American Psychological Association

Journal of Experimental Psychology: Applied

ISSN: 1076-898X

http://dx.doi.org/10.1037/xap0000272

An Evidence Accumulation Model of Perceptual Discrimination With Naturalistic Stimuli

Hector Palada
The University of Queensland

Rachel A. Searston
The University of Adelaide

Annabel Persson and Timothy Ballard
The University of Queensland

Matthew B. Thompson
Murdoch University

Evidence accumulation models have been used to describe the cognitive processes underlying performance in tasks involving 2-choice decisions about unidimensional stimuli, such as motion or orientation. Given the multidimensionality of natural stimuli, however, we might expect qualitatively different patterns of evidence accumulation in more applied perceptual tasks. One domain that relies heavily on human decisions about complex natural stimuli is fingerprint discrimination. We know little about the ability of evidence accumulation models to account for the dynamic decision process of a fingerprint examiner resolving if 2 different prints belong to the same finger or different fingers. Here, we apply a dynamic decision-making model—the linear ballistic accumulator (LBA)—to fingerprint discrimination decisions to gain insight into the cognitive processes underlying these complex perceptual judgments. Across 3 experiments, we show that the LBA provides an accurate description of the fingerprint discrimination decision process with manipulations in visual noise, speed-accuracy emphasis, and training. Our results demonstrate that the LBA is a promising model for furthering our understanding of applied decision-making with naturally varying visual stimuli.

Public Significance Statement

The article demonstrated that the linear ballistic accumulator (LBA) model was able to accurately capture decision-making processes in a fingerprint discrimination task that used naturalistic fingerprints bearing a strong resemblance to those encountered in the field. This finding was robust across manipulations of visual noise, speed-accuracy emphasis, and training. Encouragingly, the article suggests the LBA can provide further insight and understanding into decision-making processes involving naturalistic stimuli that are encountered in practice.

Keywords: evidence accumulation, linear ballistic accumulator decision models, fingerprint discrimination, perceptual expertise

Supplemental materials: http://dx.doi.org/10.1037/xap0000272.supp

Evidence accumulation models provide a detailed description of the processes underlying rapid human decision making (Donkin & Brown, 2018). The models assume that a decision-maker samples evidence from the environment until a threshold amount of evidence is accumulated, at which point an overt response is triggered. Unlike "static" decision models, such as signal detection (Green & Swets, 1966), evidence accumulation models account for choice probability and response times, as well as the interaction between the two. The appeal of an evidence accumulation approach to modeling decisions is that it can reveal novel facets of the decision process beyond raw response time, accuracy, and other static measures of performance. While these models have

Hector Palada, School of Psychology, The University of Queensland; Rachel A. Searston, School of Psychology, The University of Adelaide; Annabel Persson and Timothy Ballard, School of Psychology, The University of Queensland; Matthew B. Thompson, Discipline of Psychology, Murdoch University.

Hector Palada was supported by an Australian Government Research Training Program Scholarship. Timothy Ballard was supported by an ARC Discovery Early Career Researcher Award (DE180101340). Hector Palada and Rachel A. Searston contributed equally to this work. Timothy Ballard and Matthew B. Thompson shared senior authorship. Raw data files, experiment code, data analysis script, and output are available on the Open Science Framework (https://osf.io/kyp4b/).

Correspondence concerning this article should be addressed to Hector Palada, School of Psychology, The University of Queensland, Saint Lucia, QLD 4072, Australia. E-mail: hector.palada@uqconnect.edu.au

traditionally been applied to two-choice discrimination tasks where the stimuli vary artificially along a single dimension, such as motion or orientation detection tasks (for a review see Ratcliff & McKoon, 2008), their success has generated interest in potential applications beyond the lab. For instance, how these models fare with decisions about highly variable, multidimensional, or noisy stimuli encountered in applied areas such as air traffic control, medical diagnosis, or forensic science is unknown.

Recent efforts have shown that accumulation models can account for behavior on tasks with considerable complexity. For example, this class of model has been applied to air traffic control conflict detection tasks (Loft, Bolland, Humphreys, & Neal, 2009; Vuckovic, Kwantes, Humphreys, & Neal, 2014), unmanned aerial vehicle simulation target detection tasks (Palada, Neal, Vuckovic, Martin, Samuels, & Heathcote, 2016), and medical image decision-making tasks (Trueblood et al., 2018). These prior studies have probed the generality of evidence accumulation models using controlled stimuli captured by simple decision rules (e.g., simulated targets) to natural, high-dimensional stimuli comprising distributed features that are much more difficult to define (e.g., medical images). Extending standard response time models (e.g., the LBA and diffusion model) even further afield will help to reveal how closely each parameter tracks decision processes with different kinds of stimuli, tasks, and participant samples.

Here we apply the linear ballistic accumulator model (LBA; Brown & Heathcote, 2008) to a fingerprint discrimination task where participants are faced with accumulated evidence about naturalistic stimuli that bear a strong resemblance to forensic fingerprint evidence encountered in the field. No such model has been used to describe how people arrive at a decision about the identity or source of forensic specimens—fingerprints, handwriting, firearms, or hair—despite the analogous "evidence" accumulation nature of such forensic investigative tasks. Across three experiments, we examine whether the LBA can provide a coherent account of decision-making in a task that closely resembles the process of deciding whether a crime scene print matches a suspect or not. Our first experiment examines the ability of the LBA to account for patterns of choices and response times in a fingerprint discrimination task. Our second experiment examines the effects of standard manipulations such as speed-accuracy emphasis and stimulus difficulty to test whether the model parameters are sensitive to factors they should be, according to theory. Our third experiment examines how a training intervention affects the cognitive processes underlying finger discrimination, as reflected by the model parameters. In the following sections, we review previous literature on fingerprint discrimination, describe how fingerprints could benefit from evidence accumulation modeling, and illustrate the application of the LBA to fingerprints.

Fingerprint Discrimination

Fingerprint discrimination is done by humans, not computers. When a fingerprint is found at a crime scene a human examiner—often a police officer—compares the print to a known suspect or to a list of candidate prints. They position two prints side-by-side, physically or on a computer screen, and visually compare them to judge whether they came from the same finger or different fingers. Compared with novices, these examiners are remarkably accurate at discriminating prints (Tangen, Thompson, & McCarthy, 2011), particularly when they are highly similar in the eyes of a computer algorithm (Thompson, Tangen, & McCarthy, 2013b). Experts tend to show a conservative response bias in these experiments, such that they err on the side of making more errors that could allow a guilty person to escape detection than errors that could falsely incriminate an innocent person (Thompson, Tangen, & McCarthy, 2013a).

Other experiments on the nature of fingerprint expertise have revealed that fingerprint experts can make quick and relatively accurate decisions in noise and under time pressure compared with novices (Thompson & Tangen, 2014). They can also maintain this expertise across different fingerprint tasks that bear less of a resemblance to their daily work. For example, they can tell if two prints were left by different fingers of the same individual more accurately than novices (Searston & Tangen, 2017c), and they show even greater expertise in distinguishing common fingerprint patterns such as "loops" and "whorls" in a search task (Searston & Tangen, 2017b). However, their expertise also appears to be constrained by their specific set of experiences. For example, a longitudinal investigation of the development of expertise with fingerprints revealed that examiners' performance on a series of fingerprints tasks improved as they accumulated 12 months of formal training with fingerprints, but their performance on a series of analogous tasks with inverted face stimuli did not improve over the same period of time (Searston & Tangen, 2017a).

Signal detection models have been the predominant class of model used to understand decisions in these fingerprint discrimination experiments (Searston, Tangen, & Eva, 2016; Thompson et al., 2013a). These models have allowed researchers to examine factors that influence decision parameters such as response bias (i.e., whether people are biased toward responding "same finger" vs. "different finger") and discriminability (i.e., the ability of the decision maker to discriminate same-finger prints vs. different-finger prints). For instance, trainees with 5 weeks to 6 months experience can be more conservative in their responding (i.e., tended to say "different" more frequently) than novices, who displayed a liberal bias when comparing fingerprints (i.e., tended to say "same finger" more on the same task; Thompson et al., 2013b). Fingerprint experts demonstrate superior discriminability compared with novices irrespective of their response bias in other two-alternative choice fingerprint tasks (Searston & Tangen, 2017b, 2017c). Signal detection models have also been used to show how contextual information (e.g., case reports of crimes that vary in severity) can influence response bias on a fingerprint matching task without necessarily affecting discriminability (Searston et al., 2016).

The application of signal detection theory to fingerprint discrimination decisions has provided insights into how experts differ from novices, how context can sway people's decisions, and how people's discriminability can vary with experience. Despite the value of this model for helping to understand fingerprint discrimination, there are aspects of the underlying decision process that remain obscured, such as speed–accuracy trade-offs (Vuckovic et al., 2014). That is, signal detection models are static models that do not make use of the response time distribution and so they do not account for the dynamics of the decision process, how it might change under different conditions. An increase in hits together with a decrease in false alarms, for example, is typically interpreted as an increase in discriminability. This pattern of results, however, could also be caused by increases in the response threshold that determines when the evidence accumulated for a response alternative is sufficient to trigger an overt response (e.g., Palada, Neal, Tay, & Heathcote, 2018). Response thresholds, in turn, may change under conditions that emphasize speed and accuracy in different ways. These dynamics can be explored using evidence accumulation models, which consider both response times and accuracy.

Evidence Accumulation Models

The term "evidence accumulation model" refers to a class of cognitive model that describes the process underlying rapid human decision making. These models share the basic assumption that the decision maker accumulates evidence for response alternatives until a threshold amount of evidence is reached, at which point an overt response is triggered. Although modern evidence accumulation models often yield similar conclusions about the decision process regardless of which model is considered, the models make different assumptions about the underlying architecture (Donkin, Brown, Heathcote, & Wagenmakers, 2011).

Evidence accumulation models have developed over the last 50 years with architectural extensions added over time to account for more complex empirical observations (see Donkin & Brown, 2018 for a more detailed review). Early models assumed that a single accumulator indexed the difference in evidence for each choice followed a random walk process; response time variability was captured by the moment-to-moment change in evidence (Stone, 1960). Subsequent models used a continuous accumulation process, or a diffusion process, and added trial-to-trial variability in the evidence accumulation start point and the mean rate of evidence accumulation. These added sources of variability captured differences in response times for correct and error response times (Laming, 1968; Ratcliff, 1978). The current diffusion model has also added trial-to-trial variability in nondecision processes to account for variability in the fastest response times across conditions (Ratcliff & McKoon, 2008). Multiple accumulator models have also been proposed, which assume that evidence accumulates independently for each response alternative(s). The leaky accumulator model (Usher & McClelland, 2001) assumed response competition between accumulators and within-trial randomness. The latter assumption was subsequently omitted in the simplified ballistic accumulator model (BA; Brown & Heathcote, 2008). The LBA (Brown & Heathcote, 2008) is a further simplification of the BA model, as it assumes linear evidence accumulation (Brown & Heathcote, 2008).

The success of evidence accumulation models stems from their ability to provide insights into the latent cognitive processes that underlie choice. The models are typically applied to tasks using simple stimuli that produce rapid decision times ($<$1.5 s), such as memory recognition ("was it recently encountered?"), motion discrimination ("is it moving to the left or right?"), and lexical decisions ("is it a word or nonword?"). The pairing of evidence accumulation models with these types of tasks has facilitated our understanding of a range of issues, including sleep deprivation (Ratcliff & Van Dongen, 2011), schizophrenia (Heathcote, Suraev, Curley, Love, & Michie, 2015), and anxiety (Ho et al., 2014).

There is recent evidence that the LBA can be used to account for decisions in tasks that are inspired by applied domains where extended mean response times are common ($\sim$2.5 s; e.g., Palada et al., 2018). For example, Strickland et al. (2019) and Boag, Strickland, Heathcote, Neal, and Loft (2019) used the LBA to understand the processes underlying prospective memory in a maritime surveillance task and air traffic control task, respectively. The LBA has also been used to understand how individuals adapt to time pressure in an applied multistimulus environment (Palada et al., 2018), and in a dual-task environment with multiattribute stimuli (Palada, Neal, Strayer, Ballard, & Heathcote, 2019; Palada et al., 2016). However, these studies used artificial or controlled stimuli that were quite removed from their inspirations in the field. That is, these studies opted for a high level of control to ensure that the tasks were amenable to evidence accumulation models. For example, the maritime surveillance task used by Strickland et al. (2019) used basic target features and classification rules, and the air traffic control task used in Boag et al. (2019) only presented one pair aircraft pair per trial. In this article, we examine whether the LBA can provide a coherent account of decisions in the domain of fingerprint discrimination using a task that bears a strong resemblance to the task in the field where the stimuli are far less controlled.

The LBA model architecture is shown in Figure 1. In fingerprint discrimination, a person must decide whether two fingerprints come from the same finger or from different fingers. In this case, the model includes two evidence accumulators, each corresponding to a different response alternative (same finger vs. different finger). The LBA assumes that the level of evidence in each accumulator at the start of the decision process is sampled from a uniform distribution with a lower boundary of zero and an upper boundary of A. The rate of evidence accumulation within a trial is sampled from a normal distribution with mean v and standard deviation sv. The rate of evidence accumulation for each accumulator depends on the stimulus. For example, in cases where the two fingerprints come from the same finger, the mean rate of evidence accumulation will be higher for the accumulator corresponding to the same finger response than for the accumulator corresponding to the different finger response. All else being equal, this means that the same finger response is more likely to be triggered. A response is triggered when the level of evidence in an accumulator reaches the threshold (b) associated with that response alternative. As the two accumulators' race toward their respective response threshold, the winning accumulator will make the corresponding response (i.e., a same finger response in Figure 1). The response threshold may be the same between the different alternatives (as in Figure 1), or it may differ. Finally, the model includes a nondecision time parameter (T_{er}). Observed response times are accounted for by first partitioning the response times into nondecision time and decision time components. Nondecision time reflects the sum of times for stimulus encoding and response production. The decision time, and observed responses, are reflected by the inputs and operation of the evidence accumulation model; hence, the two evidence accumulators racing toward the threshold.

The LBA can be used to quantify the latent cognitive processes underlying decision making. The model captures response caution (i.e., the amount of information required to reach a decision) and response bias via the response threshold parameter. Lower response thresholds produce quicker, more error prone responses, because less evidence is feeding into the final decision. Higher response thresholds produce slower, but more accurate responses,

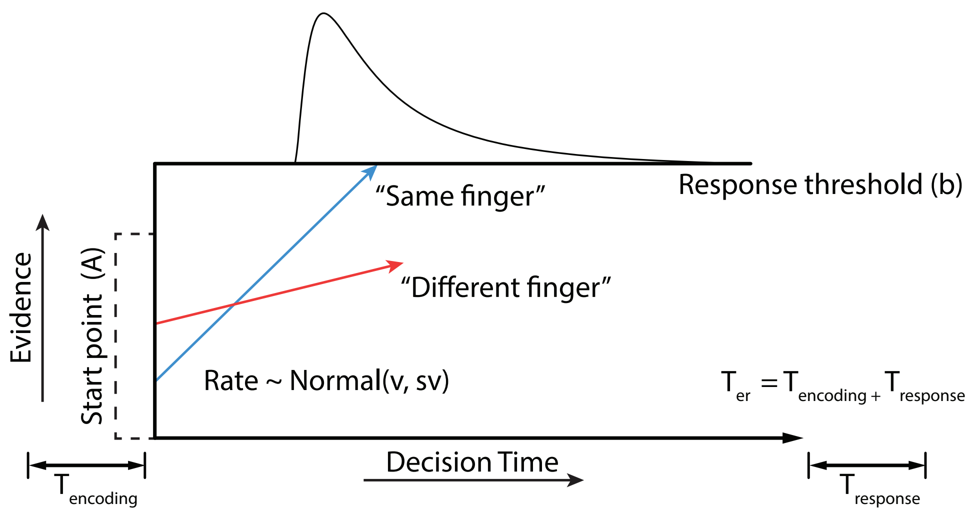

Figure 1. The standard linear ballistic accumulator model applied to fingerprint identification and associated parameters: Response threshold (*b*), mean rate (*v*), rate variability (*sv*), start-point (*A*), and nondecision time (*Ter*). See the online article for the color version of this figure.

because more evidence is being considered. Differences in thresholds between response alternatives produce a bias in favor of the response with the lower threshold, because less evidence is required to trigger that response than the competing response(s). The LBA captures the discriminability of the stimulus via the rate parameters. The mean rate between the two evidence accumulators can be used to account for changes in the speed of information processing and the rate variability parameters captures the heterogeneity in the quality of the stimulus and attention (Palada et al., 2018; Ratcliff, 1978).

Modeling Fingerprint Discrimination

We use three experiments to test the ability of the LBA to account for fingerprint discrimination decisions made by novices. Experiment 1 provides an initial test of whether the LBA is able to account for the patterns of choices and response times using an experiment in which the only manipulation is whether the stimuli were from the same finger versus different fingers. Experiment 1, to our knowledge, provides the first application of evidence accumulation models to fingerprint discrimination. Experiment 2 examines whether the parameters of the LBA are sensitive to manipulations of speed-accuracy emphasis and stimulus noise. We use these traditional benchmark manipulations to test whether the parameters of the model can be interpreted in a meaningful way, consistent with the interpretations made in studies involving other types of decisions (Donkin & Brown, 2018). Experiment 3 uses the LBA to examine the cognitive processes underlying the effects of training on fingerprint discrimination.

We analyzed data using Bayesian estimation. Under the Bayesian framework, the statistical conclusions made are not dependent on the stopping rule during data collection (Rouder, 2014). As our emphasis is in cognitive modeling, the most important issue for power is number of trials per participant rather than the number of participants (Kolossa & Kopp, 2018; Smith & Little, 2018). The number of trials per participant we used is consistent with the recommended sample size for reliable parameter estimation using the LBA (Visser & Poessé, 2017). At the same time, our sample sizes are larger than the typical "psychophysical" design discussed by Smith and Little (2018; e.g., four subjects), even though we had high power in terms of low measurement error because of a large number of trials per participant.

Experiment 1

In Experiment 1, we take the first step in examining how the LBA can be used to understand fingerprint discrimination decisions. Specifically, we test whether the model can account for the empirical patterns of choices and response times when individuals are faced with the task of discriminating whether two fingerprints originated from the same finger versus different fingers.

Method

Participants and stimuli. Thirty-six psychology undergraduate students (28 women and 8 men; mean age = 20.19 years, SD = 4.61) from The University of Queensland participated in this experiment for course credit. The experiment was approved by the University of Queensland Human Research Ethics Committee. The stimuli included 195 fingerprint trios, each consisting of a simulated crime scene print, a fully rolled print from the same finger, and a fully rolled print from a different finger. The different prints were sampled randomly for each participant from a pool of 195 prints collected from different individuals. The images were sourced from the Forensic Informatics Biometric Repository (Thompson et al., 2013a), and we cropped them to 512 × 512 pixels, isolating the print in the center of the image. Ninety-six fingerprint pairs (48 same-finger prints and 48 different-finger prints) were generated for each participant. A random sample of 48 crime-scene prints, sampled from the pool of 195, were paired with their corresponding same print, and a separate set of 48 crime-scene prints, sampled randomly from the remaining pool of 147, were paired with their corresponding different print.

Procedure. After reading an information sheet about the experiment and watching an instructional video with an example same-finger trial and different-finger trial, we presented participants with 96 pairs of fingerprints, one pair at a time. Participants were instructed to judge whether the two prints belong to the *same* finger, or two *different* fingers. They provided their judgments by pressing the "Z" or "/" key. On pressing one of the two response keys or after 10 s, the two prints disappeared, with a 1 s interval before the next pair were displayed. Participants were encouraged, in the instructional video, to respond within 10 s; not doing so recorded an "NA" response. We collected participants' keypresses and time to respond on each trial.

We randomly varied whether a "Z" keypress indicated a match (i.e., same-finger prints) or mismatch response (i.e., different-finger prints). A random selection of participants were instructed to respond by pressing "Z" if they thought the two prints were from the same finger (and "/" if they thought they were from different fingers), and the remaining participants responded by pressing "/" if they thought the two prints were from the same finger (and "Z" if they thought they were from different fingers). Each participant's particular response-key arrangement was reflected in the instructional video they viewed at the beginning of the experiment. We also displayed a 1376 × 848 images of the keyboard at the bottom center of the computer screen throughout the experiment with the labels "Same" and "Different" in bold, black text above the two corresponding response keys. The keyboard image was semitransparent, other than the two response keys and the labels (see Figure 2 for a screenshot of what participants see on a given trial).

Results

Censoring of rapid RTs was not required as the fastest observed reaction time (RT; 311ms) was sufficient to make a valid decision. Extended RTs were not censored so that we could account for the entire distribution of valid RT performance (i.e., RTs ranging 250 ms and the 10 s response deadline were included in the analysis). We removed three participants with a high nonresponse rates (>10%), where one participant had a 20% nonresponse rate, and two participants both had a 27% nonresponse rate. The remaining participants had 10% or less missing data.

Discrimination performance. We examined the effect of the stimulus type manipulation on the RTs for correct responses, the RTs for incorrect responses, and accuracy. These analyses were run using R (R Development Core Team, R Core Team, 2016), using the *brms* package to conduct Bayesian generalized linear mixed-models (GLMM; Bürkner, 2017). Priors were consistent across outcome variables (Cauchy prior; location = 0, scale = 2) for fixed effects and random intercepts; the latter was zero-truncated. Other priors used the default specifications in *brms*. We present the posterior estimates of regression coefficients (b) with two-tailed 95% credible intervals (CIs) presented in square brackets. We infer credible effects where the CI does not cross zero (Kruschke, 2014).

Correct and incorrect RTs quantiles (0.1, 0.5, and 0.9), and mean accuracy are shown in Figure 3. Stimulus type did not have a credible effect on correct RTs ($b = -0.02$ [-0.04, 0.00]; same-finger prints = 3.15 s vs. different-finger prints = 3.27 s) or on incorrect RTs ($b = 0.03$ [0.00, 0.06]; 3.64 vs. 3.43 s). There was a credible effect of stimulus type on accuracy, such that participants were less accurate when responding to same-finger prints compared with different-finger prints ($b = -0.11$ [-0.19, -0.04]; 67 vs. 71%).

LBA modeling and analysis. To apply the LBA, the researcher must specify which model parameters should be allowed to vary with experimental factors and accumulator related factors. Table 1 outlines the model parameterizations and the total number

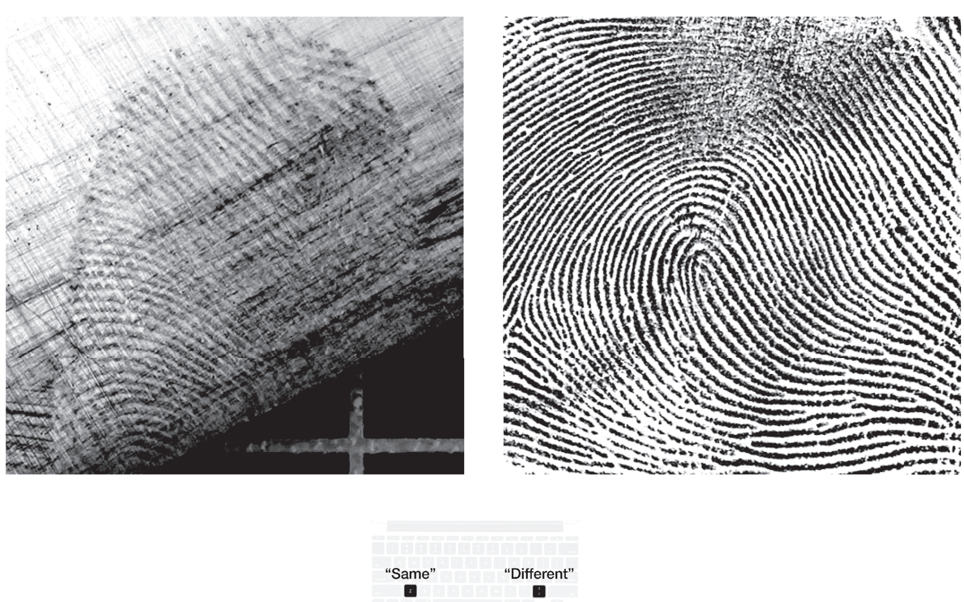

Figure 2. A screenshot of a single trial. Participants in all three experiments were presented with the same basic visual display of two fingerprints (same-finger prints in this instance) and a keyboard response map on each trial.

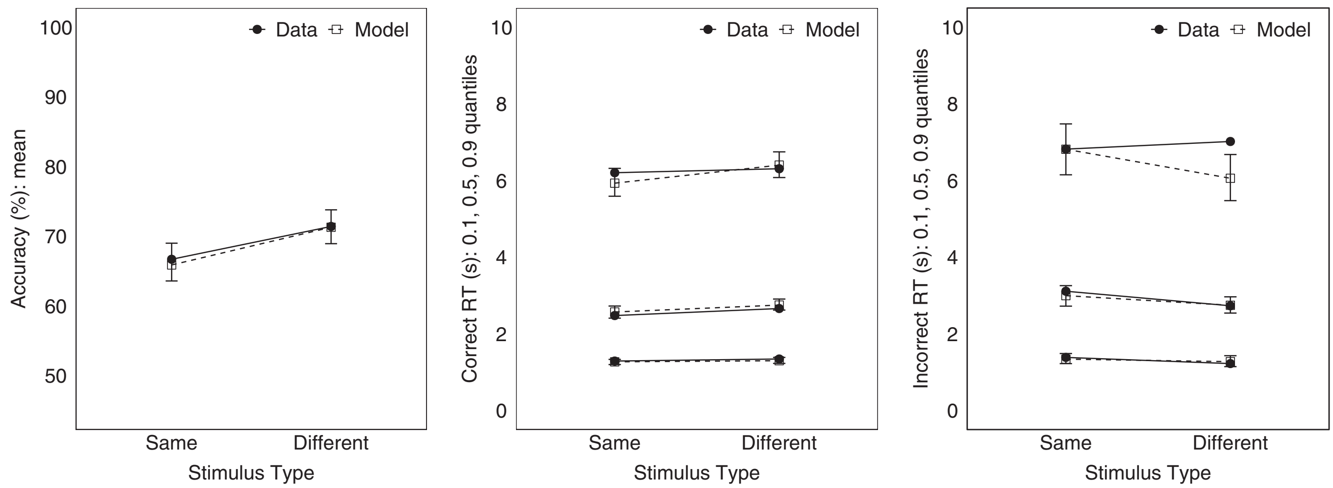

Figure 3. Experiment 1: Fits of the linear ballistic accumulator (LBA) to mean accuracy (left graph), and quantiles of correct (middle graph) and incorrect response times (right graph) of the fingerprint discrimination task. The bars show the 95% quantiles of the posterior predictives.

of parameters for the three experiments. For Experiment 1, mean rate (*v*) is allowed to vary depending on the stimulus factor (same-finger prints vs. different-fingers prints). We allow mean rate to vary with a "match" factor, which reflects whether the response that the accumulator triggers "matched" the stimulus presented in the trial. In other words, the match factor indicates whether the accumulator corresponded to the correct response for the presented stimulus. The accumulator for the correct response is referred to as the "matching" accumulator, because the response it triggers "matches" the stimulus. The accumulator for the incorrect response is referred to as the "mismatching" accumulator, because the response it triggers mismatches the stimulus. We allow rate variability (*sv*) to vary by the match and stimulus factors, with the latter hypothesized to account for heterogeneity in the features between same-finger prints and different-finger prints. We allow threshold (*b*) to vary by the response factor (same finger vs. different finger) so that we can account for any biases in responding. As is common practice (e.g., Brown & Heathcote, 2008), we express response caution as the difference between the raw threshold and the maximum starting point of evidence (denoted *B*). This provides a pure measure of caution that is not contaminated by individual differences in starting point variability. Starting point variability (*A*) and nondecision time (*t0*) are constrained to a single estimated value. To ensure that the model is identifiable, we constrain the rate variability for the mismatching accumulator for different-finger trials to 1.0 (Donkin, Brown, & Heathcote, 2009).

We implemented the LBA using a hierarchical Bayesian framework, which assumes that parameters vary across individuals but are drawn from common population distributions. The individual-level parameters were modeled using normal or truncated normal distributions. The *A*, *B*, and *sv*, parameters had a lower bound of 0 and no upper bound. The *t0* parameter had a lower bound of 0.1 and an upper bound of 1. The *v* parameters were unbounded. The population distribution for each individual-level parameter requires two population-level parameters: location and scale. The priors of the location and scale parameters were modeled using normal or truncated normal distributions. The details of the priors are provided in online supplemental materials. The posterior distributions were estimated using the differential evolution MCMC algorithm (Turner, Sederberg, Brown, & Steyvers, 2013), as implemented by the Dynamic Models of Choice package in R (Heathcote, Lin, Reynolds, Strickland, Gretton, & Matzke, 2019; see online supplemental materials for information about the number of chains and assessing convergence).

As can be seen in Figure 3, the model provides a good fit to mean accuracy for both types of stimuli. The model also provides

Table 1
Linear Ballistic Accumulator Model Parameterization for the Three Experiments

Experiment	Model	Start point (A)	Threshold (B)	Rate (v)	Rate variability (s_v)	Nondecision time (t_{er})	Total
Experiment 1	Factors	—	R	S, M	S, M	—	11
	NF	1	2	4	3	1	
Experiment 2	Factors	—	SA, R	SA, N, M, S	S, M	—	25
	NF	1	4	16	3	1	
Experiment 3	Factors	—	F, B, R	F, B, M, S	F, B, S, M	—	21
	NF	1	4	8	7	1	

Note. Experimental factors are stimulus type (S), speed-accuracy emphasis (SA), noise (N), feedback (F), and block (B). Accumulator factors are match (M) and response (R). "Total" refers the total number of parameters in the model; NF = number of factors. In Experiment 3, feedback was manipulated between-person; we fit the two feedback groups separately; therefore, feedback factor (F) does not contribute to the number of model parameters.

a good fit to the RT distribution for correct and incorrect responses. There is some evidence of underestimation of RTs for the 0.9 quantile, with data falling just outside of the credible interval of the model predictions. However, such overestimation at the higher quantiles is common when modeling complex tasks (e.g., Palada et al., 2018, 2016).

We next examined the posterior estimates of model parameters to examine how they accounted for the patterns observed in performance data, which are illustrated in Figure 4. To make inferences about parameters, and account for the uncertainty in posteriors, we create group-averaged posterior distributions by averaging every posterior sample across participants. To test for differences between parameters, we calculated the difference between parameters for each condition, posterior sample, and participant, and then averaged over participants to produce a group average posterior distribution on the difference in parameters. To test for an interaction between two factors with two levels each, we calculated the difference between levels for one factor for each of the two levels for the other factor; thus, producing two contrasts. We then calculated the difference between the two contrasts (Heathcote et al., 2019; Palada et al., 2019).

There was a credible difference in the threshold for the two response types, 0.25 [0.14, 0.37], with participants setting a lower threshold for the different finger response than the same finger response. The differences between mean rate for the matching and mismatching accumulators was credibly greater for different-finger prints than same-finger prints, 1.30 [1.10, 1.51]. There was a credible interaction between stimulus type and match factor on variability in the rate of evidence accumulation, −0.99 [−1.13, −0.84], such that for same-finger prints, the rate variability for the matching (i.e., correct) accumulator was greater than the rate variability for the mismatching (i.e., incorrect) accumulator, whereas the opposite occurred for different-finger prints, though to a lesser extent. In the LBA, greater rate variability for the matching accumulator can account for the typically observed slow error response times. In contrast, greater variability in the mismatching accumulator compared with the matching accumulator can account for relatively fast errors (Heathcote & Love, 2012), and as shown in Figure 3, errors were relatively faster for different-finger prints compared with same-finger prints.

Under the LBA architecture, greater evidence quality can be driven by a greater difference between mean rates between the two accumulators, differences in rate variability, or both. Figure 5 plots the rate of evidence accumulation using the median of the posterior estimates of mean rates and rate variability. The distributions illustrate the analogues of signal and noise distributions from signal detection theory (SDT) and, therefore, allows for an inference of the analogues of sensitivity (i.e., d'), as well as the causes of any differences in sensitivity. The graph illustrates how sensitivity was greater for different-finger prints compared with same-finger prints because of the greater differences in means, and the lesser variability for the matching accumulator, though this was slightly offset by having greater variability for the mismatching accumulator.

Overall, the greater accuracy for different-finger compared with same-finger prints was driven by the greater difference between the two accumulators, as well as the bias to respond different finger. While the bias to respond different finger would typically result in faster response times for different-finger prints than same-finger prints, as the accumulator has a shorter distance to reach the threshold, this appeared to be offset by two separate factors. First, in the case of the correct RTs, the matching (i.e., correct) accumulator of the different-finger prints had a less variable rate than the accumulator of the same-finger prints. Second, in the case of error response times, the mismatching (i.e., incorrect) accumulator of the different-finger prints had a mean rate that was lower than the accumulator of the same-finger prints.

Discussion

To our knowledge, this experiment was the first application of the LBA to examine performance observed in a fingerprint discrimination task. We found that the model provided a good fit of the observed accuracy and full distribution of correct and incorrect RTs. Moreover, the LBA revealed aspects of the decision-making processes that could not be inferred from the performance data alone.

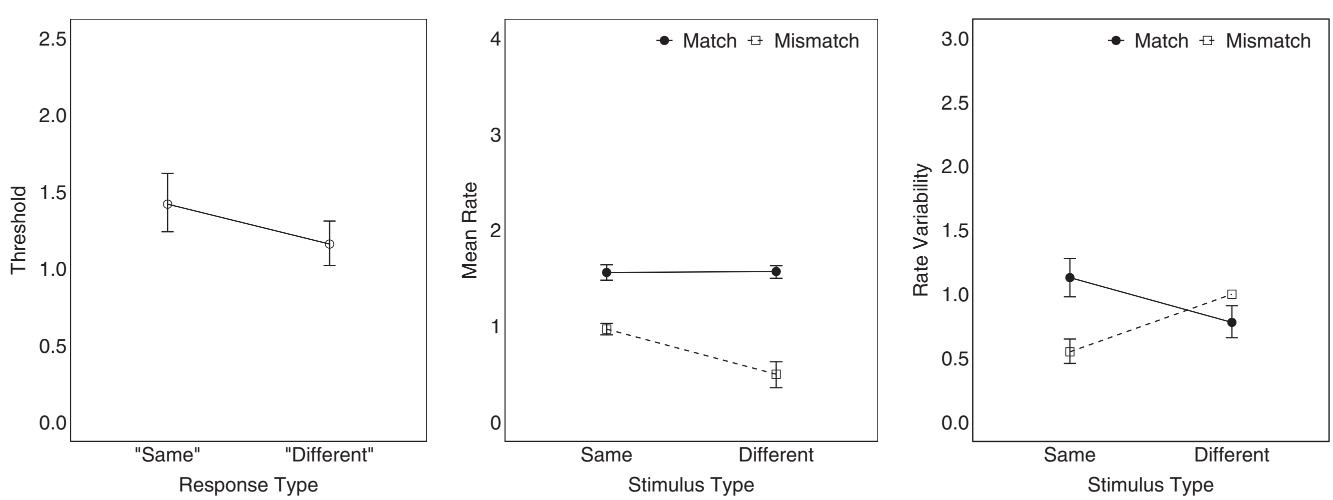

Figure 4. Experiment 1: Median parameters estimates for the linear ballistic accumulator (LBA) model, including response caution (left panel), mean rate (middle panel), and rate variability (right panel). The bars show the 95% credible intervals.

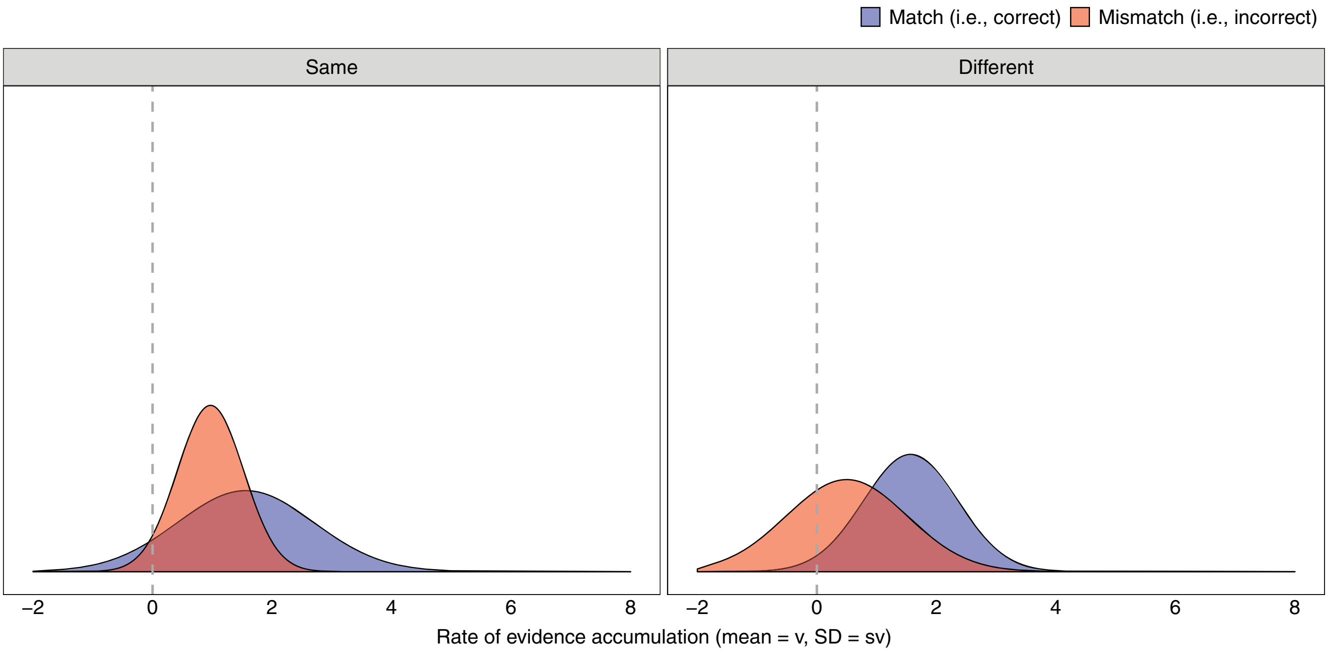

Figure 5. Experiment 1: Distribution plots of rates of evidence accumulation for matching and mismatching accumulators. Columns refer to stimulus type. Distributions were generated using posterior medians. Vertical line reflects zero truncation of rate sampling. See the online article for the color version of this figure.

Participants required more evidence before deciding that prints were left by the same finger, compared with deciding they were left by different fingers, indicating that participants were biased toward responding different. The bias to respond different is contrary to previous studies applying signal detection theory to fingerprint discrimination (Searston et al., 2016; Thompson et al., 2013a). However, this prior research used highly similar different-finger prints that were intended to be challenging for both novices and expert. We did not use highly similar different-finger prints. Rather, we used dissimilar different-finger prints by randomly pairing fingerprints so that we would have a sufficient number of different-finger trials for model fitting. As a result, the different-finger prints appeared more different than prints used in previous research. We also used simulated crime scene prints that result in considerable variability among same-finger prints because of distortion, pressure, partial contact, and surface type. Therefore, our same-finger prints tended to appear different because of the different features that were visible. These features of our design together may have contributed to the different response being more salient in our paradigm.

The model revealed that participants had greater discriminability for different-finger prints than same-finger prints. The model also revealed greater variability in the rate of evidence accumulation of the matching accumulator for same-finger prints compared with different-finger prints, which suggests that the processing of same-finger prints was more heterogenous than different-finger prints. Our use of dissimilar different-finger prints and crime-scene prints may explain why we observed these effects. Specifically, identifying different-finger prints can be done effectively and with little variability because the individual could easily identify dissimilar prints of fingerprints. In contrast, identifying same-finger prints could require additional and more variable processing because the individual needs to conduct an exhaustive evaluation of the within-finger differences to rule them out as merely contextual differences that resulted from them being crime-scene prints.

Experiment 2

Experiment 1 provided initial evidence that the LBA can account for performance in a fingerprint discrimination task. However, adequate model fit to observed data is insufficient to establish model validity. To build stronger evidence that the LBA can be used to understand fingerprint discrimination, we next examine whether the model parameters change in predictable ways in response to experimental manipulations and our theoretical understanding of their effects (Donkin & Brown, 2018). In Experiment 2, we examine the interpretability of model parameters by manipulating emphasis type (speed vs. accuracy) and noise (no noise vs. noise). We expect that emphasis type will affect parameters associated with the termination of the decision process itself (i.e., response threshold), whereas noise will affect parameters associated with the inputs to the decision process (i.e., rate of evidence accumulation). However, as we describe in the LBA modeling and analysis section, and consistent with current modeling procedures, we test for the possibility that emphasis type also affects the rate of evidence accumulation (Palada et al., 2016; Rae, Heathcote, Donkin, Averell, & Brown, 2014).

Method

Participants and stimuli. A second group of 70 psychology undergraduates (36 women and 34 men) with an average age of 19.71 (SD = 5.01) from The University of Queensland participated in Experiment 2 in return for course credit. The image-set was the same used in Experiment 1. We duplicated this entire set of images and added 20% of artificial noise or "speckle" to the duplicates, such that there was a noise and no noise "twin" of every image in the set (see Figure 6). For each participant, we then randomly sampled a total of 192 fingerprint trios, with a random half of these (96 trios) taken from the no noise set, and the other half taken from the duplicated noisy set. That is, the fingerprints were randomly and equally sampled from the noise or no noise sets for each participant. For each of the noise and no noise image sets, a random 48 crime-scene prints were paired with their corresponding same-finger print, and 48 crime-scene were paired with their corresponding different-finger print.

Procedure. The pre-experiment procedure was the same as in Experiment 1 with participants watching an instructional video describing the task and walking through an example same-finger trial and different-finger trial. We then presented participants with the 192 fingerprint pairs, one at a time, split into four blocks: Noisy prints with speed prompts, noisy prints with accuracy prompts, no noise prints with speed prompts, and no noise prints with accuracy prompts. We presented these blocks in a randomized order for each participant. For the two blocks emphasizing speed, we presented participants with "Speed Up" in black text during the intertrial interval when they responded after 5 s of exposure to the prints. Participants also viewed a short video before each of the speed blocks instructing them to respond as quickly as possible for the next series of cases. For the two blocks emphasizing accuracy, participants were instructed to respond as accurately as possible, and were presented with "Slow Down" in black text during the intertrial interval each time they responded before 5 s had elapsed. The 5 s boundary was informed by the response time data observed in Experiment 1, where 80% of responses were faster than 5 s. We used this 80% criterion as the time window for prompting fast or slowed responses. As in Experiment 1, participants were instructed to judge whether the two prints belong to the same finger or two different fingers using "Z" and "/" as response keys. The procedure for advancing through the trials was also the same as Experiment 1, with a 10 s response window and a 1 s interval between trials.

Results

We removed one participant with a nonresponse rate of 17.28%. The nonresponse rates for the remaining 69 participants ranged be-

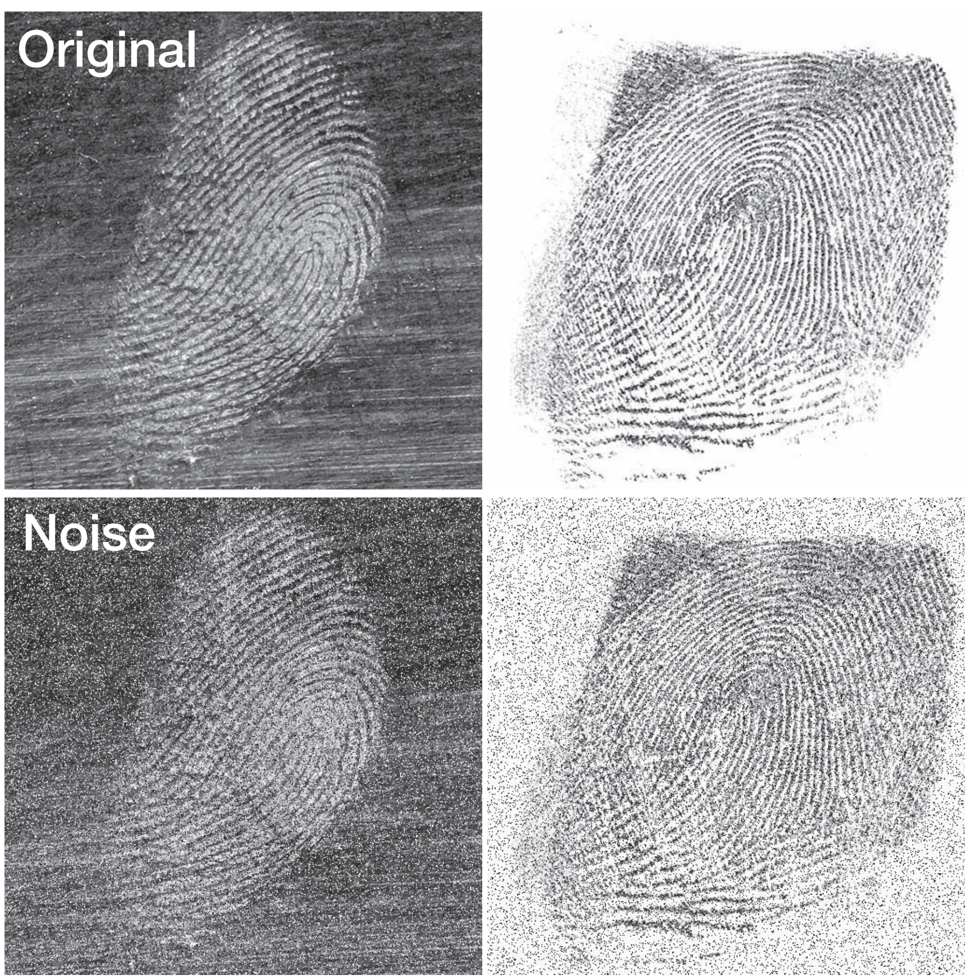

Figure 6. Experiment 2: An example same-finger prints as original without any noise (top), and with 20% noise added to the images (bottom).

tween 0.00 and 8.85%. Consistent with modeling of simple-choice tasks, we censored RTs less than 250 ms (0.75% of the data); more rapid RTs would be insufficient to make a valid decision.

Discrimination performance. We used the same statistical procedures as outlined in Experiment 1 to analyze discrimination performance (correct and incorrect RTs, and accuracy). Predictors included stimulus type, noise (no noise vs. noise), and emphasis type (speed vs. accuracy). Statistical results are shown in Table 2. Correct and incorrect RTs quantiles (0.1, 0.5, and 0.9), and mean accuracy are shown in Figure 7.

As expected, there was a credible main effect of emphasis type on correct RTs and incorrect RTs, such that responses were considerably faster under speed emphasis compared with accuracy emphasis (correct: 2.10 vs. 5.31 s; incorrect: 2.20 vs. 5.60 s). While there was a credible two-way interaction between noise and emphasis type on correct RTs, the effect was not substantial; the difference in correct RTs between no-noise and noisy prints under speed emphasis was 0.14 s, whereas the difference under accuracy emphasis was 0.08 s. Incorrect RTs were credibly faster for different-finger prints compared with same-finger prints (3.77 vs. 3.90 s). There was a credible interaction between stimulus type and noise on accuracy; for same-finger prints, accuracy declined for prints with noise compared with print without noise, whereas noise did not affect accuracy for different-finger prints. Finally, there was a credible, but weak interaction between stimulus type and emphasis-type on accuracy; whereas accuracy did not differ between emphasis types for same-finger prints, accuracy slightly improved for different-finger prints under accuracy emphasis compared with speed emphasis.

LBA modeling and analysis. Selective influence—the idea that the effect of an experimental manipulation is captured by a single model parameter—has been important for the benchmarking of evidence accumulation models. A fundamental assumption of selective influence is that the effects of speed-accuracy emphasis is captured by the changes in the response caution (i.e., threshold). However, this assumption has recently been challenged. Using basic laboratory tasks, for example, Rae et al. (2014) found that as expected, individuals reduced their response threshold under a speed emphasis. They also found an increase in mean rate under accuracy emphasis and concluded that prioritizing accuracy produces an increase in the quality of information accumulated from the stimulus. Palada et al. (2018), however, have found some evidence for selective influence in complex tasks. They show that distinct determinants of time pressure had selective influence on cognitive processes; for example, in Study 2 the number of stimuli influenced threshold, whereas the time available influenced the rate at which information was accumulated.

In Experiment 2 we test whether the parameters respond to the experimental manipulations in a theoretically coherent way. We allow emphasis type to influence both threshold and mean rate. Initially, we also sought to test whether noise would affect threshold and mean rate. However, this model was unstable, so we had to simplify the model by constraining threshold across levels of the noise manipulation. We believe this assumption is reasonable because previous research has found that noise does not influence threshold. For example, Palada et al. (2016) used a similar noise manipulation and found that it affected inputs to the decision process (i.e., mean rate), but not the response threshold. The fit of the model after making this simplifying assumption was still very good. The model parameterization was otherwise consistent with Experiment 1, such that rate variability was influenced by the match factor and stimulus type, where nondecision time and start point variability were estimated across experimental conditions (see Table 1).

We used the same modeling approach as Experiment 1. First, the model provided a good fit to mean accuracy and the distribution of correct and incorrect RTs (see Figure 7). Second, we examine the posterior estimates of model parameters to understand how the model explains the effects of noise and emphasis type manipulations on discrimination performance. The parameter estimates are shown in Figure 8. Table 3 presents the effects of the experimental manipulations and the accumulator related factors on model parameters.

Emphasis type had a credible effect on threshold, such that participants had a considerably higher threshold under accuracy emphasis compared with speed emphasis. There was a credible three-way interaction between stimulus type, noise, and emphasis type on the differences between mean rate for the matching and mismatching accumulators. Under accuracy emphasis, the difference in mean rates was greater for different-finger prints compared with same-finger prints, and this effect was stronger for prints without noise compared with prints with noise. The same effect was observed under speed emphasis for prints with noise. For prints without noise, there was no difference in mean rates between same-finger prints and different-finger prints. Finally, there was a credible two-way interaction between stimulus type and the match factor on the variability in the rate of evidence accumulation. Rate variability for the matching (i.e., correct) accumulator was greater

Table 2
Experiment 2: Bayesian Generalized Linear Mixed-Model Coefficients for Discrimination Task Performance Measures

Effect	Correct RT	Error RT	Accuracy
S	−0.01 [−0.02, 0.00]	0.02 [0.01, 0.03]	−0.13 [−0.17, −0.10]
N	0.01 [0.00, 0.02]	0.00 [−0.01, 0.01]	−0.10 [−0.14, −0.07]
SA	−0.48 [−0.49, −0.47]	−0.50 [-0.51, −0.49]	−0.04 [−0.07, 0.00]
S.N	0.01 [0.00, 0.02]	−0.01 [−0.02, 0.01]	−0.08 [−0.11, −0.04]
S.SA	−0.01 [−0.02, 0.00]	0.00 [−0.01, 0.01]	0.04 [0.00, 0.07]
N.SA	0.02 [0.01, 0.02]	0.01 [0.00, 0.03]	0.02 [−0.02, 0.05]
S.N.SA	0.01 [0.00, 0.02]	−0.01 [−0.02, 0.01]	−0.01 [−0.05, 0.03]

Note. RT = response times. Experimental effects are stimulus type (S), noise (N), and emphasis type (SA); 95% credible intervals (CIs) are in square brackets.

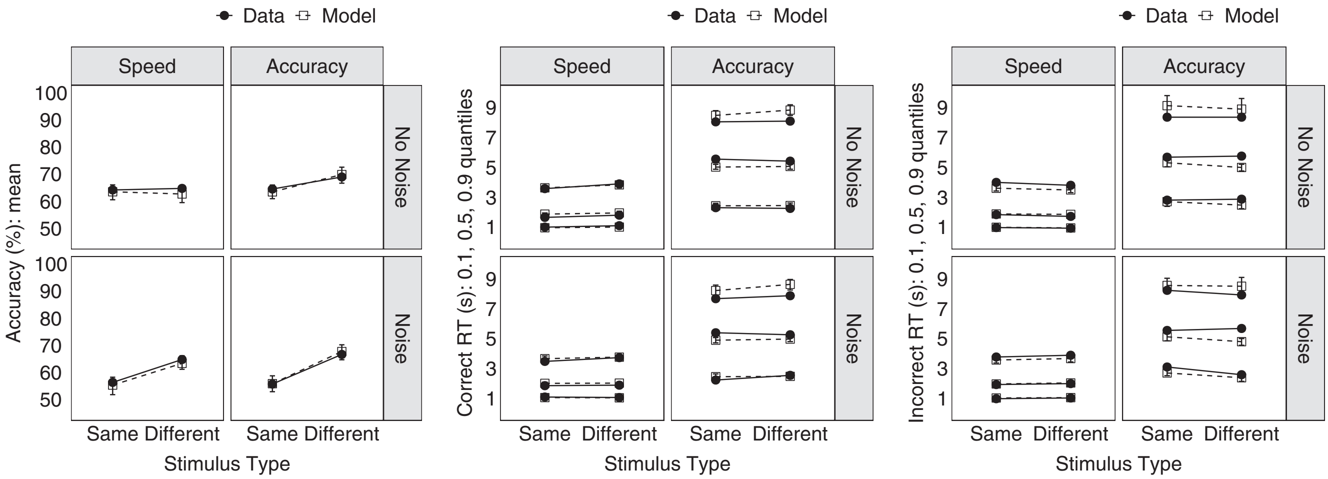

Figure 7. Experiment 2: Fits of the linear ballistic accumulator (LBA) to mean accuracy (left graph), and quantiles of correct (middle graph) and incorrect response times (right graph) of the fingerprint discrimination task. The bars show the 95% quantiles of the posterior predictives.

than the rate variability for the mismatching (i.e., incorrect) accumulator for same-finger prints, whereas opposite effect occurred for different-finger prints.

The distribution of rates of evidence accumulation is shown in Figure 9. The graph shows how sensitivity did not appear to vary across emphasis and noise conditions for different-finger prints. Unexpectedly, for same-finger prints presented without noise, sensitivity was greater for prints in the speed condition compared with the accuracy condition; this was driven by the difference in mean rates between the two accumulators. From the graphs it is clear that the decrease in sensitivity for same-finger prints presented with noise compared with prints without noise was driven by the decrease in the differences between the two accumulators.

Overall, different-finger prints had greater accuracy compared with same-finger prints when prints were presented with noise, or when prints were presented without noise and under accuracy emphasis. The model captures this by the greater difference in mean rates between the two accumulators for different-finger prints compared with same-finger prints. However, the stimuli had comparable accuracy when presented without noise and under speed emphasis. The model also captures this by the comparable difference in mean rates between the two accumulators for both stimulus types. Emphasis type mainly affected response times and had a weak effect on accuracy. The considerably faster response times under speed emphasis compared with accuracy emphasis was captured by two mechanisms: The lower threshold under speed emphasis, as well as the greater overall rate of evidence accumulation.

Discussion

In this experiment, our aim was to determine whether the LBA could provide an accurate description of the effects of emphasis type (speed or accuracy) and visual noise (no noise vs. noise) on fingerprint discrimination decisions. We examined whether the relationships between model parameters and experimental manipulations would coincide with findings from simpler choice tasks. We hypothesized that emphasis type would influence parameters concerned with the termination of the decision process itself (i.e., threshold), whereas visual noise would influence the inputs to the decision process (i.e., the rate of evidence accumulation). Our hypothesis regarding noise was supported. Our hypothesis regarding emphasis type was partially supported—in line with recent evidence (e.g., Rae et al., 2014), we found that emphasis type affected mean rate in addition to affecting threshold.

We found that emphasis type affected response caution, or the quantity of evidence required to reach a decision. Participants reduced their response caution under speed emphasis compared with accuracy emphasis to accelerate the decision process, as the accumulator has a shorter distance to travel and reaches the threshold sooner. Contrary to Experiment 1, we did not find evidence of a bias to respond different. We suspect that participants were not willing to adopt a different response bias under the experimental conditions as the bias would exacerbate the relatively poor performance for same-finger prints in this experiment, which was worse than in Experiment 1.

In line with recent studies (Rae et al., 2014), emphasis type also influenced mean rate. The overall mean rate of evidence accumulation, as indexed by the mean of the matching and mismatching accumulators was considerably higher under speed emphasis compared with accuracy emphasis. This suggests that participants processed evidence faster under the speed emphasis. This finding is consistent with Palada et al. (2018) who found that individuals increased their rate of evidence accumulation under tighter deadlines. Contrary to Rae et al. (2014), we did not find evidence that the quality of evidence, as indexed by the difference between the two accumulators, increased under accuracy emphasis compared with speed emphasis. However, we provide evidence for Rae et al.'s conclusion that whether emphasis type primarily affects the speed or quality of evidence accumulation depends on whether the manipulation largely affects responses times or accuracy, respectively.

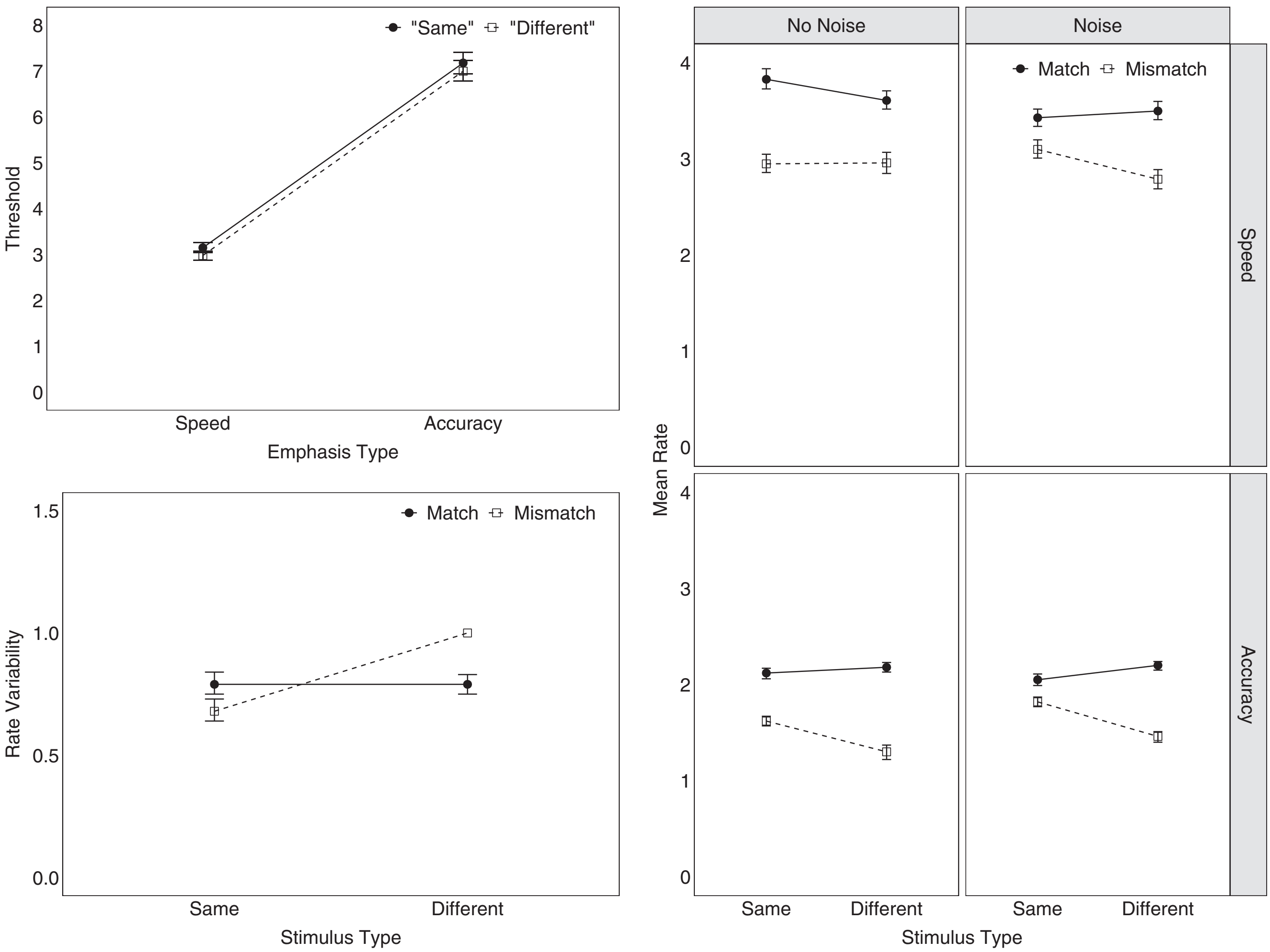

Figure 8. Experiment 2: Median parameters estimates for the linear ballistic accumulator (LBA) model, including response caution (top left panel), mean rate (right panel), and rate variability (bottom left panel). The bars show the 95% credible intervals.

As expected, we found that noise negatively influenced discriminability, though this only occurred for same-finger prints. The decrease in discriminability was driven by the difference in mean rates between the two accumulators and captured the decline in accuracy for same-finger prints presented with noise compared with prints presented without noise. In contrast, noise did not affect the discriminability of different-finger prints, which explains why no change in accuracy was observed between noise conditions. In line with our interpretation of the results from Experiment 1, we suspect that the between-finger differences inherent in the materials used in this experiment explain why we did not observe an effect of noise on different-finger discriminability. Specifically, the asymmetrical effects of noise on discriminability may have resulted from the noise manipulation being sufficiently strong to obscure similarities among same-finger prints (thereby decreasing discriminability for these prints), but not sufficiently strong to obscure differences between different-finger prints (thereby failing to reduce discriminability for these prints).

Overall, Experiment 2 suggests that the LBA model parameters can be interpreted meaningfully in the context of fingerprint discrimination. Consistent with studies using basic tasks (e.g., Rae et al., 2014), emphasis type influenced both response caution and rate of evidence accumulation (albeit in the opposite direction to what some previous research has found), whereas noise selectively influenced discriminability.

Experiment 3

In Experiment 1 and 2, we showed that the LBA can accurately capture fingerprint discrimination performance, and that model parameters can be mapped meaningfully to the underlying cognitive processes they are thought to reflect. In particular, speed-accuracy emphasis influenced response caution in a manner consistent with theory and previous research. Noise also reduced discriminability, but only for same-finger prints. In Experiment 3, we examine how novices' decision-making processes evolve over time by using the LBA to quantify any change in decision processes after a feedback training intervention. There are a number of ways that training could affect decision processes. First, feedback may alter response biases as participants learn to adjust their prior

Table 3
Experiment 2: Contrasts for Experimental, Response, and Match Factor Effects on LBA Model Parameters

Effect	Threshold (B)	Difference in mean rates (v_{match} – $v_{mismatch}$)	Rate variability (sv)
S	—	−0.38 [−0.51, −0.25]	−0.16 [−0.19, −0.12]
N	—	0.27 [0.17, 0.37]	—
SA	−4.02 [−4.20, −3.85]	−0.38 [−0.54, −0.23]	—
R	−0.17 [−0.35, 0.01]	—	—
M	—	—	−0.05 [−0.09, −0.02]
S.N	—	0.13 [−0.01, 0.28]	—
S.SA	—	−0.61 [−0.85, −0.36]	—
S.M	—	—	0.32 [0.24, 0.40]
N.SA	—	−0.28 [−0.48, −0.20]	—
R.SA	−0.01 [−0.33, 0.31]	—	—
S.N.SA	—	−0.48 [−0.77, −0.09]	—

Note. LBA = linear ballistic accumulator. Factors are stimulus type (S), noise (N), emphasis type (SA), response type (R), and match type (M); 95% CIs are in square brackets; v_{match} and v_{mismatch} refer to the values of the matching and mismatching accumulators, respectively. Dashes reflect cases where the model parameter was not allowed to vary by the effect (see Table 1 for model parameterization).

expectations of stimuli. Second, feedback may improve accuracy by leading to increased response caution. Alternatively, feedback may improve accuracy by enhancing the quality of evidence that is accumulated as the participants learn to attend to more diagnostic print features—an effect that would be mediated by the rate parameters (i.e., the mean or variability in the rate of evidence accumulation). In this experiment, we test for these possible effects by allowing model parameters to vary with group type (training vs. no training) and block (pre- vs. posttraining).

Method

Participants and stimuli. A third group of 70 psychology undergraduates (48 women and 21 men) with an average age of 20.13 (SD = 3.88) from The University of Queensland participated in Experiment 3 in return for course credit. The image-set was identical to Experiment 1, and we generated 192 fingerprint pairs (96 same-finger prints and 96 different-finger prints) for each participant using the same method. There were 32 same-finger prints and 32 different-finger prints in each of the pretest, training, and posttest conditions.

Procedure. Participants were randomly assigned to either the feedback or no-feedback condition. The pre-experiment procedure was the same as in Experiment 1 and Experiment 2; participants watched an instructional video, which consisted of the task description and an explanatory example of both same-finger prints and different-finger prints. Participants were then presented with 192 pairs of fingerprints, one at a time, in three blocks of trials. In the first block, the *pretest* phase of the experiment, 64 pairs of fingerprints were presented. In the second block, the *training* phase a new set of 64 fingerprints were presented; feedback participants were presented with "Correct" in green text, or "Incorrect" in red text, corresponding to whether they correctly classified the pair of fingerprints and no feedback participants viewed a blank screen between trials. In the third block, the *posttest* phase, a new set of 64 fingerprints were presented without feedback. As in Experiments 1 and 2, participants were instructed to classify the prints as originating from the same finger with the "Z" key, or from different fingers using the "/" key.

Results

Nonresponse rates across participants ranged between 0.00 and 5.21%. No participants were removed from analyses because none had a nonresponse rate greater than 10%. However, five responses less than 250 ms were censored.

Discrimination performance. We used the same procedure as Experiment 1 to analyze Experiment 3 performance. The experimental factors included stimulus type, block (pretraining vs. posttraining), and feedback group (feedback vs. no feedback). Results are shown in Table 4 and Figure 10. There was a credible three-way interaction on accuracy. For those in the no-feedback group, accuracy was stable across blocks for both stimuli. In contrast, for those in the feedback group, accuracy improved for same-finger prints after receiving training, whereas accuracy slightly decreased for different-finger prints. There was a credible effect of block on correct and error RTs, such that RTs were faster at posttest compared with pretest (correct RTs: 2.74 vs. 3.57 s; error RTs: 3.30 vs. 4.10 s). Error RTs were also faster for different-finger prints (3.59 s) than same-finger prints (3.80 s).

LBA modeling and analysis. We used the LBA to examine how the processes underlying fingerprint discrimination change over time and in response to feedback. The model parameterization is shown in Table 1. We allow threshold and mean rate to vary with block and training, and also allow mean rate to vary between stimuli. As we are interested in the general effects of training on the inputs to the decision processes (i.e., rate parameters), we also allow rate variability to vary with the experimental factors. The estimates of nondecision time and start point variability were constrained across experimental conditions, such that only a single value was estimated for each parameter (see Table 1). We used the same modeling approach as in Experiments 1 and 2. However, because we fit models separately for each feedback group (as feedback group was manipulated between participants), nondecision time ($t0$), start point variability (A), and rate variability (sv) also varied across feedback conditions. As can be seen in Figure 10, the model provides an excellent fit to the data for both feedback groups.

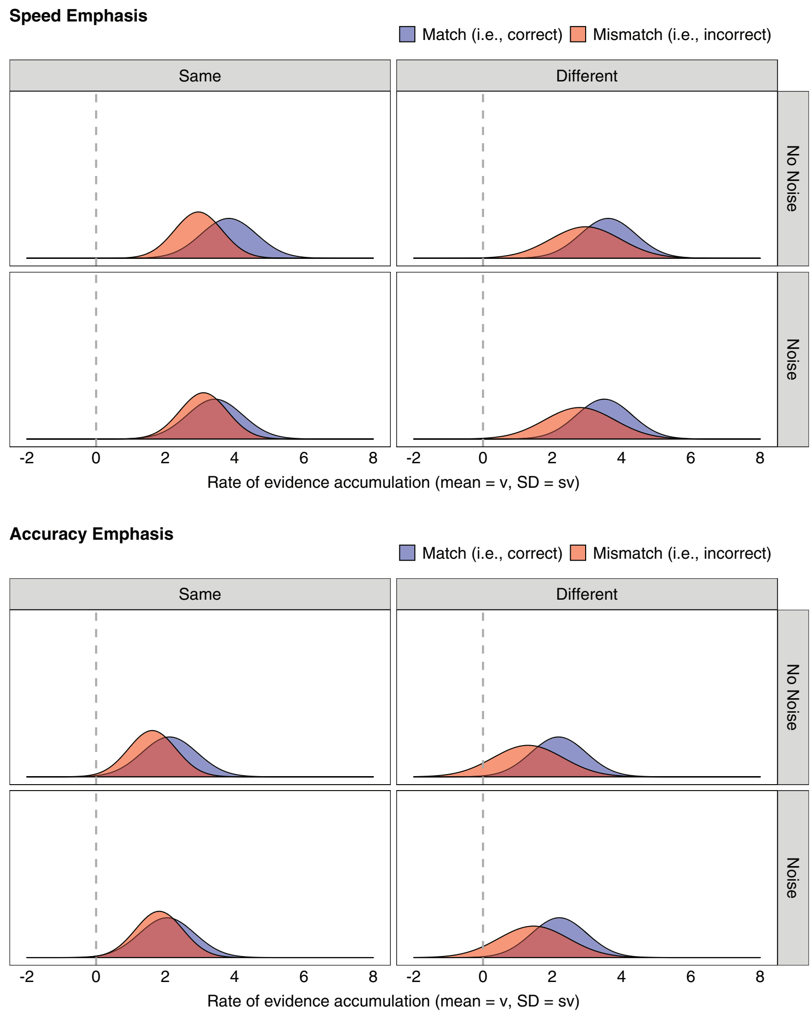

Figure 9. Experiment 2: Distribution plots of rates of evidence accumulation for matching and mismatching accumulators. The top four quadrants correspond to speed, while the bottom four quadrants correspond to accuracy emphasis. Rows alternate between prints without noise and with noise. Columns refer stimulus types. Distributions were generated using posterior medians. Vertical line reflects zero truncation of rate sampling. See the online article for the color version of this figure.

Figure 11 displays the estimates of response caution, mean rates, and variability in the rate of evidence accumulation. We used the same procedure as Experiments 1 and 2 to analyze the results, with experimental and accumulator related effects shown in Table 5. For response caution, all two-way interactions were credible, though there was an interesting trend when considering all three factors (group, block, and response type). At pretest, both groups set a lower threshold to respond different finger compared with the threshold to respond same finger. For the no-feedback group, the strength of this effect weakened at posttest, such that the threshold became more similar between the response options. For the feedback group, the effect reversed, such that participants set a higher threshold to respond same finger compared with the threshold to respond different finger. At posttest, the feedback group was also generally more cautious than the no-feedback group.

There was a credible three-way interaction between block, group, and stimulus type on the difference in mean rates. At pretest, both groups had a difference in mean rates that was higher for different-finger prints compared with same-finger prints. At posttest, after having received feedback, the difference in mean rates became more similar between the two types of stimuli, as the difference between mean rates improved for same-finger prints. For the no-feedback group, the difference in mean rates for same-finger prints did not change from pretest to posttest, whereas the difference for different-finger prints increased.

There were credible main effects of block, stimulus, and match factor on the variability in rate of evidence accumulation. Rate variability was higher at posttest compared with pretest, for different-finger prints compared with same-finger prints, and for the matching (i.e., correct) accumulator than for the mismatching (i.e., incorrect accumulator). The four-way interaction was not credible. However, Figure 11 shows that the patterns in rate variability were similar between groups at pretest, whereas there was some evidence of group differences at posttest. We tested for

Table 4
Experiment 3: Bayesian Generalized Linear Mixed-Model Coefficients for Task Performance

Effect	Correct RT	Error RT	Accuracy
S	0.01 [0.00, 0.03]	0.04 [0.02, 0.06]	−0.14 [−0.19, −0.09]
B	−0.15 [−0.16, −0.13]	−0.13 [−0.15, −0.11]	0.03 [−0.01, 0.08]
G	−0.06 [−0.12, 0.01]	−0.05 [−0.13, 0.03]	0.01 [−0.05, 0.08]
S.B	0.00 [−0.02, 0.01]	0.01 [−0.01, 0.03]	0.08 [0.03, 0.12]
S.G	0.00 [−0.01, 0.02]	0.00 [−0.02, 0.02]	0.04 [−0.01, 0.08]
B.G	−0.01 [−0.03, 0.00]	0.01 [−0.01, 0.03]	−0.01 [−0.06, 0.03]
S.B.G	−0.01 [−0.02, 0.00]	0.00 [−0.01, 0.02]	0.10 [0.05, 0.14]

Note. Experimental effects are stimulus type (S), block (B), and group type (G); 95% credible intervals (CIs) are presented in square brackets.

this by comparing posttest rate variability between accumulators for each stimulus type. For same-finger prints, the difference in rate variability between accumulators did not differ between groups (−0.05 [−0.71, 0.58]). For different-finger prints, the difference in rate variability between accumulators was greater for the feedback group compared with the nonfeedback group because of the greater match variability for the feedback group (−1.00 [−1.67, −0.34]).

The distribution of rates of evidence accumulation is shown in Figure 12. The graph shows how sensitivity was similar across groups at pretest. At posttest, any improvement in sensitivity afforded by the greater difference in means between the distributions was offset by the increase in variance. As a result, training did not appear to influence overall sensitivity. Therefore, the improvement in accuracy for same-finger prints and drop in accuracy for different-finger prints for the feedback group resulted from the reversal in response bias from pretest to posttest. The generally faster response times from pretest to posttest was captured by the increase in mean rates, as the accumulators travel fasters toward the threshold.

Discussion

In this experiment, we used the LBA to understand the cognitive processes underlying the effects of a training intervention on fingerprint discrimination. The behavioral data revealed that training improved same-finger prints accuracy compared with baseline irrespective of the considerable decrease in response times, whereas different-finger pair accuracy slightly decreased after receiving training. As expected, accuracy did not improve over trial phases for the no-feedback group, though response times decreased. Our modeling suggests that the effects of feedback on performance is mediated by a number of cognitive processes, including response caution and rate variability.

In line with Experiment 1, at pretest, both groups had a bias to respond different finger by setting a lower response threshold than the same finger response threshold. The no-feedback group maintained this response bias at the posttest phase, though to a much lesser extent. We suspect that this decline in response bias for the no-feedback group may have occurred as, over the course of the experiments, participants may have become aware of the relative frequency with which they responded different finger compared with same finger and sought to balance their response frequency between the response alternatives. The feedback group, by contrast, reversed their response bias after having received training. In this case, the feedback provided in the training phase may have made participants aware of their lower accuracy for same-finger prints compared with different-finger prints. The reversal of the response bias may reflect an attempt to compensate for this dis-

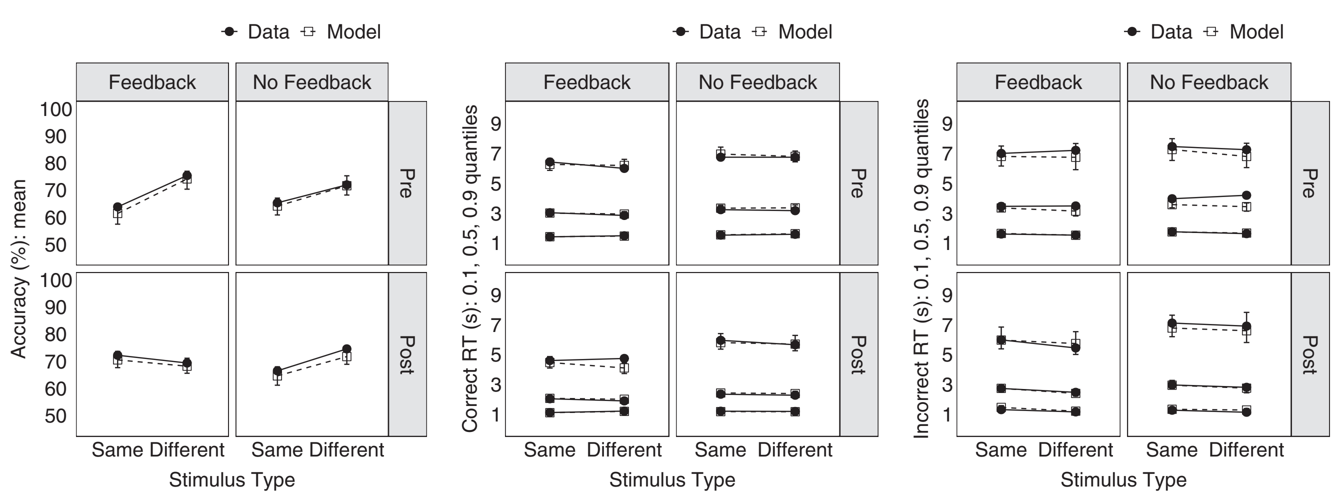

Figure 10. Experiment 3: Fits of the linear ballistic accumulator (LBA) to mean accuracy (left graph), and quantiles of correct (middle graph) and incorrect response times (right graph) of the fingerprint discrimination task. The bars show the 95% credible intervals of the posterior predictives.

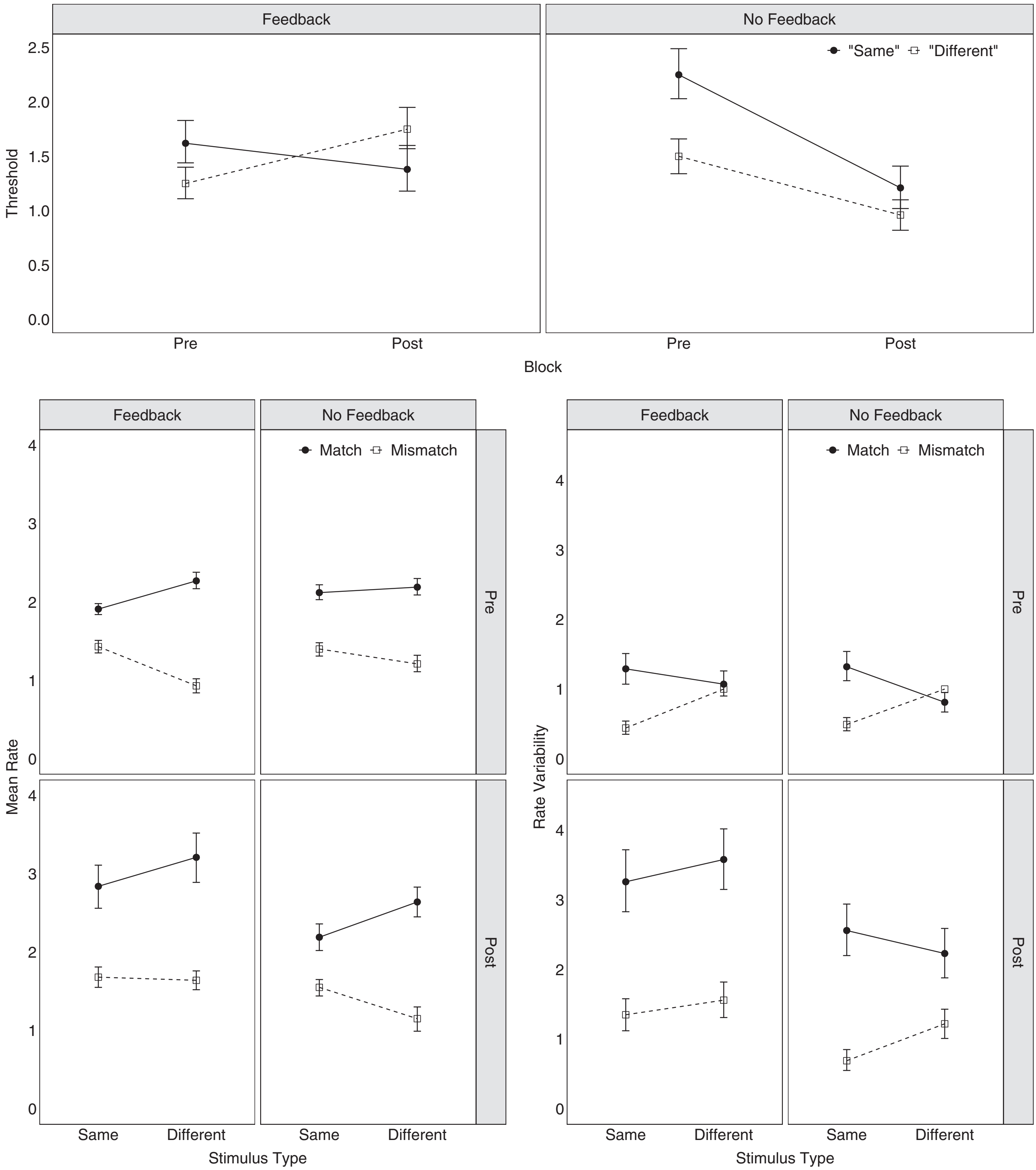

Figure 11. Experiment 3: Median parameters estimates for the linear ballistic accumulator (LBA) model, including response caution (upper-left panel), mean rate (lower-left panel), and rate variability (right panel). The bars show the 95% credible intervals.

Table 5
Experiment 3: Contrasts for Experimental, Response, and Match Factor Effects on LBA Model Parameters

Effect	Threshold (*B*)	Difference in mean rates ($v_{match} - v_{mismatch}$)	Rate variability (*sv*)
S	—	−0.63 [−0.91, −0.36]	−0.13 [−0.25, −0.02]
B	−0.62 [−0.79, −0.45]	0.29 [0.09, 0.49]	−1.12 [−1.26, −0.98]
G	0.02 [−0.18, 0.13]	−0.52 [−0.88, −0.15]	—
R	0.26 [0.13, 0.38]	—	—
M	—	—	1.04 [0.92, 1.17]
S.B	—	−0.08 [−0.39, 0.23]	—
S.G	—	−0.45 [−1.00, 0.10]	—
B.G	−1.11 [−1.45, −0.77]	−0.76 [−1.16, −0.37]	—
B.R	0.62 [0.37, 0.88]	—	—
G.R	0.50 [0.25, 0.75]	—	—
S.B.G	—	−1.06 [−1.68, −0.44]	—
R.B.G	−0.24 [−0.75, 0.25]	—	—
S.B.G.M	—	—	0.69 [−0.36, 1.75]

Note. LBA = linear ballistic accumulator. Factors are stimulus type (S), block (B), group type (G), response type (R), and match type (M); 95% credible intervals (CIs) are in square brackets; v_{match} and v_{mismatch} refer to the values of the matching and mismatching accumulators, respectively. Dashes reflect cases where the model parameter was not allowed to vary by the effect (see Table 1 for model parameterization).

crepancy in accuracy. This observation is consistent with Palada et al. (2018), who found that individuals adopted a response bias corresponding to the choice with poorer accuracy. This observation is also consistent with prior signal detection modeling of novices' fingerprint discrimination decisions (Searston et al., 2016), showing opposite patterns of response bias across experiments with and without feedback.

There was some evidence that training had an effect on the inputs to the evidence accumulation process, more specifically on the variability of evidence accumulation. Rate variability generally increased from pretest to posttest. There was a trend for this effect being stronger for the feedback group. Specifically, the rate variability for the match accumulator for different-finger prints increased after having received training, such that it became greater than the rate variability for the match accumulator of same-finger prints. It seems unlikely that this was an artifact of the stimuli, as heterogeneity was similar across pretest and posttest blocks. Rather, this pattern of results suggests that training may have led to an attempt to strategically change in the information that was attended to when discriminating prints. The process of changing strategy would have increased rate variability.[1] The change in processing also appears to have negatively influenced the effectiveness in processing different-finger prints, and this could also explain why participants reversed their response bias postfeedback from favoring a same finger finger response to a different finger finger response.

General Discussion

Our aim was to model the dynamics of decisions about whether pairs of fingerprints came from the same finger or different fingers. To better understand the underlying decision process, we modeled choices and response times on a fingerprint discrimination task using a standard evidence accumulation model—linear ballistic accumulation (Brown & Heathcote, 2008). Forensic decision-making about the source of a crime-scene trace, such as a fingerprint, relies on the perceptual sensitivity of human examiners to distinguish specimens that match from those that do not match. Signal detection models have helped to better quantify the conditions that affect forensic examiners' discriminability separately from their response bias (Searston & Tangen, 2017b, 2017c; Thompson et al., 2013a). However, forensic science is fundamentally an evidence accumulation process, and forensic examiners report taking a great deal of care and time in making their decisions so as to avoid errors. Evidence accumulation models offer an opportunity to gain new insights about the underlying cognitive processes of experts by accounting for the time course of their decisions. As a first step, we sort to test the validity of the LBA as a model of fingerprint discrimination in three experiments with novices.

The aim of Experiment 1 was to conduct an initial evaluation of the whether the LBA could account for the patterns of choice and response times on a fingerprint discrimination task. Overall, the model provided a very good account of the data, and yielded insights about the underlying decision process that would not have been clear from an inspection of the accuracy and response time data alone. Our analysis revealed, for instance, that the observed difference in accuracy between our same-finger prints and different-finger prints is likely because of the combination of two processes. First, people were biased toward classifying prints as having originated from different fingers (i.e., they had a lower threshold for the different finger response). Second, people processed information more effectively when examining prints that had originated from different fingers compared with prints from the same finger. This finding contrasts with previous studies showing a liberal response bias toward same finger decisions among novices (e.g., Searston et al., 2016; Tangen et al., 2011; Thompson et al., 2013b). Because these previous studies used highly similar distractor prints resembling those often encountered in operational contexts, this discrepancy in novices' response biases across studies may be explained by our random pairing of different-finger prints. That is, we used dissimilar different-finger prints as opposed to the similar different-finger prints used in previous research, such that our set were more distinctively different.

[1] We thank an anonymous reviewer for this insight.

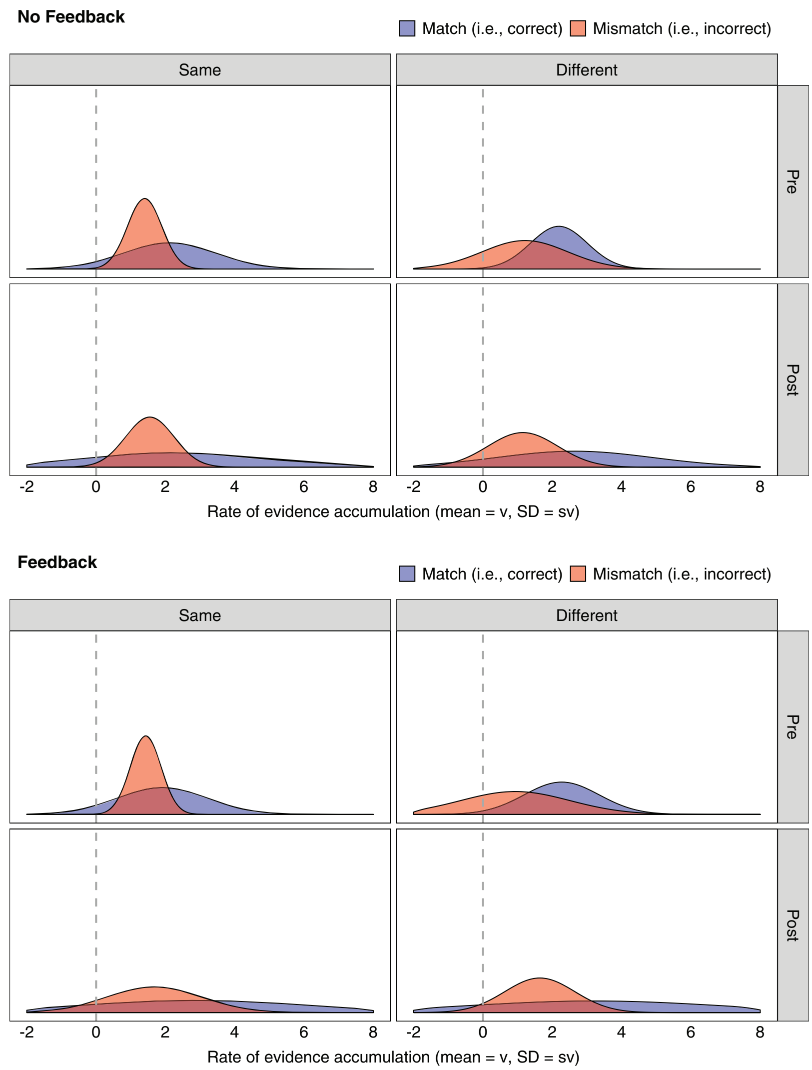

Figure 12. Experiment 3: Distribution plots of rates of evidence accumulation for matching and mismatching accumulators. The top four quadrants show the no feedback group, while the bottom four quadrants show the feedback group. Rows alternate between pre- and posttraining trials. Columns refer stimulus types. Distributions were generated using posterior medians. Vertical line reflects zero truncation of rate sampling. See the online article for the color version of this figure.

The aim of Experiment 2 was to examine how components of the decision process change when speed or accuracy are emphasized and when perceptual noise is added to the stimulus. As expected, people were more cautious in their responding (i.e., set higher response thresholds) when accuracy was emphasized compared with when speed was emphasized. We did not observe a different response bias as found in Experiment 1 and in some conditions in Experiment 3. We took this lack of difference as evidence to suggest that the bias may not be robust under more difficult or time pressured contexts, as adopting a different bias would have decreased accuracy with the same-finger prints. Indeed, this explanation is consistent with previous findings demonstrating that novice and expert examiners display a neutral response bias when discriminating fingerprints presented for just 400 ms (Searston & Tangen, 2017b).

We also replicated the finding in Experiment 1 that discriminability is higher when viewing prints that originated from different fingers. This effect was robust to the presence of noise. In contrast, and as expected, discriminability for same-finger prints decreased when noise was introduced. In line with recent evidence (e.g., Rae et al., 2014), we found that emphasis type also influenced mean rate. Because our manipulation of emphasis type affected response times considerably more than it affected accuracy, the effects of emphasis type manifested through the overall rate at which evidence was accumulated, rather than the differences in rates of evidence accumulation between the matching and mismatching accumulators.

For prints that had originated from the same finger, discriminability decreased under accuracy emphasis compared with speed emphasis. This latter result was unexpected and may highlight a unique feature of crime-scene prints. The additional processing afforded under accuracy emphasis may negatively influence the discrimination of prints of the same finger because the individual further processes contextual information (e.g., distortion or pressure) and erroneously takes this as evidence for a different-finger pair. Under speed emphasis, the individual may only process information that is at least partially diagnostic of the same-finger prints, and so accuracy is less affected compared with when accuracy is emphasized. This conclusion is consistent with Rae et al. (2014), who suggest that emphasis type may influence the kind of information accumulated during the decision process. Speed emphasis, for instance, may focus attention on the more global or distributed characteristics of fingerprints (e.g., Searston & Tangen, 2017c) that can be gleaned quickly at a glance or in noise (e.g., Thompson & Tangen, 2014). Accuracy emphasis, on the other hand, may focus attention more on the finer features such as the "minutiae" of a fingerprint. Directly examining the nature of the information that is extracted while processing fingerprints is a promising avenue for future research. This approach may involve examining how participants change the way in which they process prints (e.g., holistic vs. partial processing) or which features of the prints are attended to.

The aim of Experiment 3 was to examine how the components of the fingerprint discrimination process change in response to a training intervention where participants are given feedback regarding the correct response. The results for the preintervention block replicated Experiment 1, with people showing a bias toward discriminating prints as originating from different fingers, and also processing information more efficiently when prints came from different fingers. The training intervention had the effect of reversing this bias: Those who received the intervention showed a bias toward discriminating prints as originating from the same finger, as opposed to prints originating from different fingers, during the postintervention block. We suspect that the response bias reversed after training because participants attempted to compensate for their poorer accuracy in identifying prints as originating from the same finger. This result is consistent with prior studies on novices' fingerprint discrimination decisions showing a liberal (same finger) response bias in the absence of feedback, but a conservative (different finger) response bias in the presence of feedback (Searston et al., 2016).

Training did not improve overall discriminability. However, there was some evidence that training influenced the variability in the rate of evidence accumulation associated with making correct decisions, particularly for prints that came from different fingers. We hypothesize that training may have led individuals to change how they processed prints of prints. From training, individuals may have learned that differences in the images because of factors such as pressure, distortion, and completeness do not necessarily diagnose prints as originating from different sources. As a result, individuals may have attended to a greater range of information to inform their decisions, and this would have had a greater impact on the decision inputs to the prints originating from different fingers.

Across all three experiments, we observed an interesting difference in the variability of the rate of evidence accumulation between prints originating from the same versus different fingers. In general, the variability in the rate of evidence accumulation for the matching accumulator was greater for prints originating from the same finger than for prints originating from different fingers. We argue that the difference in rate variability is because of inherent features of the stimuli. Different-finger prints can often be identified fairly easily, by seeking out particular features that, if found, quickly diagnose the pair as originating from different fingers. This makes the rate of evidence accumulation fairly reliable, with less heterogeneity across different fingers resulting in lesser variability in the rate of evidence accumulation to make a correct decision. Indeed, such an interpretation is consistent with the original notion that rate variability captures heterogeneity in stimulus difficulty (Ratcliff, 1978), and recent evidence using heterogeneous multi-attribute stimuli (Palada et al., 2018). An experiment measuring rate variability across systematic changes in the similarity of same and different fingerprint prints could be an interesting test of this idea in the future.

Overall, our findings suggest that the LBA can accurately reproduce the empirical patterns of choices and responses times observed across the three experiments in the fingerprint discrimination task. The effects of the manipulations on the model parameters, however, raise questions about the underlying decision process that warrant further research. The pattern of results suggests that novices may attend to extraneous information that is not diagnostic in deciding whether the prints originated from the same or different fingers. Our initial evidence from Experiment 3 suggests that novices can shift this strategy, though future research should attempt to further understand what information is processed in fingerprint discrimination. In line with this possibility, our modeling suggests that the LBA can be used to describe the effects of training on fingerprint discrimination, which could be used to inform training interventions. Our results also point to the idea that the contextual information presented in crime-scene prints can induce biases, which has implications for fingerprint discrimination in the field (Dror, 2018). We are not aware of any established guidelines or thresholds for determining practical significance in studies of human performance on forensic comparison tasks, such as fingerprint discrimination. Arguably, however, any manipulation that has a credible effect (credible intervals do not contain zero) on response bias or discriminability in such tasks could be considered practically significant, because in operational contexts a single false alarm error could mean that an innocent person is wrongly identified and a single miss error could mean that a guilty person goes free (Towler et al., 2018). Small effects are important in forensic science because accuracy in the particular case matters, and interventions that can systematically increase or decrease error of either kind could have a meaningful impact. Accumulation models offer a way to capture such effects.

We believe the LBA is a fruitful avenue for understanding decisions involving naturalistic stimuli because it can model the dynamics of decision making. We have tested the model in a decision-making task that bears strong resemblance to the task in the wild, and to the materials that fingerprint examiners encounter in practice. The idiosyncratic contextual factors of the task and the stimuli did not pose a major issue for the evidence accumulation model, in that it was able to provide an accurate description of performance. Overall, our experiments continue to extend evidence supporting the wider application of evidence accumulation to applied tasks that have a close resemblance to their naturalistic inspiration (e.g., Palada et al., 2018). Further application of evidence accumulation models to fingerprint discrimination could help us better understand sources of bias in people's decision making, whether they pose a concern for discriminability, and how to calibrate them. Moreover, the evidence accumulation framework can help to understand how the decision-making process changes as people transition from novice to expert. The application of evidence accumulation models to other domains of perceptual expertise, such as forensic facial comparison (White, Phillips, Hahn, Hill, & O'Toole, 2015), could inform theory and practice in ways that were previously not possible. More broadly, further investigation of whether these models adequately characterize the underlying dynamics of all kinds of decision processes may provide theoretically meaningful insights about the cognitive processes involved in decisions about naturalistic stimuli.

References

Boag, R., Strickland, L., Heathcote, A., Neal, A., & Loft, S. (2019). Cognitive control and capacity for prospective memory in simulated air traffic control. *Journal of Experimental Psychology: General, 148,* 2181–2206. http://dx.doi.org/10.1037/xge0000599

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

Bürkner, P. C. (2017). brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software, 80,* 1–28. http://dx.doi.org/10.18637/jss.v080.i01

Donkin, C., & Brown, S. (2018). Response times and decision-making. In J. T. Wixted (Ed.), *Stevens' handbook of experimental psychology and cognitive neuroscience, Vol. 5: Methodology* (4th ed., pp. 349–382). New York, NY: Wiley. http://dx.doi.org/10.1002/9781119170174.epcn509

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

Donkin, C., Brown, S., Heathcote, A., & Wagenmakers, E.-J. (2011). Diffusion versus linear ballistic accumulation: Different models but the same conclusions about psychological processes? *Psychonomic Bulletin & Review, 18,* 61–69. http://dx.doi.org/10.3758/s13423-010-0022-4

Dror, I. E. (2018). Biases in forensic experts. *Science, 360,* 243. http://dx.doi.org/10.1126/science.aat8443

Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics*. New York, NY: Wiley.

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

Heathcote, A., & Love, J. (2012). Linear deterministic accumulator models of simple choice. *Frontiers in Psychology, 3,* 292. http://dx.doi.org/10.3389/fpsyg.2012.00292

Heathcote, A., Suraev, A., Curley, S., Love, J., & Michie, P. (2015). Decision processes and the slowing of simple choices in schizophrenia. *Journal of Abnormal Psychology, 124,* 967–974. http://dx.doi.org/10.1037/abn0000117

Ho, T. C., Yang, G., Wu, J., Cassey, P., Brown, S. D., Hoang, N., . . . Yang, T. T. (2014). Functional connectivity of negative emotional processing in adolescent depression. *Journal of Affective Disorders, 155,* 65–74. http://dx.doi.org/10.1016/j.jad.2013.10.025

Kolossa, A., & Kopp, B. (2018). Data quality over data quantity in computational cognitive neuroscience. *Neuroimage, 172,* 775–785. http://dx.doi.org/10.1016/j.neuroimage.2018.01.005

Kruschke, J. K. (2014). *Doing Bayesian data analysis: A tutorial with R, JAGS, and Stan*. London, UK: Academic Press.

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

Loft, S., Bolland, S., Humphreys, M. S., & Neal, A. (2009). A theory and model of conflict detection in air traffic control: Incorporating environmental constraints. *Journal of Experimental Psychology: Applied, 15,* 106–124. http://dx.doi.org/10.1037/a0016118

Palada, H., Neal, A., Strayer, D., Ballard, T., & Heathcote, A. (2019). Competing for cognitive resources: Measuring workload in a time pressured dual-task environment. *Journal of Experimental Psychology: Human Perception and Performance*. Advance online publication. http://dx.doi.org/10.31234/osf.io/bckat

Palada, H., Neal, A., Tay, R., & Heathcote, A. (2018). Understanding the causes of adapting, and failing to adapt, to time pressure in a complex multistimulus environment. *Journal of Experimental Psychology: Applied, 24,* 380–399. http://dx.doi.org/10.1037/xap0000176

Palada, H., Neal, A., Vuckovic, A., Martin, R., Samuels, K., & Heathcote, A. (2016). Evidence accumulation in a complex task: Making choices about concurrent multiattribute stimuli under time pressure. *Journal of Experimental Psychology: Applied, 22,* 1–23. http://dx.doi.org/10.1037/xap0000074

R Core Team. (2016). R: A language and environment for statistical computing (Version 3.3.2) [Computer software]. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.R-project.org/

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,* 1226–1243. http://dx.doi.org/10.1037/a0036801

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

Ratcliff, R., & McKoon, G. (2008). The diffusion decision model: Theory and data for two-choice decision tasks. *Neural Computation, 20,* 873–922. http://dx.doi.org/10.1162/neco.2008.12-06-420

Ratcliff, R., & Van Dongen, H. P. A. (2011). Diffusion model for one-choice reaction-time tasks and the cognitive effects of sleep deprivation. *Proceedings of the National Academy of Sciences of the United States of America, 108,* 11285–11290. http://dx.doi.org/10.1073/pnas.1100483108

Rouder, J. N. (2014). Optional stopping: no problem for Bayesians. *Psychonomic Bulletin & Review, 21,* 301–308. http://dx.doi.org/10.3758/s13423-014-0595-4

Searston, R. A., & Tangen, J. M. (2017a). The emergence of perceptual expertise with fingerprints over time. *Journal of Applied Research in Memory & Cognition, 6,* 442–451. http://dx.doi.org/10.1016/j.jarmac.2017.08.006

Searston, R. A., & Tangen, J. M. (2017b). Expertise with unfamiliar objects is flexible to changes in task but not changes in class. *PLoS ONE, 12,* e0178403. http://dx.doi.org/10.1371/journal.pone.0178403

Searston, R. A., & Tangen, J. M. (2017c). The style of a stranger: Identification expertise generalizes to coarser level categories. *Psychonomic Bulletin & Review, 24,* 1324–1329. http://dx.doi.org/10.3758/s13423-016-1211-6

Searston, R. A., Tangen, J. M., & Eva, K. W. (2016). Putting bias into context: The role of familiarity in identification. *Law and Human Behavior, 40,* 50–64. http://dx.doi.org/10.1037/lhb0000154

Smith, P., & Little, D. (2018). Small is beautiful: In defense of the small-N design. *Psychonomic Bulletin & Review, 25,* 2083–2101. http://dx.doi.org/10.3758/s13423-018-1451-8

Stone, M. (1960). Models for choice-reaction time. *Psychometrika, 25*(3), 251–260. http://dx.doi.org/10.1007/BF02289729

Strickland, L., Elliott, D., Wilson, M. D., Loft, S., Neal, A., & Heathcote, A. (2019). Prospective memory in the red zone: Cognitive control and capacity sharing in a complex, multi-stimulus task. *Journal of Experimental Psychology: Applied, 25,* 695–715. http://dx.doi.org/10.1037/xap0000224

Tangen, J. M., Thompson, M. B., & McCarthy, D. J. (2011). Identifying fingerprint expertise. *Psychological Science, 22,* 995–997. http://dx.doi.org/10.1177/0956797611414729

Thompson, M. B., & Tangen, J. M. (2014). The nature of expertise in fingerprint matching: Experts can do a lot with a little. *PLoS ONE, 9,* e114759. http://dx.doi.org/10.1371/journal.pone.0114759

Thompson, M. B., Tangen, J. M., & McCarthy, D. J. (2013a). Expertise in fingerprint identification. *Journal of Forensic Sciences, 58,* 1519–1530. http://dx.doi.org/10.1111/1556-4029.12203

Thompson, M. B., Tangen, J. M., & McCarthy, D. J. (2013b). Human matching performance of genuine crime scene latent fingerprints. *Law and Human Behavior, 38,* 84–93. http://dx.doi.org/10.1037/lhb0000051

Towler, A., White, D., Ballantyne, K., Searston, R. A., Martire, K. A., & Kemp, R. I. (2018). Are forensic scientists experts? *Journal of Applied Research in Memory & Cognition, 7,* 199–208. http://dx.doi.org/10.1016/j.jarmac.2018.03.010

Trueblood, J. S., Holmes, W. R., Seegmiller, A. C., Douds, J., Compton, M., Szentirmai, E., . . . Eichbaum, Q. (2018). The impact of speed and bias on the cognitive processes of experts and novices in medical image decision-making. *Cognitive Research: Principles and Implications, 3,* 28.

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,* 368–384. http://dx.doi.org/10.1037/a0032222

Usher, M., & McClelland, J. L. (2001). The time course of perceptual choice: The leaky, competing accumulator model. *Psychological Review, 108,* 550–592. http://dx.doi.org/10.1037/0033-295X.108.3.550

Visser, I., & Poessé, R. (2017). Parameter recovery, bias and standard errors in the linear ballistic accumulator model. *British Journal of Mathematical and Statistical Psychology, 70,* 280–296. http://dx.doi.org/10.1111/bmsp.12100

Vuckovic, A., Kwantes, P. J., Humphreys, M., & Neal, A. (2014). A sequential sampling account of response bias and speed-accuracy tradeoffs in a conflict detection task. *Journal of Experimental Psychology: Applied, 20,* 55–68. http://dx.doi.org/10.1037/xap000007

White, D., Phillips, P. J., Hahn, C. A., Hill, M., & O'Toole, A. J. (2015). Perceptual expertise in forensic facial image comparison. *Proceedings of the Royal Society B: Biological Sciences, 282,* 20151292. http://dx.doi.org/10.1098/rspb.2015.1292

Received November 25, 2018
Revision received January 8, 2020
Accepted February 22, 2020 ■