A dual-stage dual-threshold evidence accumulation theory for decision-making, motor preparation, and motor execution

Edouard Dendauw, Nathan Evans, Gordon Logan, Thibault Gajdos, Emmanuel Haffen, Djamila Bennabi, Mathieu Servant

A dual-stage dual-threshold evidence accumulation theory for decision-making, motor preparation, and motor execution

Edouard Dendauw[1,2], Nathan J. Evans[3], Gordon D. Logan[4], Thibault Gajdos[5], Emmanuel Haffen[1], Djamila Bennabi[1], and Mathieu Servant[1,2,6]

[1]Laboratoire de Recherches Intégratives en Neurosciences et Psychologie Cognitive UR 481, Université de Franche-Comté, France

[2]MSHE Ledoux USR 3124, Université de Franche-Comté, France

[3]School of Psychology, University of Queensland, Australia

[4]Department of Psychological Sciences, Vanderbilt University, USA

[5]Laboratoire de Psychologie Cognitive UMR 7286, Aix-Marseille Université, France

[6]Institut Universitaire de France

Author Note

The authors declare no conflict of interest, and wish to thank Maëva Ferry for her help in data collection. The empirical data and a codebook for processing the data are available on the Open Science Framework (https://osf.io/4unw6/). This work was funded by the European Union under the Horizon Europe programme through ERC Starting Grant n°101039226 allocated to MS. NJE was supported by an Australian Research Council Discovery Early Career Researcher Award (DE200101130). Correspondence concerning this article should be addressed to Mathieu Servant, Laboratoire de Recherches Intégratives en Neurosciences et Psychologie Cognitive UR 481, UFR Santé, bâtiment Rabelais, Université de Franche-Comté, 19 rue Ambroise Paré, 25030 Besançon Cedex, France. E-mail: mathieu.servant@univ-fcomte.fr

A dual-stage dual-threshold evidence accumulation theory for decision-making, motor preparation, and motor execution

Abstract

This article introduces an integrated and biologically-inspired theory of decision-making, motor preparation, and motor execution. The theory is formalized as an extension of the diffusion model, in which diffusive accumulated evidence from the decision-making process is continuously conveyed to motor preparation brain areas, where it is filtered out through a second accumulation processing stage. The resulting motor preparation variable is then transmitted to the response-relevant muscles when it exceeds a threshold level of activation, corresponding to the beginning of motor execution. The transmission continues until a threshold amount of force has been produced by the muscles to issue the response. We tested this dual-stage dual-threshold diffusion model by continuously probing the electrical activity of response-relevant muscles through electromyography (EMG) in four choice tasks that span a variety of domains in cognitive sciences, namely motion perception, numerical cognition, recognition memory, and lexical knowledge. The model provided a good quantitative account of behavioral and EMG data, and systematically outperformed previous models. This work represents an advance in the integration of processes involved in simple decisions, and sheds new light into the interplay between decision and motor systems.

Keywords:

Decision-making | Motor preparation | Motor execution | Diffusion model | Electromyography

Introduction

Many of our internal choices are communicated to the world, and this communication requires an interplay between decision and motor systems. For instance, the choice to vote for a candidate in a presidential election eventually results in the deposit of a ballot in a box. Deciding who to have friendships and relationships with results in concrete approach/avoidance behaviors. Choices about where to spend our money determine our consumer behavior. Decision and motor systems are also jointly engaged in many experimental cognitive tasks. For instance, recognition memory tasks, lexical decision tasks, perceptual decision tasks, numerosity judgment tasks, and conflict tasks all involve a decision between two or more options (e.g., old/new, greater/less than a quantity), each option being mapped to a specific motor plan (e.g., manual button press, saccade towards a target, vocal response). Decision and motor systems have each benefited from extensive research (for reviews, see Cisek and Kalaska, 2010; Ebbesen and Brecht, 2017; Forstmann et al., 2016; Freedman and Assad, 2016; Gold and Shadlen, 2007; Lemon, 2008; O'Connell and Kelly, 2021; Ratcliff and Smith, 2004; Ratcliff et al., 2016; Robinson, 1973; Schall, 2019; Schall and Paré, 2021; Summerfield and Parpart, 2022), and recent modeling efforts have sought to specify the relationship between them (Servant et al., 2021; Servant et al., 2015; Verdonck et al., 2020). However, as will become obvious in the next sections, current computational models fail to capture important aspects of empirical data, either at the motor preparation or at the motor execution processing levels. The present work aims to adress these shortcomings by introducing an integrated computational theory of decision-making, motor preparation, and motor execution that builds upon a dual-stage dual-threshold evidence accumulation architecture.

The manuscript is structured as follows. We will first review traditional theoretical conceptions regarding the relationship between decision and motor stages, and recent neurophysiological data that challenge them. We will then highlight shortcomings of current modeling approaches, and introduce the dual-stage dual-threshold evidence accumulation

theory. The theory will be tested against behavioral and neurophysiological data from four choice tasks that span a variety of domains in cognitive sciences, namely motion perception, numerical cognition, recognition memory, and lexical knowledge.

Deciding and acting: traditional views and challenges

A traditional view in psychology is that the motor system is engaged when the decision-maker has committed to an internal choice (Donders, 1969; Johnson-Laird, 1988; Sternberg, 1969). At a computational level, decision formation has been successfully modeled by an accumulation-to-threshold process, in which noisy evidence from our senses and memory is accumulated until a threshold quantity of cumulative evidence is attained (e.g., Bogacz et al., 2006; Evans and Wagenmakers, 2020; Laming, 1968; Ratcliff and Smith, 2004). Each accumulator is associated to a specific choice, and the accumulator that first reaches the threshold determines the choice and the duration of decision formation. The mean latency of other processing stages is represented by a residual parameter. More or less explicitely, it is generally assumed that motor latencies are included in this residual parameter, and thus do not play any role in decision formation. In other words, the motor system is engaged only once the evidence accumulation variable has reached the threshold.

A growing body of neurophysiological evidence challenges this traditional view however. Electroencephalographic (EEG) studies have identified two electrical signals that exhibit key signatures of the theoretical accumulation-to-threshold decision variable (for reviews, see O'Connell and Kelly, 2021; O'Connell et al., 2018). The first signal, termed centro-parietal positivity (CPP), reflects accumulated sensory evidence, and culminates to a threshold voltage around the time of the response. The CPP appears whenever an individual has to make a decision between two options, and shows the same temporal dynamics and spatial topography regardless of sensory and response modalities. Importantly, the CPP appears even when participants are instructed to make the decision mentally (without communicating the outcome through the motor system; O'Connell et al., 2012), or when the

stimulus-response mapping is not yet known during stimulus presentation (Twomey et al., 2016). Although the precise functional interpretation of the CPP requires additional investigations (O'Connell and Kelly, 2021), these empirical findings suggest that it may reflect a decision about alternative categories of a stimulus, with a neural generator in the parietal cortex. The second signal corresponds to effector-selective motor preparation EEG activities (de Jong et al., 1988; Gratton et al., 1988; Pfurtscheller and Lopes da Silva, 1999), such as the lateralized readiness potential or the decrease in spectral activity in the mu/beta band over the motor cortex (in case of left/right manual responses). Similar to the CPP, effector-selective EEG signals appear to reflect the theoretical accumulation-to-threshold decision variable. Although ramping electrical activities of the two classes of signals overlap in time and reach their voltage peak around the time of the response, the onset latency of effector-selective signals occurs after the onset latency of the CPP (Kelly and O'Connell, 2013). In addition, effector-selective EEG signals are absent when participants are instructed to make the decision mentally, or when the stimulus-response mapping is not yet known during stimulus presentation (O'Connell et al., 2012; Twomey et al., 2015). These results suggest that when decisions are mapped onto actions, the decision variable is represented in motor areas of the brain that prepare the response. Similar findings have been observed using magnetoencephalography (de Lange et al., 2013; Donner et al., 2009), functional resonance imaging (Filimon et al., 2013; Tosoni et al., 2008), transcranial magnetic stimulation (Klein-Flügge and Bestmann, 2012), and single-unit recordings in awake monkeys (Gold and Shadlen, 2000, 2007; Gold and Shadlen, 2003; Purcell et al., 2010; Ratcliff et al., 2003; Schall, 2019).

Another source of neurophysiological evidence that speaks against strict serial discrete processing between decision and motor stages comes from surface electromyographic (EMG) studies. EMG is a non-invasive technique that measures the electrical activity of muscles through electrodes placed on the skin surface. Recording the EMG activity of response-relevant muscles in a reaction time (RT) task (e.g., muscles of the thenar eminence for a button press with the thumb) allows researchers to partition each RT into two latencies:

a premotor time (PMT, from stimulus onset to the EMG onset of the response; see Figure 1) and a motor time (MT, from EMG onset to the response; Botwinick and Thompson, 1966; Weiss, 1965). Recent studies have shown that both mean PMT and mean MT increase as the perceptual discriminability of the stimulus decreases (Servant et al., 2021; Servant et al., 2016; Weindel et al., 2021; see also Selen et al., 2012, for similar findings obtained with a different EMG methodology based on reflex gains). These results demonstrate that EMG activity reflects a quantity that scales with sensory evidence, and suggest a flow of the decision variable down to the response-relevant muscles. The flow is not purely continuous, because EMG bursts have a discrete onset (that occurs ∼150-180 ms on average before the response for a button press with the thumb, with ∼900 gram-force required; see Servant et al., 2021).

Modeling the interplay between decision and motor systems

Servant et al. (2021) proposed a dual-threshold evidence accumulation theory to account for the aforementioned neurophysiological findings at a computational level, with a particular focus on EMG findings (for an early development of the theory in the context of conflict tasks, see Servant et al., 2015). The scope of the theory concerns decision-making problems in which choices are mapped onto actions. The theory assumes that the decision variable is transmitted to the motor structures that prepare and execute the response. Motor execution, operationally defined by the electrical excitation of response-relevant muscles, is thus determined by the same evidence accumulation variable that drives decision-making.

As reviewed in the previous section, studies have shown a representation of the decision variable in motor areas of the brain that prepare the response (such as the motor cortex for manual responses). Although several descending pathways are involved in motor control, the most direct pathway for voluntary muscle excitation involves a direct connection between corticospinal neurons (originating from the motor cortex) and α-motoneurons that depolarize muscle fibers (for reviews, see Ebbesen and Brecht, 2017; Lemon, 2008), sup-

porting the hypothesis that the evidence accumulation decision variable may be transmitted to the muscles without undergoing significant modulations. This direct pathway appears particularly involved in the control of distal extremities, and is thought to form the basis of manual dexterity in primates and humans (Bortoff and Strick, 1993; Porter, 1987). Since a subtantial proportion of research in experimental psychology involves manual responses, these neuroanatomical observations are important to consider in theory development.

Servant et al. (2015, 2021) formalized the theory (hereafter referred to as Dual-Threshold Diffusion Model DTDM) as an extension of the diffusion model (Ratcliff, 1978). The accumulation process is described by the following stochastic differential equation:

$$dx(t) = vdt + \sigma dW(t), x(0) = 0, \tag{1}$$

where $x(t)$ denotes the accumulated evidence at time t and v the drift rate. The term $\sigma dW(t)$ represents the stochastic part of the process (Gaussian white noise with mean zero and variance $\sigma^2 dt$). For simplicity, we assume an unbiased starting point $x(0)$, located at zero. In the context of a two-choice task involving left versus right manual responses, $x(t)$ continuously flows to the motor cortex. A first pair of thresholds, referred to as EMG thresholds and represented by parameter $\pm m$, operate at the motor cortex level. EMG thresholds do not terminate the evidence accumulation process, but simply act as a gating mechanism that prevents $x(t)$ from being continuously transmitted to the muscles. Specifically, $x(t)$ is transmitted to the muscles associated to the left manual response if $x(t) \geq m$, and to the muscles associated to the right manual response if $x(t) \leq -m$. $x(t)$ continues to evolve until sufficient EMG activity (and thus force) has been produced to issue the response (e.g., a button press). This is modeled by introducing another pair of thresholds, termed response thresholds (parameter $\pm r$, with $|r| > |m|$). The left response is given if $x(t) \geq r$, and the right response is given if $x(t) \leq -r$. In this framework, the difference between $|r|$ and $|m|$ is determined by the force required to respond: a larger force is associated with a larger difference. The predicted PMT corresponds to the latency at which $x(t)$ hits an EMG threshold for the last time before reaching the corresponding response threshold, and the predicted

MT corresponds to the latency at which $x(t)$ hits the response threshold minus PMT. In addition, predicted PMT and MT each incorporate residual processing components with mean duration Te and Tr respectively. At minimum, Te includes stimulus encoding processes and the corticomuscular transmission time. Tr corresponds to the electromechanical delay (time lag between muscle excitation and a measurable change in force output). This delay involves both electrochemical and mechanical processes (e.g., propagation of action potentials, force transmission along the active and passive parts of the series elastic component; e.g., Cavanagh and Komi, 1979; Hug et al., 2011). In its raw form (without between-trial variability in any of the model components and with an unbiased starting point of evidence accumulation), DTDM has five parameters: drift rate v, EMG thresholds $\pm m$, response thresholds $\pm r$, mean residual latencies Te and Tr.

Because EMG activity is determined by $x(t)$, modulations of drift rate impact both predicted PMT and MT. Consequently, DTDM predicts an increase in mean PMT and mean MT as the perceptual discriminability of the stimulus decreases, explaining empirical EMG findings reported in the previous section. Servant et al. (2021) derived other preditions from the model. First, the rising slope of rectified[1] and averaged EMG bursts should reflect drift rate, and should thus decrease as perceptual discriminability decreases. Second, for any given drift rate, the distributions of PMT and MT should exhibit a similar right-skewed shape, which should translate into an approximately linear PMT quantile-MT quantile plot. Third, DTDM predicts partial EMG bursts (i.e., EMG bursts that do not generate sufficient force to issue the response) during PMT, when $x(t)$ oscillates around an EMG threshold, due to accumulation of noise (see Equation 1, Figure 1, and supplementary Figure 1). The proportion of trials containing at least one partial EMG burst during PMT and the latency of the first partial burst should increase as drift rate decreases. Fourth, the between-trial correlation between PMT and MT should be null, due to the Markov property of the diffusion

[1] EMG rectification consists in taking the absolute value of voltages across time points (see Figure 1). The rectified EMG amplitude scales with the level of global muscle excitation (Vigotsky et al., 2018).

process (given the present, the future does not depend on the past).

To test these predictions, Servant et al. (2021) recorded the EMG activity of muscles associated with left/right manual responses in a random dot motion task. In each trial, participants had to determine the global direction (left versus right) of moving dots, and press the corresponding response button with their left or right thumb. The proportion p of dots moving coherently in the left or right signal direction, termed motion coherence, was manipulated across six levels ($p = 0, .05, .08, .12, .2, .4$), in order to modulate the perceptual difficulty of the decision. The EMG data provided evidence for each prediction [2]. In addition, fits of DTDM to the joint distributions of PMT and MT in correct and incorrect trials and to accuracy data were good, providing quantitative evidence for the model architecture.

However, Servant et al. (2021) did not attempt to fit the proportion and latency of partial EMG bursts, nor did they examine the predictive accuracy of the model with respect to these variables. Figure 2A shows the observed versus predicted proportion of correct trials containing at least one partial EMG burst during PMT (upper plot) and the mean latency of the first partial EMG burst (lower plot) averaged across subjects for each motion coherence condition. Model predictions are computed using best-fitting parameters from Servant et al. (2021), and 100,000 simulated trials per condition. DTDM strongly overestimates the proportion of correct trials containing at least one partial burst during PMT, and underestimates the average latency of the first partial burst, especially for low coherence levels. Similar results were obtained when considering both correct and incorrect trials, and by taking the median latency of the first partial burst. One could argue that a proportion of partial bursts reflects artifacts such as tonic activity (caused by a lack of relaxation of response-relevant muscles). However, this hypothesis predicts an underestimation

[2] The only apparent discrepancy concerned the between-trial correlation between PMT and MT. This correlation was slightly positive on average in the data, and there was some variability between participants, presumably due to the impact of noise on EMG onset detection. Additional simulations of DTDM showed that the model could predict a small positive correlation between PMT and MT when between-trial variability in drift rate is incorporated.

of the proportion of partial bursts by the model. These results demonstrate that the amount of within-trial noise predicted by DTDM is too large, causing too many oscillations around EMG thresholds, and suggest that we have missed a processing step.

As reviewed in the previous section, EEG studies have identified two electrical signals that exhibit key signatures of the theoretical accumulation-to-threshold decision variable, with important functional differences. The first signal (the CPP) appears to perform a decision about alternative categories of a stimulus, and is fully supramodal. We refer to this processing stage as *decision-making*. The second signal corresponds to effector-selective *motor preparation* activities. Decision-making and motor preparation EEG signals also exhibit a different sensitivity to strategic influences, as manipulations of response bias and speed-accuracy modulate motor preparation signals but not the CPP (Kelly et al., 2021; Steinemann et al., 2018). DTDM approximates decision-making and motor preparation by a single evidence accumulation diffusion process, but this assumption does not capture the lag between the two corresponding EEG signals, nor does it capture their anatomical and functional differences. The same criticism applies to the diffusion model, or to other single-stage evidence accumulation models such as the leaky competing accumulator (Usher and McClelland, 2001), the linear ballistic accumulator (Brown and Heathcote, 2008), racing diffusion models (Ratcliff et al., 2003; Tillman et al., 2020), and Poisson counter models (Ratcliff and Smith, 2004; Vickers, 1970).

Verdonck et al. (2020) recently developped a dual-stage theory as an attempt to reconcile evidence accumulation models with neurophysiological findings related to decision-making and motor preparation. The theory assumes two distinct accumulation processes. The first accumulates sensory evidence, and corresponds to the decision variable. The second takes the decision variable as a continuous input, and corresponds to the motor preparation variable. The response is executed when the motor preparation variable reaches a threshold level of activation. This dual-stage architecture for decision-making and motor preparation provides a straightforward explanation to the temporal overlap between corresponding neu-

rophysiological activities, their rise-to-threshold morphology, and the modulation of their respective accumulation rate by stimulus difficulty. It also allows for a specific influence of strategic processes on each processing stage.

The theory further assumes that the decision variable is filtered during motor preparation. The filtering process increases the signal-to-noise ratio of the decision variable, at the cost of a temporal delay between decision-making and motor preparation. The filter-related delay contributes to the observed lag between the onset latency of the two corresponding neurophysiological signals, along with a necessary transmission time between decision-making and motor preparation brain areas. At the behavioral level, stronger filtering increases the probability of giving a correct response, at the cost of longer RTs (and conversely). Consequently, the filtering process may contribute to the speed-accuracy tradeoff phenomenon (slower decisions are more accurate and conversely; for reviews, see Bogacz et al., 2010; Heitz, 2014).

Verdonck et al. (2020) formalized this theory (termed "leaky integrated threshold") as an extension of the diffusion model. Specifically, decision-making and motor preparation are modeled by variables $x(t)$ and $y(t)$ respectively. $x(t)$ is similar to Equation 1. $y(t)$ takes $x(t)$ as a continuous input, and performs a leaky accumulation of $x(t)$ according to the following differential equation:

$$dy(t) = (\beta x(t) - \lambda y(t))dt, y(0) = x(0), \tag{2}$$

where $\lambda(> 0)$ corresponds to the leak parameter and β corresponds to a scaling parameter for the input. The response is executed when the motor preparation variable $y(t)$ reaches one of two thresholds. $y(t)$ essentially acts as a smoothing filter of $x(t)$, which can be understood by looking at the solution for $y(t)$:

$$y(t) = \beta \int_{-\infty}^{t} dt' x(t') e^{\lambda(t'-t)}. \tag{3}$$

The value of the decision variable at $t - t'$ seconds before the current time t con-

tributes for $x(t')e^{\lambda(t'-t)}$ to the value of the motor preparation variable at time t. The motor preparation variable $y(t)$ thus corresponds to the weighted sum of past states $x(t')$ of the decision-making variable, with weights exponentially decreasing at rate λ (the leak parameter). When λ approaches infinity, the weights tend to zero, and the model reduces to the diffusion model. Conversely, as λ decreases, the number of past states of the decision-making variable contributing to the motor preparation variable increases. This results in a reduction of random noise (i.e., a filtering effect). Verdonck et al. (2020) further showed that for large values of t, the mean of $y(t)$ is delayed by λ^{-1} relative to the mean of $x(t)$, corrersponding to the filter-related delay. Although Verdonck and colleagues did not test the model against neurophysiological data, they showed that it provides a better account of behavioral data than the diffusion model in three datasets (a face/car discrimination task, a lexical decision task, and a random dot motion task). Two of these datasets included a speed-accuracy manipulation, which was better explained by a variation of leakage than the variation of threshold separation commonly assumed in the literature (larger separation generates slower and more accurate decisions; Bogacz et al., 2006; Ratcliff, 1978; Ratcliff and Smith, 2004).

The dual-stage model from Verdonck et al. (2020) provides a theoretical account of decision-making and motor preparation, but fails to account for properties of motor execution. This failure stems from the single-threshold assumption of the model. If the threshold operates at the motor preparation level, as in the original model definition, the choice is categorically communicated to the muscles for execution, and the model cannot account for the modulation of mean MT by stimulus discriminability and partial EMG bursts. Alternatively, one may assume that the threshold operates at the motor execution level and corresponds to the response. This hypothesis, however, would lead to a continuous activation of response-relevant muscles, at odds with the discrete nature of EMG bursts. It follows that the dual-stage model from Verdonck et al. (2020) should be combined with the dual-threshold assumption from DTDM to provide a complete theoretical account of decision-making, motor preparation, and motor execution processes. We introduce this Dual-Stage Dual-Threshold

Diffusion Model (DSDTDM) in the next section, and show how the model can capture partial EMG burst statistics.

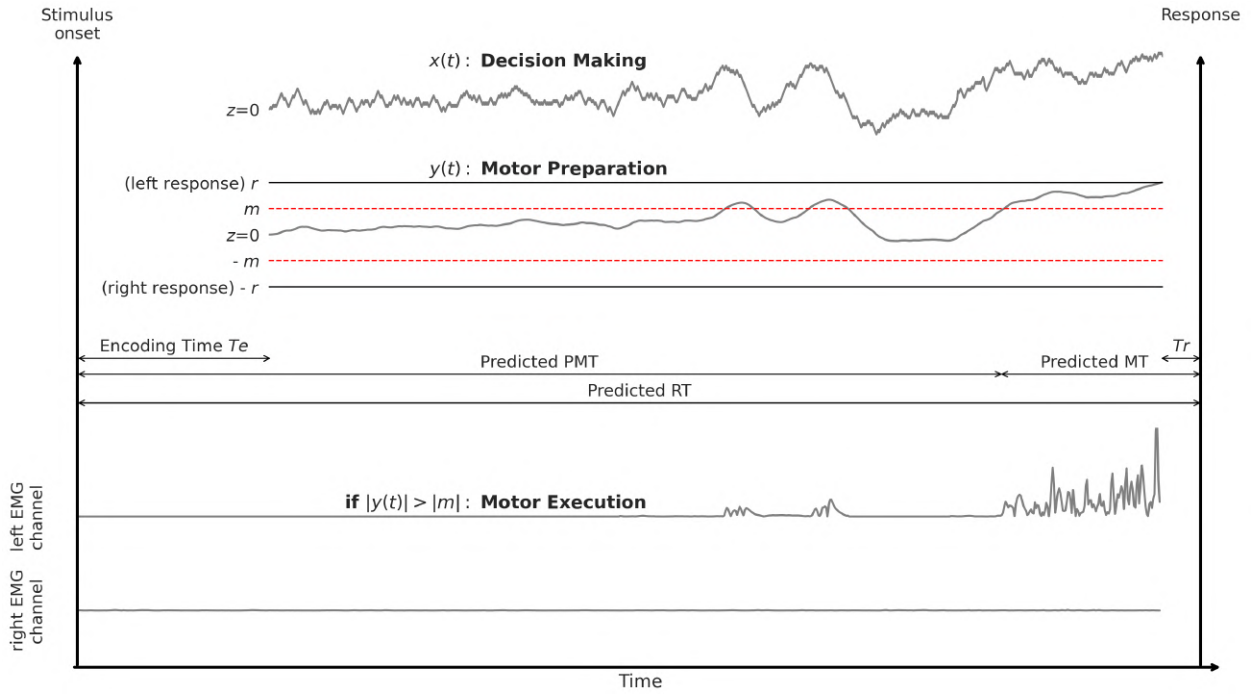
The dual-stage dual-threshold diffusion model

An illustration of DSDTDM is provided in Figure 1. Decision-making and motor preparation are modeled by $x(t)$ (Equation 1) and $y(t)$ (Equation 2) respectively. The motor preparation variable $y(t)$ is transmitted to the response-relevant muscles when $|y(t)| \geq |m|$, corresponding to the beginning of motor execution. The transmission persists until a sufficient amount of EMG activity (and thus force) has been produced to issue the response, which formally corresponds to $|y(t)| \geq |r|$. DSDTDM thus combines modeling developments from Servant et al. (2015, 2021) and Verdonck et al. (2020) to provide an integrated and biologically-inspired model of decision-making, motor preparation, and motor execution.

Within DSDTDM, the evidence filtering mechanism occuring at the motor preparation level should reduce the predicted proportion of partial EMG bursts, and increase their mean latency (due to the filter-related delay). Figure 2B shows simulations of the model with varying levels of leak λ and drift rate v. Similar to DTDM, the model predicts an increase in the proportion of correct trials containing at least one partial EMG burst during PMT (upper plot) and an increase in the averaged latency of the first partial burst as drift rate decreases (lower plot). Importantly and as predicted, the former decreases and the latter increases as the amount of leak decreases. Additional analyses of simulated data showed that DSDTDM predicts an increase in mean PMT and mean MT as drift rate decreases for each level of leak, and an approximately linear PMT quantile-MT quantile plot. Interestingly, low leak levels produce a small positive between-trial correlation between PMT and MT, especially for high drift rates (Supplementary Figure 2). This complex pattern results from two opposite forces: the Markov property of the diffusion process on the one hand (that predicts a null correlation between PMT and MT) and the filtering process on the other hand (that reduces random fluctuations and positively increases the correlation). Because

Figure 1

Architecture of the Dual-Stage Dual-Threshold Diffusion Model (DSDTDM) of decision-making, motor preparation, and motor execution



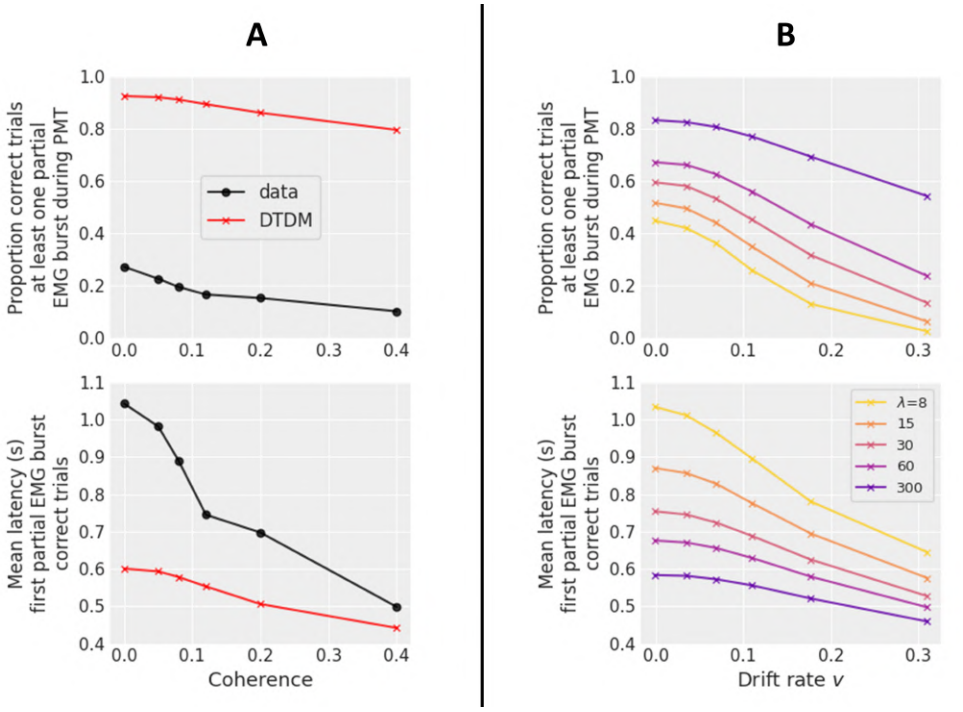
Note. The decision variable $x(t)$ accumulates evidence from the stimulus according to Equation 1. The motor preparation variable $y(t)$ takes $x(t)$ as a continuous input, and performs a leaky accumulation of $x(t)$ according to Equation 2. In the context of a choice task involving left versus right manual responses, $y(t)$ is assumed to take place at the motor cortex level. EMG and response thresholds (parameters $\pm m$ and $\pm r$ respectively) operate on $y(t)$. $y(t)$ is transmitted to the muscles associated with the left manual response if $y(t) \geq m$, and to the muscles associated with the right manual response if $y(t) \leq -m$. The transmission persists until a sufficient amount of EMG activity (and thus force) has been produced to issue the response, which formally corresponds to $|y(t)| \geq |r|$. In this example, the model predicts two partial EMG bursts in the left EMG channel before the EMG burst associated to the left response. The predicted PMT corresponds to the latency at which $y(t)$ crosses an EMG threshold for the last time before reaching the corresponding response threshold. The predicted MT corresponds to the latency at which $y(t)$ hits the response threshold minus PMT. In addition, PMT and MT each incorporate residual processing components with mean duration Te and Tr respectively. At minimum, Te includes stimulus encoding processes and the corticomuscular transmission time. Tr corresponds to the electromechanical delay.

processing components are likely variable from trial to trial (e.g., Laming, 1968; Ratcliff and Rouder, 1998), we explored the effect of between-trial variability in DSDTDM parameters. The only noticeable difference in model predictions was caused by between-trial variability

in drift rate (normally distributed with mean v and standard deviation sv). This source of variability produces a positive between-trial correlation between PMT and MT, especially for low leak and high drift rate levels (Supplementary Figure 3). It also predicts a slightly curvilinear PMT quantile-MT quantile plot (the departure from linearity increases as sv increases).

Figure 2

Partial EMG bursts statistics in a random dot motion task with varying levels of motion coherence and model predictions



Note. A: Proportion of correct trials containing at least one partial EMG burst during PMT (upper plot) and mean latency of the first partial burst (lower plot) averaged across subjects as a function of motion coherence. Observed data are shown as black dots, and DTDM predictions are shown as red crosses. B: DSDTDM simulations, with varying levels of leak λ and drift rate v. Apart from the leak parameter, simulations used best-fitting DTDM parameters averaged across subjects reported by Servant et al. (2021) and 100,000 simulated trials per condition.

Given the complexity of DSDTDM, it is difficult to guarantee that these predictions are robust across the whole (plausible) parameter space. In our opinion, a complete test of the model requires three key ingredients: (i) a quantitative fit to both behavioral

and electrophysiological data (EEG and EMG); (ii) a comparison with DTDM as a benchmark using model selection techniques; (iii) an evaluation of the fit quality of the model to both electrophysiological and behavioral data from a range of choice RT tasks that tap into different cognitive domains. The latter is important, because it will offer an assessment of the generality of the model and delineate potential boundary conditions of application. We aimed to incorporate the three ingredients in the present work, in order to provide the first attempt to jointly model decision-making, motor preparation, and motor execution processing stages. However, we restricted our analyses to behavioral and EMG data from choice tasks that tap into four different cognitive domains (motion perception, numerical cognition, recognition memory, and lexical knowledge) in order to maintain a manageable amount of electrophysiological and modeling work. Although EEG data could supplement our assessment of decision-making and motor preparation processes, the poor signal-to-noise ratio of EEG precludes a reliable evaluation of the hypothesis that the decision variable is filtered at the motor preparation level. EMG data offers a unique way to test this hypothesis, as explained previously.

Experiment 1: Motion perception

As a first step, we fit DSDTDM to the joint distributions of PMT and MT in correct and incorrect trials and to accuracy data from the left/right random dot motion task of Servant et al. (2021). The drift rate v was the only parameter free to vary across motion coherence levels. We predicted a better performance (balance between fit quality and parsimony) of DSDTDM compared to DTDM, especially with regards to partial EMG burst statistics. Arguably, there are several ways to incorporate partial burst statistics into the loss function quantifying the discrepancy between data and model predictions. We chose the following scheme for its simplicity. Accuracy data was divided into six trial types: (i) *pureC* (correct response, no partial EMG burst during PMT); (ii) *CC* (correct response, at least one partial EMG burst during PMT, first partial burst located in the correct EMG channel); (iii) *IC* (correct response, at least one partial EMG burst during PMT, first partial burst located in the incorrect EMG channel); and so forth for incorrect responses (*pureI*, *II*, *CI*). The proportion of each of these six trial types was incorporated into the loss function. Comparisons between DSDTDM and DTDM were performed with and without between-trial variability in processing components, in order to examine the robustness of findings.

Method

Critical details of the experiment are presented below, but readers are directed to Servant et al. (2021) for full details. Eighteen healthy and right-handed participants (two men; age range: 18-32; mean age: 21.1) from the University of Franche-Comté performed a random dot motion task with six levels of coherence (0, .05, .08, .12, .2, .4). In each trial, participants had to determine the global direction (leftward versus rightward) of dots, and press the corresponding response button with their left or right thumb. The force required to press each button was ∼900 gram-force. A relatively high force level was used in order to increase the separation between theoretical EMG and response bounds, and maximize the probability of detecting motion coherence effects on mean MT. The EMG activity of

response-relevant muscles (the *flexor pollicis brevis* in particular) was recorded by means of two electrodes fixed 1 cm apart on the skin of the thenar eminence of each hand. Participants performed 12 blocks of 96 trials each, with a short break between blocks. Within each block, trials were defined by a factorial combination of motion direction (left versus right) and motion coherence (six levels). All types of trials occured equally often, and were presented in a random order. Each trial started with the presentation of the random dot motion stimulus, which remained on the screen until the participant responded. A RT deadline was set to 5 s, and the interval between the response to the stimulus and the next trial was 1.5 s.

Bipolar EMG signals (sampling rate = 1024 Hz) were high-pass filtered using a 10 Hz cut-off (3^{rd} order Butterworth filter) and epoched -0.5 s to 5 s relative to stimulus onset. For each epoch, EMG burst onsets were detected using a three-step semi-automatic procedure (see Servant et al., 2021). EMG onsets could sometimes not be detected due to high tonic muscular activity (7.5% of trials on average; range 0.2-24%), and these trials were discarded from all analyses.

Models and fit procedure

DTDM and DSDTDM were coded in C, using the method and framework of Evans (2019). The fit procedure was coded in Python. The time step was set to .001 s and the diffusion coefficient σ (see Equation 1) was fixed at .1 to avoid complications arising from the scaling property of the models (see Donkin, Brown, et al., 2009). Following Verdonck et al. (2020), we further removed a redundant parameter in DSDTDM. Observe that the motor preparation process $y(t)$, the separation between EMG thresholds, and the separation between response thresholds can be multiplied by the same factor $\frac{\lambda}{\beta}$ without changing model predictions. Susbtituting $y(t)$ by $y'(t)\frac{\beta}{\lambda}$ in Equation 2 gives:

$$dy'(t) = (\lambda x(t) - \lambda y'(t))dt, y'(0) = x(0). \tag{4}$$

This simple rescaling removes parameter β. In its raw form (i.e., without between-trial variability in any of the model parameters), DSDTDM thus has 11 free parameters: one drift rate v for each of the six motion coherence levels, EMG thresholds $\pm m$, response thresholds $\pm r$, mean residual latencies Te and Tr, and the leak parameter λ). Following Servant et al. (2021), we fixed the starting point of the evidence accumulation process halfway between EMG and response thresholds, since left and right responses were equiprobable. Relaxing this contraint did not change model selection results, and had a negligible impact on the goodness-of-fit of the models. The raw DTDM has 10 free parameters (all DSDTDM parameters except λ), and the full DTDM has four additional parameters (between-trial variability in drift rate sv, starting point sz, and mean residual latencies sTe and sTr). sv corresponds to the standard deviation of a Gaussian distribution with mean v. sz, sTe, and sTr correspond to the range of a uniform distribution with mean z, Te, and Tr respectively. These distributional assumptions are directly inherited from standard applications of the diffusion model (Boehm et al., 2018; Ratcliff and Rouder, 1998; Voss et al., 2004; Wiecki et al., 2013). The full DSDTDM has one additional between-trial variability parameter, corresponding to between-trial variability in leakage (uniformally distributed with range $s\lambda$ and mean λ). A uniform distribution was chosen because we do not have any theoretical assumption about the distributional shape of variability in leakage.

The models were fit to each individual dataset by minimizing the following loss function (likelihood-ratio chi-square statistic):

$$G^2 = 2\sum_{i=1}^{6}\sum_{j=1}^{6}\sum_{k=1}^{6}\sum_{l=1}^{6} n_{ijkl} log(\frac{n_{ijkl}}{\frac{pred\_n_{ijkl}N_i}{simul\_N_i}}). \quad (5)$$

Summations over i and j extend over the 6 motion coherence levels and the six trial types (pureC, CC, IC, pureI, II, CI; see introduction section of this experiment) respectively. Summations over k and l extend over the six bins bounded by PMT quantiles (.1, .3, .5, .7, and .9) and the six bins bounded by MT quantiles (.1, .3, .5, .7, and .9) respectively[3].

[3] If subjects made a number of errors comprised between five and 10 in a given condition, a median split

The variables n_{ijkl} and $pred\_n_{ijkl}$ refer to the observed and predicted number of trials in coherence condition i, trial type j, PMT bin k, and MT bin l. Finally, the variables N_i and $simul\_N_i$ refer to the observed and simulated number of trials in coherence condition i, and *log* refers to the natural logarithm. The G^2 statistic thus characterizes the goodness-of-fit of the model to the joint distributions of PMT and MT and to the proportion of each of the six trial types. It was minimized using differential evolution (Storn and Price, 1997) and 20,000 simulated trials per condition. Observe that we did not incorporate the latency of partial bursts into the G^2 formula, in order to mitigate the potential impact of artifactual partial bursts on the fit quality of other aspects of the data. The latency of partial bursts can thus be considered as out-of-sample data, and the comparison between these data and model predictions will serve as a generalization test of the models.

Before turning to model comparison techniques, it is important to remember that DTDM is nested in DSDTDM (see general introduction): the two models are equivalent when the leak parameter λ approaches infinity. Consequently, a low best-fitting leak value would indicate that DSDTDM adds to a DTDM description of the data. The key question is whether this improvement in fit quality is sufficiently important to justify the additional complexity of the DSDTDM model. To answer this question, the G^2 was converted to both AIC and BIC model selection statistics:

$$AIC = G^2 + 2m, \tag{6}$$

$$BIC = G^2 + mlog(N), \tag{7}$$

where m corresponds to the number of free parameters, *log* corresponds to natural logarithm, and N equals the number of observations used in the G^2 computation. BIC and AIC thus both penalize for model complexity, but in a different way. Since both statistics have

---

was used to form two bins. It there were fewer than five errors, error RTs for the condition were excluded from the G^2 calculation.

advantages and drawbacks (Vrieze, 2012), we report both of them, hoping for consistency between model decisions. For each individual subject, the best model is the one associated with the smallest AIC or BIC. If 13 (or more) out of 18 subjects support one model over the other in terms of AIC or BIC (two-sided binomial test), then the result is significant.

Results

DSDTDM was associated with lower AIC and BIC statistics compared to DTDM for each of the 18 subjects of the experiment, and for both raw and full model variants (Figure 3A). The difference in AIC and BIC between raw and full models was much smaller for DSDTDM compared to DTDM, indicating that between-trial variability in DSDTDM parameters has a minor impact on model performance, contrary to DTDM. In fact, the raw DSDTDM was associated with lower AIC (BIC) statistics compared to the full DTDM for 16 (15) subjects. This analysis provides strong evidence for the superiority of DSDTDM.

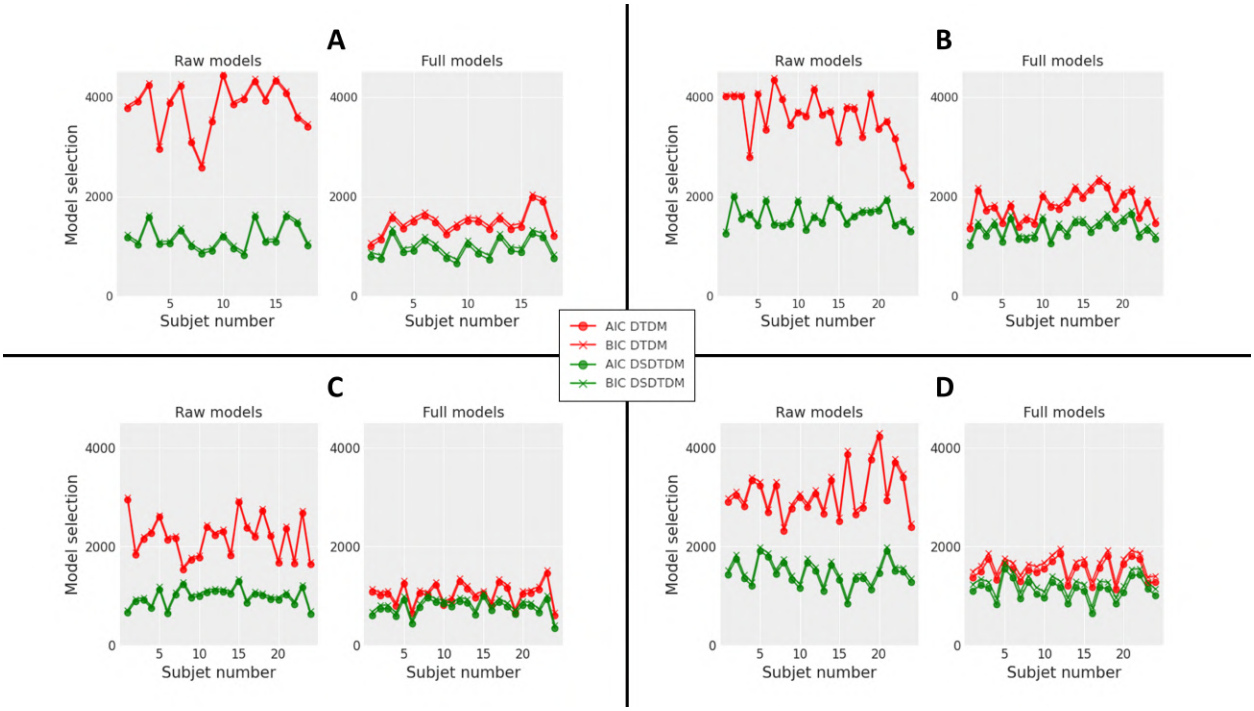
Best-fitting parameters for the full models are shown in Table 1 (main parameters) and Table 2 (between-trial variability parameters). Best-fitting parameters for the raw models are shown in Supplementary Table 1. As predicted, both raw and full DSDTDM capture the EMG data with a low level of leakage, indicating strong filtering of the evidence accumulation variable during motor preparation (model trajectories for decision-making and motor preparation variables computed from best-fitting parameters averaged across subjects are illustrated in Figure 4A). Note that the amount of between-trial variability in the best-fitting full model components was higher for DTDM compared to DSDTDM, especially for residual latencies (parameters sTe and sTr).

Figure 5 displays the goodness-of-fit of the full models to several aspects of the data. DSDTDM predictions are displayed in red, DTDM predictions in green, and the data in black. Figure 5A shows observed versus predicted mean PMT (upper plot) and mean MT (lower plot) in correct trials averaged across subjects. Figure 5B displays observed versus predicted quantile probability functions for both PMT (upper plot) and MT (lower

plot) distributions averaged across subjects. Quantile probability functions are constructed by plotting PMT or MT quantiles (y-axis) of the distributions of correct and incorrect responses for each condition against the corresponding response type proportion (x-axis). Five quantiles (.1, .3, .5, .7, .9) were chosen to provide a summary of the shape of PMT and MT distributions. If PMT and MT are uniformly distributed, the temporal separation between adjacent quantiles should be constant. If PMT and MT both exhibit a right-skewed shape, as evidenced in our previous work (Servant et al., 2021) and visible in Figure 5B, the temporal separation between .7 and .9 quantiles should be larger than the separation between .5 and .7 quantiles, the separation between .5 and .7 quantiles should be larger than the separation between .3 and .5 quantiles, and so on. Quantile probability functions thus represent a synthetic way to examine the shape of PMT and MT distributions for correct and incorrect responses, and how this shape varies across conditions (for a thorough treatment of quantile probability functions, see Ratcliff and Smith, 2004). Note that the five PMT and MT quantiles for incorrect responses in a given condition are displayed if each subject made at least 10 errors in that condition (the three easiest motion coherence conditions did not fulfill this criterion). Figure 5C shows the observed versus predicted proportion for each of the six trial types (pureC, CC, IC, pureI, II, CI) averaged across subjects. Figure 5D shows the observed versus predicted proportion of correct trials featuring at least one partial EMG burst during PMT (upper plot), and the mean latency of the first partial EMG burst averaged across subjects (lower plot). Figure 5E displays the observed versus predicted PMT quantile-MT quantile plot (computed from nine decile points) from correct trials averaged across subjects. Finally, Figure 5F shows the observed versus predicted between-trial Pearson correlation coefficient between PMT and MT in correct trials for each subject (scattered dots and crosses), as well as the correlation averaged across subjects (horizontal lines). The data shown in the lower plot of Figure 5D (mean latency of the first partial EMG burst in correct trials) and Figure 5F (between-trial Pearson correlation between PMT and MT) were not factored out in parameter estimation, and serve as a generalization test of the models.

Figure 3

Model selection statistics for Experiments 1-4



Note. Panels A-D correspond to Experiments 1-4.

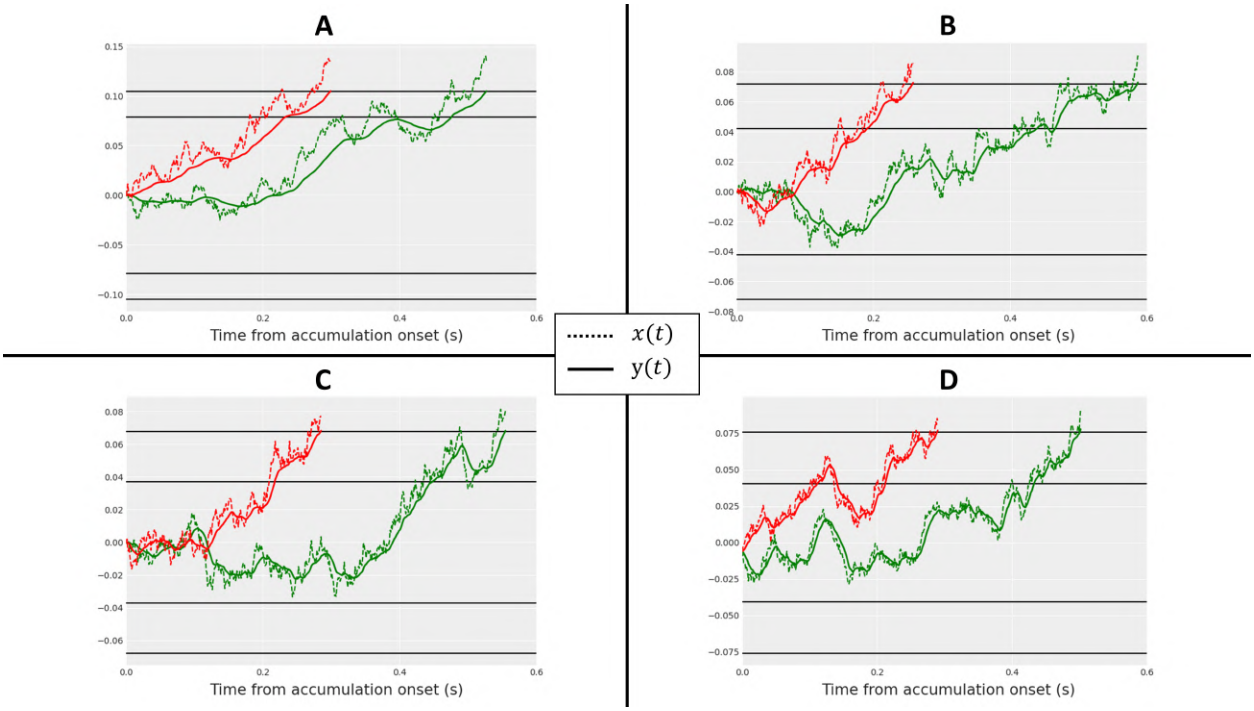
Overall, the full DSDTDM provides a good account of the data, though a few minor discrepancies are apparent. First, the model overestimates the .9 quantile of PMT distibutions as motion coherence decreases, especially for incorrect trials. Second, the model overestimates the mean latency of the first partial EMG burst in correct trials, though this discrepancy is reduced for low motion coherence levels, where the proportion of correct trials containing at least one partial EMG burst is the largest. Finally, the predicted between-trial correlation between PMT and MT for each individual subject shows less dispersion compared to observed data, likely due to noise in EMG onset detection. In addition, the model slightly overestimates the correlation for the highest motion coherence level. As discussed previously (see general introduction and supplementary Figures 2 and 3), DSDTDM predicts a positive correlation when a high drift rate is combined with a low leakage level, especially if between-trial variability in drift rate is incorporated. It is difficult to determine whether

this discrepancy is due to a problem in the DSDTDM architecture or to a misspecification of drift rate distributions, given that the model does not incorporate representational assumptions specifying how these distributions arise from the random dot motion stimuli. We will address this limitation in the next experiment.

In its raw form, DTDM grossly overestimates proportions of trials containing at least one partial EMG burst during PMT, replicating the failure of the model highlighted in the general introduction section. Since this failure was apparent in each of the four experiments presented in this paper, the raw DTDM will no longer be discussed. The full DTDM provides a better account of the six trial types (pureC, CC, IC, pureI, II, CI), though the model overestimates the proportion of CC trials as motion coherence increases. The better performance of the full DTDM comes from a smaller separation between EMG and response bounds, but this modulation has several negative consequences. Most importantly, the predicted mean MT essentially corresponds to the mean residual latency added to predicted MT (parameter Tr), and the predicted variability in MT is mostly driven by the between-trial uniform variability in this latency (sTr). As a consequence, the model strongly underestimates the effect of motion coherence on mean MT, and fails to account for the right-skewed distibution of MTs (observe the constant temporal separation between adjacent MT quantiles predicted by the model in Figure 5B, diagnostic of a uniform distribution).

Figure 4

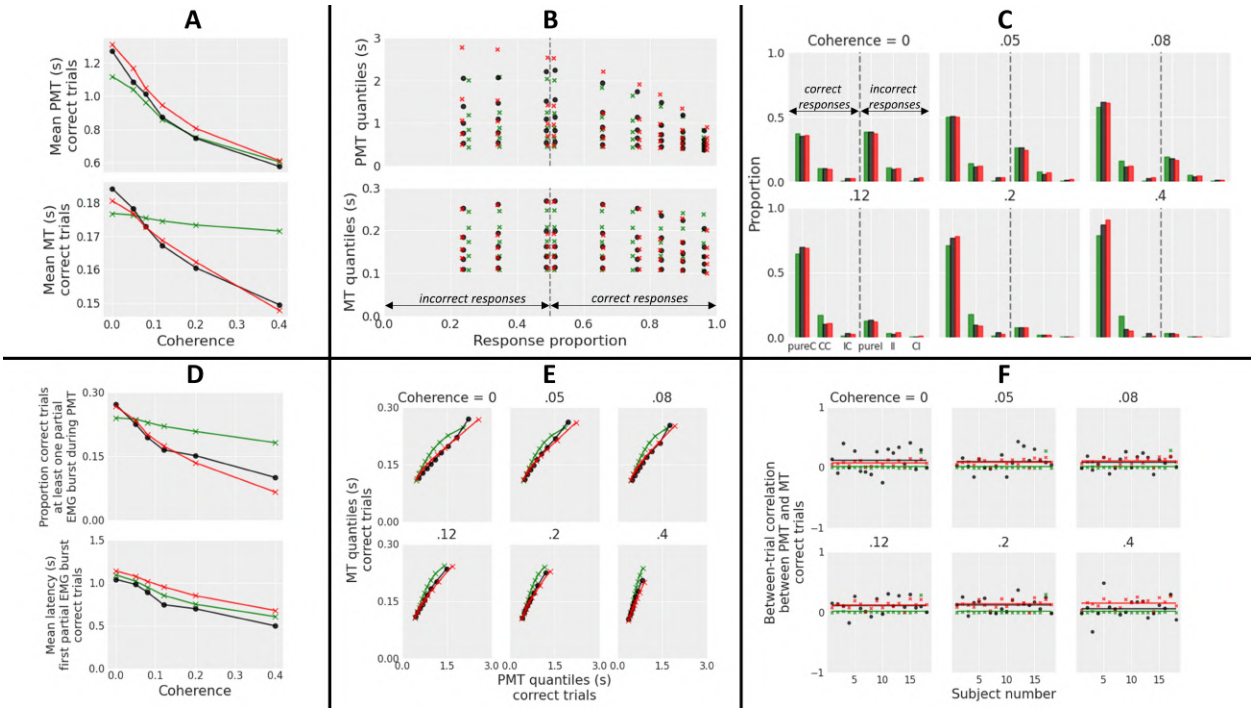
Trajectories of decision-making $x(t)$ and motor preparation $y(t)$ variables computed from the full DSDTDM using best-fitting parameters averaged across subjects from Experiments 1-4



Note. Panels A-D correspond to Experiments 1-4. For each experiment, two trials (red and green trajectories) were simulated using best-fitting DSDTDM parameters averaged across subjects, with the constraint of giving decision times in range .2-.3 s (red trajectory) and .5-.6 s (green trajectory) to facilitate comparison between experiments. Between-trial variability parameters were set to 0, and drift rate corresponded to condition coherence = 8% for Experiment 1, mean numerosity = 58 for Experiment 2, old two presentations for Experiment 3, and very low frequency words for Experiment 4. The decision-making variable ($x(t)$, dashed line) was simulated according to Equation 1, and the motor preparation variable ($y(t)$, plain line) was simulated according to Equation 4.

Figure 5

Data from a random dot motion task with varying levels of motion coherence (black) plotted against predictions from the full DSDTDM (red) and the full DTDM (green)



Note. Model predictions are computed from best-fitting parameters, using 100,000 simulated trials per condition. The six panels A-F display different aspects of the data. A: Mean PMT (y-axis, upper plot) and mean MT (y-axis, lower plot) in correct trials as a function of motion coherence (x-axis) averaged across subjects. B: Quantile probability functions averaged across subjects for each motion coherence condition, constructed by plotting PMT quantiles (.1, .3, .5, .7, .9; y-axis, upper plot) and MT quantiles (.1, .3, .5, .7, .9; y-axis, lower plot) of the distributions of correct and incorrect responses against the corresponding response type proportion (x-axis). The five PMT and MT quantiles for incorrect responses in a given condition are displayed if each subject made at least 10 errors in that condition. C: Proportion of each of the six trial types described in the introduction section of this experiment (pureC, CC, IC, pureI, II, CI) for each motion coherence condition averaged across subjects. D: Proportion of correct trials featuring at least one partial EMG burst during PMT (y-axis, upper plot) and mean latency of the first partial burst (y-axis, lower plot) as a function of motion coherence (x-axis) averaged across subjects. E: MT quantiles (y-axis) plotted against PMT quantiles (x-axis) from correct trials for each condition averaged across subjects. Quantiles are computed from nine decile points. F: Between-trial Pearson correlation coefficient between PMT and MT in correct trials for each motion coherence condition. Observed data and model predictions for each individual subject are shown as scattered black points and red crosses respectively. Observed data and model predictions averaged across subjects are shown as black and red horizontal lines respectively.

Table 1

Main parameters from the full DTDM and DSDTDM models averaged across subjects for Experiments 1-4.

Exp	Model	v_1	v_2	v_3	v_4	v_5	v_6	dc	λ	m	r	Te_1	Te_2	Te_3	Te_4	Te_5	Tr	z
1	DTDM	-0.014	0.087	0.176	0.261	0.383	0.689			0.091	0.099	0.408					0.163	
1	DSDTDM	-0.002	0.075	0.139	0.196	0.284	0.491		25.232	0.079	0.105	0.287					0.090	
2	DTDM	0.066						-0.035		0.065	0.101	0.474					0.148	0.006
2	DSDTDM	0.023						-0.011	57.842	0.042	0.072	0.339					0.106	0.000
3	DTDM	0.207	0.518	0.739	-0.656					0.045	0.069	0.544					0.139	0.001
3	DSDTDM	0.138	0.288	0.399	-0.342				73.158	0.037	0.068	0.452					0.099	0.000
4	DTDM	0.459	0.683	0.809	0.878	-0.571				0.046	0.082	0.589	0.571	0.549	0.512	0.581	0.133	-0.004
4	DSDTDM	0.277	0.393	0.460	0.517	-0.301			122.938	0.041	0.076	0.500	0.471	0.454	0.430	0.489	0.109	-0.006

Note. Experiment 1 (motion perception): parameters v_1 to v_6 correspond to drift rates for the six motion coherence levels (from 0 to .4). Experiment 3 (recognition memory): parameters v_1 to v_4 correspond to drift rates for conditions old one presentation, old two presentations, old four presentations, and new respectively. Experiment 4 (lexical knowledge): parameters v_1 to v_5 and Te_1 to Te_5 correspond to drift rates and mean residual latency added to predicted PMT for conditions very low frequency words, low frequency words, medium frequency words, high frequency words, and pseudowords respectively.

Table 2

Between-trial variability parameters from the full DTDM and DSDTDM models averaged across subjects for Experiments 1-4.

Exp	Model	sv_1	sv_2	sv_3	sv_4	sv_5	σ_1	η_0	sz	sTe	sTr	$s\lambda$
1	DTDM	0.161							0.116	0.357	0.157	
1	DSDTDM	0.158							0.094	0.216	0.073	33.507
2	DTDM						0.007	0.050	0.098	0.422	0.133	
2	DSDTDM						0.001	0.042	0.060	0.225	0.078	82.803
3	DTDM	0.439	0.427	0.426	0.274				0.065	0.327	0.113	
3	DSDTDM	0.335	0.313	0.263	0.163				0.052	0.204	0.065	86.538
4	DTDM	0.522	0.435	0.372	0.312	0.267			0.071	0.456	0.107	
4	DSDTDM	0.334	0.266	0.208	0.195	0.122			0.057	0.350	0.075	146.702

Note. Experiment 3 (recognition memory): parameters sv_1 to sv_4 correspond to between-trial variability in drift rate for conditions old one presentation, old two presentations, old four presentations, and new respectively. Experiment 4 (lexical knowledge): parameters sv_1 to sv_5 correspond to between-trial variability in drift rate for conditions very low frequency words, low frequency words, medium frequency words, high frequency words, and pseudowords respectively.

Experiment 2: Numerical cognition

Many tasks in numeracy research involve a decision between two responses based on the magnitude of some non symbolic stimulus. For example, subjects have to determine which of two arrays that are spatially separated feature the larger amount of dots, or whether an array of dots contains more blue or yellow dots. Here we use another common task in numeracy research in which subjects have to determine whether the number of dots randomly scattered in a 10×10 virtual array is greater or less than a criterion quantity (50). Performance is slower and less accurate when the difference between the number of dots and the criterion is small (e.g., 45 or 55 dots) compared to when it is large (e.g., 30 or 70 dots). Ratcliff and colleagues have demonstrated that the diffusion model captures RT distributions for correct and incorrect responses and accuracy data in this task with a variation of drift rate across numerosity conditions (e.g., Ratcliff and Childers, 2015; Ratcliff et al., 2010). Ratcliff and McKoon (2018) further showed that the modulation of drift rate could arise from an approximate number representation in which numerosities are represented as Gaussian distributions, with the mean and standard deviation of these distributions increasing linearly with numerosity (Dehaene, 2003). In this framework, the drift rate corresponds to the difference between the number of dots and the criterion, scaled by a free parameter (to account for interindividual differences in discrimination performance). Consequently, both DTDM and DSDTDM predict an increase in mean MT as the number of dots approaches the criterion, resulting in an inverted U-shaped function of numerosity (with a peak around 50). The models were fit to data and compared using the same procedure as in Experiment 1.

Method

Participants

Manipulations of numerosity produce smaller modulations of mean RT compared to motion coherence manipulations. Consequently, we increased the sample size from 18

to 24 subjects in order to maintain a reasonable amount of statistical power while keeping electrophysiologial and modeling work in manageable proportions.

Twenty-four students (five men; mean age: 20.6) from the University of Franche-Comté took part in the experiment in exchange for course credits or as volunteers. All subjects met the following criteria: being 18 to 30 years old, being right-handed, having a normal or corrected vision, and having no history of motor, psychiatric or neurological disorder. Subjects were not aware of the purpose of the experiment and provided written consent to participate. This study was approved by the ethical committee for research of the University (agreement n°CERUBFC-2022-01-18-002).

Appartus

The experiment took place in a dimly lit room. Subjects sat on a comfortable chair at a distance of 75 cm from a 34.7 × 19.5 cm LCD monitor (resolution: 1920 × 1080; framerate: 60 Hz). The experiment was programmed in python, using functions from the PsychoPy library (Peirce et al., 2019). Response buttons were identical to those used in Experiment 1. Subjects' hands were faced palm-down, resting on a supportive cushion placed on their laps in order to minimize tonic muscular activity and maximize comfort.

Stimuli

For each trial, between 31 and 70 black dots were displayed on a grey background screen, in random positions within a 10 x 10 virtual grid. Each dot was 0.24° in diameter. The horizontal and vertical separation between two adjacent dots (from center to center) was 1.15°.

Procedure

Participants were instructed to press the left button with their left thumb if they judged the number of dots displayed was less than or equal to 50, and the right button with their right thumb if they judged it was greater. Left responses to 31-50 dots and right

responses to 51-70 dots were counted as correct. Participants were told not to count the dots but instead provide a global and rapid estimation of their number. Dots remained on the screen until the participant responded. A RT deadline was set to 4 s. If participants failed to respond by then, the message “Too late! Please respond faster.” was displayed for 1.5 s. The intertrial interval was 1.5 s. Participants first completed a pratice block of 40 trials, containing each of the 40 numerosity conditions presented in a random order. A feedback on performance ("Correct response" or "Incorrect response") was displayed for 1.5 s after each response. Practice trials were not considered in the analyses. Subjects then completed 30 blocs of 40 trials with a similar structure, except that no feedback was provided after each response. Blocks were separated by self-paced breaks. The experiment lasted about an hour.

EMG recordings and signal processing

The procedure used for EMG recordings and signal processing was similar to Experiment 1, except that EMG signals were epoched -0.5 s to 4 s relative to stimulus onset. Trials with high tonic muscular activity were discarded from analyses (2.7% of trials on average; range 0-11.7%).

Models and fit procedure

The data was grouped into 8 conditions (31-35 dots; 36-40; 41-45; 46-50; 51-55; 56-60; 61-65; 66-70), represented by the mean number of dots N of each bin (33, 38, 43, 48, 53, 58, 63, 68). Following Ratcliff and McKoon (2018), we assumed that the drift rate v of DTDM and DSDTDM is a linear function of N and the criterion (50):

$$v = dc + v_1(N - 50), \tag{8}$$

where v_1 accounts for interindividual differences in discrimination performance and dc (drift criterion) accounts for interindividual variations in the representation of the criterion. Between-trial variability in drift rate sv was defined as:

$$sv = \eta_0 + \sigma_1\sqrt{N^2 + 50^2}. \tag{9}$$

All parameters were fixed across numerosity conditions. We treated the starting point z of the evidence accumulation process as a free parameter, resulting in seven (12) free parameters for the raw (full) DTDM and eight (14) free parameters for the raw (full) DSDTDM[4]. The fit procedure was identical to that used in Experiment 1, except that we modeled 'lesser than' and 'greater than' responses (instead of incorrect/correct responses, due to the starting point of the evidence accumulation process being free to vary), corresponding to lower and upper response bounds respectively. As a consequence, the six trial types considered in the fit procedure were *pureL* ('lesser than' response, no partial EMG burst during PMT), *LL* ('lesser than' response, at least one partial EMG burst during PMT, first partial burst located in the 'lesser than' EMG channel), *GL* ('lesser than' response, at least one at least one partial EMG burst during PMT, first partial burst located in the 'greater than' EMG channel), and so forth for 'greater than' responses (*pureG*, *GG*, *LG*). Model selection statistics (AIC and BIC) were computed using Equations 6 and 7. By a two-sided binomial test, if 17 out of 24 subjects support one model over the other, then the result is significant.

Results

Behavior and EMG

Anticipations (RTs < 150 ms; 0%) and trials in which participants failed to respond before the 4 s deadline (0.12%) were discarded from analyses. The data were analyzed by means of quadratic contrasts (two-sided) with numerosity as within-subjects factor and specific error terms (as recommended for within-subjects designs; e.g., Boik, 1981). Accuracy data exhibited a U-shaped function of numerosity ($t(23) = 31.36$, $p < .001$), reflecting the increased proportion of errors as numerosity approaches the criterion. Consistent with model predictions, mean RT, mean PMT, and mean MT showed an inverted U-shaped function of

[4] Ratcliff and McKoon (2018) also evaluated a model variant in which subjects solve the task by comparing the number of dots to the number of blank spaces, and found similar fits. We also found similar fits for this DTDM and DSDTDM variant.

numerosity (Figure 6A; mean RT: $t(23)$ = -10.78, $p < .001$; mean PMT: $t(23)$ = -10.75, $p < .001$; mean MT: $t(23)$ = -5.37, $p < .001$). Both the proportion of correct trials containing at least one partial EMG burst and the mean latency of the first partial burst also exhibited an inverted U-shaped function of numerosity, $t(23)$ = -9.61, $p < .001$ and $t(23)$ = -6.95, $p < .001$ respectively (Figure 6D). For each condition, PMT quantile-MT quantile plots from correct trials had an approximately linear shape, and the between-trial Pearson correlation coefficient between PMT and MT was positive and close to zero on average (with a slight initial reduction followed by a more pronouced increase as numerosity increases; Figure 6F). Overall, EMG results as a function of task difficulty are similar to those observed in the random dot motion task (Servant et al., 2021).

Model fits

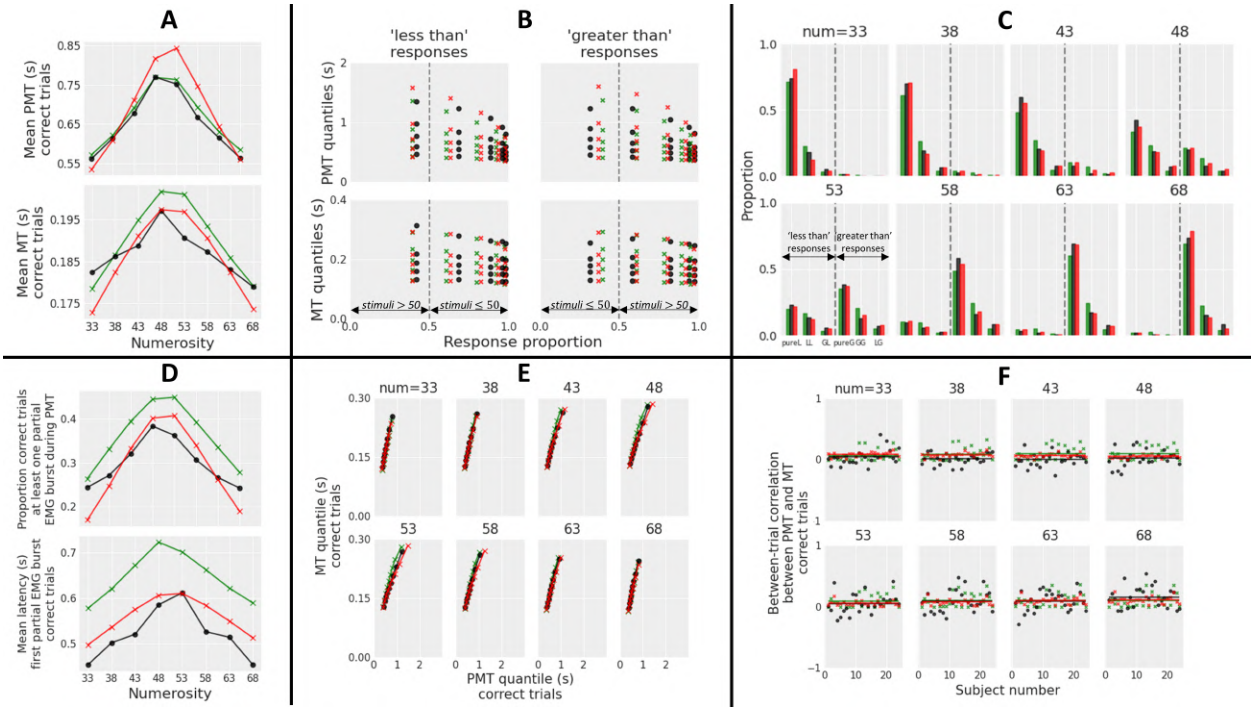
Similar to Experiment 1, DSDTDM was associated with lower AIC and BIC statistics compared to DTDM for each of the 24 subjects, and for both raw and full model variants (Figure 3B). The difference in AIC and BIC between raw and full models was much smaller for DSDTDM compared to DTDM, and the raw DSDTDM was associated with a lower AIC (BIC) compared to the full DTDM for 22 (21) subjects. This analysis provides strong evidence for the superiority of DSDTDM.

Best-fitting parameters for the full models are shown in Table 1 (main parameters) and Table 2 (between-trial variability parameters). Best-fitting parameters for the raw models are shown in Supplementary Table 1. Although the best-fitting leakage (λ) value from DSDTDM was approximately twice larger than that observed in Experiment 1, this value still implies subtantial filtering of the evidence accumulation variable during motor preparation, though with a reduced filter-related delay (for an illustration of model trajectories, see Figure 4B). The amount of between-trial variability in the best-fitting full model components was higher for DTDM compared to DSDTDM, especially for residual latencies (parameters sTe and sTr).

Figure 6 displays the goodness-of-fit of the full models to data. Compared to Experiment 1, the full DTDM provides a better account of the task difficulty effect on mean MT, thanks to a larger separation between EMG and response bounds. However, the model still fails to provide a good fit to MT quantiles, because the contribution of residual motor latencies to predicted MTs remains subtantial. In addition, the full DTDM systematically overestimates the rate of correct LL and GG trials and the mean latency of the first partial EMG burst. The full DSDTDM captures most trends of the data. The only apparent misfit is an overestimation of the right skew of PMT distributions for the most difficult conditions. Note that the model provides a reasonable account of between-trial Pearson correlation coefficients between PMT and MT across numerosity conditions, and does so with a complex combination of three ingredients: (i) a moderate and constant level of leakage across conditions, (ii) drift rates that follow a U-shaped function of numerosity, and (iii) a moderate amount of variability in drift rate that slightly increases as numerosity increases (from 0.0896 to 0.0925, computed from best-fitting parameters using Equation 9).

Figure 6

Data from a numerosity judgment task (black) and predictions from the full DSDTDM (red) and the full DTDM (green)



Note. The structure of each panel is similar to that of Figure 5. Model predictions are computed from best-fitting parameters, using 100,000 simulated trials per condition. Quantile probability functions, shown in panel B, incorporate distributions of errors if each subject made at least 10 errors in corresponding conditions. Panel C shows the proportion of each of the six trial types (pureL, LL, GL, pureG, GG, LG) for each numerosity condition averaged across subjects.

Experiment 3: Recognition memory

The diffusion model was originally developped to provide a theory of memory retrieval, and showed a good fit to behavioral data from different item recognition paradigms (Ratcliff, 1978). This finding has been replicated multiple times since (e.g., Ratcliff et al., 2004, 2010). Here we perform an EMG analysis of response-relevant muscles in a standard study-test task. During study, participants had to memorize a list of 21 words, each word being presented individually at a pace of 1 s. During test, participants were presented with 42 words (21 old and 21 new), and had to decide whether each word was old or new by pressing a left or right button. In this task, the drift rate represents the meeting point between decision-making and memory systems: it is equal to the amount of match between the test item and the memory trace. To modulate drift rate, we manipulated the number of word presentations during study (one vs. two vs. four presentations). Specifically, the drift rate should increase as the number of word presentations (and thus memory strength) increases. Consequently, both DTDM and DSDTDM predict a decrease in mean MT as the number of word presentations increases.

Although early applications of the diffusion model to recognition memory data assumed a constant between-trial variability in drift rate (parameter sv) between old and new items, there is evidence from both memory models (e.g.,Ratcliff et al., 1992; Shiffrin and Steyvers, 1997; Wixted, 2007) and diffusion model fits (e.g., Starns and Ratcliff, 2014) that the evidence entering the decision process is more variable for old than new items. One possible reason is that some old items are better learned than others (Wixted, 2007). Consequently, we let sv free to vary between conditions.

Method

Participants

Twenty-four students (six men; mean age: 19.78) from the University of Franche-Comté took part in the experiment in exchange for course credits or as volunteers. All

participants met the same inclusion criteria as for Experiment 2. Subjects were not aware of the purpose of the experiment and provided written consent to participate. This study was approved by the ethical committee for research of the University (agreement n°CERUBFC-2022-01-18-002).

Apparatus

The apparatus was identical to Experiment 2.

Stimuli

We selected a set of 1058 french words (number of letters ranging from five to eight; $M = 6.50$; $SD = 1.12$) from the Lexique database (New et al., 2004). Word frequency ranged from one to six occurences per million ($M = 2.96$, $SD = 1.14$). Words were presented in black against a grey background (font: Consolas), at the center of the screen. The height of the letters was 0.76°.

Procedure

The experiment consisted of 23 blocks (one training block with a feedback on accuracy after each trial and 22 experimental blocks without feedback), separated by self-paced breaks. Each block consisted of a study phase and a test phase. During study, participants were instructed to learn a set of 53 words made of 25 unique words. The first two words served as fillers to control for primacy effects, and the last two words served as fillers to control for recency effects. Among the 21 remaining words, seven were presented one time, seven were presented two times, and seven were presented four times. Words were presented in a random order, at a pace of 1 s. The test phase occured right after the study phase, and consisted of 42 words (21 old words and 21 new words, presented in a random order). Participants were instructed to press the left or the right button with their left or right thumb depending on whether the word was old or new (stimulus-response mapping balanced across participants). Each test word remained on screen until the participant responded, or until

a 4 s RT deadline. If participants failed to respond by then, the message "Too late! Please respond faster." was displayed for 1.5 s. The intertrial interval was 1.5 s. Words presented in a block never appeared in another block (participants were made aware of this during task instructions). In addition, the assignment of words to blocks and conditions was randomly determined. The experiment lasted about an hour and a half.

EMG recording and signal processing

The procedure used for EMG recordings and signal processing was similar to Experiment 2. Trials with high tonic muscular activity were discarded from analyses (12.4% of trials on average; range 1.6-25%).

DTDM and DSDTDM fit procedure

Drift rate and between-trial variability in drift rate were the only parameters free to vary between conditions. Similar to Experiment 2, we treated the starting point z of the evidence accumulation process as a free parameter, resulting in nine (16) parameters for the raw (full) DTDM, and 10 (18) parameters for the raw (full) DSDTDM. The fit procedure was identical to that used in the previous experiments. We modeled 'new' and 'old' responses, corresponding to lower and upper response bounds respectively. The six trial types considered in the fit procedure were *pureO* ('old' response, no partial EMG burst during PMT), *OO* ('old' response, at least one partial EMG burst during PMT, first partial burst located in the 'old' EMG channel), *NO* ('old' response, at least one at least one partial EMG burst during PMT, first partial burst located in the 'new' EMG channel), and so forth for 'new' responses (*pureN*, *NN*, *ON*).

Results

Behavior and EMG

Anticipations (RTs $<$ 150 ms; 0.004%) and trials in which participants failed to respond before the 4 s deadline (0.068%) were discarded from analyses. Performance to

old words was analyzed by means of linear contrasts (two-sided) with the number of word presentations as within-subjects factor and specific error terms. Accuracy increased as the number of word presentations increased, $t(23) = 14.14$, $p < .001$. Consistent with model predictions, mean RT, mean PMT, and mean MT decreased as word presentations increased (Figure 7A; mean RT: $t(23) = -5.72$, $p < .001$; mean PMT: $t(23) = -5.50$, $p < .001$; mean MT: $t(23) = -3.11$, $p = .011$). Note that the amplitude of the word presentation effect on mean MT ($M = 5$ ms) is smaller compared to the numerosity effect observed in Experiment 2 ($M = 10$ ms) and the motion coherence effect observed in Experiment 1 ($M = 35$ ms). The amplitude of the word presentation effect on mean PMT data ($M = 79$ ms) is also smaller compared to the numerosity effect ($M = 198$ ms) and the motion coherence effect ($M = 692$ ms). The positive correlation between the magnitude of difficulty effects on mean PMT and mean MT across tasks is consistent with the hypothesis -core to DTDM and DSDTDM- that PMT and MT are driven by a similar evidence accumulation process.

Although both the proportion of correct trials containing at least one partial EMG burst and the mean latency of the first partial burst decreased as word presentations increased (Figure 7D), only the latter reached statistical significance ($t(23) = -1.91$, $p = .069$ and $t(23) = -3.02$, $p = .006$ respectively). For each condition, PMT quantile-MT quantile plots from correct trials exhibited a slight curvilinearity (Figure 7E), and the between-trial Pearson correlation coefficient between PMT and MT was close to zero on average, with a slight decrease as word presentations increase (one presentation: $r = .07$; two presentations: $r = .04$; four presentations: $r = 0$; Figure 7F).

To compare performance between old and new items, we averaged the performance to old items across word presentation levels and ran two-sided paired sample t-tests. The only significant difference concerned accuracy data. The proportion of correct responses was higher for new than old items, $t(23) = 5.56$, $p < .001$.

Model fits

The raw DSDTDM was associated with lower AIC and BIC statistics compared to the raw DTDM for each of the 24 subjects, and the full DSDTDM was associated with lower AIC and BIC statistics compared to the full DTDM for 23 subjects (Figure 3C). The difference in AIC and BIC between raw and full model variants was smaller for DSDTDM compared to DTDM, and the raw DSDTDM was associated with a lower AIC (BIC) compared to the full DTDM for 17 (16) subjects. The pattern of model selection results is thus similar to that observed in the previous experiments, and provides strong evidence for DSDTDM.

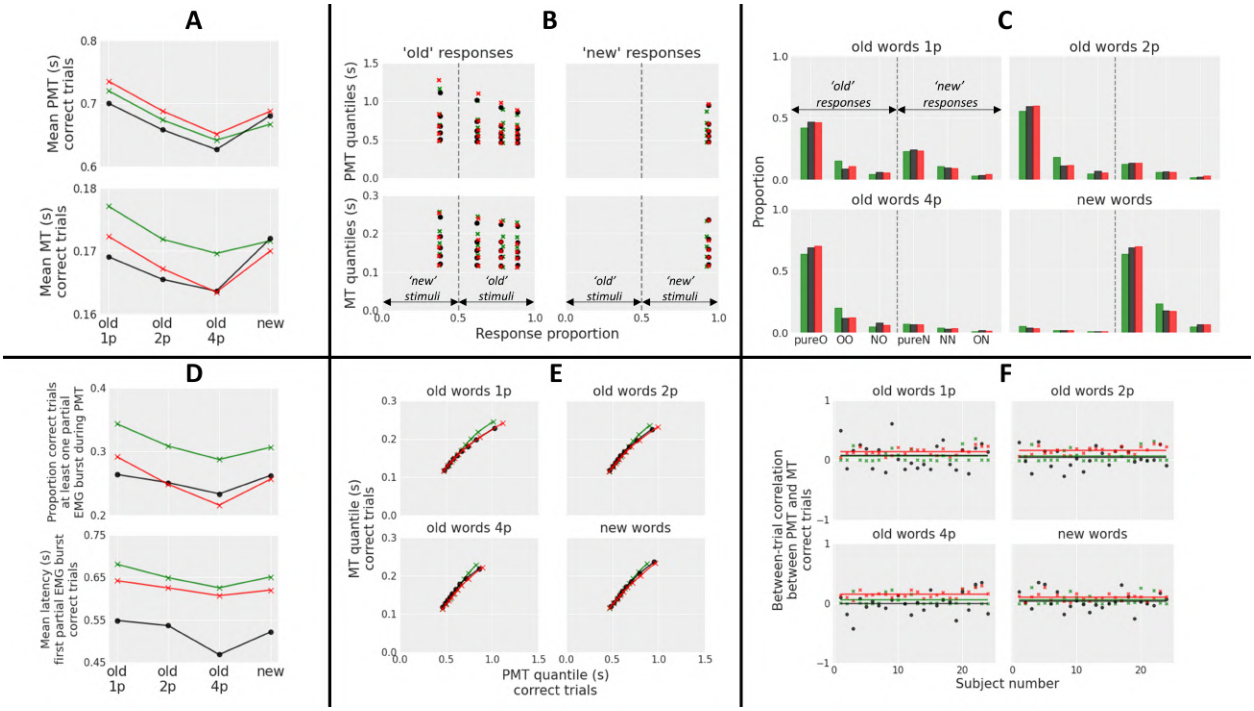
Best-fitting parameters for the full models are shown in Table 1 (main parameters) and Table 2 (between-trial variability parameters). Best-fitting parameters for the raw models are shown in Supplementary Table 1. DSDTDM captures the data with a moderate amount of leakage (λ), though the best-fitting value for the full model is a bit larger compared to Experiment 2 (implying reduced filtering and filter-related delays; for an illustration of model trajectories, see Figure 4C). The amount of between-trial variability in the best-fitting full model components was generally higher for DTDM compared to DSDTDM. Note that between-trial variability in drift rate (sv) was larger for old than new words, consistent with previous work. It also decreased as the number of word presentations increases, suggesting that evidence variability decreases as function of learning.

Figure 7 displays the goodness-of-fit of the full models to data. The full DTDM provides a poor account of MT distributions, due to the large contribution of residual motor latencies to predicted MTs. In addition, the full DTDM systematically overestimates the proportion of correct OO and NN trials and the mean latency of the first partial EMG burst. The full DSDTDM provides a good fit to data factored out in parameter estimation (though it slightly overestimates the .9 quantile of PMT distributions for old responses as the number of word presentations decreases), but shows a relatively poor generalization performance. Although the model predicts the effect of word repetitions on the mean latency

of the first partial EMG burst, it systematically overestimates this latency by about 100 ms. In addition, the model overestimates the between-trial correlation between PMT and MT, especially for old words presented two and four times during study.

Figure 7

Data from a recognition memory task (black) and predictions from the full DSDTDM (red) and the full DTDM (green)



Note. The structure of each panel is similar to that of Figure 5. Model predictions are computed from best-fitting parameters, using 100,000 simulated trials per condition. Quantile probability functions, shown in panel B, incorporate distributions of errors if each subject made at least 10 errors in corresponding conditions. Panel C shows the proportion of each of the six trial types (pureO, OO, NO, pureN, NN, ON) for each condition averaged across subjects.

Experiment 4: Lexical knowledge

The ability to recognize words is essential for reading, and the lexical decision task has been widely used to study this process. In this task, subjects have to decide whether strings of letters are words or non-words. A standard finding is that high-frequency words are recognized faster and more accurately compared to low-frequency words. Ratcliff et al. (2004) showed that the diffusion model provides a good account of performance in this task with a decrease of drift rate as word frequency decreases. Consequently, both DTDM and DSDTDM predict an increase in mean MT as word frequency decreases.

Later modeling work suggests that word frequency modulates other parameters of the diffusion model. Both Donkin, Brown, et al. (2009) and Gomez and Perea (2014) showed a variation of mean nondecision time across word frequency levels, suggesting that frequency modulates lexical access processes (that determine how much evidence the stimulus provides for each response alternative). Tillman et al. (2017) recently showed evidence for a larger between-trial variability in drift rate for words than non-words, and for high frequency compared to low frequency words, a pattern predicted by a model of lexical retrieval (Wagenmakers et al., 2004). Consequently, the mean residual latency added to predicted PMT (Te), drift rate (v), and between-trial variability in drift rate (sv) parameters were free to vary across word frequency conditions in our modeling of the data.

Method

Participants

Twenty-four students (four men; mean age: 21.00) from the University of Franche-Comté took part in the experiment in exchange for course credits or as volunteers. All participants met the same criteria as for Experiments 2 and 3. Subjects were not aware of the purpose of the experiment and provided written consent to participate. This study was approved by the ethical committee for research of the University (agreement n°CERUBFC-2022-01-18-002).

Appartus

The apparatus was identical to Experiments 2 and 3.

Stimuli

Four lists of french words (five to 12 letters) of differing frequencies (high frequency: range 10-2736 occurences per million; medium frequency: range 2-5; low frequency: range 0.5-1; very low frequency: range 0.01-0.1) were created using the Lexique database (New et al., 2004). To obtain the stimuli used in the present experiment, the lists underwent the following steps. First, all plural and feminine agreements were removed. Second, low and very-low frequency words were screened by two students, and any words they did not know were eliminated. Third, a pool of 408 words was pseudo-randomly selected from each list, with the constraint of obtaining an homogeneous number of letters across pools. For each pool, 204 words were randomly selected to become pseudowords (i.e., pronounceable non-words), created with the mulitlingual pseudoword generator Wuggy (Keuleers and Brysbaert, 2010). Pseudowords were screened by two other students to ensure that they were pronounceable and did not correspond to an existing word. This procedure resulted in four lists, each list comprising 204 words (four for practice and 200 for the experiment) and 204 pseudowords (four for practice and 200 for the experiment). The four students recruited for screening stimuli did not participate in the experiment. Statistics relative to frequency and number of letters for the final sample of words used in the experiment are provided in supplementary Table 2). Words were presented in black against a grey background (font: Consolas), at the center of the screen. The height of the letters was 0.76°.

Procedure

Participants were instructed to press the left or the right button with their left or right thumb depending on whether the stimulus was a french word or not (stimulus-response mapping balanced across participants). Participants first performed a pratice block of 32

trials (four words in each frequency level and 16 pseudowords) during which a feedback on accuracy after each trial was provided, and worked through 10 blocks of 160 trials (20 words in each frequency level and 80 pseudowords) with no feedback on accuracy. Blocks were separated by self-paced breaks. Each stimulus appeared once in the experiment. Participants were thus exposed to the same word and pseudoword stimuli. Stimuli were randomly assigned to blocks, and presented in a random order within blocks. Each trial started with the presentation of the stimulus until the participant responded, or until a 4 s RT deadline. If participants failed to respond by then, the message "Too late! Please respond faster." was displayed for 1.5 s. The intertrial interval was 1.5 s. Overall, the experiment lasted about an hour.

EMG recording and signal processing

The procedure used for EMG recordings and signal processing was similar to Experiments 2 and 3. Trials with high tonic muscular activity were discarded from analyses (10.4% of trials on average; range 0.06-27.3%).

DTDM and DSDTDM fit procedure

Drift rate (v), between-trial variability in drift rate (sv), and mean residual time added to predicted PMT (Te) were the only parameters free to vary between conditions. We treated the starting point z of the evidence accumulation process as a free parameter, resulting in 14 (22) parameters for the raw (full) DTDM, and 15 (24) parameters for the raw (full) DSDTDM. The fit procedure was identical to that used in the previous experiments. We modeled 'pseudoword' and 'word' responses, corresponding to lower and upper response bounds respectively. The six trial types considered in the fit procedure were *pureW* ('word' response, no partial EMG burst during PMT), WW ('word' response, at least one partial EMG burst during PMT, first partial burst located in the 'word' EMG channel), PW ('word' response, at least one at least one partial EMG burst during PMT, first partial burst located in the 'pseudoword' EMG channel), and so forth for 'pseudoword' responses (*pureP*, PP,

WP).

Results

Behavior and EMG

Anticipations (RTs < 150 ms; 0%) and trials in which participants failed to respond before the 4 s deadline (0.1%) were discarded from analyses. Performance to word stimuli was analyzed by means of linear contrasts (two-sided) with word frequency as within-subjects factor and specific error terms. Accuracy decreased ($t(23)$ = 12.54, p < .001) and mean RT increased ($t(23)$ = -8.36, p < .001) as word frequency decreased, reflecting the classic word frequency effect. Although mean PMT decreased as word frequency decreased ($t(23)$ = -8.70, p < .001), mean MT exhibited an unexpected inverted-U shape function of word frequency (Figure 8A). Accordingly, the planned linear contrast was not significant ($t(23)$ = 0.53, p = .60), while a post-hoc quadratic contrast reached significance ($t(23)$ = -2.24, p = .04). This inverted U-shape pattern is unlikely due to a statistical power or an EMG signal quality issue, because (i) the sample size was identical to Experiment 3, (ii) the amplitude of the word frequency effect on mean RT (M = 187 ms) was larger than the amplitude of the word presentation effect on mean RT (M = 79 ms), and (iii) EMG signal quality was approximately similar between Experiments 3 and 4, as revealed by a comparable percentage of rejected trials on average.

Both the proportion of correct trials containing at least one partial EMG burst during PMT and the mean latency of the first partial burst decreased as word frequency increased (Figure 8D, $t(23)$ = -6.13, p < .001 and $t(23)$ = -6.39, p < .001 respectively). For each condition, PMT quantile-MT quantile plots from correct trials exhibited an approximately linear shape (Figure 8E), and the between-trial Pearson correlation coefficient between PMT and MT was remarkably close to zero on average, with no apparent trend across conditions (Figure 8F).

To compare performance between word and pseudoword stimuli, we averaged the

performance to word stimuli across frequency levels and ran two-sided paired sample t-tests. The proportion of correct responses was higher for pseudowords than words ($t(23)$ = -4.24, $p < .001$). Mean RT, mean PMT, and mean MT were slower for pseudowords than words, $t(23)$ = -4.40, $p < .001$, $t(23)$ = -3.35, $p = .003$, and $t(23)$ = -2.49, $p = .02$ respectively). Finally, there was a trend for a smaller proportion of correct trials containing at least one partial EMG burst for words than pseudowords ($t(23)$ = -2.05, $p = .052$), and the mean latency of the first partial burst was faster for words ($t(23)$ = -4.21, $p < .001$).

Model fits

DSDTDM was associated with lower AIC and BIC statistics compared to DTDM for each of the 24 subjects, and for both raw and full model variants (Figure 3D). The difference in AIC and BIC between raw and full model variants was smaller for DSDTDM compared to DTDM, and the raw DSDTDM was associated with a lower AIC (BIC) compared to the full DTDM for 15 (14) subjects. The pattern of model selection results is thus similar to that observed in the previous experiments, and provides strong evidence for DSDTDM.

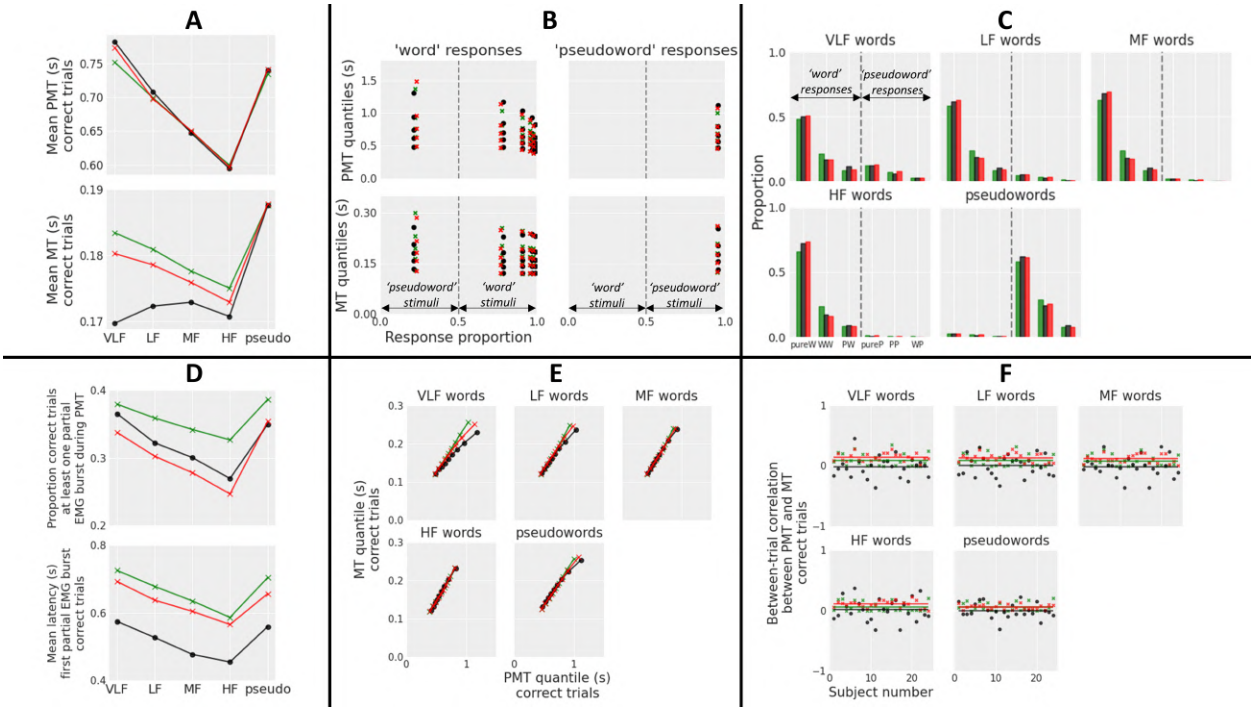
Best-fitting parameters for the full models are shown in Table 1 (main parameters) and Table 2 (between-trial variability parameters). Best-fitting parameters for the raw models are shown in Supplementary Table 1. DSDTDM captures the data with a higher level of leakage λ compared to the previous experiments (implying reduced filtering and filter-related delays; for an illustration of model trajectories, see Figure 4D). Consistent with previous work (Donkin, Brown, et al., 2009; Gomez and Perea, 2014), DSDTDM and DTDM both predict an increase in the mean residual latency parameter Te added to predicted PMT as word frequency decreases. Although evidence variability (parameter sv) was generally larger for words than pseudowords, consistent with previous work (Tillman et al., 2017; Wagenmakers et al., 2004), it increased as word frequency decreased. The latter pattern is opposite to that found by Tillman et al. (2017) using traditional diffusion model fits. The number of words for which people do not know the definition may increase as word frequency decreases,

inflating evidence variability. More generally, the amount of between-trial variability in the best-fitting full model components was generally higher for DTDM compared to DSDTDM, consistent with model fits obtained in the previous experiments.

Figure 8 displays the goodness-of-fit of the full models to data. As expected, both DSDTDM and DTDM predict an increase in predicted mean MT as word frequency decreases, and fail to capture the observed inverted U-shaped pattern. DTDM systematically underestimates the .9 quantile of PMT distributions for correct responses, and overestimates the proportion of correct WW and PP trials. DSDTDM provides a better account of PMT distributions and the six trial types (pureW, WW, PW, pureP, PP, WP). Both models overestimate the mean latency of the first partial EMG burst, and the between-trial correlation between PMT and MT.

Figure 8

Data from a lexical decision task (black) and predictions from the full DSDTDM (red) and the full DTDM (green)



Note. The structure of each panel is similar to that of Figure 5. Model predictions are computed from best-fitting parameters, using 100,000 simulated trials per condition. Quantile probability functions, shown in panel B, incorporate distributions of errors if each subject made at least 10 errors in corresponding conditions. Panel C shows the proportion of each of the six trial types (pureW ,WW, PW, pureP, PP, WP) for each condition averaged across subjects.

General discussion

To our knowledge, this work represents the first attempt to jointly model decision-making, motor preparation, and motor execution processes in choice RT tasks. DTDM, proposed by Servant et al. (2015, 2021), offers a theoretical account of decision-making and motor execution, but neglects processing properties of motor preparation. The leaky integrated threshold model, proposed by Verdonck et al. (2020), offers a theoretical account of decision-making and motor preparation, but neglects properties of motor execution. In the present work, we sought to combine DTDM and the leaky integrated threshold model to provide a complete account of the interplay between decision and motor processes. The proposed DSDTDM assumes that evidence from our senses and memory accumulates through a decision-making stage. The purpose of this stage is to perform a decision about alternative categories of a stimulus (e.g., decide whether a traffic light is red or green). When decisions are mapped onto actions (e.g., press the brake pedal if the traffic light is red), accumulated evidence from the decision stage is continuously conveyed to motor preparation brain areas, where it is filtered out through a second accumulation processing stage. The resulting motor preparation variable is then transmitted to the response-relevant muscles when it exceeds a threshold level of activation, corresponding to the beginning of motor execution. The transmission continues until a threshold amount of force has been produced by the muscles to issue the response.

We tested DSDTDM against behavioral and EMG data from four choice tasks that span a variety of domains in cognitive sciences, namely motion perception (Experiment 1), numerical cognition (Experiment 2), recognition memory (Experiment 3), and lexical knowledge (Experiment 4). Each task featured a manipulation of choice difficulty to bring additional constraints on the model. DSDTDM was evaluated in its ability to capture (i) the shape of PMT and MT distributions for correct and incorrect responses, (ii) the proportion of six trial types defined by the combination of response type (correct vs. incorrect in Experiment 1, less than vs. greater than in Experiment 2, old vs. new in Experiment 3, word

vs. pseudoword in Experiment 4), presence vs. absence of at least one partial EMG burst during PMT, and EMG channel location of the first partial burst, (iii) the mean latency of the first partial EMG burst in correct trials, (iv) the relationship between the shape of PMT and MT distributions in correct trials, (v) the between-trial Pearson correlation coefficient between PMT and MT in correct trials, and (vi) the variation of all of the above aspects of the data across difficulty conditions. Overall, DSDTDM provided a good fit to data factored out in parameter estimation (i, ii, iv, and v). The only apparent discrepancy between data and model predictions was an overestimation of the .9 quantile of PMT distributions for the most difficult experimental conditions. One way to solve this issue would be to incorporate an urgency signal to the model (Cisek et al., 2009; Ditterich, 2006; Evans, Hawkins, et al., 2020; Hawkins et al., 2015; Trueblood et al., 2021). Urgency can take the form of temporally collapsing boundaries, or a time-increasing gain applied to the incoming evidence. Both mechanisms reduce the skew of predicted RT distributions (Hawkins et al., 2015), offering a potential solution to the observed DSDTDM misfit. Although urgency signals remain controversial when considering behavioral data alone (Evans, Hawkins, et al., 2020; Hawkins et al., 2015; Ratcliff et al., 2016; Trueblood et al., 2021), neurophysiological studies have provided evidence for them at the motor preparation level in both monkeys and humans (Churchland et al., 2008; Hanks et al., 2014; Murphy et al., 2016; Steinemann et al., 2018), even when subjects are not under speed pressure (Kelly et al., 2021). Interestingly, urgency signals are not observed at the decision-making level (Kelly et al., 2021; Steinemann et al., 2018), further emphasizing the functional dissociation between decision-making and motor preparation. Fitting DSDTDM variants that incorporate urgency mechanisms is beyond the scope of the present work, and should be conducted in tandem with an electrophysiological investigation of motor preparation.

Although DSDTDM showed a good fit to data factored out in parameter estimation, it provided a mixed predictive account of the remaining data. The model captured the mean latency of the first partial EMG burst in Experiments 1 and 2, but systematically

overestimated this latency by about 100 ms in Experiments 3 and 4. We cannot exclude the possibility that a set of parameters could have better captured the partial burst latency data, had we considered these data in parameter estimation. Alternatively, variations in EMG signal quality across experiments may have contributed to this pattern of results, as the percentage of rejected trials (i.e., trials in which EMG onsets could not be detected due to high tonic muscular activity) was larger on average in Experiments 3 (12.41%) and 4 (10.4%) compared to Experiments 1 (7.5%) and 2 (2.7%). Therefore, the data from Experiments 3 and 4 might incorporate a larger amount of artifactual partial EMG bursts, caused by residual tension in response-relevant muscles.

Besides the mean latency of partial EMG bursts, DSDTDM provided a mixed predictive account of the between-trial correlation between PMT and MT. In general, model predictions showed more dispersion compared to observed data at the individual level, but this phenomenon is likely explained by noise in EMG onset detection. However, the model systematically overestimated the correlation averaged across subjects in Experiments 3 and 4, and in the easiest condition of Experiment 1. One may argue that this discrepancy between data and model predictions speaks against the model architecture, as the filtering mechanism at the motor preparation level flattens out random fluctuations of the evidence accumulation signal, and increases the predicted correlation between PMT and MT at the single-trial level. Once again, we cannot exclude the possibility that a set of parameters could have better captured the between-trial correlation between PMT and MT, had we considered these data in parameter estimation. Alternatively, it is important to remember that the filtering mechanism (regulated by the leak parameter) interacts in complex ways with drift rate and between-trial variability in drift rate (see supplementary Figures 2 and 3). This interaction is problematic because the model does not incorporate representational assumptions that specify how drift rate distributions arise from the stimuli (the same criticism applies to DTDM and the standard diffusion model). Consequently, the discrepancy between data and model predictions may stem from a misspecification of drift rate distributions.

To shed light on this problem, we incorporated representational assumptions in our modeling of the numerosity judgment data from Experiment 2. Specifically, we incorporated a front-end to the model developped by Ratcliff and McKoon (2018) that specifies how drift rate distributions arise from an approximate number representation in which numerosities are represented as Gaussian distributions, with the mean and standard deviation of these distributions increasing linearly with numerosity. This tightly constrained DSDTDM provided a good predictive account of the between-trial correlation between PMT and MT across numerosity conditions, suggesting that discrepancies between data and model predictions in the other experiments may stem from a misspecification of drift rate distributions. More generally, this finding highlights the need of considering pre-decisional processing stages when modeling post-decisional motor phenomena. This need is further highlighted by the lexical decision data from Experiment 4. The word frequency effect has been successfully modeled by assuming that word frequency modulates drift rate (Ratcliff et al., 2004), mean nondecision time (Donkin, Brown, et al., 2009; Gomez and Perea, 2014), and between-trial variability in drift rate (Tillman et al., 2017). Within the framework of DSDTDM, the decrease of drift rate as word frequency decreases should increase the predicted mean MT. Contrary to this prediction, we found an inverted U-shape relationship between mean MT and word frequency. Specifically, mean MT showed an initial increase from high frequency to medium frequency words, followed by a decrease for low and very low frequency words. At first glance, this result speaks against the architecture of DSDTDM. However, the model does not incorporate assumptions regarding how drift rate is computed in this task, so the origin of the problem is unclear. It would be useful to connect models of lexical access (e.g., Grainger, 2018; Houghton, 2018; McClelland and Rumelhart, 1981) to DSDTDM to better understand the origin of the problem.

Apart from the unexpected word frequency effect on mean MT, EMG findings were remarkably consistent across experiments, suggesting that DSDTM generalizes across cognitive domains. Both mean PMT and mean MT decreased as choice difficulty increased. In

addition, partial bursts were observed in the EMG data of each subject of each experiment. The proportion of correct trials containing at least one partial EMG burst during PMT and the mean latency of the first partial burst increased as choice difficulty increased. Interestingly, the proportion of correct trials in which the first partial EMG burst was located in the same EMG channel as the response was systematically larger than the proportion of correct trials in which the first partial EMG burst was located in the opposite EMG channel. Within the framework of DSDTDM, this finding is explained by the same mechanism that captures the relative proportion of correct and incorrect responses. Putting aside betwen-trial variability in model components, errors are produced by noise in the evidence accumulated at each time step. Although part of this noise is filtered out during motor preparation, the predicted proportion of errors is smaller than the proportion of correct responses, so long as the drift rate is not null.

Comparisons with DTDM

As predicted, DSDTDM captured the data with a relatively low level of leakage, indicating that the evidence accumulation variable is filtered at the motor preparation level. This finding suggests that the motor preparation processing stage adds to a single-stage DTDM description of the data. Model selection statistics (AIC and BIC) further showed that the additional complexity of DSDTDM was justified in light of the (large) improvement in fit quality. The AIC statistic systematically favored DSDTDM over DTDM for each of the 90 subjects that participated in the experiments, and for both raw and full model variants. The more conservative BIC statistic systematically favored DSDTDM over DTDM for 90/90 participants (raw models) and 89/90 (full models). These findings provide decisive evidence for DSDTM. Interestingly, the difference in model selection statistics between raw and full models was much larger for DTDM than for DSDTDM, suggesting that between-trial variability in DTDM components have a major impact on the fit quality of the model, contrary to DSDTM. In fact, the AIC (BIC) statistic favored the raw DSDTM over the full

DTDM for 70(66)/90 subjects. Although between-trial variability in processing components is plausible, we believe that a large contribution of between-trial variability to the fit quality of a model is problematic, as there is generally no explanation of why this variability occurs or why it has the parametric form researchers assume to represent it (see the modeling of Experiment 2 for an exception regarding between-trial variability in drift rate that arises from theoretical properties of the approximate number system). In this view, between-trial variability essentially corresponds to adding a random component to the model without any strong theoretical motivation for it rather than to improve the fit quality (Evans, Tillman, et al., 2020). Consequently, we consider our findings regarding between-trial variability as additional evidence for DSDTDM.

In its raw form, DTDM grossly overestimates the proportion of trials containing at least one partial EMG burst, especially when the first partial burst is located in the same EMG channel as the response. The full DTDM provides a better account of these proportions by placing EMG bounds very close to response bounds (Experiment 1), or by combining high drift rates with a high between-trial variability in drift rates (Experiments 2-4; see tables 1 and 2). Both processing schemes result in predicted MTs that are too fast compared to observed MTs. The model compensates this problem by increasing the contribution of residual motor latencies to predicted MTs (parameters Tr and sTr), but this compensation has three negative consequences. First, the model cannot capture large effects of choice difficulty on mean MT, such as those observed in Experiment 1. Second, the model cannot capture the right-skewed shape of MT distributions, because between-trial variability in residual motor latencies added to predicted MTs is uniformly distributed, an (arbitrary) assumption inherited from the diffusion model (Ratcliff and Rouder, 1998). Third, residual latencies added to predicted MTs are assumed to reflect the electromechanical delay (time lag between muscle excitation and force generation; see general introduction). This delay has been proposed to be in the range 30-100 ms (Cavanagh and Komi, 1979), but best-fitting full DTDM parameters show longer predicted delays (parameter Tr ranges 133-163 ms on

average across experiments). Although delays predicted by the full DSDTDM are shorter (90-109 ms), they are still located in the vicinity of the upper limit. Empirical estimates of the electromechanical delay must be taken with caution though, because there is substantial variability across studies. In fact, empirical estimates are strongly influenced by characteristics of the apparatus used to collect the data, the muscle under investigation, the task being performed, and signal processing techniques (Corcos et al., 1992; Yavuz et al., 2010). Consequently, we refrain from drawing strong conclusions based on the electromechanical delay.

Neurophysiological implementation, theoretical limitations, and possible model extensions

As detailed in the general introduction, properties of motor preparation and execution, uncovered by neurophysiological studies, are not accounted for by current single-stage evidence accumulation models such as the diffusion model (Ratcliff, 1978; Ratcliff et al., 2016), the leaky competing accumulator (Usher and McClelland, 2001), the linear ballistic accumulator (Brown and Heathcote, 2008), racing diffusion models (Ratcliff et al., 2003; Tillman et al., 2020), and Poisson counter models (Ratcliff and Smith, 2004; Vickers, 1970). In this respect, we believe that the models proposed by Servant et al. (2015, 2021), Verdonck et al. (2020), and their integration through DSDTDM constitute a major theoretical advance in the field, as they offer a mecanistic explanation to the interplay between decision and motor processes.

EEG studies have greatly contributed to our current understanding of decision-making and motor preparation in humans, as the good temporal resolution of EEG allows researchers to track evidence accumulation mechanisms at the systems level (O'Connell and Kelly, 2021). Consequently, a natural development of the present line of work would be to combine EMG and EEG recordings to supplement the assessment of hypothetical computations at decision-making and motor preparation levels. The supramodal decision-making

process of DSDTDM could be identified with the CPP, and the motor preparation process with effector-selective motor preparation activities. DSDTDM could be fit to behavioral and EMG data, and model trajectories computed from best-fitting parameters could be compared to CPP and motor preparation EEG activities in a variety of choice tasks and experimental conditions. Alternatively, EEG signals could be incorporated into the fit procedure, and some parameters could be constrained to match corresponding electrical signatures. Both methodologies would bring additional contraints to DSDTDM, and foster theoretical developments. For instance, separate accumulators could be used for each response alternative, and researchers could test for urgency signals, feedforward inhibition between inputs, independent race and lateral inhibition between accumulators, at both decision and motor preparation levels (Bogacz et al., 2006; Purcell et al., 2010; Servant et al., 2019). Kelly et al. (2021) have recently used EEG data to inform the construction and evaluation of evidence accumulation models in the context of prior-informed decisions. Although this work represents a major advance in the field, it only considered a single-stage evidence accumulation model, which does not capture anatomical, functional and temporal dissociations between decision-making and motor preparation. Additional constraints to DSDTDM could also arise from a more detailed analysis of partial EMG bursts at the motor execution level. Some trials contain more than one partial EMG burst during PMT, and these additional bursts could be considered in the modeling. In particular, the co-ocurrence of two partial bursts in left and right EMG channels would suggest some degree of independence between accumulators. We note, however, that more detailed EMG analyses entail an increased sensitivity to artifactual electrical activities.

A potential criticism of DSDTDM concerns the strategic adjustment of the leak parameter λ, as hypothesized by Verdonck et al. (2020) and suggested by the substantial variation of λ between experiments in the present work. One may argue that leakage is a structural property of neurons, and thus cannot be under strategic control. As reviewed by Usher and McClelland (2001), excitatory currents to a neuron decay with a time constant

of 5-10 ms. This time constant is too short to support evidence accumulation, and is likely counteracted by a recurrent excitation mechanism. Therefore, best-fitting leakage values in single-stage evidence accumulation models result from the combined influence of leakage and recurrent excitation, and the later could very well be under strategic control. In addition, it should be noted that the leak parameter in DSDTDM is quite different that the leak parameter in single-stage evidence accumulation models, as it regulates the amount of past states of the decision-making process that contribute to the motor preparation process. This regulation could be under strategic control.

Beyond basic mechanisms that drive the time-course of decision-making, motor preparation, and motor execution processes, we believe that future developments of DSDTDM would benefit from model-based investigations of more complex relationships between decision and motor processes. For example, decisions are often taken well before being expressed behaviorally. This scenario is involved when voters have to choose a candidate. It is currently outside the scope of DSDTDM, as the model does not specify the relationship between memory and decision/motor processes. The choice might be categorically retrieved from memory and transmitted to the motor system. Consequently, effector-selective motor preparation EEG activities and EMG signals should not be modulated by the quality of evidence. However, this hypothetical processing sheme may vary as a function of the temporal delay between decision and motor processes, and foreknowledge of the stimulus-response mapping (Twomey et al., 2016).

Another scenario that deserves additional scutiny concerns continuous movement reports. Similar to EMG findings, reaching trajectories are modulated by perceptual and cognitive factors (e.g., Buc Calderon et al., 2015; Kinder et al., 2022; Song and Nakayama, 2009; Sullivan et al., 2015). However, the application of DSDTDM to choice reaching tasks is not straightforward. Reaching movements engage a complex pattern of neuromuscular activity, making EMG recordings and analyses challenging. One way to reduce this complexity is to model reaching movements at the level of kinematic motor primitives, hypothetical build-

ing blocks that can be combined to construct motion (for reviews, see Flash and Hochner, 2005; Giszter, 2015; Latash, 2020). Despite their apparent continuity, reaching movements appear to be composed of discrete submovements. Friedman et al. (2013) hypothesized that an intermittent motor control process probes the state of accumulated evidence at discrete time points to determine submovements, and showed good fits of this model to arm movement trajectories in a variant of the random dot motion task. The relationship between this intermittent motor control process, motor preparation, and EMG activity remains to be elucidated.

To conclude, the present EMG investigations in choice RT tasks add to a growing body of behavioral and neurophysiological evidence that suggests that the motor system can have systematic effects that are computationally related to central decision processes. These effects are important to complete the story of how our choices are reflected in our actions. The proposed dual-stage dual-threshold evidence accumulation theory offers a new framework to understand this relationship.

References

Boehm, U., Annis, J., Frank, M. J., Hawkins, G. E., Heathcote, A., Kellen, D., Krypotos, A.-M., Lerche, V., Logan, G. D., Palmeri, T. J., van Ravenzwaaij, D., Servant, M., Singmann, H., Starns, J. J., Voss, A., Wiecki, T. V., Matzke, D., & Wagenmakers, E.-J. (2018). Estimating across-trial variability parameters of the Diffusion Decision Model: Expert advice and recommendations. *Journal of Mathematical Psychology*, *87*, 46–75. https://doi.org/10.1016/j.jmp.2018.09.004

Bogacz, R., Brown, E., Moehlis, J., Holmes, P., & Cohen, J. D. (2006). The physics of optimal decision making: A formal analysis of models of performance in two-alternative forced-choice tasks. *Psychological Review*, *113*(4), 700–765. https://doi.org/10.1037/0033-295X.113.4.700

Bogacz, R., Wagenmakers, E.-J., Forstmann, B. U., & Nieuwenhuis, S. (2010). The neural basis of the speed-accuracy tradeoff. *Trends in Neurosciences*, *33*(1), 10–16. https://doi.org/10.1016/j.tins.2009.09.002

Boik, R. J. (1981). A priori tests in repeated measures designs: Effects of nonsphericity [Place: Germany Publisher: Springer]. *Psychometrika*, *46*(3), 241–255. https://doi.org/10.1007/BF02293733

Bortoff, G. A., & Strick, P. L. (1993). Corticospinal terminations in two new-world primates: Further evidence that corticomotoneuronal connections provide part of the neural substrate for manual dexterity. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *13*(12), 5105–5118.

Botwinick, J., & Thompson, L. W. (1966). Premotor and motor components of reaction time. *Journal of Experimental Psychology*, *71*(1), 9–15. https://doi.org/10.1037/h0022634

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

Buc Calderon, C., Verguts, T., & Gevers, W. (2015). Losing the boundary: Cognition biases action well after action selection [Place: US Publisher: American Psychological Association]. *Journal of Experimental Psychology: General*, *144*, 737–743. https://doi.org/10.1037/xge0000087

Cavanagh, P. R., & Komi, P. V. (1979). Electromechanical delay in human skeletal muscle under concentric and eccentric contractions. *European Journal of Applied Physiology and Occupational Physiology*, *42*(3), 159–163. https://doi.org/10.1007/BF00431022

Churchland, A. K., Kiani, R., & Shadlen, M. N. (2008). Decision-making with multiple alternatives. *Nature neuroscience*, *11*(6), 693–702. https://doi.org/10.1038/nn.2123

Cisek, P., & Kalaska, J. F. (2010). Neural mechanisms for interacting with a world full of action choices. *Annual Review of Neuroscience*, *33*, 269–298. https://doi.org/10.1146/annurev.neuro.051508.135409

Cisek, P., Puskas, G. A., & El-Murr, S. (2009). Decisions in changing conditions: The urgency-gating model. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *29*(37), 11560–11571. https://doi.org/10.1523/JNEUROSCI.1844-09.2009

Corcos, D. M., Gottlieb, G. L., Latash, M. L., Almeida, G. L., & Agarwal, G. C. (1992). Electromechanical delay: An experimental artifact. *Journal of Electromyography and Kinesiology: Official Journal of the International Society of Electrophysiological Kinesiology*, *2*(2), 59–68. https://doi.org/10.1016/1050-6411(92)90017-D

Dehaene, S. (2003). The neural basis of the Weber-Fechner law: A logarithmic mental number line. *Trends in Cognitive Sciences*, *7*(4), 145–147. https://doi.org/10.1016/s1364-6613(03)00055-x

de Jong, R., Wierda, M., Mulder, G., & Mulder, L. J. (1988). Use of partial stimulus information in response processing [Place: US Publisher: American Psychological Association]. *Journal of Experimental Psychology: Human Perception and Performance*, *14*(4), 682–692. https://doi.org/10.1037/0096-1523.14.4.682

de Lange, F. P., Rahnev, D. A., Donner, T. H., & Lau, H. (2013). Prestimulus oscillatory activity over motor cortex reflects perceptual expectations. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *33*(4), 1400–1410. https://doi.org/10.1523/JNEUROSCI.1094-12.2013

Ditterich, J. (2006). Stochastic models of decisions about motion direction: Behavior and physiology. *Neural Networks: The Official Journal of the International Neural Network Society*, *19*(8), 981–1012. https://doi.org/10.1016/j.neunet.2006.05.042

Donders, F. C. (1969). On the speed of mental processes. *Acta Psychologica*, *30*, 412–431. https://doi.org/10.1016/0001-6918(69)90065-1

Donkin, C., Brown, S., Heathcote, A., & Andrews, S. (2009). Non-Decision Time Effects in the Lexical Decision Task. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *31*. Retrieved August 4, 2022, from https://escholarship.org/uc/item/07q9n3tq

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

Donner, T. H., Siegel, M., Fries, P., & Engel, A. K. (2009). Buildup of choice-predictive activity in human motor cortex during perceptual decision making. *Current biology: CB*, *19*(18), 1581–1585. https://doi.org/10.1016/j.cub.2009.07.066

Ebbesen, C. L., & Brecht, M. (2017). Motor cortex — to act or not to act? [Number: 11 Publisher: Nature Publishing Group]. *Nature Reviews Neuroscience*, *18*(11), 694–705. https://doi.org/10.1038/nrn.2017.119

Evans, N. J. (2019). A method, framework, and tutorial for efficiently simulating models of decision-making. *Behavior Research Methods*, *51*(5), 2390–2404. https://doi.org/10.3758/s13428-019-01219-z

Evans, N. J., Hawkins, G. E., & Brown, S. D. (2020). The role of passing time in decision-making [Place: US Publisher: American Psychological Association]. *Journal of Ex-*

perimental Psychology: Learning, Memory, and Cognition, *46*, 316–326. https://doi.org/10.1037/xlm0000725

Evans, N. J., Tillman, G., & Wagenmakers, E.-J. (2020). Systematic and random sources of variability in perceptual decision-making: Comment on Ratcliff, Voskuilen, and McKoon (2018). *Psychological Review*, *127*(5), 932–944. https://doi.org/10.1037/rev0000192

Evans, N. J., & Wagenmakers, E.-J. (2020). Evidence accumulation models: Current limitations and future directions. *The Quantitative Methods for Psychology*, *16*, 73–90.

Filimon, F., Philiastides, M. G., Nelson, J. D., Kloosterman, N. A., & Heekeren, H. R. (2013). How Embodied Is Perceptual Decision Making? Evidence for Separate Processing of Perceptual and Motor Decisions. *The Journal of Neuroscience*, *33*(5), 2121–2136. https://doi.org/10.1523/JNEUROSCI.2334-12.2013

Flash, T., & Hochner, B. (2005). Motor primitives in vertebrates and invertebrates. *Current Opinion in Neurobiology*, *15*(6), 660–666. https://doi.org/10.1016/j.conb.2005.10.011

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. https://doi.org/10.1146/annurev-psych-122414-033645

Freedman, D. J., & Assad, J. A. (2016). Neuronal Mechanisms of Visual Categorization: An Abstract View on Decision Making. *Annual Review of Neuroscience*, *39*, 129–147. https://doi.org/10.1146/annurev-neuro-071714-033919

Friedman, J., Brown, S., & Finkbeiner, M. (2013). Linking cognitive and reaching trajectories via intermittent movement control. *Journal of Mathematical Psychology*, *57*(3), 140–151. https://doi.org/10.1016/j.jmp.2013.06.005

Giszter, S. F. (2015). Motor primitives–new data and future questions. *Current Opinion in Neurobiology*, *33*, 156–165. https://doi.org/10.1016/j.conb.2015.04.004

Gold, J. I., & Shadlen, M. N. (2000). Representation of a perceptual decision in developing oculomotor commands. *Nature*, *404*(6776), 390–394. https://doi.org/10.1038/35006062

Gold, J. I., & Shadlen, M. N. (2007). The neural basis of decision making. *Annual Review of Neuroscience*, *30*, 535–574. https://doi.org/10.1146/annurev.neuro.29.051605.113038

Gold, J. I., & Shadlen, M. N. (2003). The Influence of Behavioral Context on the Representation of a Perceptual Decision in Developing Oculomotor Commands [Publisher: Society for Neuroscience Section: ARTICLE]. *Journal of Neuroscience*, *23*(2), 632–651. https://doi.org/10.1523/JNEUROSCI.23-02-00632.2003

Gomez, P., & Perea, M. (2014). Decomposing encoding and decisional components in visual-word recognition: A diffusion model analysis. *Quarterly Journal of Experimental Psychology (2006)*, *67*(12), 2455–2466. https://doi.org/10.1080/17470218.2014.937447

Grainger, J. (2018). Orthographic processing: A 'mid-level' vision of reading: The 44th Sir Frederic Bartlett Lecture. *Quarterly Journal of Experimental Psychology (2006)*, *71*(2), 335–359. https://doi.org/10.1080/17470218.2017.1314515

Gratton, G., Coles, M. G., Sirevaag, E. J., Eriksen, C. W., & Donchin, E. (1988). Pre- and poststimulus activation of response channels: A psychophysiological analysis. *Journal of Experimental Psychology. Human Perception and Performance*, *14*(3), 331–344. https://doi.org/10.1037//0096-1523.14.3.331

Hanks, T., Kiani, R., & Shadlen, M. N. (2014). A neural mechanism of speed-accuracy tradeoff in macaque area LIP. *eLife*, *3*, e02260. https://doi.org/10.7554/eLife.02260

Hawkins, G. E., Forstmann, B. U., Wagenmakers, E.-J., Ratcliff, R., & Brown, S. D. (2015). Revisiting the evidence for collapsing boundaries and urgency signals in perceptual decision-making. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *35*(6), 2476–2484. https://doi.org/10.1523/JNEUROSCI.2410-14.2015

Heitz, R. P. (2014). The speed-accuracy tradeoff: History, physiology, methodology, and behavior. *Frontiers in Neuroscience*, *8*, 150. https://doi.org/10.3389/fnins.2014.00150

Houghton, G. (2018). Action and perception in literacy: A common-code for spelling and reading. *Psychological Review*, *125*(1), 83–116. https://doi.org/10.1037/rev0000084

Hug, F., Gallot, T., Catheline, S., & Nordez, A. (2011). Electromechanical delay in biceps brachii assessed by ultrafast ultrasonography. *Muscle & Nerve*, *43*(3), 441–443. https://doi.org/10.1002/mus.21948

Johnson-Laird, P. N. (1988). *The Computer and the Mind: An Introduction to Cognitive Science* [Google-Books-ID: Tf5gRFgVuegC]. Harvard University Press.

Kelly, S. P., Corbett, E. A., & O'Connell, R. G. (2021). Neurocomputational mechanisms of prior-informed perceptual decision-making in humans. *Nature Human Behaviour*, *5*(4), 467–481. https://doi.org/10.1038/s41562-020-00967-9

Kelly, S. P., & O'Connell, R. G. (2013). Internal and External Influences on the Rate of Sensory Evidence Accumulation in the Human Brain. *Journal of Neuroscience*, *33*(50), 19434–19441. https://doi.org/10.1523/JNEUROSCI.3355-13.2013

Keuleers, E., & Brysbaert, M. (2010). Wuggy: A multilingual pseudoword generator [Place: US Publisher: Psychonomic Society]. *Behavior Research Methods*, *42*(3), 627–633. https://doi.org/10.3758/BRM.42.3.627

Kinder, K. T., Buss, A. T., & Tas, A. C. (2022). Tracking flanker task dynamics: Evidence for continuous attentional selectivity. *Journal of Experimental Psychology. Human Perception and Performance*, *48*(7), 771–781. https://doi.org/10.1037/xhp0001023

Klein-Flügge, M. C., & Bestmann, S. (2012). Time-dependent changes in human corticospinal excitability reveal value-based competition for action during decision processing. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *32*(24), 8373–8382. https://doi.org/10.1523/JNEUROSCI.0270-12.2012

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

Latash, M. L. (2020). On Primitives in Motor Control. *Motor Control*, *24*(2), 318–346. https://doi.org/10.1123/mc.2019-0099

Lemon, R. N. (2008). Descending pathways in motor control. *Annual Review of Neuroscience*, *31*, 195–218. https://doi.org/10.1146/annurev.neuro.31.060407.125547

McClelland, J. L., & Rumelhart, D. E. (1981). An interactive activation model of context effects in letter perception: I. An account of basic findings [Place: US Publisher: American Psychological Association]. *Psychological Review*, *88*, 375–407. https://doi.org/10.1037/0033-295X.88.5.375

Murphy, P. R., Boonstra, E., & Nieuwenhuis, S. (2016). Global gain modulation generates time-dependent urgency during perceptual choice in humans [Number: 1 Publisher: Nature Publishing Group]. *Nature Communications*, *7*(1), 13526. https://doi.org/10.1038/ncomms13526

New, B., Pallier, C., Brysbaert, M., & Ferrand, L. (2004). Lexique 2: A new French lexical database. *Behavior Research Methods, Instruments, & Computers: A Journal of the Psychonomic Society, Inc*, *36*(3), 516–524. https://doi.org/10.3758/bf03195598

O'Connell, R. G., Dockree, P. M., & Kelly, S. P. (2012). A supramodal accumulation-to-bound signal that determines perceptual decisions in humans. *Nature Neuroscience*, *15*(12), 1729–1735. https://doi.org/10.1038/nn.3248

O'Connell, R. G., & Kelly, S. P. (2021). Neurophysiology of Human Perceptual Decision-Making. *Annual Review of Neuroscience*, *44*, 495–516. https://doi.org/10.1146/annurev-neuro-092019-100200

O'Connell, R. G., Shadlen, M. N., Wong-Lin, K., & Kelly, S. P. (2018). Bridging Neural and Computational Viewpoints on Perceptual Decision-Making. *Trends in Neurosciences*, *41*(11), 838–852. https://doi.org/10.1016/j.tins.2018.06.005

Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., Kastman, E., & Lindeløv, J. K. (2019). PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods*, *51*(1), 195–203. https://doi.org/10.3758/s13428-018-01193-y

Pfurtscheller, G., & Lopes da Silva, F. H. (1999). Event-related EEG/MEG synchronization and desynchronization: Basic principles. *Clinical Neurophysiology: Official Journal of*

the International Federation of Clinical Neurophysiology, *110*(11), 1842–1857. https://doi.org/10.1016/s1388-2457(99)00141-8

Porter, R. (1987). The Florey Lecture, 1987 - Conticomotoneuronal projections: Synaptic events related to skilled movement [Publisher: Royal Society]. *Proceedings of the Royal Society of London. Series B. Biological Sciences*, *231*(1263), 147–168. https://doi.org/10.1098/rspb.1987.0039

Purcell, B. A., Heitz, R. P., Cohen, J. Y., Schall, J. D., Logan, G. D., & Palmeri, T. J. (2010). Neurally constrained modeling of perceptual decision making [Place: US Publisher: American Psychological Association]. *Psychological Review*, *117*(4), 1113–1143. https://doi.org/10.1037/a0020311

Ratcliff, R., Sheu, C. F., & Gronlund, S. D. (1992). Testing global memory models using ROC curves. *Psychological Review*, *99*(3), 518–535. https://doi.org/10.1037/0033-295x.99.3.518

Ratcliff, R. (1978). A theory of memory retrieval. *Psychol. Rev*, *85*(2), 59–108.

Ratcliff, R., Cherian, A., & Segraves, M. (2003). A comparison of macaque behavior and superior colliculus neuronal activity to predictions from models of two-choice decisions. *Journal of Neurophysiology*, *90*(3), 1392–1407. https://doi.org/10.1152/jn.01049.2002

Ratcliff, R., & Childers, R. (2015). Individual Differences and Fitting Methods for the Two-Choice Diffusion Model of Decision Making. *Decision (Washington, D.C.)*, *2015*. https://doi.org/10.1037/dec0000030

Ratcliff, R., & McKoon, G. (2018). Modeling numerosity representation with an integrated diffusion model. *Psychological Review*, *125*(2), 183–217. https://doi.org/10.1037/rev0000085

Ratcliff, R., & Rouder, J. N. (1998). Modeling Response Times for Two-Choice Decisions. *Psychological Science*, *9*(5), 347–356. Retrieved September 27, 2019, from https://www.jstor.org/stable/40063319

Ratcliff, R., & Smith, P. L. (2004). A comparison of sequential sampling models for two-choice reaction time. *Psychological Review*, *111*(2), 333–367. https://doi.org/10.1037/0033-295X.111.2.333

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

Ratcliff, R., Thapar, A., & McKoon, G. (2004). A diffusion model analysis of the effects of aging on recognition memory [Place: Netherlands Publisher: Elsevier Science]. *Journal of Memory and Language*, *50*(4), 408–424. https://doi.org/10.1016/j.jml.2003.11.002

Ratcliff, R., Thapar, A., & McKoon, G. (2010). Individual differences, aging, and IQ in two-choice tasks. *Cognitive Psychology*, *60*(3), 127–157. https://doi.org/10.1016/j.cogpsych.2009.09.001

Robinson, D. A. (1973). Models of the saccadic eye movement control system. *Kybernetik*, *14*(2), 71–83. https://doi.org/10.1007/BF00288906

Schall, J. D. (2019). Accumulators, Neurons, and Response Time. *Trends in neurosciences*, *42*(12), 848–860. https://doi.org/10.1016/j.tins.2019.10.001

Schall, J. D., & Paré, M. (2021). The unknown but knowable relationship between Pre-saccadic Accumulation of activity and Saccade initiation. *Journal of Computational Neuroscience*, *49*(3), 213–228. https://doi.org/10.1007/s10827-021-00784-7

Selen, L. P. J., Shadlen, M. N., & Wolpert, D. M. (2012). Deliberation in the motor system: Reflex gains track evolving evidence leading to a decision. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *32*(7), 2276–2286. https://doi.org/10.1523/JNEUROSCI.5273-11.2012

Servant, M., Logan, G. D., Gajdos, T., & Evans, N. J. (2021). An integrated theory of deciding and acting. *Journal of Experimental Psychology. General*. https://doi.org/10.1037/xge0001063

Servant, M., Tillman, G., Schall, J. D., Logan, G. D., & Palmeri, T. J. (2019). Neurally constrained modeling of speed-accuracy tradeoff during visual search: Gated accumulation of modulated evidence. *Journal of Neurophysiology*, *121*(4), 1300–1314. https://doi.org/10.1152/jn.00507.2018

Servant, M., White, C., Montagnini, A., & Burle, B. (2015). Using Covert Response Activation to Test Latent Assumptions of Formal Decision-Making Models in Humans. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *35*(28), 10371–10385. https://doi.org/10.1523/JNEUROSCI.0078-15.2015

Servant, M., White, C., Montagnini, A., & Burle, B. (2016). Linking Theoretical Decision-making Mechanisms in the Simon Task with Electrophysiological Data: A Model-based Neuroscience Study in Humans. *Journal of Cognitive Neuroscience*, *28*(10), 1501–1521. https://doi.org/10.1162/jocn_a_00989

Shiffrin, R. M., & Steyvers, M. (1997). A model for recognition memory: REM—retrieving effectively from memory [Place: US Publisher: Psychonomic Society]. *Psychonomic Bulletin & Review*, *4*(2), 145–166. https://doi.org/10.3758/BF03209391

Song, J.-H., & Nakayama, K. (2009). Hidden cognitive states revealed in choice reaching tasks. *Trends in Cognitive Sciences*, *13*(8), 360–366. https://doi.org/10.1016/j.tics.2009.04.009

Starns, J. J., & Ratcliff, R. (2014). Validating the unequal-variance assumption in recognition memory using response time distributions instead of ROC functions: A diffusion model analysis. *Journal of Memory and Language*, *70*, 36–52. https://doi.org/10.1016/j.jml.2013.09.005

Steinemann, N. A., O'Connell, R. G., & Kelly, S. P. (2018). Decisions are expedited through multiple neural adjustments spanning the sensorimotor hierarchy. *Nature Communications*, *9*(1), 3627. https://doi.org/10.1038/s41467-018-06117-0

Sternberg, S. (1969). The discovery of processing stages: Extensions of Donders' method. *Acta Psychologica*, *30*, 276–315. https://doi.org/10.1016/0001-6918(69)90055-9

Storn, R., & Price, K. (1997). Differential Evolution – A Simple and Efficient Heuristic for global Optimization over Continuous Spaces. *Journal of Global Optimization*, *11*(4), 341–359. https://doi.org/10.1023/A:1008202821328

Sullivan, N., Hutcherson, C., Harris, A., & Rangel, A. (2015). Dietary self-control is related to the speed with which attributes of healthfulness and tastiness are processed. *Psychological Science*, *26*(2), 122–134. https://doi.org/10.1177/0956797614559543

Summerfield, C., & Parpart, P. (2022). Normative Principles for Decision-Making in Natural Environments. *Annual Review of Psychology*, *73*, 53–77. https://doi.org/10.1146/annurev-psych-020821-104057

Tillman, G., Osth, A. F., van Ravenzwaaij, D., & Heathcote, A. (2017). A diffusion decision model analysis of evidence variability in the lexical decision task. *Psychonomic Bulletin & Review*, *24*(6), 1949–1956. https://doi.org/10.3758/s13423-017-1259-y

Tillman, G., Van Zandt, T., & Logan, G. D. (2020). Sequential sampling models without random between-trial variability: The racing diffusion model of speeded decision making. *Psychonomic Bulletin & Review*, *27*(5), 911–936. https://doi.org/10.3758/s13423-020-01719-6

Tosoni, A., Galati, G., Romani, G. L., & Corbetta, M. (2008). Sensory-motor mechanisms in human parietal cortex underlie arbitrary visual decisions [Number: 12 Publisher: Nature Publishing Group]. *Nature Neuroscience*, *11*(12), 1446–1453. https://doi.org/10.1038/nn.2221

Trueblood, J. S., Heathcote, A., Evans, N. J., & Holmes, W. R. (2021). Urgency, leakage, and the relative nature of information processing in decision-making [Place: US Publisher: American Psychological Association]. *Psychological Review*, *128*(1), 160–186. https://doi.org/10.1037/rev0000255

Twomey, D. M., Kelly, S. P., & O'Connell, R. G. (2016). Abstract and Effector-Selective Decision Signals Exhibit Qualitatively Distinct Dynamics before Delayed Perceptual

Reports. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *36*(28), 7346–7352. https://doi.org/10.1523/JNEUROSCI.4162-15.2016

Twomey, D. M., Murphy, P. R., Kelly, S. P., & O'Connell, R. G. (2015). The classic P300 encodes a build-to-threshold decision variable. *The European Journal of Neuroscience*, *42*(1), 1636–1643. https://doi.org/10.1111/ejn.12936

Usher, M., & McClelland, J. L. (2001). The time course of perceptual choice: The leaky, competing accumulator model. *Psychological Review*, *108*(3), 550–592. https://doi.org/10.1037/0033-295x.108.3.550

Verdonck, S., Loossens, T., & Philiastides, M. G. (2020). The Leaky Integrating Threshold and its impact on evidence accumulation models of choice response time (RT). *Psychological Review*. https://doi.org/10.1037/rev0000258

Vickers, D. (1970). Evidence for an accumulator model of psychophysical discrimination. *Ergonomics*, *13*(1), 37–58. https://doi.org/10.1080/00140137008931117

Vigotsky, A. D., Halperin, I., Lehman, G. J., Trajano, G. S., & Vieira, T. M. (2018). Interpreting Signal Amplitudes in Surface Electromyography Studies in Sport and Rehabilitation Sciences [Publisher: Frontiers]. *Frontiers in Physiology*, *8*. https://doi.org/10.3389/fphys.2017.00985

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

Vrieze, S. I. (2012). Model selection and psychological theory: A discussion of the differences between the Akaike information criterion (AIC) and the Bayesian information criterion (BIC) [Place: US Publisher: American Psychological Association]. *Psychological Methods*, *17*(2), 228–243. https://doi.org/10.1037/a0027127

Wagenmakers, E.-J., Steyvers, M., Raaijmakers, J. G. W., Shiffrin, R. M., van Rijn, H., & Zeelenberg, R. (2004). A model for evidence accumulation in the lexical decision task.

Cognitive Psychology, *48*(3), 332–367. https://doi.org/10.1016/j.cogpsych.2003.08.001

Weindel, G., Anders, R., Alario, F.-X., & Burle, B. (2021). Assessing model-based inferences in decision making with single-trial response time decomposition. *Journal of Experimental Psychology. General*. https://doi.org/10.1037/xge0001010

Weiss, A. D. (1965). THE LOCUS OF REACTION TIME CHANGE WITH SET, MOTIVATION, AND AGE. *Journal of Gerontology*, *20*, 60–64. https://doi.org/10.1093/geronj/20.1.60

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

Wixted, J. T. (2007). Dual-process theory and signal-detection theory of recognition memory. *Psychological Review*, *114*(1), 152–176. https://doi.org/10.1037/0033-295X.114.1.152

Yavuz, S. U., Sendemir-Urkmez, A., & Türker, K. S. (2010). Effect of gender, age, fatigue and contraction level on electromechanical delay. *Clinical Neurophysiology: Official Journal of the International Federation of Clinical Neurophysiology*, *121*(10), 1700–1706. https://doi.org/10.1016/j.clinph.2009.10.039

Supplemental materials for:

A dual-stage dual-threshold evidence accumulation theory for decision-making, motor preparation, and motor execution

Edouard Dendauw[1,2], Nathan J. Evans[3], Gordon D. Logan[4], Thibault Gajdos[5], Emmanuel Haffen[1], Djamila Bennabi[1], and Mathieu Servant[1,2,6]

[1]Laboratoire de Recherches Intégratives en Neurosciences et Psychologie Cognitive UR 481, Université de Franche-Comté, France

[2]MSHE Ledoux USR 3124, Université de Franche-Comté, France

[3]School of Psychology, University of Queensland, Australia

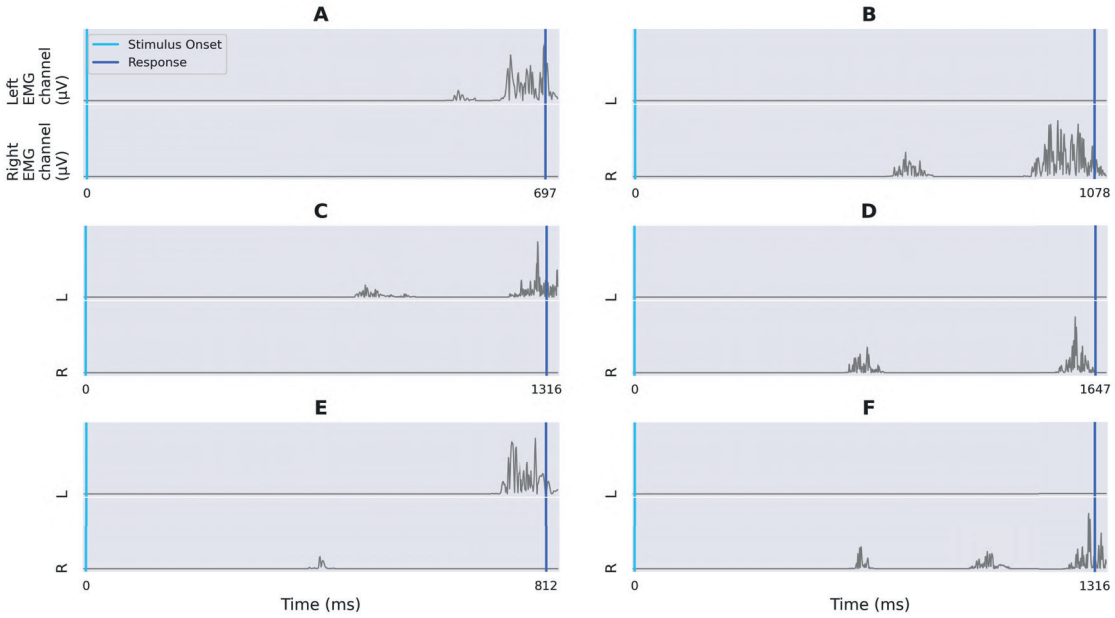
[4]Department of Psychological Sciences, Vanderbilt University, USA

[5]Laboratoire de Psychologie Cognitive UMR 7286, Aix-Marseille Université, France

[6]Institut Universitaire de France

Supplementary Figure 1

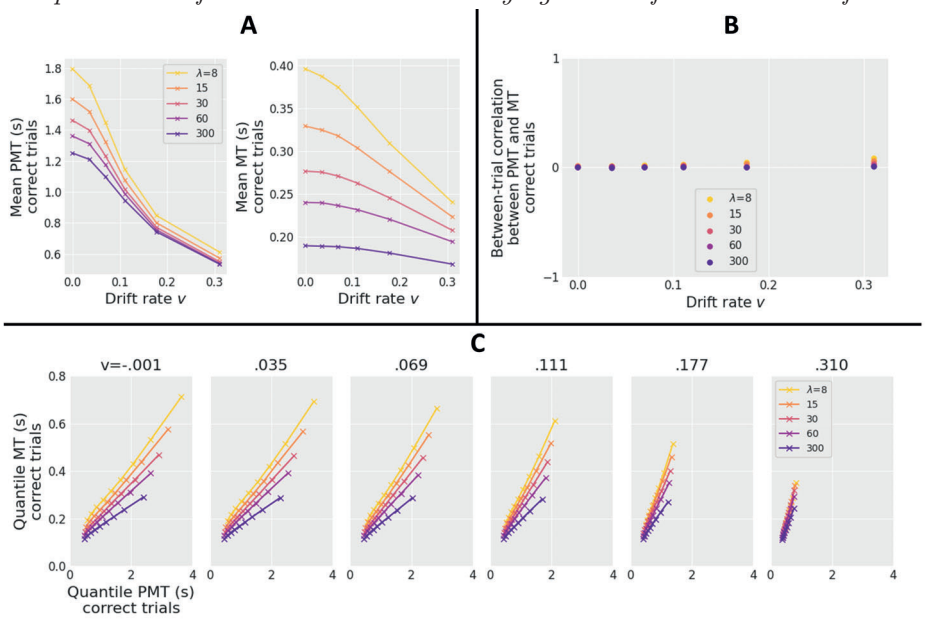
Empirical illustrations of trials containing at least one partial EMG burst during PMT



Note. Panels A, B, C and D show trials that contain one partial EMG burst in the same EMG channel as the response. Panel E shows a trial with a partial EMG burst in the opposite EMG channel as the response. Panel F shows a trial with two partial EMG bursts in the same EMG channel as the response.

Supplementary Figure 2

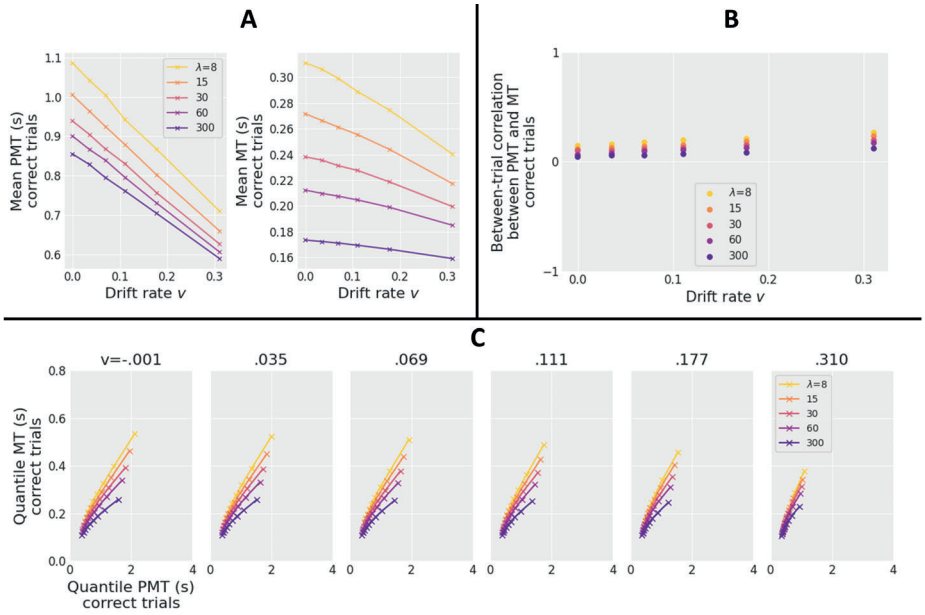
Additional predictions from DSDTDM with varying levels of leak λ and drift rate v.



Note. Apart from the leak parameter, simulations used best-fitting DTDM parameters averaged across subjects reported by Servant et al. (2021) and 100,000 simulated trials per condition. A) Predicted mean PMT and mean MT in correct trials. B) Predicted between-trial correlation between PMT and MT in correct trials. C) Predicted PMT quantile-MT quantile plot in correct trials.

Supplementary Figure 3

Predictions from DSDTDM with varying levels of leak λ and drift rate v, and with between-trial variability in drift rate (normally distributed with mean v and standard deviation sv)



Note. Apart from the leak parameter, simulations used best-fitting DTDM parameters averaged across subjects reported by Servant et al. (2021) and 100,000 simulated trials per condition. *sv* was fixed at .2. A) Predicted mean PMT and mean MT in correct trials. B) Predicted between-trial correlation between PMT and MT in correct trials. C) Predicted PMT quantile-MT quantile plot in correct trials.

Table 1

Parameters from the raw DTDM and DSDTDM models averaged across subjects for Experiments 1-4.

Exp	Model	v_1	v_2	v_3	v_4	v_5	v_6	dc	λ	m	r	Te_1	Te_2	Te_3	Te_4	Te_5	Tr	z
1	DTDM	-0.076	0.125	0.170	0.230	0.310	0.465			0.132	0.182	0.221					0.074	
1	DSDTDM	-0.002	0.035	0.066	0.104	0.155	0.249		12.544	0.056	0.080	0.191					0.066	
2	DTDM	0.020						-0.026		0.058	0.107	0.293					0.089	0.004
2	DSDTDM	0.015						-0.016	35.937	0.039	0.071	0.267					0.084	0.002
3	DTDM	0.363	0.588	0.628	-0.594					0.259	0.329	0.188					0.067	-0.001
3	DSDTDM	0.083	0.155	0.223	-0.216				35.690	0.043	0.072	0.367					0.080	0.001
4	DTDM	0.438	0.472	0.534	0.583	-0.445				0.225	0.291	0.171	0.167	0.166	0.159	0.180	0.073	-0.006
4	DSDTDM	0.111	0.176	0.233	0.282	-0.180			52.543	0.043	0.078	0.337	0.330	0.323	0.305	0.369	0.084	-0.006

Note. Experiment 1 (motion perception): parameters v_1 to v_6 correspond to drift rates for the six motion coherence levels (from 0% to 40%). Experiment 3 (recognition memory): parameters v_1 to v_4 correspond to drift rates for conditions old 1 presentation, old 2 presentations, old 4 presentations, and new respectively. Experiment 4 (lexical knowledge): parameters v_1 to v_5 and Te_1 to Te_5 correspond to drift rates and mean residual latency added to predicted PMT for conditions very low frequency words, low frequency words, medium frequency words, high frequency words, and pseudowords respectively.

Table 2

Statistics relative to frequency and number of letters for the sample of words used in Experiment 4

	Occurences per million		Letters	
	Mean	SD	Mean	SD
High frequency words	77.26	142.37	7.05	1.70
Medium frequency words	3.32	0.89	7.03	1.66
Low frequency words	0.72	0.15	7.05	1.70
Very low frequency words	0.04	0.03	7.05	1.69

References

Servant, M., Logan, G. D., Gajdos, T., & Evans, N. J. (2021). An integrated theory of deciding and acting. *Journal of Experimental Psychology. General*. https://doi.org/10.1037/xge0001063