Supplementary Material

S1. Additional procedural details for training on the random dot motion task

The experiment began with three sets of practice blocks designed to familiarize participants with the speed and accuracy requirements of the task. Each set was composed of a series of short blocks of 16 trials. Blocks within each set were repeated until the participant reached a criterion level of performance (i.e., accuracy requirements, speed requirements – see below). At the end of each block, the mean response time and error rate for that block were provided as summary feedback, which indicated if the participant had met the speed and accuracy demands of the block. This information was presented by color-coding the numbers in which the mean response time and error rate were presented (red for too high, green for low enough).

The first set of practice blocks focused on decision accuracy. Participants were instructed to focus solely on the accuracy of their responses, ensuring that they could perceive the coherent motion in the display. Participants received positive feedback when their mean response time was faster than 1000ms and error rate was below 20% in a block.

The second practice set shifted the emphasis to response speed. Participants began the set with a response deadline of 600ms; positive feedback was given if the mean response time in a block was faster than 600ms. If the participant met this response deadline, the next practice block decreased the deadline to 550ms. This process was repeated with subsequent deadlines of 470ms and 400ms for those able to achieve the new deadline. As the focus here was on speed, an error rate of 40% or lower per block was met with positive feedback. The second set was used to identify an appropriate response deadline on an individual basis, since some factors contributing to the observed response time are related to the duration of non-decision components (e.g., movement speed). We employed this procedure since the patients were more or less impaired in executing the response because the dorsal part of the striatum

is motor-output related; if a fixed deadline were used for all participants, the effect of the time pressure on the decision process itself would differ across individuals.

The final practice set randomized the two types of instructions from one trial to the next, mimicking the procedure used in the main experiment. Each trial began with a cue that instructed the participant to respond quickly or accurately. Feedback for the mixed practice blocks was presented as a 2x2 matrix. The top row indicated whether the mean response time and error rate were low enough in the speed trials, where the response deadline was set to the lowest deadline the participant was able to meet during the speed practice blocks, and the error rate set to 40%. The bottom row indicated whether the mean response time and error rate were low enough in the accuracy trials, with a response time limit of 1500ms and an error rate limit of 20%. When the participant met the response time and error rate requirements of the third practice set they proceeded to the main experiment.

S2. Linear Ballistic Accumulator (LBA) model analysis

S2.1. Background to the LBA

Decision phenomena – such as the speed-accuracy tradeoff (SAT) – have been most profitably studied using cognitive process models. Cognitive process models decompose observed variables, such as choices and response times, into latent components of processing of deeper psychological interest, such as processing efficiency and response caution. The most successful class of cognitive process models of decision-making in neuroscience and psychology are known as sequential sampling models (for overview, see Forstmann, Ratcliff, & Wagenmakers, 2016). There are a range of sequential sampling models that differ in minor details (e.g., Brown & Heathcote, 2008; Ratcliff, 1978; Usher & McClelland, 2001; Van Zandt, Colonius, & Proctor, 2000), however, all assume that decisions are made through a gradual process of sampling noisy information from the environment. The sampled

information is integrated into an evidence counter that tracks support for the response alternatives until the counter reaches a pre-determined threshold value, triggering a choice.

To provide a deeper account of the psychological processes underlying performance in our experiment, we analyzed the data using the Linear Ballistic Accumulator model (LBA; Brown & Heathcote, 2008, see Figure 2B of the main text), but note that the conclusions follow from similar sequential sampling models (cf. Donkin, Brown, Heathcote, & Wagenmakers, 2011). The LBA model assumes that evidence favoring each response alternative is represented by the activation state of independent accumulators. For each of these accumulators, the level of activation at the beginning of the decision process is drawn from a uniform distribution. From that point on, evidence accumulates linearly over time. Once one of the accumulators reaches the decision threshold, the corresponding alternative is selected. The predicted response time is the time taken for the winning accumulator to reach threshold plus some fixed non-decision time, reflecting aspects such as the time taken to encode the stimulus and produce a motor response.

The LBA model measures the efficiency of information processing with a parameter known as the *drift rate*, which indicates the average speed of evidence accumulation. If a task is easy, or if a participant is highly skilled, the correct alternative will have a much higher drift rate than the incorrect alternative(s). A larger drift rate leads to rapid evidence accumulation resulting in both fast and correct decisions, on average. The difference between the drift rate for the correct and incorrect alternatives, scaled by the variance of the drift rate, is referred to as the sensitivity. This can be compared to the sensitivity measure *d'* in the signal detection theory framework, except that the LBA measure of sensitivity incorporates response time distributions as well as response accuracy. When the task is difficult, or a participant is poorly skilled, there is a smaller difference between drift rates for correct and

incorrect responses, meaning sensitivity is low. This results in slower decisions and a larger probability of committing errors, on average.

The LBA model measures cautiousness in decision-making with a parameter known as the *response threshold*, which indicates the amount of evidence required to trigger a response. Response caution is often parameterized as a transformed version of the response threshold parameter; the average distance that a process must travel to the response threshold. By changing the level of response caution, the LBA accounts for the SAT. In particular, for a given drift rate, high levels of response caution lead to slower responses with a low likelihood of errors, and low levels of response caution lead to faster responses with a greater likelihood errors.

The parameters of the LBA and related sequential sampling models – such as the drift rate and response threshold – make unique predictions about patterns of response probability and the shape of correct and incorrect distributions of response times. In this way, we can estimate the particular combination of model parameters that provide the best fit to the observed data (for tutorials and methods, see Donkin, Brown, & Heathcote, 2011; Vandekerckhove & Tuerlinckx, 2007; Voss, Nagler, & Lerche, 2013). Because sequential sampling models expose factors that define the complete shape of response time distributions for correct and erroneous responses, they provide a more informative account of underlying psychological processes than descriptive statistics. Moreover, sequential sampling models offer an elegant tool to analyze the potentially complex effect of an experimental manipulation, providing a way to account for variations in response time distributions and error rates with a single parameter. For instance, manipulating the probability of a particular stimulus or the magnitude with which a particular alternative is rewarded influences the starting values in the drift diffusion model, while manipulating stimulus discriminability selectively changes the drift rate (Voss, Ruthermond, & Voss, 2004). Most relevant to the

SAT, manipulating the instructions to emphasize response speed or response accuracy influences the threshold parameter; the level of response caution (e.g., Forstmann et al., 2008).

S2.2. Details of the LBA analysis reported in the main text

The LBA model was evaluated using a hierarchical Bayesian approach, which is ideally suited to compare various outcome measures between two groups and assess whether the weight of evidence suggests the performance of the groups differ (supporting an alternative hypothesis) or are equivalent (supporting the null hypothesis). The data from each participant were assumed to be generated from the LBA model, with parameters unique to that participant, but drawn from separate distributions for the patient and control groups (Turner, Sederberg, Brown, & Steyvers, 2013). For each participant, three parameters were allowed to take on different values in the speed-emphasis and accuracy-emphasis conditions: the threshold parameter (b) and the drift rate parameters for the correct and error accumulators (v_c and v_e). The choice to freely estimate drift rates and response thresholds in the speed- and accuracy-emphasis conditions reflects the current best practice in the field (cf. Cassey, Heathcote, & Brown, 2014; Rae, Heathcote, Donkin, Averell, & Brown, 2014). Two other parameters, the maximum value of the start point distribution (A) and the non-decision time ($t0$) were assumed to have the same value for the speed- and accuracy-emphasis conditions. The drift variance parameter (s) was set to 1 in all conditions to provide a scaling factor; without such a scaling parameter the model is underdetermined (Donkin, Brown, & Heathcote, 2009). To correct for partial guessing responses, we excluded trials with responses faster than 300ms and assumed 5% contaminant responses in the predicted distributions. For these contaminant responses, the response times were taken from a uniform distribution (0 - 1200ms) with a response accuracy of 50% (Ratcliff & Tuerlinckx, 2002).

To ensure our assumption that response thresholds and correct and error drift rates differed across the speed- and accuracy-emphasis conditions, we also tested two reduced models. The first reduced model freely estimated a separate response threshold in the speed- and accuracy-emphasis conditions but assumed common correct and error drift rates over conditions (i.e., no effect of stimulus difficulty). The second reduced model assumed a common response threshold in the speed- and accuracy-emphasis conditions but freely estimated separate correct and error drift rates over conditions (i.e., no effect of speed vs. accuracy instructions). We compared the three models using the Deviance Information Criterion (DIC; Spiegelhalter, Best, Carlin & Van der Linde, 2002), a Bayesian measure that balances model parsimony with goodness of fit. The model with the lowest DIC from the set of models under comparison provides the best account of the data, where differences in information criteria greater than 10 provide strong evidence in favor of the lowest-DIC model (Burnham & Anderson, 2004; Raftery, 1995). Both reduced models had substantially higher DICs than the model that freely estimated response threshold and correct and error drift rates in the speed- and accuracy-emphasis conditions (163 and 161 vs. 142, respectively). Accordingly, we focus on the parameter estimates of the full model in the main text.

To assess the impact of the speed and accuracy instructions, we defined two variables from the parameter estimates of the most parsimonious LBA model: response caution and sensitivity. Response caution was defined as the average distance that a process must travel to the threshold (b-$A/2$); as such, high values of caution correspond to the situation where the participant accumulates more information before making a response. This definition of caution incorporates variability in the starting position of evidence accumulation. The relative amount of start point noise should influence cautiousness; that is, for an equivalent response threshold, more/less start point noise should, on average, correspond to more/fewer errors and faster/slower responses. Response caution can also be defined in other ways. For example,

Forstmann et al. (2008) expressed response caution as *b/A*. We settled on *b-A/2* as it has the most natural interpretation, the average distance to threshold, and is the most analogous to the diffusion decision model's boundary separation parameter. Sensitivity was defined as the scaled difference between the drift rate for correct and erroneous responses ($(v_c - v_e)/s$).

The difference between patient and control groups was modeled by allowing the two groups to have different group-level distributions. We imposed weak prior distributions on the parameters of group-level distributions (truncated normal distributions for the mean parameters, and gamma distributions for the variance parameters, shown in Table S1). The prior distributions were identical for the patient and control groups, and also for the speed-emphasis and accuracy-emphasis conditions, thus ensuring that any observed posterior differences were driven by the data. No explicit prior distributions were placed on the response caution and sensitivity variables since those variables are transformed from parameter estimates of the model (response caution: starting point and response threshold; sensitivity: correct and error drift rates), and hence do not have prior distributions that are independent of the LBA model parameters. Instead, their prior distributions were determined by the prior distributions on the estimated parameters (*b* and *A*). Samples were drawn from the joint posterior distribution over all participant-level parameters and group-level parameters using differential evolution Markov chain Monte-Carlo methods with 100 chains and 5,000 samples, with the default settings (Turner et al., 2013). We initialized the chains with random draws from the prior distributions, and discarded the first 4,000 samples. We used the last 1,000 samples to obtain estimates of the stable distributions for the four conditions (2 groups x 2 instructions).

Differences in the group-level distributions were assessed using odds (for a similar approach, see Hawkins, Hayes, & Heit, 2016; Mittner et al., 2014). Odds are computed as $x/(1-x)$, where x is the probability that a sample from one distribution is larger than a sample

from another distribution. Odds are conceptually similar to the Bayes factor, a common way of quantifying evidence in Bayesian analyses, though the two differ in their assumptions about the role of prior distributions in assessing the weight of evidence for an effect. While odds themselves are informative, they can also be categorized as indicating positive evidence (>3:1), substantial evidence (>10:1), strong evidence (>30:1), or decisive evidence (>100:1) (cf. Jeffreys, 1961).

S2.3. Group-level differences in LBA parameter estimates

Table S2 shows a complete list of the group-level differences in LBA model parameters, and Table S3 shows individual participant parameter estimates. In the main text we reported the group-level differences of primary interest to the study hypotheses; a main effect where patients had lower response caution than controls but similar sensitivity to perceptual information. We also observed that patients had a larger non-decision time than controls, which suggests that they may have at least partial motor execution deficits. This is consistent with the role of the dorsal striatum in controlling motor-output (Draganski et al., 2008). Our finding that patients differ to controls in their motor movement time does not cloud the interpretation of the response caution effect, since caution and motor movement reflect different component processes of decision-making. The two processes may have been confounded in analyses of raw response time data, which highlights the importance of using cognitive models to decompose and understand behavioral data.

S3. Simulation study to test the effect of sample size on estimated odds

We conducted a simulation study to investigate whether our sample size allowed us to reliably detect differences between controls and patients. The simulation study calculated the likelihood of the odds observed in real data under two hypotheses:

1. The *null hypothesis* that the parameters of all participants were sampled from a common population-level distribution. To simulate data from this hypothesis we first re-fit the real data from controls and patients under the assumption that all participant-level parameters were drawn from common population-level distributions. We then randomly sampled parameters for synthetic participants from these common population-level posterior distributions.
2. The *alternative hypothesis* that the parameters of controls and patients were sampled from different population-level distributions. To simulate data from this hypothesis we randomly sampled parameters for synthetic control and patient participants from the population-level posterior distributions estimated from the controls and patients in real data, respectively.

We generated 1,000 synthetic data sets from each of the null and alternative hypotheses. Each synthetic data set was exactly the same size as the real data: 5 control and 5 patient participants, each with the same number of trials per condition as completed in real data. The model reported in the main text was independently fit to each synthetic data set, using the same methods as for the real data. We used the posterior distributions calculated from each fit to calculate the odds that controls had a larger value of each parameter than patients, using identical methods as applied to real data. We collated the estimated odds from each synthetic data set to generate a distribution of odds expected under the null and alternative hypotheses, separately for each model parameter. These distributions are shown with log scaling in Figure S1. The distributions of log odds are very close to Gaussian, so we summarize them with Gaussian distributions, which are also shown in Figure S1. We used the parameters of the best-fitting Gaussian distributions to calculate the likelihood of the log odds observed in real data under the null and alternative hypotheses.

For some parameters there were insufficient data to discriminate the null and alternative hypotheses. For example, the correct and error drift rates in the accuracy condition and the maximum of the start point distribution are just as likely, or more likely, under the null hypothesis than the alternative hypothesis. This is consistent with the results reported in the main text and supplementary material, where the observed odds for these parameters, reported in Table S2, were close to 1. This can be seen in Figure S1 where panels for these parameters have largely overlapping distributions for the null and alternative hypotheses.

For those parameters where there was some discriminative ability, the effects were consistent with the results reported in the main text. For example, the observed odds that patients had lower response caution than controls in the speed and accuracy conditions were both more likely under the alternative hypothesis than the null hypothesis, shown in Figure S1. The simulation study thus verifies that there was sufficient information in the sample size we collected to discriminate between patients and controls on the parameters of key interest to our study hypothesis.

Table S1. Prior distributions for group-level parameters of the LBA model.

Parameter	Truncated normal mean	Truncated normal SD	Gamma shape	Gamma rate
Max starting point	2	2	1	1
Threshold	2	2	1	1
Non-decision time	.5	.5	.2	.2
Correct drift	3	3	1	.5
Error drift	2	2	1	.5

Note. The prior distributions of the group level parameters, with the population mean parameters defined by the mean and standard deviation of truncated (positive-only) normal distributions, and the population standard deviation parameters defined by the shape and rate (i.e., inverse scale) parameters of gamma distributions.

Table S2. Cognitive modeling results for group-level contrasts.

Parameter	Condition	Odds (X-to-1)	Patients	Controls
Max start point	Both	2.7	**2.56**	1.85
Response threshold	Speed	27	.39	**1.39**
	Accuracy	69	1.05	**2.30**
Correct drift rate	Speed	3.7	4.41	**5.09**
	Accuracy	1.8	4.40	**4.69**
Error drift rate	Speed	9.3	2.81	**4.68**
	Accuracy	1.7	2.21	**2.97**
Non-decision time (s)	Both	28	**.22**	.13
Sensitivity	Speed	3.9	**1.61**	.42
	Accuracy	1.0	**2.19**	1.72
Response caution	Speed	6.3	1.66	**2.31**
	Accuracy	21	2.32	**3.22**

Group level effects for each LBA model parameter and the higher-order measures of sensitivity and response caution, shown separately for the speed- and accuracy-emphasis conditions (where appropriate; parameters that were constrained to a common value across conditions are marked as "Both"). Effects are expressed as the odds for the patient group having a larger value than the control group, or vice versa. The rightmost two columns present posterior mean parameter estimates for the two groups, with the higher value in boldface. Parameter estimates for the two response thresholds refer to the distance from the maximum of the start point distribution to the response threshold. Drift variance parameter is not included since this was set to 1 in all conditions to serve as a scaling factor.

Table S3. Cognitive modeling results for individual participants.

	Patients					Controls				
Parameter	1	2	3	4	5	1	2	3	4	5
Max start point	2.92	2.79	2.28	2.25	2.63	3.10	1.15	1.09	2.11	2.99
Response threshold - Speed	0.08	0.62	0.50	0.50	0.65	1.62	1.25	1.22	1.48	1.46
Response threshold - Accuracy	1.29	1.19	1.04	0.77	1.11	2.22	2.87	3.75	1.81	1.69
Correct drift rate - Speed	3.97	4.59	4.46	4.75	4.47	5.04	4.97	5.18	5.16	5.25
Correct drift rate - Accuracy	5.42	3.52	4.64	4.29	4.60	4.82	5.35	5.17	4.40	3.99
Error drift rate - Speed	0.99	4.29	3.59	4.13	4.01	4.52	4.79	4.88	4.71	4.77
Error drift rate - Accuracy	0.46	2.87	2.92	3.33	4.09	4.20	2.19	3.04	3.16	3.28
Non-decision time (s)	0.28	0.23	0.22	0.21	0.19	0.13	0.12	0.13	0.13	0.14

Posterior mean parameter estimates of individual participants for each LBA model parameter. Higher-order measures of sensitivity and response caution are not shown here, as they were post-computed rather than directly estimated from data (shown in Table S2). Parameter estimates for the two response thresholds refer to the distance from the maximum of the start point distribution to the response threshold. Drift variance parameter is not included since this was set to 1 in all conditions to serve as a scaling factor.

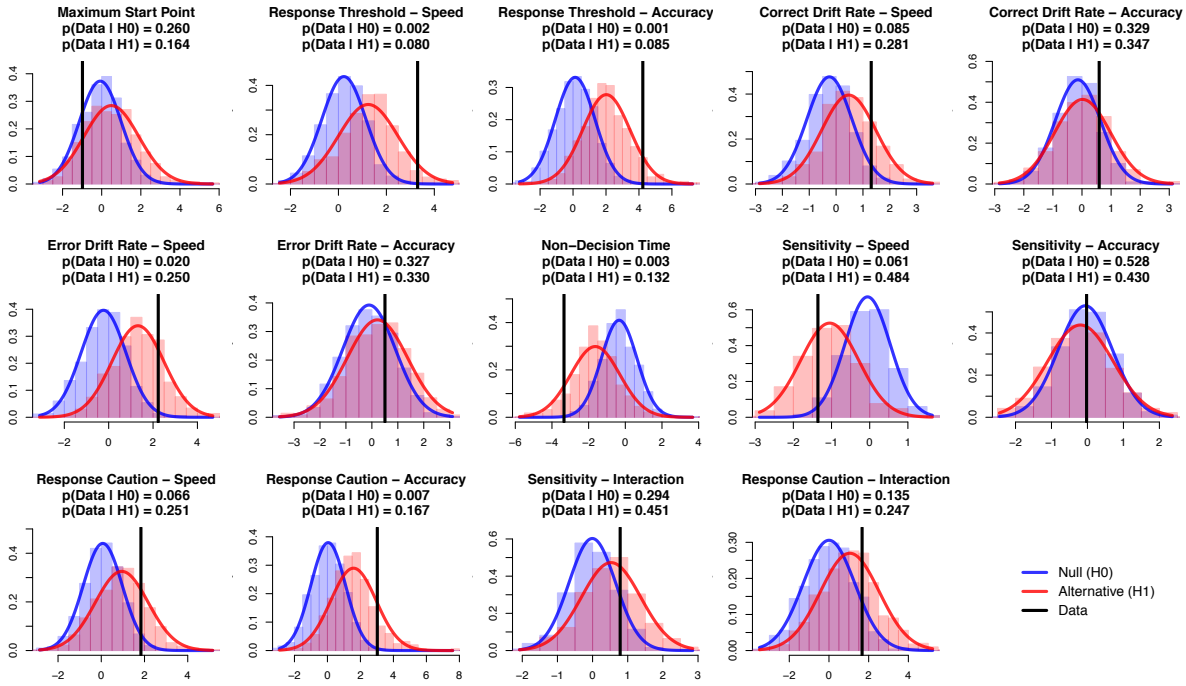

Figure S1. Density of the log odds estimated from real data under the null and alternative hypotheses. Blue and red histograms show the distribution of log odds estimated from the null and alternative hypotheses in the simulation study, respectively, separately for each parameter. Blue and red curves show best-fitting Gaussian distributions to the null and alternative distributions of log odds, respectively. Vertical black lines show the log odds estimated from real data. The density of the observed log odds under the null and alternative distributions is shown above each panel. Log odds greater than 0 (i.e., odds greater than 1) indicate evidence that controls had a larger value of the parameter than patients.

Supplementary Material References

Brown, S. D., & Heathcote, A. (2008). The simplest complete model of choice response time: linear ballistic accumulation. *Cognitive Psychology*, *57*, 153–178.

Burnham, K. P., & Anderson, D. R. (2004). Multimodel inference: Understanding AIC and BIC in model selection. *Sociological Methods & Research, 33*, 261–304.

Cassey, P., Heathcote, A., & Brown, S. D. (2014). Brain and behavior in decision-making. *PLoS Computational Biology, 10*. doi:10.1371/journal.pcbi.1003700

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

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

Donkin, C., 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.

Draganski, B., Kherif, F., Klöppel, S., Cook, P. A., Alexander, D. C., Parker, G. J. M., … Frackowiak, R. S. J. (2008). Evidence for segregated and integrative connectivity patterns in the human basal ganglia. *The Journal of Neuroscience, 28*, 7143–7152.

Forstmann, B. U., Dutilh, G., Brown, S., Neumann, J., von Cramon, D. Y., Ridderinkhof, K. R., & Wagenmakers, E.-J. (2008). Striatum and pre-SMA facilitate decision-making under time pressure. *Proceedings of the National Academy of Sciences of the United States of America*, *105*, 17538–17542.

Forstmann, B. U., Ratcliff, R., & Wagenmakers, E.-J. (2016). Sequential sampling models in cognitive neuroscience: Advantages, applications, and extensions. *Annual Review of Psychology, 67*, 641–666.

Hawkins, G. E., Hayes, B. K., & Heit, E. (2016). A dynamic model of reasoning and memory. *Journal of Experimental Psychology: General, 145*, 155–180.

Mittner, M., Boekel, W., Tucker, A., Turner, B. M., Heathcote, A., & Forstmann, B. U. (2014). When the brain takes a break: A model-based analysis of mind wandering. *The Journal of Neuroscience, 34*, 16286–16295.

Raftery, A. E. (1995). Bayesian model selection in social research. *Sociological Methodology, 25*, 111–163.

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

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

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., & Van der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society B, 64*, 583–639.

Turner, B., Sederberg, P. B., Brown, S. D., & Steyvers, M. (2013). A method for efficiently sampling from distributions with correlated dimensions. *Psychological Methods*, *18*, 368–383.

Usher, M., & McClelland, J. L. (2001). The time course of perceptual choice: The leaky, competing accumulator model. *Psychological Review*, *108*, 550–592.

Van Zandt, T., Colonius, H., & Proctor, R. W. (2000). A comparison of two response time models applied to perceptual matching. *Psychonomic Bulletin & Review, 7*, 208–256.

Vandekerckhove, J., & Tuerlinckx, F. (2007). Fitting the Ratcliff diffusion model to experimental data. *Psychonomic Bulletin & Review*, *14*, 1011–1026.

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