moderate the effect of drift rate and non-decision time on confidence. Only in Experiment 1b did our results show a significant moderation effect of condition on the relationship between confidence and non-decision time, suggesting the effect of non-decision time on confidence was smaller (less negative) in hard-to-perceive condition, $b=3.68$, $t=2.69$, $p=.035$, $\eta_p^2=0.54$. To further explore whether non-decision time could predict confidence in both easy- and hard-to-perceive conditions of Experiment 1b, we regressed confidence on both drift rate and non-decision time separately for easy- and hard-to-perceive condition. Results indicated that the effect of non-decision time on confidence was statistically significant in both conditions, bs < -8.64, ts > 2.98, ps < 0.05, $\eta_p^2>0.55$. None of the other moderation effects reached statistical significance, $|b|$ < 0.63, ts < 1.71, ps > 0.12, $\eta_p^2<0.28$.

Effects of drift rate and non-decision time on confidence accuracy

To investigate whether the reliance of drift rate $\boldsymbol{v}$ affected confidence accuracy, we performed 2 (Correlation type: zero-order vs. semi-partial) × 2 (Condition: easy vs. hard-to-perceive) ANOVA on Kendall's τ_b correlation to compare zero-order correlation and semi-partial correlation when the effect of $\boldsymbol{v}$ on confidence was partialled out, and examine whether experimental condition interacted with the type of correlation. Results from both Experiments 1a and 1b revealed that the main effect of correlation type was statistically significant, and the correlation coefficient was lower when the effect of $\boldsymbol{v}$ was partialled out, Exp 1a: $F(1, 30)=168.75$, $p<.001$, $\eta_p^2=0.85$; Exp 1b: $F(1, 29)=79{,}38$, $p<.001$, $\eta_p^2=0.73$. There was no main effect of experimental condition, or the interaction between correlation type and condition, $F<3.23$, $p>.08$, $\eta_p^2<0.11$ (see Table 6).

Similarly, we conducted 2 (Correlation type: zero-order vs. semi-partial) × 2 (Condition: easy vs. hard-to-perceive) ANOVA to examine the effect of non-decision time $\boldsymbol{T_{er}}$ on confidence accuracy. We found that zero-order Kendall's τ_b correlation was reliably lower than the semi-partial correlation when the effect of $\boldsymbol{T_{er}}$ on confidence was partialled out, Exp 1a: $F(1, 30)=10.20$, $p=.003$, $\eta_p^2=0.25$; Exp 1b: $F(1, 29)=18.15$, $p<.001$, $\eta_p^2=0.39$. The main effect of experiment condition and the interaction were not statistically detectable in any experiment, $F<3.32$, $p>.07$, $\eta_p^2<0.11$ (see Table 6). Thus, our results replicated those from the analysis on previous published experiments: the reliance of confidence on drift rate increased confidence accuracy, while basing confidence on non-decision time decreased confidence accuracy.

Mediating role of non-decision time in the relationship between the ease of perception and confidence

Finally, we examined whether non-decision time T_{er} could mediate the effect of the ease of perception on retrospective confidence ratings. In both Experiments 1a and 1b, there was a statistically detectable indirect effect of the ease of perception on confidence through non-decision time, Exp 1a: b = -0.14, Z=2.19, p=.029; Exp 1b: b = -0.20, Z=2.34, p=.020. These results indicated that non-decision time reliably contributed to the effect of the ease of perception on confidence: difficulty in perceiving the stimuli could lead to long non-decision time, resulting in low confidence. However, there was also a trend for the positive direct effect of the ease of perception on confidence. This direct effect was close to significance in Experiment 1a, b=0.16, t=1.90, p=.072, η_p^2=0.15, and statistically significant in Experiment 1b, b=0.19, t=2.44, p=.025, η_p^2=0.25 (see Fig. 5). Thus, there might be another path for the effect of the ease of perception on confidence, in which more difficulty in perceiving the stimuli was associated with higher confidence.

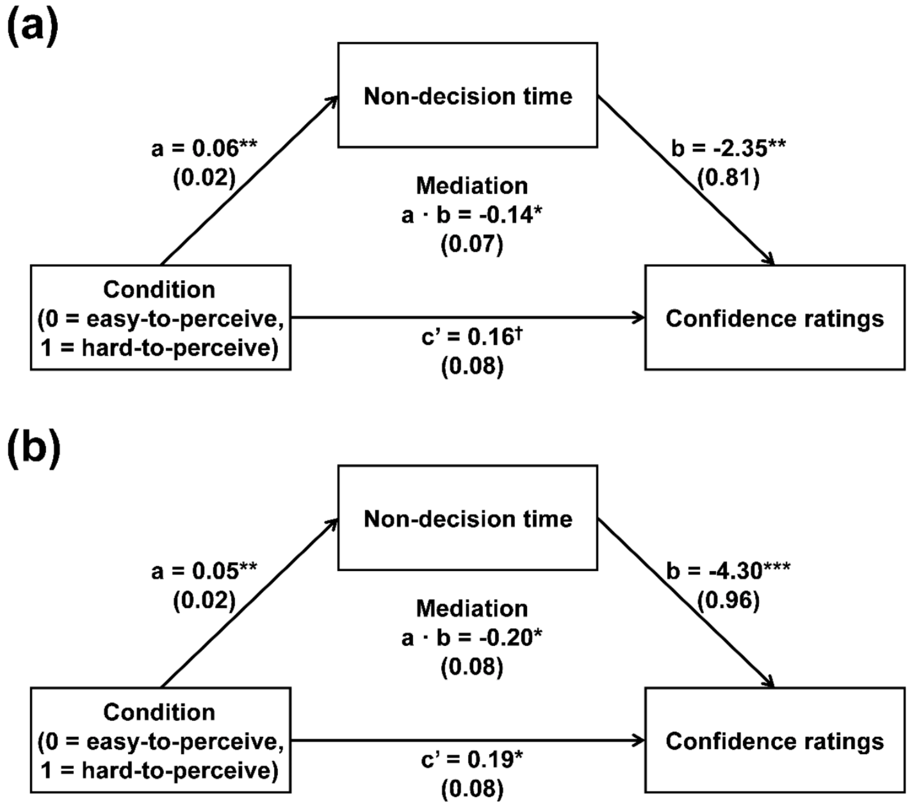

Fig. 5 The mediation effect of non-decision time T_{er} on the relationship between experimental condition and retrospective confidence ratings in Experiments 1a (a) and 1b (b). Standard errors are reported in parentheses

Discussion

Experiments 1a and 1b replicated the results from our analysis on previous published experiments. Specifically, Experiments 1a and 1b showed that retrospective confidence ratings were higher when response time in recognition test was shorter. Although response time could reliably predict memory performance in recognition test, the reliance of confidence on response time went beyond the effect of memory performance. In addition, higher confidence ratings were associated with higher drift rate and shorter non-decision time across trials. We also found that basing confidence on drift rate increased confidence accuracy, while giving confidence ratings based on non-decision time negatively affected confidence accuracy.

Furthermore, Experiments 1a and 1b implemented different manipulations of the ease of perception for the words presented in recognition test, and revealed that the ease of perceiving the words significantly affected non-decision time. Participants took longer time to recognize the content of words when perception of words was more difficult, leading to longer overall non-decision time (Yang et al., 2018). However, the boundary separation and drift rate did not reliably differ between easy- and hard-to-perceive condition, suggesting the ease of perception did not affect the retrieval process during recognition test.

Although results in the two experiments revealed a statistically significant effect of ease of perception on non-decision time, the overall response time in recognition test did not reliably differ between easy- and hard-to-perceive conditions. One possibility is that overall response time might not be sensitive to perceptual manipulation because it might also be influenced by other cognitive process. Previous studies have shown that while font size of stimuli affects the response time in continuous identification task which mainly measures the speed of perception, self-paced study time did not differ across different font sizes because study time also depends heavily on conceptual processing (Chang & Brainerd, 2022; Mueller et al., 2014; Yang et al., 2018). Similarly, overall response time in recognition test is closely related to the time required for memory retrieval, and might not reliably reflect the difference in ease of perception across experimental conditions. Furthermore, it is also possible that participant might limit the maximum amount of time they were willing to spend on each trial during recognition, or adjusted their decision boundary as time went on (Ackerman, 2014; Ratcliff et al., 2016). Future studies should try to explore these possibilities.

In the two experiments, the influence of perceptual manipulation on mean confidence did not reach statistical significance, even when non-decision time was reliably affected by the ease of perception. This result was different from those in some previous studies, which reveal that increasing the ease of perception often leads to prediction of higher overall performance in memory test (e.g., Busey et al., 2000; Undorf et al., 2017; Yang et al., 2018; Yue et al., 2013; but see Sungkhasettee et al., 2011; Thompson et al., 2013 for different results). Further mediation analysis indicated that the relationship between ease of perception and retrospective confidence ratings in Experiments 1a and 1b was complicated, and there might be two opposite effects of the ease of perception on confidence ratings. One is the indirect effect through non-decision time: difficulty in perceiving the stimuli increased non-decision time, which was utilized as one of the cues during retrospective confidence process, and led to low confidence ratings. This result is consistent with previous studies suggesting that high level of difficulty in stimulus identification reduces the speed of perceptual process, which then makes people predict they will have low memory performance (Undorf et al., 2017; Wang et al., 2020; Yang et al., 2018).

The other effect of perceptual manipulation on retrospective confidence revealed in the mediation analysis was the direct effect in the opposite direction, in which difficulty in stimulus perception was related to high confidence. One possibility is that the amount of effort required in recognition test might affect retrospective confidence ratings. For example, in the study by Turner et al., (2021), participants gave higher retrospective confidence in perceptual decision making task when they invested more motor effort in reporting the decision. Turner et al. suggest that high level of effort invested into the decision might be interpreted post-hoc as a signal that the decision is likely to be correct. Similarly, more difficulty in stimulus perception requires more effort for recognizing the words, which might then be interpreted as higher probability of choosing the correct answer.

General discussion

Previous studies consistently showed that higher retrospective confidence ratings were associated with shorter response time in memory test (Kelley & Lindsay, 1993; Koriat & Ackerman, 2010; Robinson et al., 1997; Siedlecka et al., 2019). Researchers often interpret response time as a measure of retrieval fluency, which is an important cue utilized in retrospective confidence process (Koriat et al., 2008; Koriat & Ackerman, 2010; Robinson et al., 1997). However, DDM, which is a widely applied model concerning decision making, indicates that response time in recognition memory test contains both a decision (i.e., retrieval) and a non-decision component, and does not solely reflect retrieval fluency (Ratcliff & McKoon, 2008). Few previous studies have investigated whether retrospective confidence is affected by the speed of both retrieval and non-decision process in recognition test. Thus, the current study aimed to fill this gap.

Based on the analysis of data from six published experiments, our results revealed that higher retrospective confidence ratings were associated with both higher speed of accumulating memory evidence (i.e., drift rate) and shorter non-decision time in recognition test. Furthermore, the reliance of confidence on drift rate strengthened the positive relationship between confidence and memory performance, while the reliance of confidence on non-decision time negatively affected confidence accuracy. We then manipulated the ease of perceiving the stimuli in Experiments 1a and 1b to directly influence non-decision time in recognition test. Results indicated that more difficulty in stimulus perception led to longer non-decision time, which could then result in lower retrospective confidence ratings. However, there might also be another opposite effect of the ease of perception on retrospective confidence, in which difficulty in stimulus perception increased confidence ratings.

Consistent with our hypothesis, retrospective confidence ratings in recognition test were positively related to drift rate across trials. In addition, the reliance of confidence on drift rate could partly account for the positive relationship between confidence and memory performance. According to DDM, drift rate directly reflects the speed of retrieving information from memory storage in recognition test, and affects both the time required for memory retrieval and actual memory performance (Ratcliff, 1978; Ratcliff & McKoon, 2008). Thus, compared with the total response time, drift rate should be a better measure of retrieval fluency in recognition test. Our results supported previous theoretical accounts concerning retrospective confidence ratings, which suggest that retrieval fluency in memory test is one of the key factors affecting confidence process (Koriat et al., 2008; Koriat & Ackerman, 2010).

We also observed that retrospective confidence ratings were affected by non-decision time in recognition test. Furthermore, Experiments 1a and 1b revealed that the ease of perception could affect confidence ratings through the change of non-decision time (although there might be another direct effect on confidence in the opposite direction). These results suggested that besides retrieval fluency, the speed of non-decision process might also be an important cue utilized during confidence process. In addition, our data analysis indicated that the reliance of confidence on non-decision time could negatively affect confidence accuracy. This result is consistent with the assumption of DDM: non-decision time only influences the total amount of response time but not memory performance (Ratcliff & McKoon, 2008). Thus, basing retrospective confidence on non-decision time would make confidence ratings unable to appropriately reflect the variation of memory performance across trials.

Although our results indicated the influence of non-decision time on retrospective confidence ratings, there is no clear conclusion about why non-decision time is utilized as a cue during confidence process. Experiments 1a and 1b suggest that the effect of non-decision time on retrospective confidence ratings may be due to the processing experience obtained from non-decision process such as perception: participants could experience the variation in the ease of perception across trials, which might lead to change in the subjective feeling of knowing the answer, and then affect retrospective evaluation of memory performance (cf. Koriat et al., 2004, 2008). However, there is another possibility that non-decision time may affect confidence due to people' prior beliefs about memory process (Mueller et al., 2014; Mueller & Dunlosky, 2017). For example, participants might hold a belief that long response time is more likely to be associated with incorrect response in recognition test. Based on this belief, participants might give lower confidence when they observed longer response time during recognition, regardless of whether the increasing of response time came from retrieval or non-decision process. Future studies should try to examine the role of prior beliefs in the effect of non-decision time on confidence ratings.

The current study illustrated the advantage of decomposing response time in recognition test based on DDM. For example, when investigating the influence of the ease of perception on metamemory, previous studies often look at the change in total response time to validate the perceptual manipulation (Thompson et al., 2013; Yang et al., 2018). However, in our Experiments 1a and 1b, while the effect of perceptual manipulation on total response time was not statistically detectable, non-decision time in DDM reliably differ between easy- and hard-to-perceive condition. These results suggest that compared with total response time, non-decision time is more sensitive to (and should more directly reflect) the manipulations of the ease of perception. In addition, decomposing the response time based on DDM can also reflect which specific process (retrieval or non-decision process) in recognition test is affected by experimental manipulations (Ratcliff & McKoon, 2008). We suggest that future studies should use DDM to decompose response time and to directly quantify the effect of retrieval fluency and non-decision time on confidence ratings.

There have been a few studies exploring the potential effect of the ease of perception on retrospective confidence ratings (Busey et al., 2000; Thompson et al., 2013). For example, Busey et al., (2000) revealed that when face pictures were used as materials in recognition test, lower luminance of pictures were associated with lower retrospective confidence. In contrast, Thompson et al., (2013) showed that while difficulty in perceiving the words (induced by irregular font style) increased response time in reasoning tasks, retrospective confidence was not affected by the ease of perception. The main difference between the current study and previous studies is that previous studies did not examine whether non-decision time mediated the influence of the ease of perception on confidence, and did not reveal two opposite effects of perceptual manipulation as in our Experiments 1a and 1b. Future studies should further investigate whether the ease of perceiving the stimuli affects confidence with different materials and manipulations, and explore the role of non-decision time in this effect.

The current study used DDM to examine how participants monitored and evaluated their memory performance based on cues (retrieval fluency and non-decision time) obtained from recognition test. Future studies could consider using DDM to investigate the cognitive mechanisms underlying metamemory control process. For example, previous theories have provided different assumptions regarding how people decide when to stop their study during self-regulated learning. The Discrepancy Reduction Model suggests that people set a constant criterion of learning, and they continue to study until the perceived degree of learning exceeds the criterion (Nelson & Narens, 1990). Thus, more difficult items generally require longer study time. In contrast, A Region of Proximal Learning Model indicates that people's learning criterion relies on their perseverance, and they spend more time studying easier unknown items than the most difficult ones (Metcalfe & Kornell, 2005). A third theory, the Diminishing Criterion Model, assumes that learning criterion reduces with the increase of study time (Ackerman, 2014). In the framework of DDM, the differences between the three theories above may be represented by the setting of decision boundary: the boundary for the decision about when to stop learning should be fixed in Discrepancy Reduction Model, while the decision boundary may depend on item difficulty according to A Region of Proximal Learning Model. On the other hand, the Diminishing Criterion Model may be related to collapse boundary DDM, which suggests the decision boundaries for different choices collapse over time (Ratcliff et al., 2016). Future studies should try to use quantitative model comparison based on DDM to examine which theory best fits people's self-regulated learning behavior.

In order to investigate the effect of both drift rate $\boldsymbol{v}$ and non-decision time $\boldsymbol{T_{er}}$ on retrospective confidence, the current study estimated the value of $\boldsymbol{v}$ and $\boldsymbol{T_{er}}$ separately in each trial, and regressed trial-by-trial confidence ratings on both $\boldsymbol{v}$ and $\boldsymbol{T_{er}}$ at the same time. Thus, the reliability of the results in this study depends on whether the effect of single-trial $\boldsymbol{v}$ and $\boldsymbol{T_{er}}$ on confidence could be independently estimated in the model. To address this question, we first combined data from all of the eight experiments (including six published experiments and Experiments 1a-1b) to examine the relationship between single-trial estimation of $\boldsymbol{v}$ and $\boldsymbol{T_{er}}$. Results from multilevel linear regression model revealed that $\boldsymbol{v}$ could be negatively predicted by $\boldsymbol{T_{er}}$ within each participant (and vice versa), bs < -0.01, ts > 3.53, ps < 0.005, η_p^2 > 0.63, suggesting the estimated values of single-trial $\boldsymbol{v}$ and $\boldsymbol{T_{er}}$ were correlated. We then performed a data simulation (see Section S4 in Supplemental Materials), which indicated that there tended to be a small-to-moderate negative correlation between the estimated values of single-trial $\boldsymbol{v}$ and $\boldsymbol{T_{er}}$ when the true values of the two parameters did not correlate with each other. This result suggests that the estimation of single-trial $\boldsymbol{v}$ and $\boldsymbol{T_{er}}$ might not be fully independent. However, further analysis showed that the overall effect of $\boldsymbol{v}$ and $\boldsymbol{T_{er}}$ on confidence ratings could be successfully recovered in data simulation. Thus, although we need to be cautious when interpreting single-trial estimated value of $\boldsymbol{v}$ and $\boldsymbol{T_{er}}$, the overall positive effect of $\boldsymbol{v}$ and negative effect of $\boldsymbol{T_{er}}$ on confidence ratings found in the current study should be reliable.

In the current study, we only examined the effect of the ease of perception on non-decision time. Other factors, such as familiarity of words (Hanczakowski, Pasek, et al., 2013; Kelley & Lindsay, 1993), could also affect the time required for perceiving the words. In addition, according to DDM, non-decision time contains not only the perception of stimuli, but also the execution of response (Ratcliff & McKoon, 2008). Thus, the ease of motoric response may also affect non-decision time in recognition test (Susser et al., 2017; Susser & Mulligan, 2015). Furthermore, motoric response speed has been shown to affect retrospective confidence in perceptual decision tasks (Palser et al., 2018; Patel et al., 2012). Future studies should explore whether familiarity and motoric response can affect retrospective confidence ratings in recognition test through non-decision time.

Another limitation of the current study is that we only examined the effect of drift rate and non-decision time on retrospective confidence in 2-choice recognition test, but not in other test formats such as recall test or multiple-choice test with more than 2 options (Nelson & Narens, 1990; Robinson et al., 1997). However, DDM can only be applied to 2-choice tasks, and is not suitable for more complicated test format (Ratcliff & McKoon, 2008). Future studies should try to decompose the response time in multiple-choice and recall test based on other computational models (such as the linear ballistic accumulator model for multiple-choice test; see Brown & Heathcote 2008), and investigate whether drift rate and non-decision time could affect retrospective confidence ratings.

Conclusions

Retrospective confidence ratings in recognition test are affected by both drift rate which represents retrieval fluency, and non-decision time which reflects the required time for cognitive process other than memory retrieval. Furthermore, the reliance of confidence on drift rate could partly account for the positive relationship between confidence and memory performance, while basing confidence on non-decision time negatively affects confidence accuracy.

Acknowledgements This study was supported by the Natural Science Foundation of China (32171045, 32000742).

Data and code availability.

Data and analysis code for the current study are available at Open Science Framework (https://osf.io/kd3f8/).

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

References

Ackerman, R. (2014). The diminishing criterion model for metacognitive regulation of time investment. *Journal of Experimental Psychology: General* (143 vol., pp. 1349–1368). American Psychological Association. 3https://doi.org/10.1037/a0035098

Anderson, J. R. (1981). Interference: The relationship between response latency and response accuracy. *Journal of Experimental Psychology: Human Learning and Memory*, 7(5), 326–343. https://doi.org/10.1037/0278-7393.7.5.326

Brown, S. D., & Heathcote, A. (2008). The simplest complete model of choice response time: Linear ballistic accumulation. *Cognitive Psychology*, 57(3), 153–178. https://doi.org/10.1016/J.COGPSYCH.2007.12.002

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

Chang, M., & Brainerd, C. J. (2022). Association and dissociation between judgments of learning and memory: A Meta-analysis of the font size effect. *Metacognition and Learning*, 1–34. https://doi.org/10.1007/S11409-021-09287-3

Coltheart, M. (1981). The MRC psycholinguistic database. *The Quarterly Journal of Experimental Psychology Section A*, 33(4), 497–505. https://doi.org/10.1080/14640748108400805

Cousineau, D. (2005). Confidence intervals in within-subject designs: A simpler solution to Loftus and Masson's method. *Tutorials in Quantitative Methods for Psychology*, 1(1), 42–45. https://doi.org/10.20982/tqmp.01.1.p042

Diemand-Yauman, C., Oppenheimer, D. M., & Vaughan, E. B. (2011). Fortune favors the bold (and the italicized): Effects of disfluency on educational outcomes. *Cognition*, 118(1), 111–115. https://doi.org/10.1016/j.cognition.2010.09.012

Dougherty, M. R., Scheck, P., Nelson, T. O., & Narens, L. (2005). Using the past to predict the future. *Memory & Cognition*, 33(6), 1096–1115. https://doi.org/10.3758/BF03193216

Dougherty, M. R., Robey, A. M., & Buttaccio, D. (2018). Do metacognitive judgments alter memory performance beyond the benefits of retrieval practice? A comment on and replication attempt of Dougherty, Scheck, Nelson, and Narens (2005). *Memory & Cognition*, *46*(4), 558–565. https://doi.org/10.3758/s13421-018-0791-y

Enders, C. K., & Tofighi, D. (2007). Centering Predictor Variables in Cross-Sectional Multilevel Models: A New Look at an Old Issue. *Psychological Methods*, 12(2), 121–138. https://doi.org/10.1037/1082-989X.12.2.121

Fleming, S. M., Ryu, J., Golfinos, J. G., & Blackmon, K. E. (2014). Domain-specific impairment in metacognitive accuracy following anterior prefrontal lesions. *Brain*, 137(10), 2811–2822. https://doi.org/10.1093/brain/awu221

Geller, J., & Peterson, D. (2021). Is this going to be on the test? Test expectancy moderates the disfluency effect with sans forgetica. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. https://doi.org/10.1037/XLM0001042

Gelman, A., & Rubin, D. B. (1992). Inference from Iterative Simulation Using Multiple Sequences. *Statistical Science*, 7(4), 457–472. https://doi.org/10.1214/ss/1177011136

Hanczakowski, M., Pasek, T., Zawadzka, K., & Mazzoni, G. (2013a). Cue familiarity and 'don't know' responding in episodic memory tasks. Journal of Memory and Language, 69(3), 368–383. https://doi.org/10.1016/J.JML.2013a.04.005

Hanczakowski, M., Zawadzka, K., Pasek, T., & Higham, P. A. (2013b). Calibration of metacognitive judgments: Insights from the underconfidence-with-practice effect. Journal of Memory and Language, 69(3), 429–444. https://doi.org/10.1016/J.JML.2013b.05.003

Hayes, A. F., & Rockwood, N. J. (2020). Conditional Process Analysis: Concepts, Computation, and Advances in the Modeling of the Contingencies of Mechanisms. *American Behavioral Scientist*, 64(1), 19–54. https://doi.org/10.1177/0002764219859633

Hu, X., Liu, Z., Chen, W., Zheng, J., Su, N., Wang, W. … Luo, L. (2017). Individual differences in the accuracy of judgments of learning are related to the gray matter volume and functional connectivity of the left mid-insula. *Frontiers in Human Neuroscience*, 11, 399. https://doi.org/10.3389/fnhum.2017.00399

Hu, X., Luo, L., & Fleming, S. M. (2019). A role for metamemory in cognitive offloading. *Cognition*, 193, 104012. https://doi.org/10.1016/J.COGNITION.2019.104012

Kantner, J., & Lindsay, D. S. (2014). Cross-situational consistency in recognition memory response bias. *Psychonomic Bulletin & Review*, 21(5), 1272–1280. https://doi.org/10.3758/S13423-014-0608-3

Kelley, C. M., & Lindsay, D. S. (1993). Remembering Mistaken for Knowing: Ease of Retrieval as a Basis for Confidence in Answers to General Knowledge Questions. *Journal of Memory and Language*, 32(1), 1–24. https://doi.org/10.1006/JMLA.1993.1001

Kim, S. (2015). ppcor: An R Package for a Fast Calculation to Semi-partial Correlation Coefficients. *Communications for Statistical Applications and Methods*, 22(6), 665–674. https://doi.org/10.5351/CSAM.2015.22.6.665

Koriat, A. (1997). Monitoring one's own knowledge during study: A cue-utilization approach to judgments of learning. *Journal of Experimental Psychology: General*, 126(4), 349–370. https://doi.org/10.1037/0096-3445.126.4.349

Koriat, A., Bjork, R. A., Sheffer, L., & Bar, S. K. (2004). Predicting one's own forgetting: The role of experience-based and theory-based processes. *Journal of Experimental Psychology: General*, 133(4), 643–656. https://doi.org/10.1037/0096-3445.133.4.643

Koriat, A., Nussinson, R., Bless, H., & Shaked, N. (2008). Information-based and experience-based metacognitive judgments: Evidence from subjective confidence. In J. Dunlosky & R. A. Bjork (Eds.), *Handbook of Metamemory and Memory* (pp. 117–135). https://doi.org/10.4324/9780203805503

Koriat, A., & Ackerman, R. (2010). Choice latency as a cue for children's subjective confidence in the correctness of their answers. *Developmental Science*, 13(3), 441–453. https://doi.org/10.1111/J.1467-7687.2009.00907.X

Lee, M. D., & Wagenmakers, E. J. (2013). *Bayesian cognitive modeling: A practical course*. Cambridge University Press. https://doi.org/10.1017/CBO9781139087759

Masson, M. E., & Rotello, C. M. (2009). Sources of bias in the Goodman-Kruskal gamma coefficient measure of association: implications for studies of metacognitive processes. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 35(2), 509–527. https://doi.org/10.1037/a0014876

McCurdy, L. Y., Maniscalco, B., Metcalfe, J., Liu, K. Y., de Lange, F. P., & Lau, H. (2013). Anatomical coupling between distinct metacognitive systems for memory and visual perception. *Journal of Neuroscience*, 33(5), 1897–1906. https://doi.org/10.1523/JNEUROSCI.1890-12.2013

Metcalfe, J., & Finn, B. (2008). Evidence that judgments of learning are causally related to study choice. *Psychonomic Bulletin and Review*, 15(1), 174–179. https://doi.org/10.3758/PBR.15.1.174

Metcalfe, J., & Kornell, N. (2005). A Region of Proximal Learning model of study time allocation. *Journal of Memory and Language*, 52(4), 463–477. https://doi.org/10.1016/j.jml.2004.12.001

Mueller, M. L., Dunlosky, J., Tauber, S. K., & Rhodes, M. G. (2014). The font-size effect on judgments of learning: Does it exemplify fluency effects or reflect people's beliefs about memory? *Journal of Memory and Language*, 70, 1–12. https://doi.org/10.1016/j.jml.2013.09.007

Mueller, M. L., & Dunlosky, J. (2017). How beliefs can impact judgments of learning: Evaluating analytic processing theory with beliefs about fluency. *Journal of Memory and Language*, 93, 245–258. https://doi.org/10.1016/j.jml.2016.10.008

Murayama, K., Usami, S., & Sakaki, M. (2020). *Summary-statistics-based power analysis: A new and practical method to determine sample size for mixed-effects modelling*. https://doi.org/10.31219/OSF.IO/6CER3

Nelson, T. O., & Narens, L. (1990). Metamemory: A theoretical framework and new findings. In H. B. Gordon (Ed.), *Psychology of Learning and Motivation: Vol. Volume 26* (pp. 125–173). Academic Press. https://doi.org/10.1016/S0079-7421(08)60053-5

Nunez, M. D., Vandekerckhove, J., & Srinivasan, R. (2017). How attention influences perceptual decision making: Single-trial EEG correlates of drift-diffusion model parameters. *Journal of Mathematical Psychology*, 76, 117–130. https://doi.org/10.1016/J.JMP.2016.03.003

Palmer, E. C., David, A. S., & Fleming, S. M. (2014). Effects of age on metacognitive efficiency. *Consciousness and Cognition*, 28, 151–160. https://doi.org/10.1016/j.concog.2014.06.007

Palser, E. R., Fotopoulou, A., & Kilner, J. M. (2018). Altering movement parameters disrupts metacognitive accuracy. *Consciousness and Cognition*, 57, 33–40. https://doi.org/10.1016/J.CONCOG.2017.11.005

Patel, D., Fleming, S. M., & Kilner, J. M. (2012). Inferring subjective states through the observation of actions. *Proceedings of the Royal Society B: Biological Sciences*, *279*(1748), 4853–4860. https://doi.org/10.1098/RSPB.2012.1847

Plummer, M. (2003). JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling. *Proceedings of the 3rd International Workshop on Distributed Statistical Computing*, 20–22. https://doi.org/10.1.1.13.3406

Rahnev, D., Desender, K., Lee, A. L. F., Adler, W. T., Aguilar-Lleyda, D., Akdoğan, B. … Zylberberg, A. (2020). The Confidence database. *Nature Human Behaviour*, 4(3), 317–325. https://doi.org/10.1038/s41562-019-0813-1

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

Ratcliff, R., McKoon, G., & Gomez, P. (2004). A Diffusion Model Account of the Lexical Decision Task. *Psychological Review*, 111(1), 159–182. https://doi.org/10.1037/0033-295X.111.1.159

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

Ratcliff, R., & Tuerlinckx, F. (2002). Estimating parameters of the diffusion model: Approaches to dealing with contaminant reaction times and parameter variability. *Psychonomic Bulletin & Review*, 9(3), 438–481. https://doi.org/10.3758/BF03196302

Ratcliff, R. (2013). Parameter variability and distributional assumptions in the diffusion model. *Psychological Review*, 120(1), 281–292. https://doi.org/10.1037/A0030775

Ratcliff, R., Smith, P. L., Brown, S. D., & McKoon, G. (2016). Diffusion Decision Model: Current Issues and History. *Trends in Cognitive Sciences*, 20(4), 260–281. https://doi.org/10.1016/J.TICS.2016.01.007

Robey, A. M., Dougherty, M. R., & Buttaccio, D. R. (2017). Making Retrospective Confidence Judgments Improves Learners' Ability to Decide What Not to Study. *Psychological Science*, 28(11), 1683–1693. https://doi.org/10.1177/0956797617718800

Robey, A. M., Castillo, C., Ha, J., Kerlow, M., Tesfa, N., & Dougherty, M. (2021). Generalizing the effect of type of metacognitive judgment on restudy decisions. *Metacognition and Learning*. https://doi.org/10.1007/s11409-021-09274-8

Robinson, M. D., Johnson, J. T., & Herndon, F. (1997). Reaction time and assessments of cognitive effort as predictors of eyewitness memory accuracy and confidence. *Journal of Applied Psychology*, 82(3), 416–425. https://doi.org/10.1037/0021-9010.82.3.416

Rosner, T. M., Davis, H., & Milliken, B. (2015). Perceptual blurring and recognition memory: A desirable difficulty effect revealed. *Acta Psychologica*, 160, 11–22. https://doi.org/10.1016/J.ACTPSY.2015.06.006

Sadeghi, S., Ekhtiari, H., Bahrami, B., & Ahmadabadi, M. N. (2017). Metacognitive deficiency in a perceptual but not a memory task in methadone maintenance patients. *Scientific Reports*, 7(1), 1–8. https://doi.org/10.1038/s41598-017-06707-w

Schmidt, C., Reyes, G., Barrientos, M., Langer, Á. I., & Sackur, J. (2019). Meditation focused on self-observation of the body impairs metacognitive efficiency. *Consciousness and Cognition*, 70, 116–125. https://doi.org/10.1016/j.concog.2019.03.001

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(4), 712–723. https://doi.org/10.1037/xlm0000608

Singmann, H., & Kellen, D. (2020). An introduction to mixed models for experimental psychology. In D. H. Spieler, & E. Schumacher (Eds.), *New Methods in Cognitive Psychology*. Routledge. https://doi.org/10.4324/9780429318405-2

Smith, R. L., Ager, J. W., & Williams, D. L. (1992). Suppressor Variables in Multiple Regression/Correlation. *Educational and Psychological Measurement*, 52(1), 17–29. https://doi.org/10.1177/001316449205200102

Su, Y. S., & Yajima, M. (2021). *R2jags: Using R to Run 'JAGS". R package version 0.7-1.* 'https://cran.r-project.org/package=R2jags

Sungkhasettee, V. W., Friedman, M. C., & Castel, A. D. (2011). Memory and metamemory for inverted words: Illusions of competency and desirable difficulties. *Psychonomic Bulletin and Review*, 18(5), 973–978. https://doi.org/10.3758/S13423-011-0114-9/FIGURES/3

Susser, J. A., & Mulligan, N. W. (2015). The effect of motoric fluency on metamemory. *Psychonomic Bulletin and Review*, 22, 1014–1019. https://doi.org/10.3758/s13423-014-0768-1

Susser, J. A., Panitz, J., Buchin, Z., & Mulligan, N. W. (2017). The motoric fluency effect on metamemory. *Journal of Memory and Language*, 95, 116–123. https://doi.org/10.1016/J.JML.2017.03.002

Thompson, V. A., Turner, J. A. P., Pennycook, G., Ball, L. J., Brack, H., Ophir, Y., & Ackerman, R. (2013). The role of answer fluency and perceptual fluency as metacognitive cues for initiating analytic thinking. *Cognition*, 128(2), 237–251. https://doi.org/10.1016/J.COGNITION.2012.09.012

Turner, W., Angdias, R., Feuerriegel, D., Chong, T. T. J., Hester, R., & Bode, S. (2021). Perceptual decision confidence is sensitive to forgone physical effort expenditure. *Cognition*, 207, 104525. https://doi.org/10.1016/J.COGNITION.2020.104525

Undorf, M., Zimdahl, M. F., & Bernstein, D. M. (2017). Perceptual fluency contributes to effects of stimulus size on judgments of learning. *Journal of Memory and Language*, 92, 293–304. https://doi.org/10.1016/j.jml.2016.07.003

Voss, A., Rothermund, K., & Voss, J. (2004). Interpreting the parameters of the diffusion model: An empirical validation. *Memory & Cognition*, 32(7), 1206–1220. https://doi.org/10.3758/BF03196893

Voss, A., Voss, J., & Lerche, V. (2015). Assessing cognitive processes with diffusion model analyses: a tutorial based on fast-dm-30. *Frontiers in Psychology*, 6, 336. https://doi.org/10.3389/FPSYG.2015.00336

Vuorre, M., & Bolger, N. (2018). Within-subject mediation analysis for experimental data in cognitive psychology and neuroscience. *Behavior Research Methods*, 50(5), 2125–2143. https://doi.org/10.3758/s13428-017-0980-9

Vuorre, M., & Metcalfe, J. (2021). Measures of relative metacognitive accuracy are confounded with task performance in tasks that permit guessing. *Metacognition and Learning*. https://doi.org/10.1007/S11409-020-09257-1. Advance online publication

Wabersich, D., & Vandekerckhove, J. (2014). Extending JAGS: A tutorial on adding custom distributions to JAGS (with a diffusion model example). *Behavior Research Methods*, 46(1), 15–28. https://doi.org/10.3758/S13428-013-0369-3

Wang, Z., Yang, C., Zhao, W., & Jiang, Y. (2020). Perceptual Fluency Affects Judgments of Learning Non-analytically and Analytically Through Beliefs About How Perceptual Fluency Affects Memory. *Frontiers in Psychology*, 11, 552824. https://doi.org/10.3389/fpsyg.2020.552824

Weidemann, C. T., & Kahana, M. J. (2016). Assessing recognition memory using confidence ratings and response times. *Royal Society Open Science*, 3, 150670. https://doi.org/10.1098/RSOS.150670

Whelan, R. (2008). Effective Analysis of Reaction Time Data. *The Psychological Record*, 58, 475–482. https://doi.org/10.1007/BF03395630

Wiecki, T. V., Sofer, I., & Frank, M. J. (2013). HDDM: Hierarchical Bayesian estimation of the Drift-Diffusion Model in Python. *Frontiers in Neuroinformatics*, 7, 14. https://doi.org/10.3389/FNINF.2013.00014

Yang, C., Huang, T. S. T., & Shanks, D. R. (2018). Perceptual fluency affects judgments of learning: The font size effect. *Journal of Memory and Language*, 99, 99–110. https://doi.org/10.1016/j.jml.2017.11.005

Yang, C., Yu, R., Hu, X., Luo, L., Huang, T. S. T., & Shanks, D. R. (2021). How to assess the contributions of processing fluency and beliefs to the formation of judgments of learning: methods and pitfalls. *Metacognition and Learning*, 16, 319–343. https://doi.org/10.1007/s11409-020-09254-4

Yoon, H. D., Shin, M., & Jeon, H. A. (2021). The critical role of interference control in metaphor comprehension evidenced by the drift–diffusion model. *Scientific Reports*, 11(1), 1–16. https://doi.org/10.1038/s41598-021-98351-8

Yue, C. L., Castel, A. D., & Bjork, R. A. (2013). When disfluency is-and is not-a desirable difficulty: The influence of typeface clarity on metacognitive judgments and memory. *Memory and Cognition*, 41(2), 229–241. https://doi.org/10.3758/s13421-012-0255-8

Zawadzka, K., & Higham, P. A. (2015). Judgments of learning index relative confidence, not subjective probability. *Memory and Cognition*, 43(8), 1168–1179. https://doi.org/10.3758/s13421-015-0532-4

Zhang, Z., Zyphur, M. J., & Preacher, K. J. (2009). Testing multilevel mediation using hierarchical linear models: Problems and solutions. *Organizational Research Methods*, 12(4), 695–719. https://doi.org/10.1177/1094428108327450