UNIVERSITY OF AMSTERDAM

UvA-DARE (Digital Academic Repository)

What Happens After a Fast Versus Slow Error, and How Does It Relate to Evidence Accumulation?

Damaso, K.A.M.; Williams, P.G.; Heathcote, A.

DOI
10.1007/s42113-022-00137-2

Publication date
2022

Document Version
Final published version

Published in
Computational Brain and Behavior

License
CC BY

Link to publication

Citation for published version (APA):
Damaso, K. A. M., Williams, P. G., & Heathcote, A. (2022). What Happens After a Fast Versus Slow Error, and How Does It Relate to Evidence Accumulation? *Computational Brain and Behavior*, *5*(4), 527-546. https://doi.org/10.1007/s42113-022-00137-2

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: https://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, P.O. Box 19185, 1000 GD Amsterdam, The Netherlands. You will be contacted as soon as possible.

UvA-DARE is a service provided by the library of the University of Amsterdam (https://dare.uva.nl)

Download date:16 Oct 2026

Computational Brain & Behavior (2022) 5:527–546
https://doi.org/10.1007/s42113-022-00137-2

ORIGINAL PAPER

What Happens After a Fast Versus Slow Error, and How Does It Relate to Evidence Accumulation?

Karlye A. M. Damaso[1] · **Paul G. Williams**[1,2] · **Andrew Heathcote**[1]

Accepted: 22 April 2022 / Published online: 23 November 2022
© The Author(s) 2022

Abstract

It has traditionally been assumed that responding after an error is slowed because participants try to improve their accuracy by increasing the amount of evidence required for subsequent decisions. However, recent work suggests a more varied picture of post-error effects, with instances of post-error speeding, and decreases or no change in accuracy. Further, the causal role of errors in these effects has been questioned due to confounds from slow fluctuations in attention caused by factors such as fatigue and boredom. In recognition memory tasks, we investigated both post-error speeding associated with instructions emphasising fast responding and post-error slowing associated with instructions emphasising the accuracy of responding. In order to identify the causes of post-error effects, we fit this data with evidence accumulation models using a method of measuring post-error effects that is robust to confounds from slow fluctuations. When the response-to-stimulus interval between trials was short, there were no post-error effect on accuracy and speeding and slowing were caused by differences in non-decision time (i.e. the time to encode choice stimuli and generate responses). In contrast, when the interval was longer, due to participants providing a confidence rating for their choice, there were also effects on the rate of evidence accumulation and the amount of evidence required for a decision. We discuss the implications of our methods and results for post-error effect research.

Keywords Post-error effects · Post-error slowing · Evidence accumulation models

"What does a man do after he makes an error?" (Rabbitt & Rodgers, 1977). This question has driven a burgeoning area of research since it was first asked over 40 years ago. The classic answer has been that following an error a person takes more care by slowing response time (RT) — an empirical regularity known as *post-error slowing* (Rabbitt, 1966).

Slowing responses to avoid errors relies on a speed-accuracy tradeoff mechanism (Schouten & Bekker, 1967), a concept central to evidence accumulation models (EAMs; for review, see Donkin & Brown, 2018). EAMs describe the process of (usually rapid) choice, assuming evidence favouring different options accumulates over time until one reaches a threshold amount and the associated response is triggered. EAMs are assessed by their ability to accommodate the probabilities of correct and error responses and the shapes of corresponding RT distributions. Model parameters are able to provide a coherent and psychologically meaningful characterization of the often-complex relationships between RT and accuracy and a range of benchmark effects, including emphasis placed on speed versus accuracy, and factors that affect choice difficulty (e.g. Brown & Heathcote, 2008; Donkin & Brown, 2018; Leite & Ratcliff, 2010; Luce, 1986).

EAM parameters represent latent cognitive variables characterising the decision-making process and its inputs. These include the rate evidence is accumulated (*drift rate*), the amount of evidence required to make a decision (*threshold*) (which quantifies response caution), and the difference between the starting point of accumulation and the threshold for different choice options (*response bias*). They also provide an estimate of the total amount of time it takes for non-decision processes like stimulus encoding and response execution — *non-decision time*. To capture the full pattern of findings from choice experiments, particularly the speed of correct versus error responses, some EAMs such as the drift diffusion model (DDM, Ratcliff, 1978; Ratcliff &

✉ Andrew Heathcote
a.j.heathcote@uva.nl

[1] Department of Psychology, University of Amsterdam, Amsterdam, North Holland, The Netherlands

[2] Health Point, Mingara, Tumbi Umbi, New South Wales, Australia

McKoon, 2008) and the linear ballistic accumulator (LBA, Brown & Heathcote, 2008) also incorporate parameters that allow random variability from trial-to-trial in response bias and drift rate. Under an EAM framework, the classic account of post-error adjustments corresponds to a post-error increase in response caution, causing more accurate and slower responding.

In this paper, we use EAM-based analyses to investigate both the classic account and newer proposals that aim to explain post-error effects. Our aim is to explain the consequence of making an error on subsequent performance and explain them in terms of latent cognitive variables operationalized through EAM model parameters. In the next section, we briefly review the classic and newer accounts, and in the following section, we review previous applications of EAMs to investigate these accounts, all of which have used the DDM. We particularly focus on the work of Dutilh et al. (2013), who control for potential confounds to interpreting response accuracy as causal (i.e. consequential rather than merely associated) for differences in post-error versus post-correct performance by analysing only a subset of trials. We then review recent work showing that EAMs with between-trial variability in drift rates predict two types of errors that occur with different frequencies depending on whether decision speed or accuracy is emphasised and empirical work showing that these two types of errors have different post-error consequences.

We then report the results of an EAM analysis that uses the trial-subset selection approach as Dutilh et al. (2013) to identify the causes of post-error slowing in two recognition memory experiments. Both experiments manipulate speed vs. accuracy emphasis, with the first having a short response to stimulus interval (Rae et al., 2014) and the second a much longer intervals because after making a recognition decision, participants rated their response confidence (Osth et al., 2017). Our analysis differs from Dutilh et al.'s in two ways. First, we used models with between trial drift rate variability so that we can account for different types of errors. In contrast, Dutilh et al. used a version of the DDM with no between trial variability (Wagenmakers Wagenmakers et al., 2007) that is unable to account for different types of errors. Second, as well as fitting the DDM, as this has been exclusively used in previous work on post-error effects, we also fit the LBA. This allowed us to compare the relative ability of the two models to fit post-error effects and to determine if they provide the same or different accounts of the causes of these effects.

To foreshadow, we made two main findings. First, the LBA provided a more accurate account than the DDM, leading us to focus on the explanation of post-error effects provided by the LBA. Second, we observed that post-error effects and their underlying causes as identified by the LBA depend on the interactive effects of instructions emphasising fast vs. accurate responding and what occurs between each trial. Although the aim of this paper, and the scope data we analyse, do not allow us to adjudicate between the DDM and LBA in general, potential reasons for the differences between the models in the present data are addressed in the discussion.

Theories of Post-Error Performance

The *classic account* explains post-error slowing as adaptive control of error rates by increasing response caution (Rabbitt, 1966). However, this account is increasingly contested (e.g. Dutilh et al., 2012a). One drawback that has been acknowledged for some time (Laming, 1979) is that it cannot accommodate the *decrease* in post-error accuracy that is commonly found with easy discriminations and very short inter-trial intervals (0.02–0.22 ms). The classic account of post-error adjustments also fails to accommodate more recent findings that post-error *speeding* occurs in some contexts (Damaso et al., 2020; Notebaert et al., 2009; Purcell & Kiani, 2016; Williams et al., 2016). Several alternative accounts have been offered that attempt to accommodate the various combinations of post-error adjustments for speed and accuracy (for reviews, see Alexander & Brown, 2010; Ullsperger et al., 2014; van Veen & Carter, 2006). Wessel (2018) categorized these accounts as a function of whether they predicted post-error accuracy increases ("adaptive") or decreases ("maladaptive"). Adaptive theories assume errors trigger processes that ultimately aim to prevent future errors. Maladaptive theories assume processes triggered by errors have deleterious effects on subsequent cognitive processing. A limitation of Wessel's taxonomy is that it considers adaptivity only in terms of accuracy even though making optimal decisions sometimes requires balancing speed and accuracy (Bogacz et al., 2006).

According to Wessel's (Wessel, 2018) taxonomy, the classic account of post-error slowing (Rabbitt, 1966) is an example of an adaptive theory and is arguably the most prominent. Support for the classic account is abundant. For example, where post-error improvements in accuracy occur, its degree correlates with the magnitude of post-error slowing (e.g. Hajcak et al., 2003), and the magnitude of post-error slowing is associated with error awareness and conscious detection of errors (e.g. Nieuwenhuis et al., 2001). Another early adaptive theory suggested that the commission of an error leads to a delay in the start of evidence accumulation on subsequent trials (Laming, 1968). This prevents the sampling of non-discriminative evidence before the task-relevant stimulus is presented, avoiding errors due to the bias induced by such premature sampling. In line with this suggestion, neuroscientific evidence supports post-error

tuning of sensory processing (e.g. Danielmeier et al., 2011; King et al., 2010).

The most prominent maladaptive theory is possibly the *orienting* account (Notebaert et al., 2009). It suggests that the usual rarity of errors makes them novel, which elicits an automatic orientation of attention. This re-orientation is argued to slow responses and reduce resources available for subsequent processing, leading to a decrement in accuracy on the next trial. Consistent with this explanation, in a task where errors are more common than correct responses, post-error speeding has been found, reverting to the usual post-error slowing pattern when errors were rare (Notebaert et al., 2009). Because errors were always associated with a subsequent decrement in accuracy, the orienting account is considered maladaptive by Wessel's taxonomy — even in the case where the decrement in accuracy is linked to post-error speeding.

Several recent accounts have suggested that errors can elicit affective responses and that the presence and subsequent processing of this effect impacts upon both the resources available to complete a task and the way in which the task is completed (e.g. Dutilh et al., 2013; White et al., 2010; Williams et al., 2016). The specific affective response dictates whether the triggered process increases or reduces accuracy. Dutilh et al. (2013) suggested that participants experience negative affect following an error, and subsequently, they are more hesitant on future trials in order to reduce error rates and thus mitigate the unpleasant affect. Williams et al. (2016) suggested that participants may become despondent following an error, particularly if previous efforts at error remediation have been unsuccessful and/or led to very slow responses. They then react by becoming reckless following an error, prioritising speed over accuracy.

Although Dutilh et al.'s (2013) and Williams et al.'s (2016) accounts both focus on affect as a key behavioural determinant, they predict quite different accuracy outcomes. As defined by Wessel (2018), they might be considered adaptive and maladaptive, respectively. However, this may not always be the case, such as when rewards associated with making correct decisions are only available for a limited period of time, and so slow decisions come at an opportunity cost. Bogacz et al. (2006) studied such "reward-rate maximization" paradigms from the perspective of a simplified version of the DDM with no between-trial variability (Stone, 1960). In this simple DDM, accuracy always approaches 100% with increasing thresholds because the effect of the only source of noise causing errors, moment-to-moment variability in the evidence total, is reduced to zero by increasing accumulation. However, the cost in time for improving accuracy also increases rapidly, so that if thresholds are set too high, it can become optimal to lower them to speed up responding if that provides more opportunities for future correct responses. More broadly, considering both speed and accuracy makes it clear an adaptive vs. maladaptive classification depends on what quantity is to be optimized, leaving room for individual differences, particularly where it is unclear to participants what quantity, if anything, they should focus on.

Explaining Post-Error Performance with Evidence Accumulation Modelling

Dutilh et al. (2012a) used the full DDM to compare competing explanations of post-error slowing, linking each explanation to specific EAM parameters. In data from a lexical decision task, they found that the observed pattern of post-error adjustments — slowing with an increase in accuracy — could be attributed exclusively to an increase in threshold. This provided support for the classic account of post-error slowing. White et al. (2010) also reported evidence of post-error increases in threshold when modelling data from a recognition memory task with the full DDM, but also found a post-error decrease in non-decision time. A threshold increase was reported by Purcell and Kiani (2016) applying the simple DDM to a random-dot motion task with both human and monkey participants and by Schiffler et al. (2017) in a visual search task with a very large sample of people using a full DDM, but assuming all participants had identical trial-to-trial variability parameters. Schiffler et al. also reported a post-error decrease in non-decision time, and both they, Purcell and Kiani, reported a post-error decrease in drift rates. The drift rate decreases could be due to a reduction in the attentional resources available to complete task relevant behaviour, possibly due to processing related to detecting errors endogenously or through exogenous feedback (Ullsperger et al., 2014), or integration of affective consequences into the evidence used for subsequent decisions (Greifeneder et al., 2010; Roberts & Hutcherson, 2019).

A weakness of many of the EAM analyses just reviewed is that they use the traditional approach to measuring post-error slowing, which compares all post-correct RTs to all post-error RTs. That is, these EAM analyses fit post-error effects by allowing different parameters for these two sets of trials. This approach is vulnerable to errors being more frequent during periods simultaneously associated with RT effects, such as slow time scale fluctuations in tiredness or boredom (Dutilh et al., 2012b). It can, therefore, be confounded and so over- or underestimate the causal effect of an error on post-error responding.

Dutilh et al. (2012b) proposed a potential remedy to these problems, a *robust method* of measuring post-error effects by pairing post-error trials with their pre-error counterparts (that were also post-correct). As trials that originate in closely contiguous locations within the time series of stimuli and responses are compared, the effects of slow fluctuations

Table 1 Average evidence-accumulation modelling results for post-error effects. Note that all studies provided accuracy feedback. Note also that White et al. (2010) reported only the difference between drift rates for each choice, not the rates themselves, so post-error effects on rates could not be assessed

Study and task	Post-error accuracy	Response instructions	Response-to-stimulus interval	Model type	Threshold	Drift rate	Non-decision time
White et al. (2010) Recognition Memory	Decrease	Both speed and accuracy	0.75 s	DDM	Increase (for higher anxiety group)	Not reported	Decrease
Dutilh et al. (2012a) Lexical Decision	Low frequency words: increase High frequency words: no difference	Both speed and accuracy	0.5 s	DDM	Increase	Not significant	Not significant
Dutilh et al. (2013) Experiment 1 Random Dot Motion	No difference	Alternated speed/accuracy emphasis across blocks	Individualised to participants.	Simple DDM	Increase (Aged only)	Decrease (Aged only)	Increase
Experiment 2 Lexical Decision	No difference	Both speed and accuracy	Correct: 0.15 s Incorrect: 0.8 s	Simple DDM	Increase.	Decrease.	Increase (Aged only)
Purcell and Kiani (2016) Random Dot Motion	No difference	Both speed and accuracy.	Short: 0.35–0.8 s Long: 2.4–3.3 s	Simple DDM	Increase	Decrease	Not significant
Schiffler et al., 2017Visual Search	Decrease	Unreported.	0.5s	DDM	Increase	Decrease	Decrease

are controlled. The robust approach to selecting post-correct and post-error trials can be used in fitting EAMs, but it has the drawback that it limits the amount of data available, which can make parameter estimates of more complex EAMs unreliable. To address this issue, Dutilh et al. (2013) fit a simple DDM to robustly selected trials using Wagenmakers et al.'s (2007) "EZ" method, which has been shown to have good estimation properties in detecting experimental effects (van Ravenzwaaij et al., 2017). Their results differed between tasks (random-dot motion and lexical decision) and between older participants and younger controls, as summarized in Table 1 (along with the results for the five other EAM experiments analysed with non-robust fitting methods). Generally, thresholds increased, drift rates decreased, and non-decision time increased. The latter finding contrasts with White et al. (2010) and Schiffler et al. (2017), who both found a post-error decrease in non-decision time effects, but all interpreted these effects as being due to an affective component, hesitancy with respect to the increases and impulsivity with respect to the decreases.

Ratcliff (2008) pointed out a number of potential problems with EZ estimation, including biases in the values of the subset of DDM parameters it uses due to being unable to accommodate common patterns or results. Most germane here, the simple DDM underlying the EZ method predicts that error and correct responses have exactly the same RT distribution, which is rarely, if ever, observed. Although Dutilh et al. (2013) showed the simple DDM provided a good fit to RT distributions, those distributions did not separate correct and error responses, and so would not reveal the inevitable misfit. Not only does that mean the model did not provide an accurate description of their data, but also, as we now discuss, the differences in the speed of error and correct responses can be particularly crucial to understanding post-error effects. As such, even with the robust method of calculating post-error slowing, an analysis based on the EZ model is of potentially limited utility.

Two Types of Errors

The inconsistency of post-error adjustments across tasks and the resulting EAM parameter estimates have led researchers to suspect that there is likely variability between individuals and tasks in post-error adjustments (Dutilh et al., 2012a; Dutilh et al., 2013; Purcell & Kiani, 2016). Little attention, however, has been focused on the variability in errors themselves, even though Laming (1979) noted that there could be potential consequences for post-error effects from the fact

that errors can sometimes be faster and sometimes slower than correct responses. Recently, Damaso et al. (2020) examined the role of fast versus slow errors in post-error effects, distinguishing two error types: *response speed* errors, which are typically faster than correct responses and occur in very easy tasks and/or with instructions that emphasise fast responding, and *evidence quality* errors, which are typically slower than correct responses and occur predominantly in more difficult tasks and/or those that emphasise accuracy.

Damaso et al.'s (2020) analysis was motivated from the perspective of EAMs that assume trial-to-trial variability in parameters. Variability in the initial bias (i.e. starting point) of accumulation in a DDM was introduced by Laming (1968) to account for fast errors. Ratcliff (1978) highlighted the role of variability in drift rates in explaining slow errors. Ratcliff and Rouder (1998) showed how the two types of between-trial variability can work together to explain the full range of relationships that are observed between error and correct RT. The same interplay occurs between bias and rate variability in enabling the LBA to also provide a comprehensive account of these effects (Brown & Heathcote, 2008). Both models assume that the two categories of errors occur together in most experimental conditions, with their relative proportions varying systematically as a function of whether the choice is difficult or easy and whether the speed or accuracy of responses is emphasized.

Importantly for the traditional view of post-error slowing, increasing thresholds is only effective in reducing response-speed errors, by integrating out random start-point biases in both the LBA and DDM, and in the DDM the effects of moment-to-moment accumulation variability as well. For evidence-quality errors, the situation is very different. Their cause, between trial rate variability, corresponds to either faulty evidence or a faulty interpretation of the evidence that cannot be ameliorated by continued accumulation. Indeed, such variability can result in cases where evidence systematically favours the wrong response, so that an increased threshold makes an evidence-quality error more rather than less likely; in such cases, lower thresholds, and hence faster rather than slower responses, can promote accuracy.

Damaso et al. (2020) analysed data from two recognition memory experiments (Rae et al., 2014; Osth et al., 2017), both of which featured conditions that differentially emphasized the speed or accuracy of responses. They explored the possibility that different types of errors might also have different types of post-error adjustments. They utilised the robust method of calculating post-error slowing and compared the results to another approach, the *matched method* (Hajcak & Simons, 2002). Because errors typically differ in speed from correct responses, it is possible that post-error effects may be confused with effects that arise from the speed of previous responses (e.g. a homeostatic tendency to speed up after a slow response or slow down after a fast response). To account for this, the matched method compares post-error and post-correct responses for pairs of correct and error responses that are closely matched on RT.

Damaso et al. (2020) found that under speed emphasis, there was a preponderance of fast, response-speed errors that were followed by post-error slowing. In contrast, in accuracy-emphasis conditions they found a preponderance of slow evidence quality errors that were followed by post-error speeding. They also found this pattern reflected in the overall distribution of error responses — specifically they found more post-error slowing for the fastest 50% of responses and more post-error speeding in the slowest 50% of responses. These effects in RT were significant when using the robust, matched, and standard methods of calculating post-error effects for data from Rae et al. (2014) and the matched and standard methods for Osth et al. (2017), although results from the robust method were also in the expected direction. Consistent with previous studies (see Table 1), Damaso et al. also observed either no change (Rae et al., 2014) or a decrease (Osth et al., 2017) in post-error accuracy compared to post-correct accuracy.

Based upon their findings, Damaso et al. (2020) suggested that in the speed condition, the occurrence of post-error slowing with no increase in accuracy might be explained by a combination of adaptive (e.g. classic) and maladaptive (e.g. orienting) accounts. However, they noted that no existing explanation could account for post-error speeding in the accuracy conditions. As outlined above, in order to accommodate this finding, it is critical that adaptivity of processes be considered with reference to both speed and accuracy. They suggested speeding may be the norm for slow evidence-quality errors. It is possible that this might have gone unnoticed as almost all previous post-error studies employed (1) rapid-response tasks that promote response-speed errors and/or (2) task instructions that emphasis speed and accuracy of responses simultaneously. They reasoned that in conditions promoting evidence-quality errors a threshold increase has little effect on accuracy because the remaining errors mainly arise from the evidence entering the decision process. In this instance, following errors participants may become frustrated, prioritise speed over accuracy, and subsequently respond faster (and potentially less accurately) — a phenomenon they called *post-error recklessness* (Williams et al., 2016). In support of this, Purcell and Kiani (2016) observed that two participants (one human and one monkey) went against their average findings, showing post-error speeding accompanied by a decrease in threshold which was sufficient to overcome the slowing effect of the decrease in drift rate that was also observed.

Evidence Accumulation Model Analysis

To investigate the cause or causes of the post-error slowing and post-error speeding reported by Damaso et al. (2020), we applied an EAM analysis to trials they used in their robust analysis, making our analysis comparable to that of Dutilh et al.

(2013). However, in contrast to Dutilh et al., and motivated by the key role it plays, we used EAMs capable of accounting for error speed, the full DDM, and the LBA. Our use of the LBA was also motivated by its ability (along with other racing accumulator models) to represent two types of mean drift rate effect (Boag et al. 2019a; Boag et al. 2019b). The first type, which we refer to as rate *quality*, is the difference between the mean rate for the matching accumulator (i.e. the accumulator that corresponds to the stimulus; e.g. the old accumulator for an old test item) and the mismatching accumulator (e.g. the new accumulator for an old item), which mostly influences accuracy (i.e. a greater accuracy for a greater difference). The second, which we refer to as rate *quantity*, is the overall level of the mean rates (conveniently represented by the average over accumulators). Quantity directly influences the overall speed of responding, with faster response when it is larger, but it can also impact on accuracy. In particular, increased rate quantity acts similarly to a decrease in threshold. For the DDM, in contrast, mean rates affect both accuracy and speed simultaneously and in the same direction (i.e. more accurate and faster for larger mean rates). This gives mean rate effects in the LBA greater flexibility than mean rate effects in the DDM, but at the cost of requiring twice as many parameters.

Because the robust method made reliable estimation more difficult by reducing the number of trials per participant, at least for those who were more accurate (106–336 for Rae et al., 2014, and 196–720 for Osth et al., 2017), we fit the models using hierarchical Bayesian estimation, as described in Heathcote et al. (2019) (for details, see the "Data Analysis Overview" section below). This increases the reliability of our estimates by taking advantage of the mutual constraint among relatively large number of participants in each data set (47 and 46, respectively). It quantifies uncertainty in a way that guards against inferences being overly optimistic (Rouder & Haaf, 2019) and also takes into account differences in measurement noise among participants due to different numbers of trials. In light of the different numbers of parameters required for DDM and LBA models and their different functional forms, we used the DIC model selection criterion (Spiegelhalter et al., 2014), which weighs model fit against model complexity related to both number of parameters and functional form. As well as using DIC to compare the DDM and LBA, we also used it to select among different parameterizations within each type of model that also differed in complexity because they allowed different combinations of parameters to be affected by last-trial accuracy (i.e. correct vs. error).

For brevity, we refer to the Rae et al.'s (2014) data set as Experiment 1 and the Osth et al.'s (2017) data set as Experiment 2. Both experiments had a common design, with participants studying a list of words and then pressing one of two buttons to indicate if a series of single test items had been studied (*old*) or not (*new*); no feedback was given as to response accuracy in either case. For half of the lists, instructions emphasised response speed, and for the other half, they emphasised response accuracy. The studies differed in that there was only a brief response-to-stimulus interval for Experiment 1, whereas Experiment 2's participants followed their recognition response by a confidence rating (low or high). We took advantage of the latter design feature, and the larger number of trials in this case, to examine the results of fits where trials were further broken down by whether the choice for the last trial was given a low or high confidence rating. We reasoned that participants were likely to be aware that they had made an error if they gave a low confidence rating, and so effects related to error detection and affective reactions would be more likely to occur. Before reporting the results of the EAM analyses, we briefly describe each model and its parameters, with a summary provided in Table 2 (for a more comprehensive treatment, see Donkin & Brown, 2018).

The DDM belongs to a class of models that bases decisions on accumulating differences in evidence for two options where the difference can vary from moment to moment during the accumulation process. A response is triggered when the accumulated evidence differences reach a threshold either above (for one option) or below (for the other option) the starting point of accumulation. The DDM has four key parameters: the mean rate of accumulation or drift rate v; response caution, quantified by separation between upper and lower boundaries (i.e. thresholds) a; a priori bias, quantified by mean start point z (no bias corresponds to $z = a/2$); and non-decision time, T_{er}. This model also includes parameters that quantify between-trial variability (Ratcliff & Tuerlinckx, 2002), which is assumed to be normal for the drift rate (with standard deviation s_v), and uniformly distributed for the start point (with width s_z and centred on z) and non-decision time (with width s_t centred on T_{er}).

The LBA belongs to the class of independent race models, where each choice option has a separate evidence accumulator. Each accumulator has a threshold, and the choice is determined by the first accumulator to reach its threshold (b). The LBA assumes that each accumulator accrues evidence independently and linearly. That is, the rate of accumulation is constant during each decision, with all variability caused by trial to trial variations. The rate on each trial is drawn independently for each accumulator from a normal distribution with a mean drift rate (v) and trial-to-trial variability (s_v). The start point of evidence accumulation is independently drawn for each accumulator from the interval 0-A with uniform probability. Large gaps between A and the threshold ($B = b - A$) are indicative of high response caution that integrates out errors due to random biases caused by start-point variability. As in the DDM, RT is the sum of decision time and non-decision time. Here, as is typical in most applications, non-decision time is assumed not to vary from trial to trial, although uniform variability can be used just as it is with the DDM (e.g. Heathcote & Hayes, 2012).

Table 2 Parameter notation and definitions for the diffusion decision model (DDM) and the linear ballistic accumulator (LBA)

Model	Parameter	Description
DDM	v and s_v	The mean and standard deviation of normally distributed between trial variability in accumulation rates
	a	Response caution, the location of the upper response threshold (the lower threshold is at zero)
	z and s_z	Mean response bias (no bias corresponds to accumulation starting at $z = a/2$) and the width of uniformly distributed between trial bias variability
	T_{er} and s_t	Mean non-decision time and the width of uniformly distributed between trial non-decision time variability
LBA	v and s_v	As for the DDM, but independent for each accumulator. Average v across accumulators indexes *quantity* and v for the accumulator matching the stimulus minus v for the mismatching accumulator indexes *quality*
	A	The width of uniformly distributed start-point variability, independent for each accumulator
	B	The distance between maximum start-point variability (A) and the response threshold (b) for each accumulator (so $b = B + A$). Response bias corresponds to a difference in B between accumulators
	T_{er}	As for the DDM, but with no between-trial variability

Data Analysis Overview

Before reporting the detailed results, we provide an overview of the goals and setup of our analysis methods. Note that we report results in units of seconds (s) where appropriate and do not provide units for rate and threshold model parameters as they are scaled relative to a unit defined by a standard deviation, either for the mismatching accumulator rate for the LBA or for the DDM to moment-to-moment variability.

Posterior samples were obtained using a Differential Evolution Markov-Chain Monte-Carlo (DE-MCMC: Turner et al., 2013). We used the R implementation described in Heathcote et al. (2019) assuming independent normal hyper-distributions for all parameters. Exponential priors with a mean of one for hierarchical standard deviation parameters and normal priors for hierarchical means (truncated to positive values for start-point noise and threshold parameters and truncated to the range 0.1–1 s for non-decision time) were chosen to be only weakly informative. Plots of posterior densities over prior densities confirmed this was the case. For the normal priors, start-point noise, thresholds, and rate standard deviation parameters had a mean of one, and non-decision time parameters a mean of 0.3 s. Mean rate parameters that matched the stimulus had a prior mean of three and those that mismatched the stimulus a mean of one. Sampling was run with three times as many chains as model parameters. After thinning by a factor of 10, 250 iterations of each chain were retained for subsequent analysis after convergence was obtained, as assessed visually and by $\hat{R} < 1.1$ (Brooks & Gelman, 1998). *R* code for model fitting is supplied along with the data used for fitting at https://osf.io/4uc9z/.

Results

We first briefly summarize the significance of effects on manifest (i.e. accuracy and mean RT) measures (see supplementary materials for details), which are in line with those originally reported by Damaso et al. (2020) and consistent between Experiments 1 and 2 for all but the effects of last trial type (correct or error) on accuracy. RT was faster under speed than accuracy instructions (1: 0.54 s vs. 0.76 s; 0.66 s vs. 0.85 s), and accuracy was less for speed than accuracy instructions (1: 70.5 vs.77.4%; 2: 65% vs. 72.5%). Error responses were faster than correct responses under speed emphasis, but slower under accuracy emphasis (respective margins, 1: 0.02 s vs. 0.033 s; 2: 0.058 s vs. 0.011 s).

Under speed emphasis, there was post-error slowing, and under accuracy emphasis, there was post-error speeding (Experiment 1, both 0.018 s; Experiment 2, speed 0.014 s vs. accuracy 0.048 s). The inconsistency between experiments is that in Experiment 1, there was no evidence for a difference in accuracy after correct vs. error trials, whereas in Experiment 2, accuracy was significantly lower after an error than after a correct response (67.3% vs. 69.2%), but there was no evidence that this effect differed with speed versus accuracy emphasis. Unique to Experiment 2 were the effect of the confidence of the last response. Responses were slower when the last trial received a low than high confidence rating (0.771 s vs. 0.757 s), and this effect interacted with instructions, being negligible for speed emphasis but 0.021 s under accuracy emphasis.

Differences between correct and error speed were also larger after a low than high confidence response. However, no interactions of last trial accuracy with last-trial confidence were significant in mean RT. Although the main effect of last confidence rating on accuracy did not achieve significance, it did interact significantly with post error/correct and instructions, although only just at the 0.05 level. Under speed emphasis, accuracy was 4.2% higher after a correct high confidence response than after a correct low confidence response, whereas the difference was reversed by 0.5% after an error. Under accuracy emphasis, there was little difference in either case (1.3% and 1.9% favouring post-correct, respectively).

Table 3 Model variant definition. N non-decision time (T_{er}), R mean rate (v), and T threshold (a in the DDM, B in the LBA). Each model corresponds to a hypothesis about the effect of whether the last trial was correct or incorrect, it has no effect on any model parameter, and it affects only non-decision time (N), mean rates (R) or thresholds (T), or all but one of these, or all three

Model variant	Variant parameterization
No last-trial effect	For this (and all other variants) instructions affect N, R, T, and bias (z in the DDM, difference in B for the LBA), and stimulus affects only R
N	Non-decision time different for last-trial error vs. correct responses
R	Mean rate (for the LBA for both accumulators) different for last-trial error vs. correct responses
T	Threshold (a in the DDM, B in the LBA) different for last-trial error vs. correct responses
R, N	Both rate and non-decision time vary with last trial type
T, N	Both threshold and non-decision time vary with last trial type
T, R	Both threshold and rate vary with last trial type
T, R, N	Threshold, rate, and non-decision time vary with last trial type

Model Specification, Selection, and Fit

We fit the same eight types of models (summarized in Table 3) to each experiment, seven that allowed last-trial accuracy (correct vs. error) to affect different combinations of parameters, and one with no affect one any parameter. The most complex model allowed last trial to affect thresholds, rates and non-decision time, three of intermediate complexity allowed only two such effects, and the remaining three allowed only one effect. To avoid overfitting, we performed model selection based on DIC and then confirmed graphically that the selected model provided an adequate fit. Having established that the selected model provided an accurate representation of the data, we then interpreted the associated parameter estimates.

We also fit the data from Experiment 2 versions of the seven models that allowed for a last-trial effect while further breaking it down according to its associated confidence (low vs. high). For these "confidence models", both types of last-trial effect were assumed to occur together, so we can describe these models in the same way as the models that ignore confidence.

For all DDM and LBA models, we allowed the instruction manipulation to affect thresholds and response bias and also the mean drift rate and non-decision time. Stimulus (i.e. old vs. new test items) affected only mean rates. Following the findings of Rae et al. (2014), we would expect that thresholds, rate quality, and non-decision time would be lower under speed emphasis, and rate quantity higher.

The most complex DDM model had 22 parameters (4 threshold, 4 bias, 8 mean rate, 4 non-decision time, 1 start-point noise, and 1 rate standard deviation parameter), whereas the most complex LBA model had 30 parameters due to the need to specify different mean rate parameters for each accumulator (16 in total). Otherwise counts of parameter types are as for the DDM, with the rate standard deviation for the accumulator that mismatched the test item fixed at one to make the model identifiable (Donkin et al., 2009). The value for the matching accumulator was estimated. As is conventional, identifiability for the DDM was obtained by fixing the diffusion coefficient to one. Simpler DDM models ranged from 20 to 14 parameters and simpler LBAs ranged from 28 to 18.

Table 4 displays model selection results, where the variant with a zero entry is favoured and larger values indicate worse results. In every case, the model with no last-trial effect was clearly inferior (note that for the confidence models, this is the same as for the fit ignoring confidence and so that DIC value is replicated in the table), consistent with the significant last-trial effects found in manifest analyses. The advantage was greater in Experiment 2, indicating that the last trial had a stronger effect there than in Experiment 1, again consistent with the weaker effects of last trial in manifest analyses of Experiment 1 vs. Experiment 2.

For the DDM, the same 18 parameter model was selected for both experiments, where last-trial accuracy affected threshold and non-decision time. For the LBA, non-decision time alone wins for Experiment 1, whereas the most complex model wins for Experiment 2. For Experiment 1, both models indicate non-decision time performs best of any single effect model, with threshold second, detracting slightly from DIC for the LBA and improving it by the same margin for the DDM. For Experiment 2, the DDM's preference for including the threshold in addition to non-decision time increases. For the LBA in Experiment 2, non-decision time again does best of the single effect models but also detracts least when it is moved from the selected model with all effects; the threshold effect appears to have the most important role, and mean rate effects the second most important role, in the selected model.

When DIC was used to compare the DDM and LBA, the LBA was selected by a large margin, even when models with the same number of parameters are compared. Given the clear and consistent advantage displayed in the present data, we focus on the LBA in the remainder of the paper, with corresponding DDM results reported in supplementary

Table 4 DIC model selection results (smaller values preferred), where the minimum value in each column is subtracted from every value in the column so a 0 entry indicates the selected model. LBAc indicates the confidence model. Actual values for the selected DDM models are 1362 and 26620, and for the selected LBA models −1915 and 20649 for Experiments 1 and 2, respectively, with the selected LBAc model for Experiment 2 being 20692. Values in brackets are the disadvantage in DIC for the best model within a column relative to the best model for the overall experiment. For model variants, an effect of last trial on threshold is indicated by a T, mean rate by an R, and non-decision time by an N

	Experiment 1		Experiment 2		
Model variant	DDM	LBA	DDM	LBA	LBAc
T, R, N	141	151	81	0	67
T, R	979	248	1981	42	0 (43)
T, N	0 (3277)	51	0 (5971)	96	245
R, N	244	145	245	215	238
T(hreshold)	836	81	1929	401	89
R(ate)	1438	266	2227	1149	1250
N(on-decision time)	51	0	215	166	93
No last-trial effect	1262	195	3272	1305	1305

materials. Given these results, we only fit the LBA to trials broken down by last confidence response in Experiment 2. In this case, Table 4 shows that the variant with threshold and rate effects but without a non-decision time effect was selected, with threshold clearly having the strongest effect, including being preferred as a single cause. However, the best confidence model was slightly worse, by 43 DIC units, than the best LBA model that ignored confidence, so from here, we refer to the latter model simply as the best model.

Figure 1 shows the fits of the selected LBA models to the accuracy data. It shows that the model captures the lack of effect of last trial on accuracy in Experiment 1 and in Experiment 2 the post-error decrease in accuracy and interactive effect of confidence in the speed condition. Post-error accuracy is clearly decreased when the last response was given with high confidence and slightly increased after a low-confidence response. Figure 2 displays LBA fits to RT 10th, 50th (median) and 90th percentiles for both correct and error responses. The model LBA captures the pattern of post-error slowing under speed emphasis and post-error speeding under accuracy emphasis and the relative invariance of this pattern with last-response confidence. In general, the model fits the data well, with most data points falling within 95% credible intervals, with misses mainly for rarer errors in the accuracy condition, particularly for the slowest (i.e. 90th percentile) false alarms (i.e. incorrect responses to new stimuli). Fits to RT distribution broken down by last-trial confidence in Experiment 2 are of similar quality (see supplementary materials).

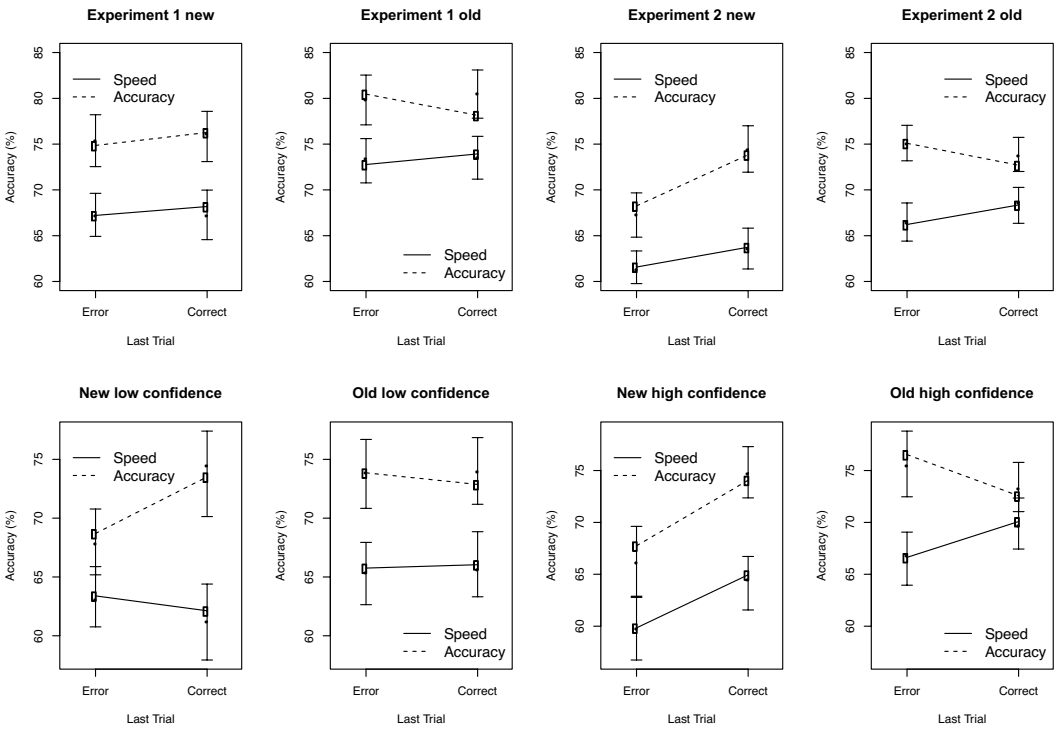

Fig. 1 LBA fit to accuracy for new and old test items (panels), speed vs. accuracy instructions (lines), and last trial type (*x*-axis). The second row gives results for Experiment 2 broken down by last-response confidence. Observed values are given by large circles and joined by lines. Fits are indicated by solid points (posterior medians) with 95% credible intervals

Note that in supplementary materials, we show that when we simulated data form the selected models and fit them with the same procedures as applied to the real data, DIC correctly identified the selected model variants for both the LBA and DDM, giving us confidence in the veracity of our conclusions. Supplementary materials also show that high correlations between generating and estimated parameters for the selected models, except for the DDM start-point noise and rate variability parameters, were generally poorly recovered. Recovery of the other parameters for the DDM was unbiased, and there was no evidence of bias for the LBA model for Experiment 1, but for Experiment 2 start-point noise and rate parameters were slightly over-estimated, and thresholds underestimated. For LBA fits to Experiment 2 data broken down by last-trial confidence, there was no evidence of bias for rate parameters, but the bias in the

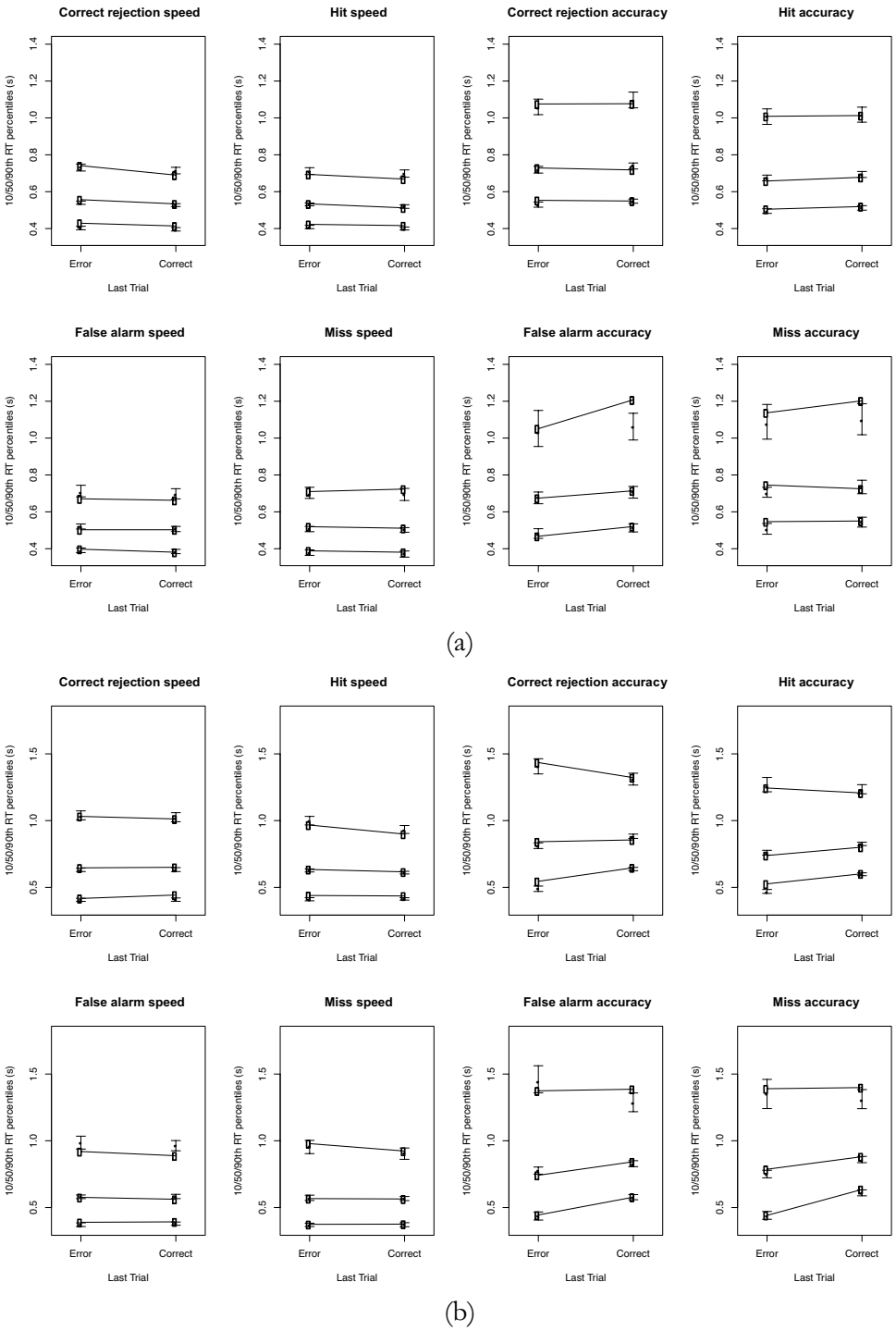

Fig. 2 LBA fits to RT distribution represented by 10th, 50th and 90th percentiles **a** Experiment 1, **b** Experiment 2 broken down by responses to new items (correct rejections and false alarms) and old items (hits and misses) and speed vs. accuracy instructions (panels). Observed values are given by large circles and joined by lines. Fits are indicated by solid points (posterior medians) with 95% credible intervals

Table 5 Non-decision estimates after trials after an error minus trials after a correct response. Tail probabilities test if the post-error effect is different from zero

Experiment	Instruction	Post-error effect	Median [95% CI]	Tail probability
1	Speed	Slowing	0.011s [0.008–0.014]	$p < .001$
	Accuracy	Speeding	−0.021s [0.013–.028]	$p < .001$
2	Speed	Slowing	0.009s [0–0.018]	$p = .022$
	Accuracy	Speeding	−0.050s [0.037–0.064]	$p < .001$

start-point noise and threshold parameters remained, and there was also overestimation of non-decision time parameters. As correlations between observed and recovered parameters were high, and interest focuses on parameter differences, we do not believe these biases affect our inferences about the causes of post-error effects in the next section.

LBA Model Parameters

We used Bayesian 95% credible intervals and p values (Klauer, 2010; Matzke et al., 2015) to test differences in parameters between conditions. This involved taking the difference between parameter values (or functions of them, like averages across conditions) for each iteration in each chain for each participant and then averaging these values over participants[1]. Credible intervals (provided below in square brackets) were estimated by the range between 2.5 and 97.5 percentiles of the resulting distribution and p values by the proportion that falls above or below zero. For ease of interpretation, we report p values as "tail probabilities", corresponding to the side that occurs least (e.g. if most differences are negative, we report the proportion that are positive), so small p values indicate reliable differences (e.g. $p < .025$ corresponds to a two-tailed .05 criterion). We first report results for the models with the best overall DIC in each experiment. We then report results for the best confidence model.

Selected Models

As shown in Table 5, in both experiments, non-decision time was slower after a correct response than after an error under speed instructions and faster under accuracy instruction. In Experiment 2, there was a general shift towards post-error speeding, particularly under accuracy instructions, but the interaction remained, as shown in Fig. 3.

The remaining parameter analyses pertain to Experiment 2 only, with results also plotted in Fig. 3. We report mean rate results in terms of quality (i.e. the difference in rates between the accumulator that matches and mismatches the stimulus) and quantity (i.e. the rate over accumulators), and threshold results in terms of caution, the average over accumulators of the gap from the top of the start-point noise distribution to the threshold. A value of zero indicates the least possible caution as thresholds must fall above the start-point noise distribution in the LBA. As shown in Table 6, average caution was higher under accuracy than speed emphasis, as were both rate quality and quantity.

In terms of last-trial accuracy effects, Table 7 shows that under speed emphasis, participants were slightly *less* cautious after an error than after a correct response and much more so under accuracy emphasis. As shown in Fig. 3, there was no threshold difference between speed and accuracy emphasis for post-error trials, with the last-trial effect due to increased caution after a correct response.

Last-trial effects on rate quality were relatively weak, with a trend under speed emphasis for greater quality when the last trial was correct than when it was an error, and no indication of a difference under accuracy emphasis. In contrast, last-trial effects on rate quantity were more marked, with larger values after correct trial than error trials, particularly under accuracy emphasis. Fig. 3 depicts the rate effects broken down by stimulus type. Differences in quality effects between old and new stimuli were negligible under speed emphasis, but under accuracy emphasis, they were very large. As Fig. 3 shows, there was a cross-over interaction, with quality less after a correct response for old stimuli, by 0.55 [0.42–0.66], and greater for new stimuli, by 0.49 [0.36–0.62], $p < .001$. Quantity effects were in the same direction but larger for old than new stimuli, both under speed emphasis, by 0.12 [0.05–0.18], $p < .001$, and under accuracy emphasis, by 0.28 [0.2–0.36], $p < .001$.

Best Confidence Model

Parameters for the best confidence model are shown in Fig. 4. Generally, the patterns of effect were similar after high and low confidence responses. However, in several cases, the magnitude of the difference between post-correct and post-error trials varied, particularly under accuracy emphasis. We focus on reporting these variations.

In terms of caution, after a low confidence response, there was little difference between post-correct and post-error trials under speed emphasis, 0.01 [−0.1–0.4], $p = .16$, whereas under accuracy emphasis, it increased greatly, 0.53

[1] Note that this approach is analogous to a fixed-effects analysis; we did not perform this analysis with population level estimates because our population model did not account for random participant effects.

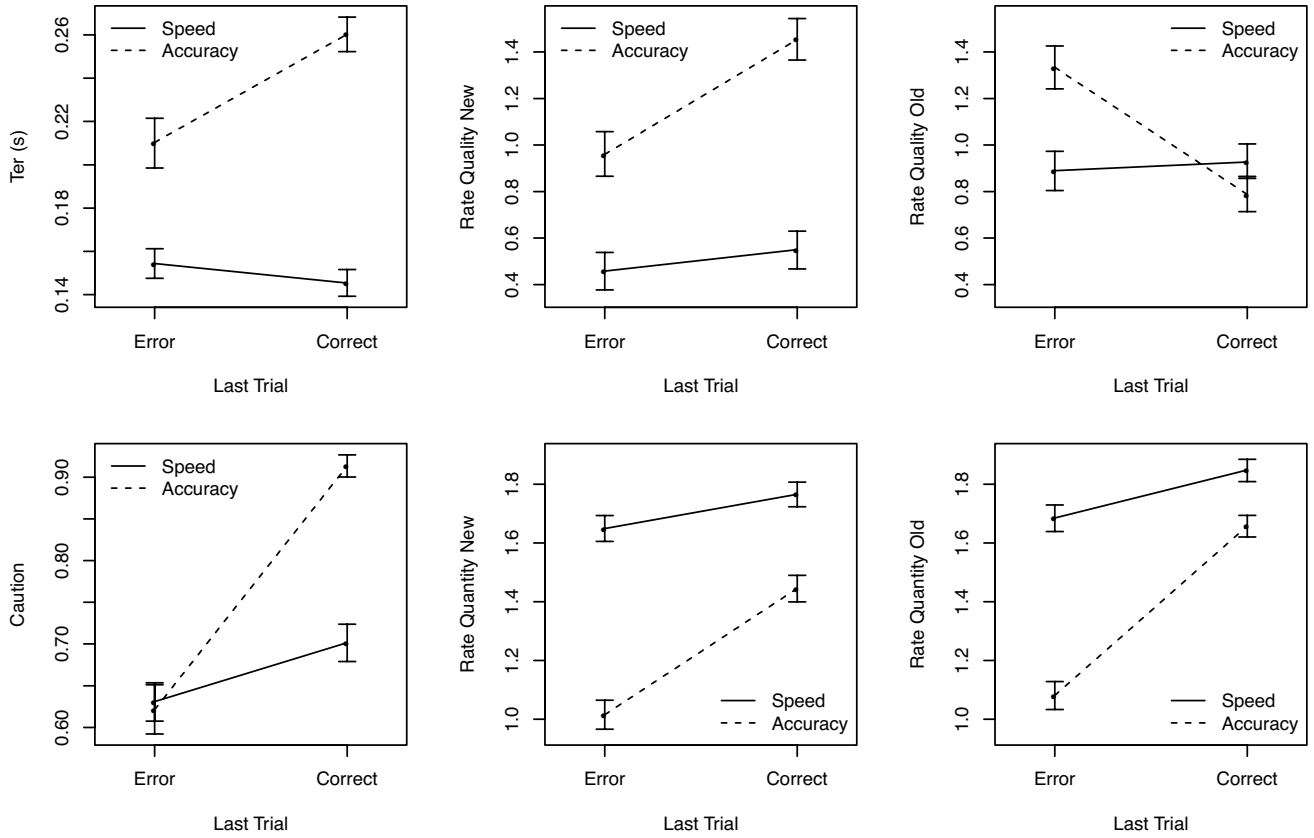

Fig. 3 Parameters for the best LBA model for Experiment 2. Solid points are posterior medians with 95% credible intervals. The *x*-axis specifies whether the response to the last trial was an error or correct. Lines correspond to speed or accuracy emphasis conditions. Caution is the average of thresholds over accumulators, and quantity is the average of accumulation rates over accumulators. Quality is the rate for the accumulator that matches the stimulus minus the rate for the accumulator that mismatches the stimulus. Rate parameters are broken down by whether the stimulus is old (previously studied) or new (not studied). See Table 3 for definitions of non-decision time (T_{er}), thresholds, and accumulation rates

Table 6 Experiment 2 parameter estimates. Tail probabilities are for the difference between speed and accuracy emphasis

Parameter	Speed	Accuracy	Tail probability
Caution	0.77 [0.75–0.78]	0.67 [0.65–0.68]	$p < .001$
Rate quality	1.13 [1.09–1.18	0.71 [0.67–0.74]	$p < .001$
Rate quantity	1.74, [1.71–1.76]	1.3, [1.27–1.33]	$p < .001$

[0.50–0.56], $p < .001$. After a high confidence response, the effect under speed emphasis was slightly reduced but still present, 0.05 [0.02–0.08], $p < .001$, whereas under accuracy emphasis, it was again increased greatly, 0.57 [0.53–0.60], $p < .001$. Recall that the confidence model did not have a non-decision time effect; the changes in the magnitude of the threshold effect under accuracy emphasis appear to have absorbed the shift in RT associated with non-decision time.

Table 7 Parameter estimates after trials after an error minus trials after a correct response. Tail probabilities test if the post-error effect is different from zero

Parameter	Speed	Tail probability	Accuracy	Tail probability
Caution	−0.07 [0.04–0.10]	$p < .001$	−0.29 [0.26–0.33]	$p < .001$
Rate quality	−0.065 [−0.13–0.002]	$p = .028$	−0.03 [−0.05–0.1]	$p = .26$
Rate quantity	−0.14 [0.1–0.19]	$p < .001$	−0.5 [0.46–0.55]	$p < .001$

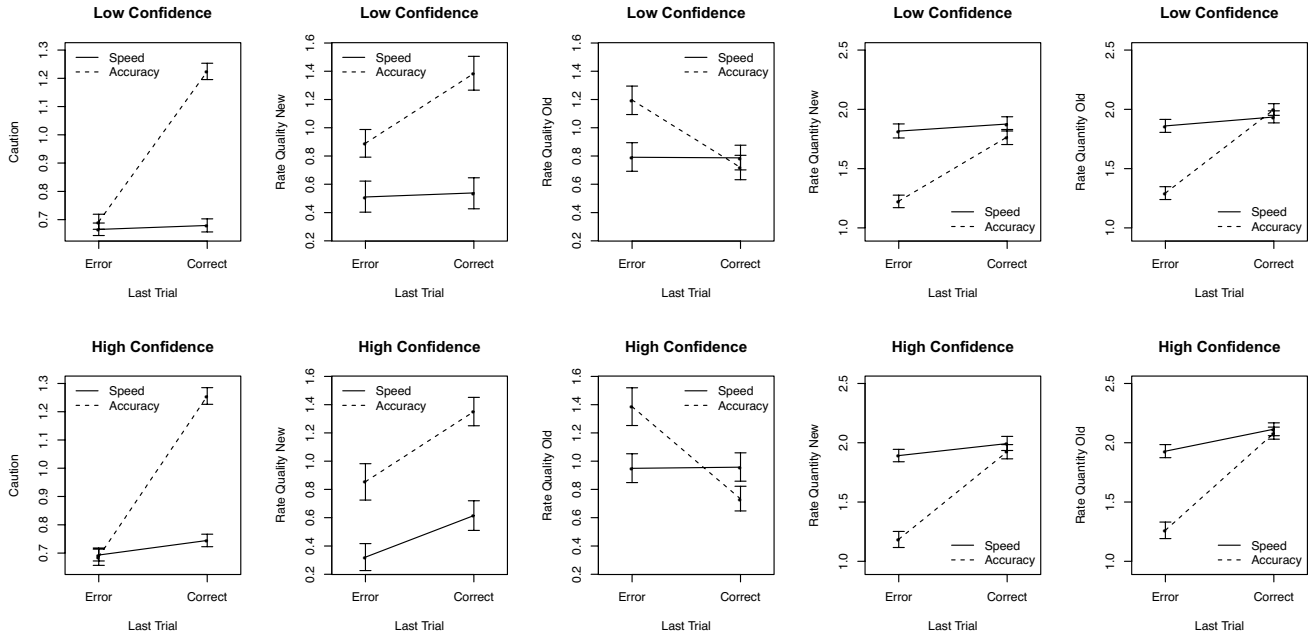

Fig. 4 Parameters for the selected LBA confidence model for Experiment 2. Solid points are posterior medians with 95% credible intervals. The *x*-axis specifies whether the response to the last trial was an error or correct. Lines correspond to speed or accuracy emphasis conditions. Panels are broken down by whether the last response received a low or high confidence rating. Caution is the average of thresholds over accumulators and quantity is the average of accumulation rates over accumulators. Quality is the rate for the accumulator that matches the stimulus minus the rate for the accumulator that mismatches the stimulus. Rate parameters are broken down by whether the stimulus is old (previously studied) or new (not studied). See Table 3 for definitions of non-decision time (T_{er}), thresholds, and accumulation rates

Rate effects were generally smaller after a low confidence response than after a high confidence response. In terms of rate quality, this clearly occurred for new items under speed emphasis, by 0.13 [0.03–0.1], $p = .006$, and for old items under accuracy emphasis, there was a smaller trend, 0.09 [−0.02–0.19], $p = .05$. In terms of rate quantity, attenuation of the last trial effect was more pervasive, occurring for all but new items under speed emphasis. For old items under speed emphasis, the attenuation was by 0.06 [0.01–0.1], $p = .013$. For accuracy emphasis, it was by 0.1 [0.04–0.16], $p < .001$, and by 0.06 [.01–0.11], $p = .104$, for new and old items, respectively.

Parameter Roles in Explaining Post-Error Effects

Model selection is informative in terms of identifying which types of model parameters are required to explain post-error effects. However, for Experiment 2, where more than one type was implicated, the question arises of how they each contribute and potentially tradeoff in explaining observed effects on mean RT and accuracy in the selected model. To investigate this issue, we simulated from each model of interest where we fixed all parameters to their estimated values except that we removed differences in one or more parameter types between post-error and post-correct trials by setting them both to their average value. We then noted the magnitude of the post-error effects predicted by the remaining parameters that differed between post-correct and post-error trials. Note that we did not base this analysis on the non-selected models reported in Table 4, as for them the remaining parameters will change to compensate for the omitted parameter types, and so this would not reveal their role in the selected model (Strickland et al., 2018). We based our analysis on predictions about mean RT and accuracy in the form of "posterior predictives" calculated by randomly drawing 100 parameter sets from the posterior of each participant, simulating data sets based on each draw with the same number of observations and design as for the corresponding participant, and then averaging results over participants.

For the best model, we examined eight restricted versions. These included three that allowed only a single effect, of non-decision time, thresholds, and mean rates, and three that allowed pairwise combinations of these effects. We also examined special cases of the rate-only model that allowed only rate quality or rate quantity effects. For the best confidence model, which omitted non-decision time effects, we examined threshold only and rate only versions as well as the rate quality only and rate quantity only special cases of the latter model. For each case, we tabulated the

average predicted mean RT and accuracy post-error versus post-correct effects with results reported in supplementary materials and summarized here. Note that only threshold and rate effects can explain accuracy as non-decision time affects only mean RT. Note also that, with the exception of the effect of non-decision time on mean RT, the effects of each parameter are not independent of the values of the other parameters (e.g. the magnitude of the effect of the same difference in a given parameter type can vary depending on the values of the other parameters) so the effects for each variant will not necessarily add up to the overall effect. Full results are provided in supplementary materials.

Rate effects explained the lion's share of the best model's prediction of reduced accuracy after an error. This was particularly the case under speed emphasis, whereas under accuracy emphasis, the lowered threshold after an error played a greater role, accounting for a little less than half of the effect. When the rate effect was broken down into quality and quantity, reduced quality explained the majority under speed emphasis, whereas increased quantity explained the majority under accuracy emphasis.

More complicated tradeoffs explained mean RT effects. The decreased in post-error thresholds when acting alone led to substantial post-error speeding under both speed and accuracy emphasis. In contrast, the decrease in rate quantity led to even more substantial post-error slowing. When taken together, the result is a small post-error slowing effect under speed emphasis (by 0.009 s) and greater post-error slowing under accuracy emphasis (by 0.024 s). The non-decision time effects given in Table 5 then slightly increase the post-error slowing under speed emphasis and completely overcoming it in the accuracy condition, so that the full model shows the post-error speeding evident in the data.

For the confidence model, under speed emphasis, the much larger post-error drop in accuracy after a high than low confidence response was due to a larger drop in quality. Under accuracy emphasis, the more similar drops in post-error accuracy after low and high confidence responses were explained by both increased rate quantity and decreased thresholds, the same pattern of parameters we found when confidence was ignored.

Discussion

The aim of our analysis was to identify the consequence of making an error on subsequent performance and explain them in terms of latent cognitive variables operationalized in EAM parameters. In an attempt to differentiate consequences from mere associations, our analysis used Dutilh et al.'s (2012a) robust method of selecting closely contiguous post-correct and post-error responses so that we could rule out mediation of performance differences by common causes such as slow time scale fluctuations in fatigue (e.g. slower post-error responses caused by errors occurring in periods of generally slow responding). In the data that we examined from two recognition memory experiments, Damaso et al. (2020) showed that the consequences of an error depended both on whether the speed or accuracy of responses was emphasised in task instructions and on the type of error. They identified two types of errors: response-speed errors, a rushed response that is on average faster than a correct response and could have been more accurate if it were made more slowly; and evidence-quality errors, occurring on average slower than a correct response and caused by erroneous evidence where accuracy is not improved by slower responding. Hence, our analysis used EAMs that are capable of accommodating differences in the speed of correct and error responses — the full DDM (Ratcliff & McKoon, 2008) and the LBA (Brown & Heathcote, 2008) — through the inclusion of between-trial variability in response bias and the mean rate of evidence accumulation.

We examined three EAM parameters that have been most prominently claimed to play a role in explaining post-error effects: mean evidence accumulation rates, thresholds, and non-decision time (see Table 1). We allowed one or more of these parameters to vary between post-error and post-correct trials. The LBA provided a good fit to most aspects of the data from the first experiment, where there was only a very short interval between the binary recognition response — previously studied ("old") or not ("new") — and the next test stimulus. Last-trial accuracy effects, in the form of post-error slowing in the speed-emphasis condition and post-error speeding in the accuracy-emphasis condition with no reliable effects on error rates, were best explained by the model in terms of corresponding differences in non-decision time (i.e. shorter after an error in the speed-emphasis condition and longer after an error in the accuracy emphasis-condition).

The LBA also provided a good fit to most aspects of Experiment 2, where there was a much longer interval between the recognition response and the next stimulus as participants made a second, high versus low confidence, response. In addition to the same pattern of RT effects seen in the first experiment, errors were also less common after a correct response under both speed and accuracy emphasis. The LBA again provided a good fit to most aspects of the data, but the best model required differences in all three parameter types to do so. The pattern of non-decision time effects was the same as for Experiment 1. In addition, response caution (i.e. the average of new and old accumulator thresholds) was lower after a correct than error response, with a larger difference under accuracy than speed emphasis. The same pattern held for new items in terms of rate quantity (i.e. the average over accumulators) and rate quality (i.e. the difference between the accumulator that matches and that mismatches the stimulus). For old items, the same pattern

held for rate quantity, but rate quality was greater after an error under accuracy emphasis and did not differ much under speed emphasis.

In contrast, the best DDM models for both experiments explained post-error effects in terms of thresholds (larger after an error under both speed and accuracy emphasis) and non-decision time (faster after an error, but only under accuracy emphasis). These results are difficult to reconcile with the lack of a post-error effect on accuracy in Experiment 1, as an increased threshold after an error predicts an increase in accuracy. The likely reason is that, in contrast to the LBA, the DDM displayed clear qualitative misfits (see supplementary materials).

It is unclear why the LBA performed better than the DDM in the present data. Across the many tasks in which these DDM and LBA have been compared in the literature, no consistent pattern of superiority has emerged. For example, Hawkins and Heathcote's (2021) comparison of fits of the DDM and LBA to four perceptual and lexical tasks found each was selected an equal number of times by DIC. The fact that we used recognition memory data is also unlikely to be the reason. For example, White et al. (2010) provided evidence of good fits of the full DDM to their recognition memory data, with non-significant χ^2 tests of misfit. The disadvantage of DDM was not due to including an account of last-trial effects, as the disadvantage for DDM models with no last trial affect relative to LBA models with no last trial effect increased in both experiments. It was also not due to fitting only trials selected by Dutilh et al.'s (2013) robust analysis. In supplementary materials, we report fits of the DDM and LBA models in Table 4 to original (i.e. non-robust) data and the disadvantage remained, although interestingly both models agreed in favouring the most complex model variants, with post-error effects on non-decision times, drift rates, and thresholds.

One possible reason the DDM's misfit here is that in the present data, there is a dissociation between effects of the overall magnitude of the inputs that determine accumulation rates and the ability of these inputs to discriminate between choices, which is naturally handled by rate quantity and quality in the LBA. Recently, Ratcliff et al. (2018) showed that the standard DDM is unable to accommodate such dissociations in a perceptual choice task but showed that a modified DDM could accommodate them by increasing drift rate variability in proportion to the mean drift rate. Future work might examine whether the modified DDM affords better fit, and if so whether it provides a different account of the causes of post-error effects. For the present, we focus on the implications of our LBA results for the causes of post-error effects.

However, before doing so, we draw two conclusions from the comparison of our LBA and DDM results. First, clearly the model, and its ability to fit the data, matters. Therefore, we believe it is important to not simply assume one model is correct, and even more so that fit should always be closely examined to determine if a model provides an adequate descriptive account before basing any conclusions on its parameters. Second, both models support, at the latent level, Dutilh et al.'s (2013) conclusion that post-error analyses can be confounded by slow fluctuations that have correlated effects on accuracy and speed. When these effects were not controlled, both the DDM and LBA supported post-error effects on all parameters, whereas when they were controlled by Dutilh et al.'s robust methods, many of these effects dropped out. Hence, we recommend the use of the robust method in all future model-based analyses of post-error effects.

Explaining Post-Error Consequences

In summary, we found effects on RT that directly corresponded to changes in non-decision time: slowing after an error under speed emphasis but speeding under accuracy emphasis. The same pattern held both when there was little delay between trials and when the delay was longer due to participants rating their decision confidence. Only in the latter case was there an effect on accuracy, which reduced after an error. Under accuracy emphasis, this was equally explained by a reduction in thresholds and an increase in rate quantity after an error. Under speed emphasis, it was explained by a reduction in rate quality after an error. These rate and threshold effects traded off, so that the net effect was a small degree of post-error slowing. When non-decision time was added, it increased the slowing under speed emphasis and reversed it under accuracy emphasis, resulting in observed post-error speeding.

The marked difference in post-error effects and their causes between the two experiments is likely due to the greater time available between trials for differential processing to emerge in Experiment 2, as well as the additional confidence response potentially facilitating reflection on the accuracy of the last response. When there was minimal time in Experiment 1 the effects on RT were best explained purely by non-decision time. Dutilh et al.'s (2013) attribution of increased non-decision time following an error to "irrelevant processing", such as overcoming disappointment (Rabbitt & Rodgers, 1977), could explain our finding of post-error slowing under speed emphasis. To explain our finding of decreased non-decision time following an error under accuracy emphasis, this affective explanation could be extended, based on Williams et al.'s (2016) concept of post-error recklessness and the understanding that in this condition, participants might have had some awareness that they were less able to avoid errors by modulating their decision process. Alternatively, under speed emphasis,

participants may have delayed the onset of accumulation in order to reduce errors caused by processing evidence before it is fully encoded (Laming, 1968). The decrease under accuracy emphasis might be based on an appreciation that a delay offered little or no such gain. In terms of Wessel's (Wessel, 2018) accuracy-based taxonomy all but the mechanism suggested by Laming (1968) are maladaptive. However, when time-cost becomes a factor the speeding associated with recklessness can be advantageous. With this consideration, it is possible to provide an entirely adaptive framing to all of our non-decision time results. The same pattern of non-decision time effects was seen in Experiment 2, except with a general shift to greater post-error speeding. This suggests the longer time available led to greater recklessness or perhaps a greater emphasis on time cost.

Our results suggested that the increased time available for differential processing after correct versus error responses in Experiment 2 was associated with both threshold and rate effects. These effects worked together to reduce post-error accuracy but worked in opposition in determining post-error RT. Post-error speeding due to lower post-error thresholds was counteracted by even greater slowing due to reduced rate quantity, so the net effect was mild slowing. Both the rate and threshold effects were bigger under accuracy than speed emphasis, and so was the net slowing. However, under accuracy emphasis speeding due to a decrease in non-decision time reversed the outcome to produce post-error speeding. Conversely, under speed emphasis the small slowing effect of non-decision time reinforced the post-error slowing.

Clearly our threshold results are inconsistent with the classical account of post-error slowing as a way to control errors (Rabbitt, 1966) and might, therefore, be considered maladaptive from some perspectives. However, they could also be considered adaptive from a time-cost perspective, particularly as they were largest under accuracy emphasis where threshold changes are likely to be least effective (as response-speed errors that can be avoided by a threshold increase are rarer, see Damaso et al., 2020). The larger effect under accuracy emphasis is also consistent with an affective account where increased frustration leads to increased post-error recklessness. More generally, the fact that threshold effects emerged in Experiment 2 when there was more time before the next trial is consistent with evidence-accumulation thresholds changing relatively slowly (Donkin et al., 2011).

Our rate effects are consistent with the orienting account (Notebaert et al., 2009) whereby post-error processing consumes resources that otherwise would have led to higher evidence accumulation rates on the next trial. The fact that the rate effect was greater under accuracy emphasis is consistent with errors being rarer and, hence, causing a greater orienting response. Another possibility is that performance monitoring processes associated with detecting an error (Ullsperger et al., 2014) and reacting to an error (e.g. by changing the threshold for the subsequent trial) consume attentional resources that remain depleted for a sufficient time to reduce rates on the subsequent trial. Again, the emergence of rate effects in Experiment 2 in association with threshold effects is consistent with this view to the degree that these processes take time. Similarly, the rate effects are consistent with the integration of the affective consequence of an error into the evidence accumulation process (Greifeneder et al., 2010; Roberts & Hutcherson, 2019), suggesting that this detracts from its overall level. One possibility is that rate quantity effects are mediated by a general arousal mechanism that increases gain in a non-specific manner and affects evidence accumulation equally for the correct and error responses (Boag et al., 2019a). Interestingly, under accuracy emphasis, increased error rates after an error were mediated by increased rate quantity. However, under speed emphasis, they were mediated by reduced rate quality. This suggests different/multiple mechanisms affecting participants' ability to encode evidence that effectively differentiates between new and old items.

In Experiment 2, we also sought further insight into the explanations related to orienting and other error-detection related processes. This was facilitated via analysis of performance contingent on whether the previous response was associated with a low confidence rating, indicating that an error was more likely to have been detected. Consistent with increased post-error processing, responding was slower after a low than high confidence response, although only under accuracy emphasis. Although last-trial confidence did not interact with any last-trial RT effects, there was an exaggerated reduction in post-error accuracy following a low confidence response under accuracy emphasis, but not under speed emphasis. Surprisingly, rate effects were actually reduced after a low-confidence response, and there was little difference in threshold effects as a function of confidence. Neither observation is consistent with mediation by an error-detection process. Although certainly not definitive, these results clearly do not support a relationship between low confidence and error detection as having a bearing on threshold and rate effects.

Conclusions and Future Directions

Our modelling results are consistent with some of the six previous studies summarized in Table 1, but contrast with others. Some of these discrepancies may be due to all but one of the previous studies not using the

robust trial selection method, all using the DDM, and half using the simple DDM (including the study that utilised the robust method). Importantly, the simple DDM is in principle incapable of accounting for the differences between correct and error RT that Damaso et al. (2020) showed which could play an important role in post-error effects. However, it is useful to review the commonalities and differences among results in order to consider where future research may most profitably focus.

Our results are consistent with the four studies in Table 1 that found drift rate effects, as all reported the same post-error decrease that we found. Hence, this finding appears to generalise across the DDM analyses used in Table 1 and the LBA analysis used here. One point of contrast is that our DDM analysis found no rate effects, although that is consistent with two of the studies in Table 1 that report similar findings. It is possible the discrepancy for the other studies is task related: like us, White et al. (2010) studied recognition memory and found no post-error rate effects with DDM analysis. Post-error non-decision time effects occurred in four cases in Table 1, with two decreases and two increases. We also found increases and decreases, but they were systematically related to speed and accuracy emphasis, respectively. In three of the cases, it is possible that instruction emphasis may have also mediated some of the discrepancy in previous results — how participants interpreted/misinterpreted response instructions that emphasised both speed and accuracy of responses (or one study where instructions were unreported) may be relevant. However, Dutilh et al. (2013) manipulated speed versus accuracy instructions in one of their experiments using a random-dot motion task and found decreased non-decision time after an error in both cases. Indeed, in marked contrast to our findings, they found no interactions between this manipulation and post-error effects on parameters. One possibility is that this was due to the inability of the simple DDM to accommodate the effects of the relative speeds of correct and error responses caused by this manipulation. Alternatively, it could also be due to task differences. In light of these considerations, it would be useful in future studies of post-error effects to include a speed versus accuracy emphasis manipulation with a variety of tasks and analyse the results with a model capable of accommodating the effects of this manipulation.

The largest discrepancy is that all studies in Table 1 reported an increase in post-error thresholds, whereas we found a decrease. A likely reason is that our thresholds effects were associated with the LBA, whereas all of theirs were associated with the DDM. When we fit the DDM, we also found increased post-error thresholds, but we also found that the DDM provided an inadequate fit. As previously suggested, it is also possible that the full DDM could be augmented to better accommodate data like ours through a more flexible approach to drift-rate variability. In any case, we do not recommend using the simple DDM or more restricted versions of the full DDM, such as the one implemented in the HDDM software (Wiecki et al., 2013) and applied by Schiffler et al. (2017) to post-error effects. These models are unable to account the difference in the speed of error vs. correct responses or, in the latter case, individual variation in this difference, and so are clearly not appropriate for research focusing on errors.

Future research might also address several limitations of the work presented here. First, our analysis was restricted to a recognition memory task, and clearly some of the discrepancies between our results and those in Table 1 could be due to task differences. Post-error effects occur very broadly across a wide variety of tasks, and our results suggest that even within a particular task their manifestations, and the underlying causes, can vary markedly depending on task demands. Rather than assuming observed post-error effects will be homogenous (e.g. only post-error slowing) and have only a single cause (e.g. increased response caution), we propose that future research should explore the potential diversity of effects on speed and accuracy across tasks and manipulations and use a variety of modelling approaches that are open to a range of underlying causes.

To better understand the origins of strong differences we found between Experiments 1 and 2, it would be useful to systematically manipulate response to stimulus interval with and without giving a confidence rating to see what separate effects they have. Hajcak and Simons' (2002) matched trial-selection method might be used to control for potential confounds between the speed of the previous trial and whether it was correct or error response. We did not do so here because we wanted to follow Dutilh et al. (2013) in using the robust method and so make our results more comparable. Future work might also consider post-error effects on other EAM parameters, such as those quantifying between-trial variability. Dutilh et al. (2012a) found that between-trial variability was slightly higher post-error, but because these parameters are harder to estimate, the effects were not reliable even though their data had a very large number of trials per participant. Similarly, our motivation for focusing on only thresholds, mean rates, and non-decision time parameters was that they are easier to estimate, which was a pressing issue because of our use of the robust method. Additionally, our review of the literature also indicated these parameters were of most interest.

Another candidate which is easier to estimate is response bias. Dutilh et al. (2012a) found a consistent increase in bias to "word" responses post-error in their lexical decision task,

but noted this result had no clear interpretation. We did not address response bias or indeed the closely related concept of stimulus bias (White & Poldrack, 2014), mainly because this would have increased the number and complexity of models that we would have had to fit, with the latter effect again straining our ability to obtain reliable estimates given limited trial numbers. A potential direction for future research that might address this and other limitations of the present approach is to attempt to move beyond simply dividing trials into post-correct and post-error categories, but rather to model dynamic adaptation of these and other parameters on a trial-by-trial basis (e.g. Gunawan et al., in press). This has the potential to provide a more comprehensive account of not only error-related sequential effects but also sequential effects of other types (e.g. stimulus and response related).

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s42113-022-00137-2.

Availability of Data and Material https://osf.io/4uc9z/

Code Availability https://osf.io/4uc9z/

Author Contribution Rationale development and scholarly research was conducted KD, PW, and AH.
Composition and writing were shared by KD and AH.
PES analysis was conducted by PW.
Modelling in R was conducted by AH.
Interpretation was led by AH and shared by KD and PW.
Editing was shared by KD, PW, and AH.

Funding KD is supported by a Department of Education and Training Australian Postgraduate Award Scholarship. AH is supported by an Australian Research Council Discovery project (DP120102907).

Declarations

Ethics Approval Data used in this study comes from two published studies.

Consent to Participate Not applicable

Consent for Publication Not applicable

Conflict of Interest The authors declare no competing interests.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

References

Alexander, W. H., & Brown, J. W. (2010). Computational models of performance monitoring and cognitive control. *Topics in Cognitive Science, 2*(4), 658–677. https://doi.org/10.1111/j.1756-8765.2010.01085.x

Boag, R., Strickland, L., Loft, S., & Heathcote, A. (2019a). Strategic attention and decision control support prospective memory in a complex dual-task environment. *Cognition, 191*, 1–24. https://doi.org/10.1016/j.cognition.2019.05.011

Boag, R., Strickland, L., Heathcote, A., Neal, A., & Loft, S. (2019b). Cognitive control and capacity for prospective memory in simulated air traffic control. *Journal of Experimental Psychology: General, 148*, 2181–2206.

Bogacz, R., Brown, E., Moehlis, J., Holmes, P., & Cohen, J. D. (2006). The physics of optimal decision making: A formal analysis of models of performance in two-alternative forced-choice tasks. *Psychological Review, 113*(4), 700.

Brooks, S. P., & Gelman, A. (1998). General methods for monitoring convergence of iterative simulations. *Journal of Computational and Graphical Statistics, 7*(4), 434–455.

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

Damaso, K., Williams, P., & Heathcote, A. (2020). Different types of errors and post-error changes. *Psychonomic Bulletin & Review*.

Danielmeier, C., Eichele, T., Forstmann, B. U., Tittgemeyer, M., & Ullsperger, M. (2011). Posterior medial frontal cortex activity predicts post-error adaptations in task-related visual and motor areas. *The Journal of Neuroscience, 31*(5), 1780–1789. https://doi.org/10.1523/jneurosci.4299-10.2011

Donkin, C., Brown, S. D., & Heathcote, A. (2009). The overconstraint of response time models: Rethinking the scaling problem. *Psychonomic Bulletin & Review, 16*(6), 1129–1135. https://doi.org/10.3758/pbr.16.6.1129

Donkin, C., Brown, S. D., & Heathcote, A. (2011). Drawing conclusions from choice response time models: a tutorial using the linear ballistic accumulator. *Journal of Mathematical Psychology, 55*, 140–151.

Donkin, C. B., & Brown, S. D. (2018). Response times and decision-making. In E.-J. Wagenmakers (Ed.), *Stevens' Handbook of Experimental Psychology and Cognitive Neuroscience* (4th ed., Vol. 5)

Dutilh, G., Vandekerckhove, J., Forstmann, B. U., Keuleers, E., Brysbaert, M., & Wagenmakers, E.-J. (2012a). Testing theories of post-error slowing. *Attention, Perception, & Psychophysics, 74*(2), 454–465. https://doi.org/10.3758/s13414-011-0243-2

Dutilh, G., van Ravenzwaaij, D., Nieuwenhuis, S., van der Maas, H. L. J., Forstmann, B. U., & Wagenmakers, E.-J. (2012b). How to measure post-error slowing: A confound and a simple solution. *Journal of Mathematical Psychology, 56*(3), 208–216. https://doi.org/10.1016/j.jmp.2012.04.001

Dutilh, G., Forstmann, B. U., Vandekerckhove, J., & Wagenmakers, E.-J. (2013). A diffusion model account of age differences in posterror slowing. *Psychology and Aging, 28*(1), 64–76. https://doi.org/10.1037/a0029875

Greifeneder, R., Bless, H., & Pham, M. T. (2010). When do people rely on affective and cognitive feelings in judgment? A review. *Personality and Social Psychology Review, 15*(2), 107–141. https://doi.org/10.1177/1088868310367640

Gunawan, D. E., Hawkins, G., Kohn, R., Tran, M. N., & Brown, S. D. (in press). Time-evolving psychological processes over repeated decisions. *Psychological Review.*

Hajcak, G., & Simons, R. F. (2002). Error-related brain activity in obsessive–compulsive undergraduates. *Psychiatry Research, 110*(1), 63–72. https://doi.org/10.1016/S0165-1781(02)00034-3

Hajcak, G., McDonald, N., & Simons, R. F. (2003). To err is autonomic: Error-related brain potentials, ANS activity, and post-error compensatory behavior. *Psychophysiology, 40*(6), 895–903. https://doi.org/10.1111/1469-8986.00107

Heathcote, A., & Hayes, B. (2012). Diffusion versus linear ballistic accumulation: Different models for response time with different conclusions about psychological mechanisms? *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale, 66*(2), 125–136. https://doi.org/10.1037/a0028189

Heathcote, A., Lin, Y.-S., Reynolds, A., Strickland, L., Gretton, M., & Matzke, D. (2019). Dynamic models of choice. *Behavior Research Methods, 51*(2), 961–985. https://doi.org/10.3758/s13428-018-1067-y

King, J. A., Korb, F. M., von Cramon, D. Y., & Ullsperger, M. (2010). Post-error behavioral adjustments are facilitated by activation and suppression of task-relevant and task-irrelevant information processing. *The Journal of Neuroscience, 30*(38), 12759–12769. https://doi.org/10.1523/jneurosci.3274-10.2010

Klauer, K. C. (2010). Hierarchical multinomial processing tree models: A latent–trait approach. *Psychometrika, 75*, 70–98.

Laming, D. R. J. (1968). *Information theory of choice-reaction times*. Academic Press.

Laming, D. R. J. (1979). Choice reaction performance following an error. *Acta Psychologica, 43*(3), 199–224. https://doi.org/10.1016/0001-6918(79)90026-X

Leite, F. P., & Ratcliff, R. (2010). Modeling reaction time and accuracy of multiple-alternative decisions. *Attention, Perception, & Psychophysics, 72*(1), 246–273. https://doi.org/10.3758/app.72.1.246

Luce, R. D. (1986). *Response times: Their role in inferring mental organization*. Oxford University Press, Clarendon Press.

Matzke, D., Dolan, C. V., Batchelder, W. H., & Wagenmakers, E.-J. (2015). Bayesian estimation of multinomial processing tree models with heterogeneity in participants and items. *Psychometrika, 80*, 205–235.

Nieuwenhuis, S., Ridderinkhof, K. R., Blom, J. O. S., Band, G. P. H., & Kok, A. (2001). Error-related brain potentials are differentially related to awareness of response errors: Evidence from an antisaccade task. *Psychophysiology, 38*(5), 752–760. https://doi.org/10.1111/1469-8986.3850752

Notebaert, W., Houtman, F., Opstal, F. V., Gevers, W., Fias, W., & Verguts, T. (2009). Post-error slowing: An orienting account. *Cognition, 111*(2), 275–279. https://doi.org/10.1016/j.cognition.2009.02.002

Osth, A. F., Bora, B., Dennis, S., & Heathcote, A. (2017). Diffusion vs. linear ballistic accumulation: Different models, different conclusions about the slope of the zROC in recognition memory. *Journal of Memory and Language, 96*, 36–61. https://doi.org/10.1016/j.jml.2017.04.003

Purcell, B. A., & Kiani, R. (2016). Neural mechanisms of post-error adjustments of decision policy in parietal cortex. *Neuron, 89*(3), 658–671. https://doi.org/10.1016/j.neuron.2015.12.027

Rabbitt, P. M. A. (1966). How old and young subjects monitor and control responses for accuracy and speed. *Journal of Experimental Psychology, 71*, 264–272.

Rabbitt, P., & Rodgers, B. (1977). What does a man do after he makes an error? An analysis of response programming. *Quarterly Journal of Experimental Psychology, 29*(4), 727–743. https://doi.org/10.1080/14640747708400645

Rae, B., Heathcote, A., Donkin, C., Averell, L., & Brown, S. (2014). The hare and the tortoise: Emphasizing speed can change the evidence used to make decisions. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 40*(5), 1226.

Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review March, 85*(2), 59–108.

Ratcliff, R. (2008). The EZ diffusion method: Too EZ? *Psychonomic Bulletin & Review, 15*(6), 1218–1228. https://doi.org/10.3758/PBR.15.6.1218

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., & Rouder, J. N. (1998). Modeling response times for two-choice decisions. *Psychological Science, 9*(5), 347–356.

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., Voskuilen, C., & Teodorescu, A. (2018). Modeling 2-alternative forced-choice tasks_ Accounting for both magnitude and difference effects. *Cognitive Psychology, 103*, 1–22. https://doi.org/10.1016/j.cogpsych.2018.02.002

Roberts, I. D., & Hutcherson, C. A. (2019). Affect and decision making: Insights and predictions from computational models. *Trends in Cognitive Sciences, 23*(7), 602–614. https://doi.org/10.1016/j.tics.2019.04.005

Rouder, J. N., & Haaf, J. M. (2019). A psychometrics of individual differences in experimental tasks. *Psychonomic Bulletin and Review, 26*(2), 452–467. https://doi.org/10.3758/s13423-018-1558-y

Schiffler, B. C., Bengtsson, S. L., & Lundqvist, D. (2017). The sustained influence of an error on future decision-making. *Frontiers in Psychology, 8*(1077). https://doi.org/10.3389/fpsyg.2017.01077

Schouten, J. F., & Bekker, J. A. M. (1967). Reaction time and accuracy. *Acta Psychologica, 27*, 143–153. https://doi.org/10.1016/0001-6918(67)90054-6

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., & van der Linde, A. (2014). The deviance information criterion: 12 years on. *Journal of the Royal Statistical Society: Series B (Statistical Methodology), 76*(3), 485–493. https://doi.org/10.1111/rssb.12062

Stone, M. (1960). Models for choice-reaction time. *Psychometrika, 25*(3), 251–260.

Strickland, L., Loft, S., Remington, R. W., & Heathcote, A. (2018). Racing to remember: A theory of decision control in event-based prospective memory. *Psychological Review, 125*, 851–887.

Turner, B. M., Sederberg, P. B., Brown, S. D., & Steyvers, M. (2013). A method for efficiently sampling from distributions with correlated dimensions. *Psychological Methods, 18*(3), 368–384.

Ullsperger, M., Danielmeier, C., & Jocham, G. (2014). Neurophysiology of performance monitoring and adaptive behavior. *Physiological Reviews, 94*(1), 35–79. https://doi.org/10.1152/physrev.00041.2012

van Ravenzwaaij, D., Donkin, C., & Vandekerckhove, J. (2017). The EZ diffusion model provides a powerful test of simple empirical effects. *Psychonomic Bulletin & Review, 24*, 547–556. https://doi.org/10.3758/s13423-016-1081-y

van Veen, V., & Carter, C. S. (2006). Error detection, correction, and prevention in the brain: A brief review of data and theories. *Clinical EEG and Neuroscience, 37*(4), 330–335. https://doi.org/10.1177/155005940603700411

Wagenmakers, E.-J., Van Der Maas, H. L. J., & Grasman, R. P. P. P. (2007). An EZ-diffusion model for response time and accuracy. *Psychonomic Bulletin & Review, 14*(1), 3–22. https://doi.org/10.3758/BF03194023

Wessel, J. R. (2018). An adaptive orienting theory of error processing. *Psychophysiology, 55*(3), e13041. https://doi.org/10.1111/psyp.13041

White, C. N., & Poldrack, R. A. (2014). Decomposing bias in different types of simple decisions. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 40*(2), 385–398. https://doi.org/10.1037/a0034851

White, C. N., Ratcliff, R., Vasey, M. W., & McKoon, G. (2010). Using diffusion models to understand clinical disorders. *Journal of Mathematical Psychology, 54*(1), 39–52. https://doi.org/10.1016/j.jmp.2010.01.004

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

Williams, P., Heathcote, A., Nesbitt, K., & Eidels, A. (2016). Post-error recklessness and the hot hand. *Judgment and Decision Making, 11*(2), 174–184.

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