



**HAL**  
open science

# The Gated Cascade Diffusion Model: An Integrated Theory of Decision Making, Motor Preparation, and Motor Execution

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

► **To cite this version:**

Edouard Dendauw, Nathan Evans, Gordon Logan, Emmanuel Haffen, Djamila Bennabi, et al.. The Gated Cascade Diffusion Model: An Integrated Theory of Decision Making, Motor Preparation, and Motor Execution. *Psychological Review*, In press, 10.1037/rev0000464 . hal-03924556v4

**HAL Id: hal-03924556**

**<https://hal.science/hal-03924556v4>**

Submitted on 8 Jan 2024 (v4), last revised 17 Jan 2024 (v5)

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

**The gated cascade diffusion model: An integrated theory of  
decision-making, motor preparation, and motor execution**

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

<sup>1</sup>University of Franche-Comté, LINC, 25000 Besançon, France

<sup>2</sup>University of Franche-Comté, CNRS, MSHE, 25000 Besançon, France

<sup>3</sup>Ludwig Maximilian University of Munich, Department of Psychology, 80799 Munich,  
Germany

<sup>4</sup>Vanderbilt University, Department of Psychological Sciences, TN 37240-7817 Nashville,  
USA

<sup>5</sup>Aix-Marseille University, CNRS, LPC, 13331 Marseille, France

<sup>6</sup>Institut Universitaire de France

© 2023, American Psychological Association.

**This paper is not the copy of record and may not exactly replicate the final,  
authoritative version of the article.**

**Please do not copy or cite without authors' permission.**

**The final article will be available, upon publication, via its DOI: 10.1037/rev0000464**

### Author Note

The authors declare no conflict of interest, and wish to thank Jeff Starns and two anonymous reviewers for constructive comments on earlier versions of this manuscript. Model codes, empirical data, and a codebook for processing empirical data are available on the Open Science Framework (<https://osf.io/4unw6/>). This work was funded by the European Union (ERC Starting Grant ITDA n°101039226 allocated to MS). Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Research Council. Neither the European Union nor the granting authority can be held responsible for them. Some of the data and ideas appearing in this manuscript have been presented at the 2022 Annual Meeting of the Psychonomic Society and the 22nd ESCOP Conference. NJE was supported by an Australian Research Council Discovery Early Career Researcher Award (DE200101130). TG and MS share the senior authorship. 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

© 2023, American Psychological Association.

**This paper is not the copy of record and may not exactly replicate the final, authoritative version of the article.**

**Please do not copy or cite without authors' permission.**

**The final article will be available, upon publication, via its DOI: 10.1037/rev0000464**

**The gated cascade diffusion model: An integrated theory of  
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 areas of the brain that prepare the response, where it is smoothed by a mechanism that approximates a Kalman-Bucy filter. The resulting motor preparation variable is gated prior to reaching agonist muscles until it exceeds a particular level of activation. We tested this gated cascade diffusion model by continuously probing the electrical activity of the response agonists through electromyography 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 electromyographic 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 | Kalman-Bucy filter | 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 an election eventually results in the deposit of a ballot in a box (or pushing a button on a voting machine). 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., 2015, 2021; Verdonck et al., 2020). However, as will become obvious in the next sections, current 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 address these shortcomings by introducing an integrated theory of decision-making, motor preparation, and motor execution. The theory builds upon a gated cascade evidence accumulation architecture and incorporates a filtering mechanism at the motor preparation level, for which we provide a computational foundation.

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 integrated theory.

## 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; Logan and Cowan, 1984; Sternberg, 1969). The decision process produces a discrete result, indicating which response to prepare and execute. This view still persists in contemporary decision-making models, according to 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; Tillman et al., 2020). 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 choice is then passed on to the motor system, and does not carry any information about the strength of the evidence.

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 (*i.e.*, 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 agonist muscles in a reaction time (RT) task (e.g., the *flexor pollicis brevis* 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; 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 agonist muscles. The flow is not purely continuous, because EMG bursts have a discrete onset (that occurs  $\sim$ 150-180 ms on average before the response for a button press with the thumb, with  $\sim$ 900 gram-force required; see Servant et al., 2021).

### **Modeling the interplay between decision and motor systems**

Servant et al. (2021, 2015) proposed a dual-threshold diffusion model to account for the aforementioned neurophysiological findings, with a particular focus on EMG findings. The theory concerns two-choice decisions that are mapped onto actions, and assumes that the decision variable is continuously transmitted to motor areas of the brain that prepare the response (such as premotor and primary motor cortices for body movements). Through continuous flow, some of the work usually done in forming motor commands can be done during decision formation, providing an advantage in terms of processing time (Eriksen and Schultz, 1979; McClelland, 1979; Shadlen and Kiani, 2013). In addition, this architecture offers substantial flexibility to motor control by allowing for real-time modulations and revisions of evolving motor commands based on incoming evidence (Nakayama et al., 2023; Resulaj et al., 2009; Stone et al., 2022). This flexibility appears particularly important in real-life settings, as individuals are constantly dealing with a changing environment. For instance, the affordance competition hypothesis states that the brain continuously processes sensory information to determine an ensemble of possible actions, while simultaneously gathering information to select among these actions (Cisek, 2007).

The theory further assumes that the transmission of information from motor preparation to agonist muscles is continuous, but regulated by a gate. The gate determines the minimum level of accumulated evidence required to pass on the decision variable to muscle fibers, and is presumably mediated by the basal ganglia system (hypothesized to act as a gate-keeping mechanism for the execution of motor plans; e.g., Frank, 2011; Hikosaka, 2007; Mink, 1996). The gate might serve two main purposes. First, it prevents low levels of accumulated evidence from exciting muscle fibers. Low levels of accumulated evidence are associated with a low likelihood that the decision is



correct, so gating these activations prevents unnecessary muscular activity. Second, the gate offers a shield against unwanted behaviors. These purposes are consistent with the gating function of the basal ganglia system: Patients with basal ganglia disorders often encounter difficulties initiating purposeful movements and exhibit involuntary movements such as tremors and chorea (Hikosaka, 2007; Mink, 1996).

Formally, the decision variable follows a diffusion process (Ratcliff, 1978):

$$dx(t) = vdt + \sigma dW(t), \quad x(0) = x_0, \quad (1)$$

where  $x(t)$ ,  $v$ ,  $\sigma$ , and  $W(t)$  respectively correspond to the accumulated evidence at time  $t$ , the drift rate, the diffusion coefficient, and a Brownian motion. Parameter  $x_0$  represents the starting point of the process. If there is no bias for a particular response,  $x_0 = 0$ .

The gate was originally formalized as a threshold (termed 'EMG threshold') superimposed on  $x(t)$ . Here we propose a different interpretation (though mathematically equivalent) of the gate as a constant inhibition. This interpretation is consistent with the tonic inhibitory control of the basal ganglia over motor areas (Hikosaka, 2007; Mink, 1996), and allows for a clearer description of inputs to muscle fibers. Specifically, in the context of a choice task involving left vs. right manual responses, inputs to left and right muscle fibers (variables  $z_L(t)$  and  $z_R(t)$  respectively), are defined as follows:

$$\begin{cases} z_L(t) &= \max(-x(t) - g, 0) \\ z_R(t) &= \max(x(t) - g, 0), \end{cases} \quad (2)$$

where parameter  $g$  corresponds to gating inhibition. Variables  $z_L(t)$  and  $z_R(t)$  are classically referred to as neural drive. The electrical excitation of muscle fibers, measured by EMG, starts when the neural drive becomes positive. The full-wave rectified EMG signal<sup>1</sup> can be interpreted as a noisy approximation of the neural drive to the area of muscle over which the electrodes are placed (Dideriksen and Farina, 2019; Farina et al.,

---

<sup>1</sup> Full-wave rectification consists in taking the absolute value of voltages across time points (so all negative voltages become positive). This transformation requires the EMG signals to be centered around zero.

2010; Vigotsky et al., 2018). For this reason, we will represent full-wave rectified EMG signals (instead of raw signals) in the present work.

Researchers in biomechanics have evidenced a strong association between the neural drive, the resulting electrical excitation of muscle fibers, and force production (for a review, see Vigotsky et al., 2018)<sup>2</sup>. Consequently, the model assumes that the response (e.g., a button press) is issued when a particular level of neural drive has been produced, which is modeled by applying a threshold (parameter  $r$ ) on  $z_L(t)$  and  $z_R(t)$ . A left response is issued if variable  $z_L(t)$  first reaches  $r$  whereas a right response is issued if variable  $z_R(t)$  first reaches  $r$ . We refer to  $r$  as response threshold. This parameter depends on the force required to give the response, properties of agonist muscles, and muscular variables specific to each individual (e.g., strength and endurance).

This processing architecture makes EMG predictions (such as distributions of PMT and MT) that are strictly identical to the dual-threshold diffusion model proposed by Servant et al. (2021, 2015). The only difference concerns the interpretation of the gate as a constant inhibition (instead of a threshold), which translates into a minor parametric change<sup>3</sup> and prompts a renaming of the model. We refer to it as *gated cascade diffusion model* (GC) to emphasize the main processing components (diffusion decision variable, continuous flow, and gate).

In each trial, the predicted PMT corresponds to the latency between the onset of evidence accumulation and the time at which  $z_L(t)$  or  $z_R(t)$  becomes positive. Note, however, that the latter event can occur more than one time during a trial, due to noisy fluctuations of the accumulated evidence variable  $x(t)$  around the gate. This phenomenon allows the model to predict a well-established EMG phenomenon termed partial EMG burst (Coles et al., 1985; Servant et al., 2015). A partial EMG burst corresponds to a

---

<sup>2</sup> Technically, the neural drive refers to the electrical activity of  $\alpha$ -motoneurons that innervate muscle fibers. This activity triggers an excitation-contraction coupling at the level of muscle fibers, leading to force production (Calderón et al., 2014). EMG measures the electrical excitation of muscle fibers, which is a precursor to force generation.

<sup>3</sup> Parameters  $g$  and  $r$  respectively correspond to parameters  $m$  and  $r - m$  in the dual-threshold diffusion model.

small burst of muscular excitation that sometimes occurs during PMT, and that does not generate sufficient force to issue the response (for empirical illustrations, see Figure A1 in Appendix A). The predicted PMT thus corresponds to the latency between accumulation onset and the time  $t_g$  at which  $z_L(t)$  or  $z_R(t)$  becomes positive for the last time before reaching the response threshold, consistent with the empirical definition (Figure 1). The predicted MT corresponds to the latency at which  $z_L(t)$  or  $z_R(t)$  hits the response threshold minus  $t_g$ . 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  $Tr$  includes 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., Cavannah and Komi, 1979). In its raw form (without between-trial variability in any of the model components and with an unbiased starting point of evidence accumulation  $x_0 = 0$ ), GC has five free parameters: drift rate  $v$ , gating inhibition  $g$ , response threshold  $r$ , mean residual latencies  $Te$  and  $Tr$ .

Because muscular excitation is determined by  $x(t)$ , modulations of drift rate impact both predicted PMT and MT. Consequently, GC 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 predictions from the model. First, for any given drift rate, distributions of PMT and MT should exhibit a similar right-skewed shape, which should translate into an approximately linear PMT quantile-MT quantile plot. Second and as mentioned before, GC sometimes predicts partial EMG bursts during PMT. The proportion of trials containing at least one partial burst and the latency of the first partial burst should increase as drift rate decreases. Third, for any given drift rate, the between-trial correlation between PMT and MT should be null, due to the Markov property of the diffusion 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<sup>4</sup>. In addition, fits of GC 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.

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. GC 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. These results demonstrate that the amount of within-trial noise predicted by GC is too large, causing too many oscillations around the gate. To solve this problem, one may be tempted to decrease the diffusion coefficient (parameter  $\sigma$  in Equation 1) that regulates the amplitude of within-trial noise. However, the diffusion coefficient is fixed at an arbitrary value to satisfy a scaling property within the model (Ratcliff, 1978). This means that adjusting

---

<sup>4</sup> 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 GC showed that the model could predict a small positive correlation between PMT and MT for a given drift rate level when between-trial variability in drift rate is incorporated.

the value of  $\sigma$  –either to a different fixed value, or as a free parameter– has no impact on model predictions, as any modulation will be counteracted by a proportional modulation of the other model parameters to produce identical predictions. This analysis suggests that a processing step is missing in GC.

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). GC 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 (Ratcliff, 1978), or to other 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 hypothesized that the evidence accumulation decision variable is continuously transmitted to motor areas of the brain that prepare the response, similar to Servant et al. (2021, 2015). They further assumed that the decision variable is filtered during motor preparation. Formally, the evidence accumulation variable  $x(t)$  follows a diffusion process identical 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 the following differential equation:

$$dy(t) = \lambda(x(t) - y(t))dt, \quad y(0) = x_0, \quad (3)$$

where  $\lambda > 0$  corresponds to the leak parameter. The response is executed when  $y(t)$  reaches one of two thresholds. As can be seen by solving Equation 3, the motor preparation variable is obtained by applying a smoothing filter to  $x(t)$ :

$$y(t) = \lambda \int_{-\infty}^t dt' x(t') e^{\lambda(t'-t)}. \quad (4)$$

The value of the decision variable at  $t - t'$  seconds before the current time  $t$  contributes  $x(t')e^{\lambda(t'-t)}$  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  $\lambda$  (the leak parameter). When  $\lambda$  approaches infinity (high leak), all weights on past states tend to zero, and the model reduces to the diffusion model (Ratcliff, 1978). Conversely, as  $\lambda$  decreases (low leak), 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. The amount of smoothing at the motor preparation level thus increases as  $\lambda$  decreases. Verdonck et al. (2020) further showed that for large values of  $t$ , the mean of  $y(t)$  is delayed by  $\lambda^{-1}$  relative to the mean of  $x(t)$ , corresponding to the filter-related delay. A modulation of  $\lambda$  thus produces a modulation of the speed and accuracy of the decision.

This cascade evidence accumulation architecture for decision-making and motor preparation, termed leaky integrated threshold model (LIT) by the authors, provides a straightforward explanation to the partial temporal overlap between corresponding neurophysiological activities, their rise-to-threshold morphology, and the modulation of their respective accumulation rate by stimulus difficulty. It also allows for specific strategic influences at each processing stage. Although Verdonck et al. (2020) did not test LIT against neurophysiological data, they showed that it provides a better account of behavioral data than the diffusion decision 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 decision thresholds commonly assumed in the literature.

There are two theoretical issues with the LIT framework. First, the rationale for the model is not clear. The diffusion model is known to implement an optimal strategy,

in the sense that it minimizes the expected decision time for a given expected accuracy level (Bogacz et al., 2006). Consequently, it is not clear why the motor preparation process would average past accumulated evidence (containing less information) with current accumulated evidence (containing more information). The second issue concerns motor execution. 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 EMG bursts. The integrated theory of decision-making, motor preparation, and motor execution introduced in the next section provides a solution to both problems. We further illustrate how the theory can capture partial EMG burst statistics.

### **An integrated theory of decision-making, motor preparation, and motor execution**

We first propose a computational foundation (Marr, 1982) for the motor preparation process. The brain is a noisy information processing system (Shadlen and Newsome, 1994), and the decision variable is likely corrupted by noise during its continuous transmission to motor preparation areas. Therefore, an important goal of motor preparation might be to recover the original decision variable from noise. Formally, the corrupted decision variable  $\tilde{x}$  received by the motor preparation process can be defined as:

$$\tilde{x}(t) = x(t) + \xi U(t), \quad \tilde{x}(0) = x_0, \quad (5)$$

where  $\xi U(t)$  corresponds to white Gaussian noise with standard deviation  $\xi$  added during the transmission process. It is well known that the Kalman-Bucy filter provides the optimal solution to this problem, in the sense that the Kalman-Bucy filtered process  $\tilde{y}$  minimizes the mean squared prediction error  $\mathbb{E}[(x(t) - \tilde{y}(t))^2]$  (Kalman and Bucy, 1961; Øksendal, 2003).  $\tilde{y}$  satisfies the following differential equation (see Appendix B for the

mathematical derivation):

$$d\tilde{y}(t) = \frac{\sigma}{\xi} \tanh\left(\frac{\sigma}{\xi}t\right) (\tilde{x}(t) - \tilde{y}(t))dt + vdt, \quad \tilde{y}(0) = x_0. \quad (6)$$

The term  $\frac{\sigma}{\xi} \tanh\left(\frac{\sigma}{\xi}t\right)$  corresponds to the so-called Kalman gain, and determines the amount of smoothing of  $\tilde{x}(t)$  needed to optimally recover  $x(t)$  from noise.

Equation 6 poses an important challenge for the motor preparation system: it requires knowledge of the ratio between  $\sigma$  (diffusion coefficient) and  $\xi$  (amplitude of transmission noise) as well as the drift rate  $v$ , which is implausible. After all, if the motor preparation system had knowledge of the drift rate, no decision-making mechanism would be necessary. Consequently, these parameters may be replaced by priors under strategic control. Let parameter  $\lambda'$  denotes the prior on the ratio between  $\sigma$  and  $\xi$ , and parameter  $v'$  the prior on the drift rate. If there is no bias towards a particular stimulus/response, a reasonable prior for  $v'$  is  $v' = 0$ . Equation 6 thus becomes:

$$d\tilde{y}(t) = \lambda' \tanh(\lambda't) (\tilde{x}(t) - \tilde{y}(t))dt, \quad \tilde{y}(0) = x_0, \quad (7)$$

Equation 7 can be viewed as the best estimate of the decision variable  $x(t)$  that the motor preparation system can plausibly provide, given available information<sup>5</sup>. It turns out the motor preparation process  $y(t)$  from LIT is very similar to  $\tilde{y}(t)$ . Note that the gain  $\lambda' \tanh(\lambda't)$  quickly increases to its asymptotic value  $\lambda'$ , and therefore:

$$d\tilde{y}(t) \approx \lambda'(\tilde{x}(t) - \tilde{y}(t))dt, \quad \tilde{y}(0) = x_0. \quad (8)$$

Comparing Equations 3 and 8, the main difference between  $\tilde{y}(t)$  and  $y(t)$  concerns the input to motor preparation:  $\tilde{y}(t)$  takes the corrupted decision variable  $\tilde{x}(t)$  as input, instead of the decision variable  $x(t)$  (see Appendix B for additional mathematical details). The prior  $\lambda'$  corresponds to the leak parameter  $\lambda$ , and can be interpreted in a similar way. To summarize, the filter at the motor preparation level can be viewed as an attempt

---

<sup>5</sup> Equation 7 is still optimal in the sense that it minimizes the mean squared prediction error  $\mathbb{E}[(x(t) - \tilde{y}(t))^2]$ , but optimality here must be understood under the assumption that parameters  $v$ ,  $\sigma$ , and  $\xi$  are replaced by priors  $v'$  and  $\lambda'$ . It can be contrasted with the optimal filter applied by an ideal observer – who would use the actual values of  $v$ ,  $\sigma$  and  $\xi$  (see Appendix C).



to recover the original decision variable from noise. This mechanism has an important consequence that will become clear in the next section: it prevents unnecessary muscular activity triggered by random noise in the accumulated evidence, and thus appears complementary to the action of the gate in stabilizing motor control.

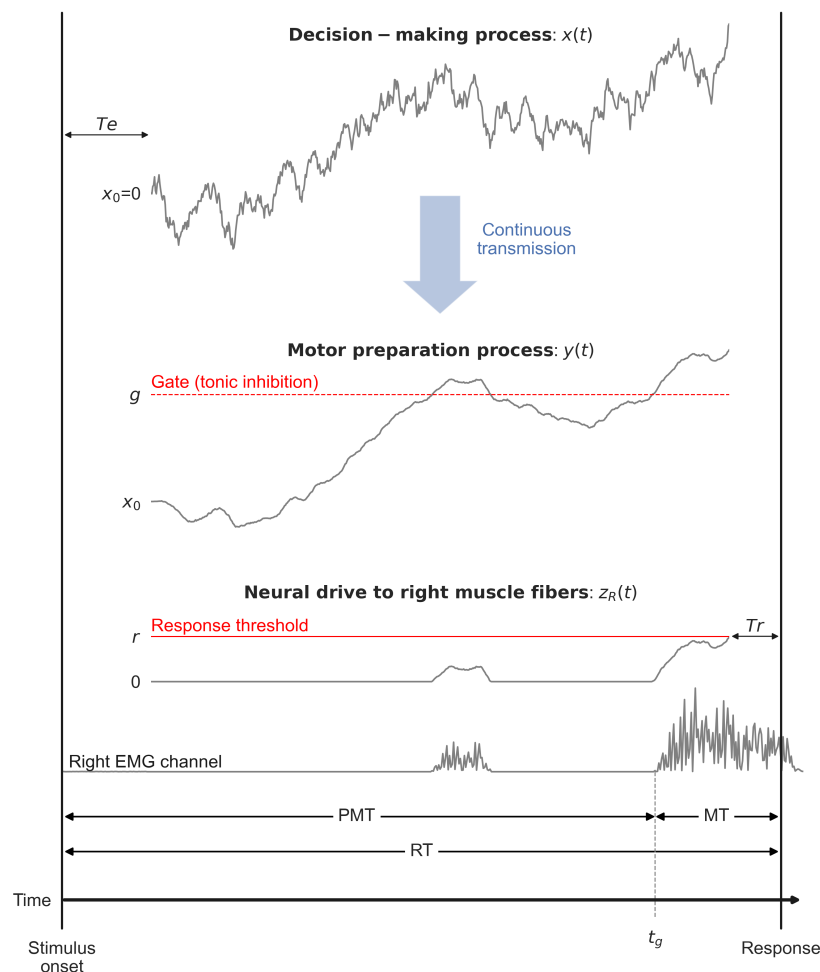
The above analysis suggests that a complete theory of decision-making, motor preparation, and motor execution may require a combination of GC and LIT models. We refer to this new processing architecture as GCF, with the letter 'F' indicating that the decision variable is filtered at the motor preparation level. An illustration of the model is provided in Figure 1. Decision-making and motor preparation are modeled by  $x(t)$  (Equation 1) and  $y(t)$  (Equation 3) respectively. We chose  $y(t)$  instead of  $\tilde{y}(t)$  to save one free parameter (the amplitude of transmission noise  $\xi$ ) and reduce the risk of tradeoffs between parameters. This choice does not have any impact on our main conclusions, since model variants incorporating  $y(t)$  vs.  $\tilde{y}(t)$  provide a comparable fit performance to data (see Appendix C), consistent with the above analysis. Gating inhibition now operates on  $y(t)$ , so inputs to left and right muscle fibers (variables  $z_L(t)$  and  $z_R(t)$  respectively) are defined as follows:

$$\begin{cases} z_L(t) &= \max(-y(t) - g, 0) \\ z_R(t) &= \max(y(t) - g, 0). \end{cases} \quad (9)$$

The response (left vs. right) is determined by the variable that first hits the response threshold  $r$ , similar to GC.

### Model simulations

Within GCF, the smoothing mechanism 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  $\lambda$  and drift rate  $v$ . Similar to GC, 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  $\lambda$  decreases. Additional analyses of simulated data showed that GCF

**Figure 1***Architecture of GCF*

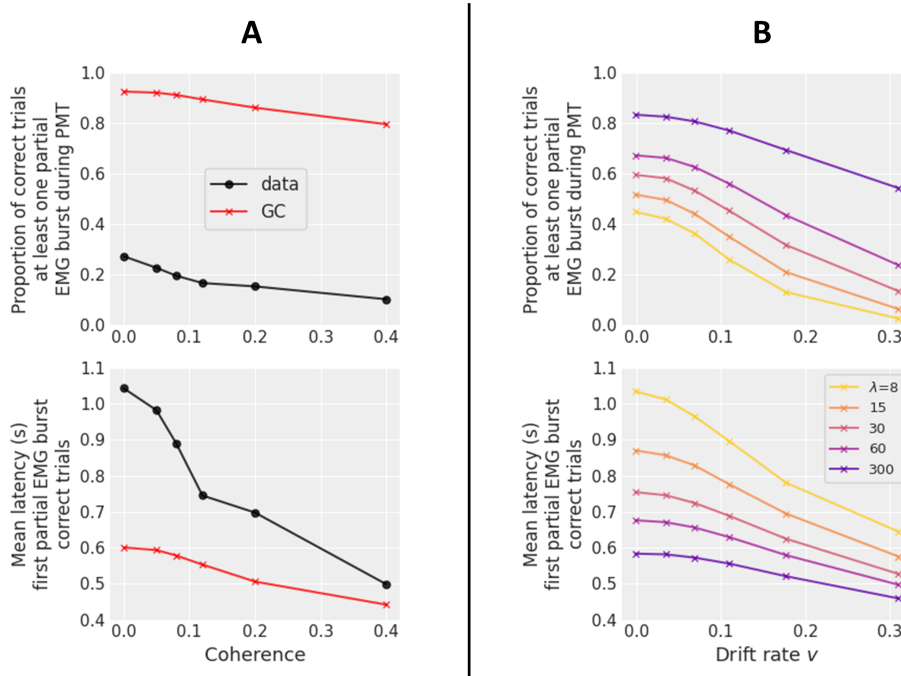
*Note.* In this trial, the model predicts a right response and a partial EMG burst in the right EMG channel. See text for details.

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 (Figure D1 in Appendix D). 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 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 GCF

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 (Figure D2 in Appendix D). 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 GC predictions are shown as red crosses. B: GCF simulations, with varying levels of leak  $\lambda$  and drift rate  $v$ . Apart from the leak parameter, simulations used best-fitting GC parameters averaged across subjects reported by Servant et al. (2021) and 100,000 simulated trials per condition.

Given the complexity of GCF, 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 GC 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 forces researchers to apply a low-pass filter to smooth the signal, obscuring the degree to which the brain may perform such filtering.

### Experiment 1: Motion perception

As a first step, we fit GCF 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). Following Palmer et al. (2005), we assumed a linear relationship between motion coherence and drift rate. We predicted a better performance (balance between fit quality and parsimony) of GCF compared to GC, 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 were 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 GC and GCF 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 sampling rate of the response device was 1000 Hz, and the force required to press each button was  $\sim$ 900 gram-force. The EMG activity of response agonists (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 occurred 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<sup>rd</sup> 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). Trials with a high level of noise were discarded from analyses (7.5% of trials on average; range 0.2-24%).

### *Models and fit procedure*

GC and GCF 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 to provide the same granularity as the behavioral and EMG data, and the diffusion coefficient  $\sigma$  (see Equation 1) was fixed at .1 to satisfy a scaling property within the models (see general introduction). Following Servant et al. (2021), we fixed the starting point of the

decision-making process at  $x_0 = 0$ , since left and right responses were equiprobable<sup>6</sup>. Consequently, we modeled incorrect/correct responses instead of left/right responses (with negative evidence favoring the incorrect response and positive evidence favoring the correct response). In its raw form (i.e., without between-trial variability in any of the model parameters), GCF has 6 free parameters: the slope  $k$  of the linear relationship relating motion coherence to drift rate, gating inhibition  $g$ , response threshold  $r$ , mean residual latencies  $Te$  and  $Tr$ , and the leak parameter  $\lambda$ . The raw GC has 5 free parameters (all GCF parameters except  $\lambda$ ), and the full GC has four additional parameters (between-trial variability in drift rate  $sv$ , starting point  $sx_0$ , and residual latencies  $sTe$  and  $sTr$ ).  $sv$  corresponds to the standard deviation of a Gaussian distribution with mean  $v$ .  $sx_0$ ,  $sTe$ , and  $sTr$  correspond to the range of a uniform distribution with mean  $x_0$ ,  $Te$ , 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 GCF has one additional between-trial variability parameter, corresponding to between-trial variability in leakage (uniformly distributed with range  $s\lambda$  and mean  $\lambda$ ). A uniform distribution was chosen because we do not have any theoretical assumption about the distributional shape of variability in leakage. All free parameters were constrained to be  $\geq 0$ , and were not allowed to vary between motion coherence conditions. Parameters  $sx_0$ ,  $sTe$ ,  $sTr$ , and  $s\lambda$  were further constrained to not exceed 180% of  $g$ ,  $Te$ ,  $Tr$ , and  $\lambda$  respectively. The models were fit to each individual dataset by minimizing the following loss function (likelihood-ratio chi-square statistic):

$$G^2 = 2 \sum_{i=1}^6 \sum_{j=1}^6 \sum_{k=1}^6 \sum_{l=1}^6 n_{ijkl} \log \left( \frac{n_{ijkl}/n_i}{n'_{ijkl}/n'_i} \right). \quad (10)$$

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,

---

<sup>6</sup> Relaxing this constraint did not change model selection results, and had a negligible impact on the goodness-of-fit of the models.

and .9) and the six bins bounded by MT quantiles (.1, .3, .5, .7, and .9) respectively<sup>7</sup>. The variables  $n_{ijkl}$  and  $n'_{ijkl}$  refer to the observed and simulated number of trials in coherence condition  $i$ , trial type  $j$ , PMT bin  $k$ , and MT bin  $l$ . Finally, the variables  $n_i$  and  $n'_i$  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 EMG 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 note that GC is nested in GCF: the two models are equivalent when the leak parameter  $\lambda$  approaches infinity. Consequently, a low best-fitting leak value would indicate that GCF adds to a GC description of the data. The key question is whether this improvement in fit quality is sufficiently important to justify the additional complexity of GCF. To answer this question, the  $G^2$  was converted to both AIC and BIC model selection statistics:

$$AIC = G^2 + 2m, \quad (11)$$

$$BIC = G^2 + m\log(N), \quad (12)$$

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 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

---

<sup>7</sup> If subjects made a number of errors comprised between five and 10 in a given condition, a median split was used to form two bins. If there were fewer than five errors, error PMTs and MTs for the condition were excluded from the  $G^2$  calculation.

associated with the smallest AIC or BIC. If 14 (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

GCF was associated with lower AIC and BIC statistics compared to GC 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 GCF compared to GC, indicating that between-trial variability in GCF parameters has a minor impact on model performance, contrary to GC. In fact, the raw GCF was associated with lower AIC (BIC) statistics compared to the full GC for 16 (16) subjects. This analysis provides strong evidence for the superiority of GCF.

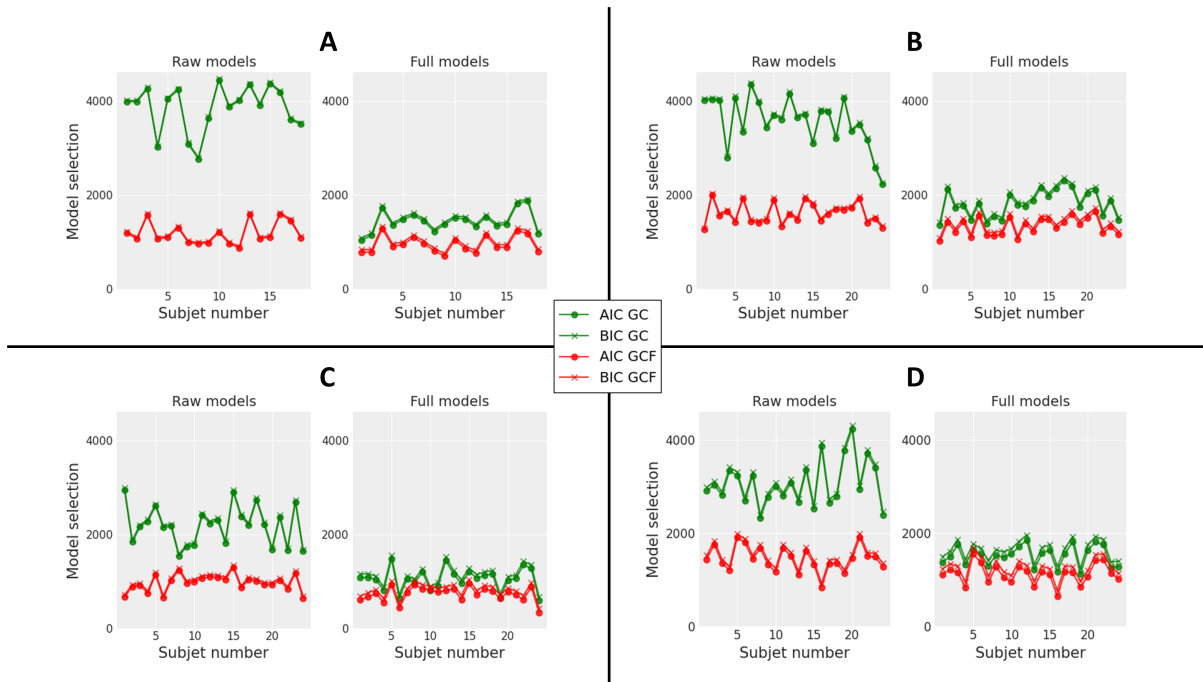
Best-fitting parameters for the full models are shown in Table 1 (main parameters) and Table 2 (between-trial variability parameters). Best-fitting parameters for the raw models are shown in Table C1 in Appendix C. As predicted, both raw and full GCF capture the EMG data with a low level of leakage, 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 GC compared to GCF, 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. GCF predictions are displayed in red, GC 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 by 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 concise 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. 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 used to constrain parameter estimation, and serve as a generalization test of the models.

Overall, the full GCF provides a good account of the data, though two minor discrepancies are apparent. First, the model overestimates the .9 quantile of PMT distributions as motion coherence decreases, especially for incorrect trials. Second, the predicted between-trial correlation between PMT and MT for each individual subject shows less dispersion compared to observed data, but this phenomenon is likely due to noise in EMG onset detection. In addition, GCF slightly overestimates the correlation

**Figure 3***Model selection statistics for Experiments 1-4.*

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

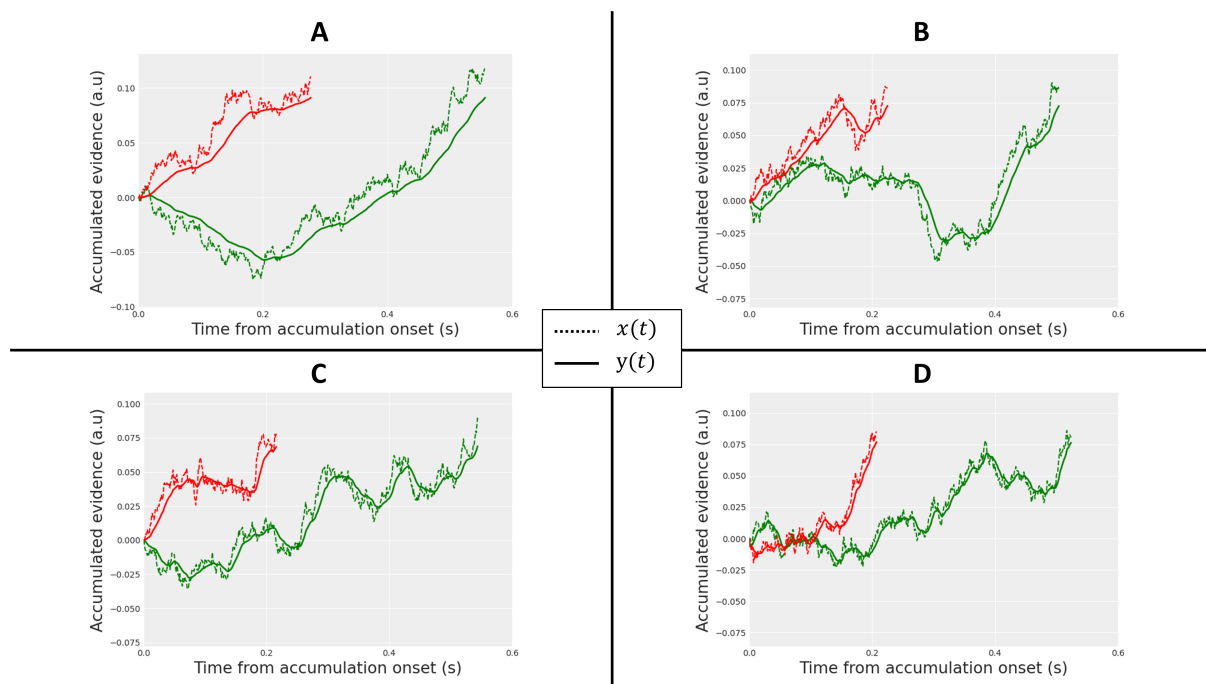
for the highest motion coherence level. As discussed previously (see general introduction and Appendix D), the model predicts a positive correlation when a high drift rate is combined with a low leakage level, especially if between-trial variability in drift rate is incorporated.

In its raw form, GC 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 GC will no longer be discussed. The full GC 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 GC comes from a much smaller response threshold  $r$ , but this modulation has several negative consequences. Most importantly, the predicted mean MT essentially corresponds to parameter  $Tr$ , which is implausible from a physiological perspective, and the predicted variability in MT is mostly driven by parameter  $sTr$ . The

model thus strongly underestimates the effect of motion coherence on mean MT, and fails to account for the right-skewed distribution 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**

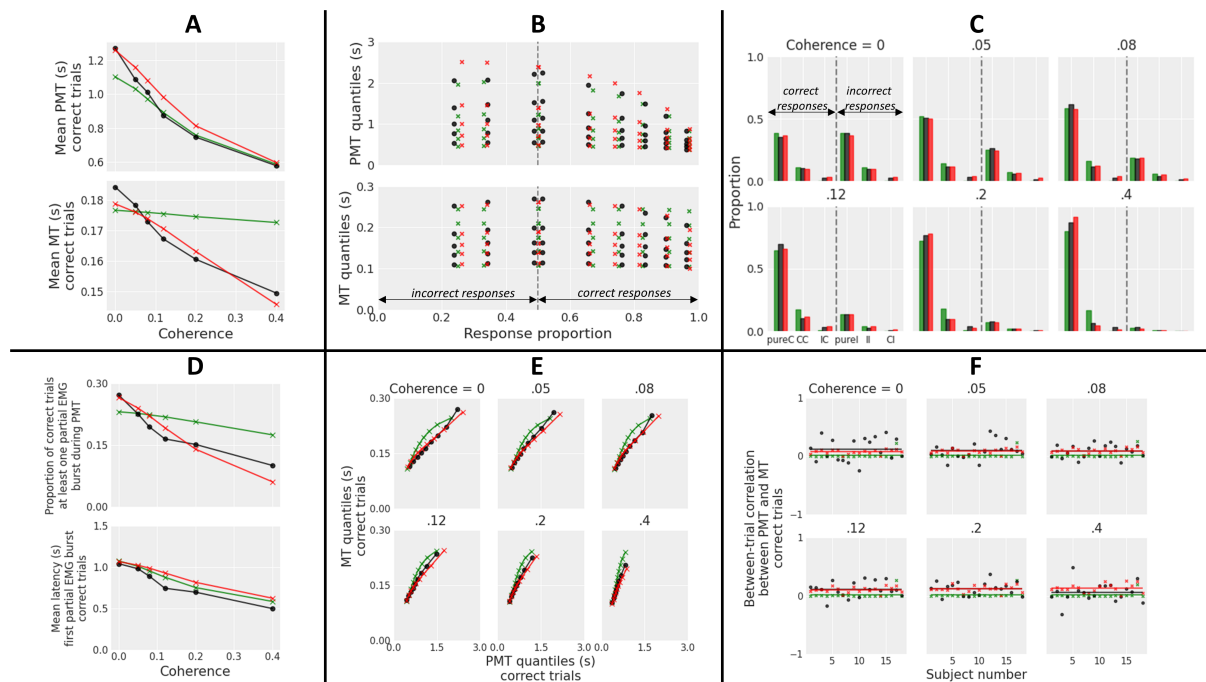
*Trajectories of decision-making  $x(t)$  and motor preparation  $y(t)$  variables computed from the full GCF using best-fitting parameters averaged across subjects from Experiments 1-4.*



*Note.* Panels A-D correspond to Experiments 1-4. For each experiment, two correct trials (red and green trajectories) were simulated using best-fitting parameters from the full GCF averaged across subjects, with the constraint of giving latencies between accumulation onset and response threshold crossing 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 words studied twice 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 3. Variations in the amount of smoothing at the motor preparation level between tasks are due to variations in the leak parameter  $\lambda$ . These variations are not surprising because parameter  $\lambda$  is under strategic control. a.u: arbitrary units.

**Figure 5**

Data from a random dot motion task with varying levels of motion coherence (black) plotted against predictions from the full GCF (red) and the full GC (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 (coherence conditions 0, .05, and .08 fulfilled this requirement). 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 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 condition. Observed data and model predictions for each individual subject are shown as scattered dots and crosses. Observed data and model predictions averaged across subjects are shown as horizontal lines.

Table 1

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

Exp	Model	$v_1$	$v_2$	$v_3$	$v_4$	$v_5$	$k$	$dc$	$\lambda$	$g$	$r$	$Te_1$	$Te_2$	$Te_3$	$Te_4$	$Te_5$	$Tr$	$x_0$
1	GC						1.534			0.091	0.005	0.353					0.169	
1	GCF						1.183		21.471	0.069	0.022	0.283					0.093	
2	GC	0.066						-0.035		0.065	0.036	0.474					0.148	0.006
2	GCF	0.023						-0.011	57.842	0.042	0.030	0.339					0.106	0.000
3	GC	0.207	0.518	0.739	-0.656					0.045	0.024	0.544					0.139	0.001
3	GCF	0.138	0.288	0.399	-0.342				73.158	0.037	0.031	0.452					0.099	0.000
4	GC	0.459	0.683	0.809	0.878	-0.571				0.046	0.036	0.589	0.571	0.549	0.512	0.581	0.133	-0.004
4	GCF	0.277	0.393	0.460	0.517	-0.301			122.938	0.041	0.035	0.500	0.471	0.454	0.430	0.489	0.109	-0.006

Note. Experiment 3 (recognition memory): parameters  $v_1$  to  $v_4$  correspond to drift rates for conditions old words studied one time, old words studied two times, old words studied four times, and new words 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 GC and GCF models averaged across subjects for Experiments 1-4.*

Exp	Model	$sv_1$	$sv_2$	$sv_3$	$sv_4$	$sv_5$	$\sigma_1$	$\eta_0$	$sx_0$	$sTe$	$sTr$	$s\lambda$
1	GC	0.134							0.085	0.286	0.168	
1	GCF	0.117							0.074	0.203	0.076	29.423
2	GC						0.007	0.050	0.098	0.422	0.133	
2	GCF						0.001	0.042	0.060	0.225	0.078	82.803
3	GC	0.439	0.427	0.426	0.274				0.065	0.327	0.113	
3	GCF	0.335	0.313	0.263	0.163				0.052	0.204	0.065	86.538
4	GC	0.522	0.435	0.372	0.312	0.267			0.071	0.456	0.107	
4	GCF	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 words studied one time, old words studied two times, old words studied four times, and new words 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 (range: 31-70) randomly scattered in a  $10 \times 10$  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., 31 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 GCF and GC 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). Methodological details regarding the experiment and the modeling of the data are provided in Appendix E.

## Results

### *Behavior and EMG*

The data from 24 subjects were 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 of each bin (33, 38, 43, 48, 53, 58, 63, 68). They 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). Anticipations (RTs < 150 ms; 0%) and trials in which participants failed to respond before the 4 s deadline (0.12%) were discarded from analyses.

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 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 pronounced 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***

The fit procedure was identical to that used in Experiment 1, except that we treated the starting point  $x_0$  of the decision-making process as a free parameter. We modeled 'lesser than' and 'greater than' responses (with negative evidence favoring the 'lesser than' response, and positive evidence favoring the 'greater than' response). 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*).

Similar to Experiment 1, GCF was associated with lower AIC and BIC statistics compared to GC 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 GCF compared to GC, and the raw GCF was associated with a lower AIC (BIC) compared to the full GC for 21 (22) subjects. This analysis provides strong evidence for the superiority of GCF.

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 Table C1 in Appendix C. Although the best-fitting leakage ( $\lambda$ ) value from GCF was larger than that observed in Experiment 1, this value still implies substantial smoothing 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 compo-

nents was higher for GC compared to GCF, especially for residual latencies (parameters  $sTe$  and  $sTr$ ).

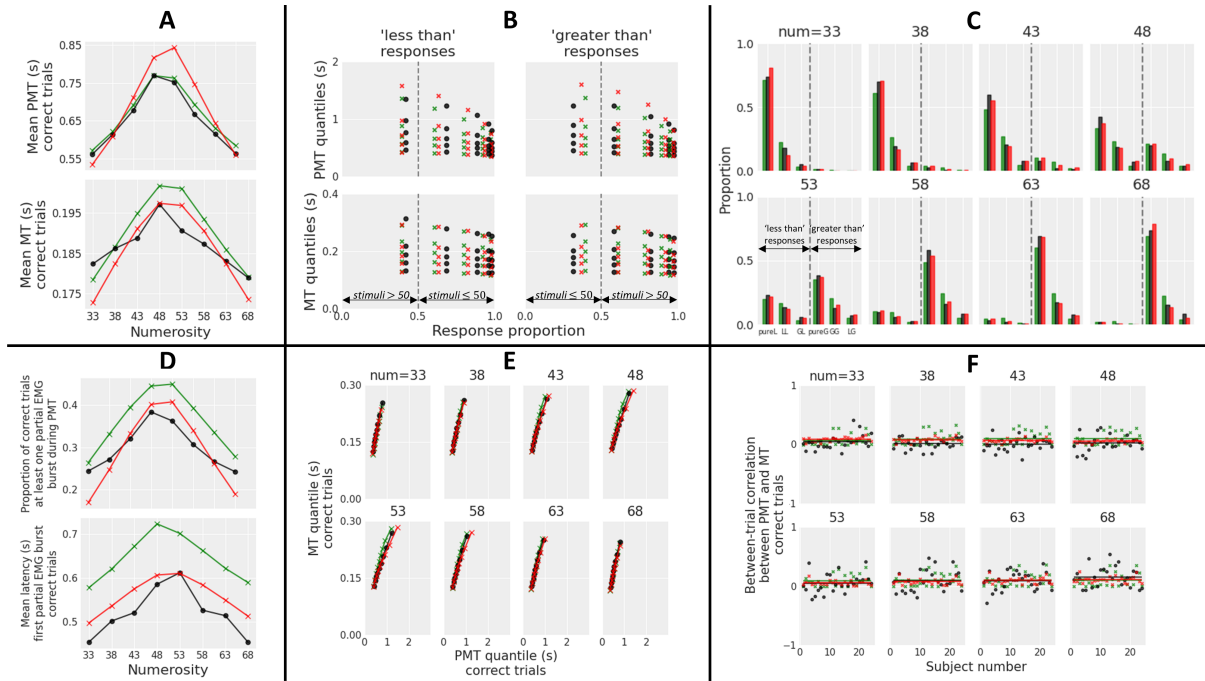
Figure 6 displays the goodness-of-fit of the full models to data. Compared to Experiment 1, the full GC provides a better account of the task difficulty effect on mean MT, thanks to a larger response threshold. However, the model still fails to provide a good fit to MT quantiles, because the contribution of residual motor latencies to predicted MTs remains substantial. In addition, the full GC systematically overestimates the rate of correct LL and GG trials and the mean latency of the first partial EMG burst. The full GCF 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 E2 in Appendix E).

### Experiment 3: Recognition memory

The diffusion model was originally developed 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 words, each word being presented individually at a pace of 1 s. During test, studied words were intermixed with non-studied words, and participants 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 repetitions during study. Specifically, each word was studied one time, two times, or four times. The drift rate should increase as the number of repetitions (and

**Figure 6**

Data from a numerosity judgment task (black) and predictions from the full GCF (red) and the full GC (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 PMT and MT quantiles of incorrect responses in a given condition if each subject made at least 10 errors (numerosity conditions 48 and 53 fulfilled this requirement). 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.

thus memory strength) increases. Consequently, both GC and GCF predict a decrease in mean MT as memory strength 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. Methodological details regarding the experiment and the modeling of the data are provided in Appendix E.

## Results

### *Behavior and EMG*

Twenty-four subjects completed the experiment. 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 memory strength (words studied one time, two times, or four times) as within-subjects factor and specific error terms.

Accuracy increased as memory strength increased,  $t(23) = 14.14$ ,  $p < .001$ . Consistent with model predictions, mean RT, mean PMT, and mean MT decreased as memory strength 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 memory strength 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 memory strength 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 GC and GCF- 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 memory strength 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). The between-trial Pearson correlation coefficient between PMT and MT was close to zero on average, and slightly decreased as memory strength increased (words studied one time:  $r = .07$ ; words studied two times:  $r = .04$ ; words studied four times:  $r = 0$ ; Figure 7F).

To compare performance between old and new items, we averaged the performance to old items across memory strength 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 fit procedure was identical to that used in the previous experiments. We modeled 'new' and 'old' responses (with negative evidence favoring the 'new' response and positive evidence favoring the 'old' response). 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*).

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

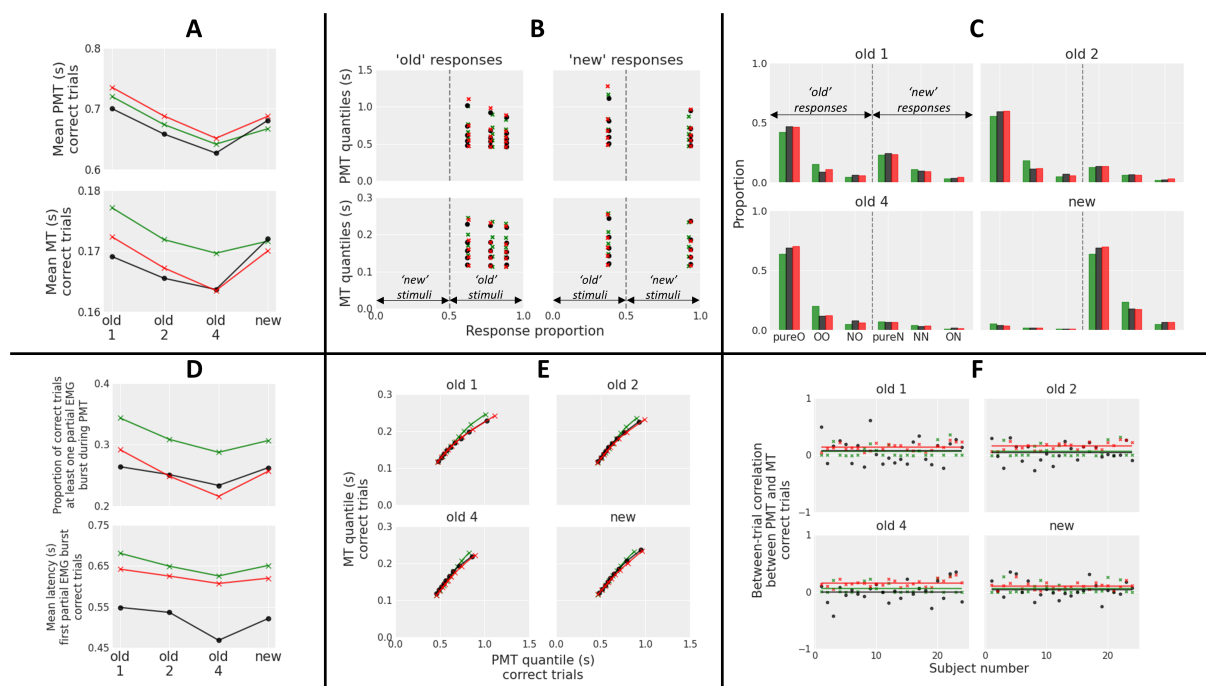
Best-fitting parameters for the full models are shown in Table 1 (main parameters) and Table 2 (between-trial variability parameters). Best-fitting parameters for the raw models are shown in Table C1 in Appendix C. GCF captures the data with a moderate amount of leakage ( $\lambda$ ), though the best-fitting value for the full model is a bit larger compared to Experiment 2 (implying reduced smoothing 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 GC compared to GCF. Note that between-trial variability in drift rate ( $sv$ ) was larger for old than new words, consistent with previous work. It also decreased as memory strength increased, suggesting that evidence variability decreases as function of learning.

Figure 7 displays the goodness-of-fit of the full models to data. The full GC

provides a poor account of MT distributions, due to the large contribution of residual motor latencies to predicted MTs. In addition, the full GC systematically overestimates the proportion of correct OO and NN trials and the mean latency of the first partial EMG burst. The full GCF provides a good fit to data used to constrain 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 memory strength 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 studied two and four times during study.

### Figure 7

Data from a recognition memory task (black) and predictions from the full GCF (red) and the full GC (green).



*Note.* Conditions 'old 1', 'old 2', and 'old 4' refer to old words studied one time, two times, and four times respectively. 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 PMT and MT quantiles of incorrect responses in a given condition if each subject made at least 10 errors (condition old words studied once fulfilled this requirement). 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 GC and GCF 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 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. Methodological details regarding the experiment and the modeling of the data are provided in Appendix E.

## Results

### *Behavior and EMG*

Twenty-four subjects completed the experiment. 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 (very low, low, medium, high) 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 memory strength 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*

The fit procedure was identical to that used in the previous experiments. We modeled 'pseudoword' and 'word' responses (with negative evidence favoring 'pseudoword' responses and positive evidence favoring 'word' responses). 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*).

GCF was associated with lower AIC and BIC statistics compared to GC 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 GCF compared to GC, and the raw GCF was associated with a lower AIC (BIC) compared to the full GC for 14 (15) subjects. The pattern of model selection results is thus similar to that observed in the previous experiments, and provides strong evidence for GCF.

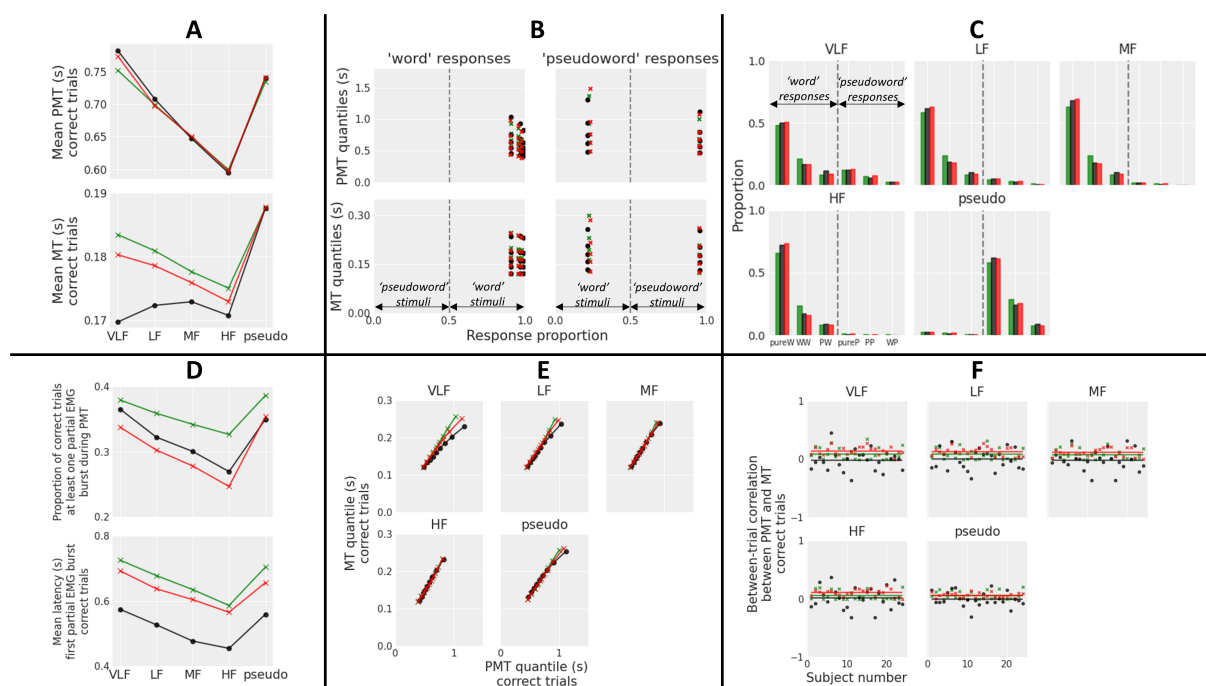
Best-fitting parameters for the full models are shown in Table 1 (main parameters) and Table 2 (between-trial variability parameters). Best-fitting parameters for the raw models are shown in Table C1 in Appendix C. GCF captures the data with a higher level of leakage  $\lambda$  compared to the previous experiments (implying reduced smoothing and filter-related delays; for an illustration of model trajectories, see Figure 4D). Consistent with previous work (Donkin et al., 2009; Gomez and Perea, 2014), GCF and GC 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

GC compared to GCF, 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 GCF and GC predict an increase in predicted mean MT as word frequency decreases, and fail to capture the observed inverted U-shaped pattern. GC systematically underestimates the .9 quantile of PMT distributions for correct responses, and overestimates the proportion of correct WW and PP trials. GCF 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 GCF (red) and the full GC (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 PMT and MT quantiles of incorrect responses in a given condition if each subject made at least 10 errors (condition very low frequency words fulfilled this requirement). Panel C shows the proportion of each of the six trial types (pureW, WW, PW, pureP, PP, WP) for each condition averaged across subjects.

**Comparison between the neural drive to muscle fibers predicted by GCF  
and full-wave rectified EMG signals**

As a final evaluation of GCF, we compared the predicted neural drive to muscle fibers (variables  $z_L(t)$  or  $z_R(t)$ ; see Equation 9) with full-wave rectified EMG signals. As mentioned in the general introduction, the full-wave rectified EMG signal can be interpreted as a noisy approximation of the neural drive to the area of muscle over which the electrodes are placed, and should thus scale with the predicted neural drive computed from best-fitting parameters. We restricted this analysis to EMG data from Experiments 1 and 2, in which the effects of experimental manipulations on MT were the largest and well-accounted for by the model.

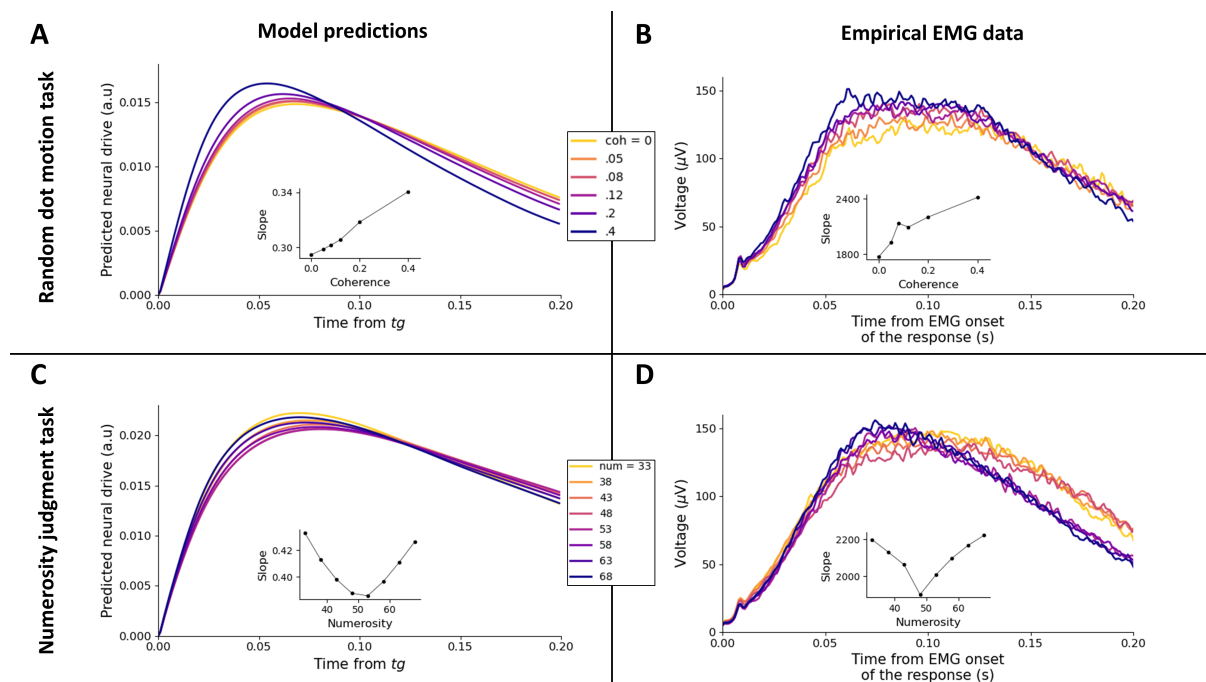
Figures 9A and 9C show the predicted neural drive to muscle fibers associated with the correct response averaged across correct trials and subjects for the random dot motion task and the numerosity judgment task respectively. Model trajectories are time-locked on  $t_g$  (latency at which  $z_L(t)$  or  $z_R(t)$  becomes positive for the last time before reaching the response threshold; see Figure 1). For each task, subject and condition, 1,000 correct trials were simulated using best-fitting parameters from the full GCF. In each simulated trial, the neural drive was assumed to decay at an arbitrary linear rate of -0.15 units/second after hitting the response threshold. Trajectories were then time-locked on  $t_g$  and averaged. For each task, the rising slope of the predicted neural drive decreases as difficulty increases, reflecting dynamics of the underlying motor preparation signal.

Figures 9B and 9D show full-wave rectified EMG signals from agonist muscles associated with the correct response averaged across correct trials and subjects from the random dot motion task and the numerosity judgment task respectively. Signals are time-locked to the EMG onset of the response. Consistent with model predictions, the rising slope of EMG signals decreases as difficulty increases. A linear contrast computed on the rising slope (estimated by linear regression in the 0 to 50 ms window) in the random dot motion task was highly significant ( $t(27) = 4.89, p < .001$ ), and so was the quadratic contrast in the numerosity judgment task ( $t(23) = 3.66, p < .001$ ). This analysis provides

additional evidence for GCF.

### Figure 9

*Neural drive to muscle fibers associated with the correct response predicted by GCF and corresponding full-wave rectified EMG signals averaged across correct trials and subjects for each condition of the random dot motion task (Experiment 1) and the numerosity judgment task (Experiment 2)*



*Note.* Panels A and B show model predictions and empirical EMG data respectively in the random dot motion task. Panels C and D show model predictions and empirical EMG data respectively in the numerosity judgment task. Insets represent the rising slope of signals in each condition, estimated by linear regression in the 0 to 50 ms window. a.u.: arbitrary units.

## 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. The proposed GCF assumes a continuous flow of the evidence accumulation decision variable to agonist muscles. The model further incorporates a smoothing mechanism at the motor preparation level, and a gate which regulates the flow of information from motor preparation to muscle fibers. This architecture offers substantial flexibility to motor control by allowing for real-time adjustments of motor commands based on incoming evidence,

while simultaneously shielding the system against unwanted behaviors and preventing unnecessary muscular activity. The smoothing mechanism at the motor preparation level may also reflect an attempt to recover the decision variable from noise that can corrupt it during the transmission process, and can be seen as an approximation of a Kalman-Bucy filter.

We tested GCF 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. GCF 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, 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, (vi) the neural drive to muscle fibers, and (vii) the variation of all of the above aspects of the data across difficulty conditions. Overall, GCF provided a good fit to data used to constrain parameter estimation (i, ii, and iv). 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, and Brown, 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 GCF misfit. Although urgency signals remain controversial when considering behavioral data alone (Evans, Hawkins, and Brown, 2020; Glickman and Usher, 2019; Glickman et al., 2022; 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), 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 GCF 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 GCF showed a good fit to data used to constrain 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 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.

Besides the mean latency of partial EMG bursts, GCF 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 Appendix D). This interaction is problematic because GCF variants used in Experiments 3 and 4 do not incorporate representational assumptions that specify how drift rate distributions arise from the stimuli. In addition, although the linear relationship between motion coherence and drift rate provided a good fit to data from Experiment 1 (see also Palmer et al., 2005; Ratcliff and McKoon, 2008), a more complex representational assumption has recently been proposed for the random dot motion task (Smith and Lilburn, 2020). Consequently, discrepancies between data and model predictions may stem from a misspecification of drift rate distributions.

More generally, these findings highlight 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 et al., 2009; Gomez and Perea, 2014), and between-trial variability in drift rate (Tillman et al., 2017). Within the framework of GCF, 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 GCF. 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 GCF to shed light on this issue.

Apart from the unexpected word frequency effect on mean MT, EMG findings were remarkably consistent across experiments, suggesting that GCF generalizes across

cognitive domains. Both mean PMT and mean MT increased as choice difficulty increased. The increase of mean MT as choice difficulty increased was caused by a decrease in the rising slope of the neural drive, as suggested by our analysis of full-wave rectified EMG signals. This important aspect of the data was nicely captured by GCF, because the neural drive predicted by the model reflects evidence-dependent dynamics of the underlying motor preparation signal. Partial bursts were also 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 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 burst was located in the opposite EMG channel. Within the framework of GCF, this finding is explained by the same mechanism that captures the relative proportion of correct and incorrect responses. Putting aside between-trial variability in model components, errors are produced by noise in the evidence accumulated at each time step. Although part of this noise is flattened 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 GC**

As predicted, GCF captured the data with a relatively low level of leakage, indicating that the evidence accumulation variable is smoothed at the motor preparation level. This finding suggests that the smoothing mechanism adds to a GC description of the data. Model selection statistics (AIC and BIC) further showed that the additional complexity of GCF was justified in light of the (large) improvement in fit quality. Both AIC and BIC statistic favored GCF over GC for 90/90 participants (raw models) and 89/90 (full models). These findings provide decisive evidence for GCF. Interestingly, the difference in model selection statistics between raw and full models was much larger for GC than for GCF, suggesting that between-trial variability in GC components have a major impact on the fit quality of the model, contrary to GCF. In fact, the AIC (BIC) statistic favored the raw GCF over the full GC for 67(70)/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. 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, and Wagenmakers, 2020). Consequently, we consider our findings regarding between-trial variability as additional evidence for GCF.

In its raw form, GC grossly overestimated the proportion of trials containing at least one partial EMG burst, especially when the first partial burst was located in the same EMG channel as the response. The full GC provided a better account of these proportions by using a very small response bound (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). However, both processing schemes resulted in predicted MTs that were too fast compared to observed MTs. The model compensated this problem by increasing the contribution of residual motor latencies to predicted MTs (parameters  $Tr$  and  $sTr$ ), but this compensation had two negative consequences. First, the model was not able to capture large effects of choice difficulty on mean MT, such as those observed in Experiment 1. Second, the model was not able to 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 decision model (Ratcliff and Rouder, 1998).

### **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 RT 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), their development and integration through GCF constitute a major theoretical advance in the field, as they offer a mechanistic explanation to the interplay between decision and motor processes.

The diffusion decision process has been the object of several neurocomputational characterizations, and has been hypothesized to arise from recurrent loops within neural networks (Smith and McKenzie, 2011; Wang, 2002; Wong and Wang, 2006). Interestingly, recurrent neural networks have also been used to approximate Kalman-Bucy filters (Denève et al., 2007), paving the way for a joint characterization of decision and motor preparation processes at the neural circuit level. At the systems level, we believe that future empirical tests of GCF would benefit from a combination of EEG and EMG recordings. Although the application of a low-pass filter during EEG signal processing precludes a precise test of the Kalman-Bucy filter hypothesis at the motor preparation level, averaged model trajectories at decision-making and motor preparation levels computed from best-fitting parameters could be compared to CPP and effector-selective motor preparation EEG activities respectively. Some parameters of the model could also be constrained to match corresponding electrical signatures (Kelly et al., 2021).

Additional constraints to GCF could also arise from a more detailed analysis of partial EMG bursts at the motor execution level. Some trials contain more than one partial burst during PMT, and these additional bursts could be considered in the modeling. In particular, the co-occurrence 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 potentially artifactual electrical activities. In addition, partial EMG bursts might be followed by a refractory period. Consequently, the analysis and modeling of trials with multiple partial bursts represent an important challenge for future work.

Beyond basic mechanisms that drive the time-course of decision-making, motor preparation, and motor execution processes, we believe that future developments of GCF

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 GCF, 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 scheme 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 scrutiny 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 GCF 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 building 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 GCF 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>
- 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>
- 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>
- Calderón, J. C., Bolaños, P., & Caputo, C. (2014). The excitation–contraction coupling mechanism in skeletal muscle. *Biophysical Reviews, 6*(1), 133–160. <https://doi.org/10.1007/s12551-013-0135-x>
- Cavanagh, P. R., & Komi, P. V. (1979). Electromechanical delay in human skeletal muscle under concentric and eccentric contractions. *European Journal of Applied Phys-*

- iology 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. (2007). Cortical mechanisms of action selection: The affordance competition hypothesis. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 362(1485), 1585–1599. <https://doi.org/10.1098/rstb.2007.2054>
- 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>
- Coles, M. G. H., Gratton, G., Bashore, T. R., Eriksen, C. W., & Donchin, E. (1985). A psychophysiological investigation of the continuous flow model of human information processing [Place: US Publisher: American Psychological Association]. *Journal of Experimental Psychology: Human Perception and Performance*, 11(5), 529–553. <https://doi.org/10.1037/0096-1523.11.5.529>
- 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](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>
- Denève, S., Duhamel, J.-R., & Pouget, A. (2007). Optimal sensorimotor integration in recurrent cortical networks: A neural implementation of Kalman filters. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *27*(21), 5744–5756. <https://doi.org/10.1523/JNEUROSCI.3985-06.2007>
- Dideriksen, J. L., & Farina, D. (2019). Amplitude cancellation influences the association between frequency components in the neural drive to muscle and the rectified EMG signal [Publisher: Public Library of Science]. *PLOS Computational Biology*, *15*(5), e1006985. <https://doi.org/10.1371/journal.pcbi.1006985>
- 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](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>
- 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>

- Eriksen, C. W., & Schultz, D. W. (1979). Information processing in visual search: A continuous flow conception and experimental results. *Perception & Psychophysics*, *25*(4), 249–263. <https://doi.org/10.3758/bf03198804>
- 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 Experimental 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.
- Farina, D., Holobar, A., Merletti, R., & Enoka, R. M. (2010). Decoding the neural drive to muscles from the surface electromyogram. *Clinical Neurophysiology: Official Journal of the International Federation of Clinical Neurophysiology*, *121*(10), 1616–1623. <https://doi.org/10.1016/j.clinph.2009.10.040>
- 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>
- Frank, M. J. (2011). Computational models of motivated action selection in corticostriatal circuits. *Current Opinion in Neurobiology, 21*(3), 381–386. <https://doi.org/10.1016/j.conb.2011.02.013>
- 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>
- Glickman, M., Moran, R., & Usher, M. (2022). Evidence integration and decision confidence are modulated by stimulus consistency. *Nature Human Behaviour, 6*(7), 988–999. <https://doi.org/10.1038/s41562-022-01318-6>
- Glickman, M., & Usher, M. (2019). Integration to boundary in decisions between numerical sequences. *Cognition, 193*, 104022. <https://doi.org/10.1016/j.cognition.2019.104022>
- 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 [Pub-

- lisher: 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>
- Hikosaka, O. (2007). GABAergic output of the basal ganglia. *Progress in Brain Research*, *160*, 209–226. [https://doi.org/10.1016/S0079-6123\(06\)60012-5](https://doi.org/10.1016/S0079-6123(06)60012-5)
- 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>
- Kalman, R. E., & Bucy, R. S. (1961). New Results in Linear Filtering and Prediction Theory. *Journal of Basic Engineering*, *83*(1), 95–108. <https://doi.org/10.1115/1.3658902>

- 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>
- Logan, G. D., & Cowan, W. B. (1984). On the ability to inhibit thought and action: A theory of an act of control [Place: US Publisher: American Psychological Association]. *Psychological Review*, *91*(3), 295–327. <https://doi.org/10.1037/0033-295X.91.3.295>
- Marr, D. (1982). *Vision: A computational investigation into the human representation and processing of visual information*. W.H. Freeman.

- McClelland, J. L. (1979). On the time relations of mental processes: An examination of systems of processes in cascade [Place: US Publisher: American Psychological Association]. *Psychological Review*, *86*(4), 287–330. <https://doi.org/10.1037/0033-295X.86.4.287>
- 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>
- Mink, J. W. (1996). The basal ganglia: Focused selection and inhibition of competing motor programs. *Progress in Neurobiology*, *50*(4), 381–425. [https://doi.org/10.1016/s0301-0082\(96\)00042-1](https://doi.org/10.1016/s0301-0082(96)00042-1)
- Murphy, P. R., Boonstra, E., & Nieuwenhuis, S. (2016). Global gain modulation generates time-dependent urgency during perceptual choice in humans. *Nature Communications*, *7*, 13526. <https://doi.org/10.1038/ncomms13526>
- Nakayama, K., Moher, J., & Song, J.-H. (2023). Rethinking Vision and Action [\_eprint: <https://doi.org/10.1146/annurev-psych-021422-043229>]. *Annual Review of Psychology*, *74*(1), 59–86. <https://doi.org/10.1146/annurev-psych-021422-043229>
- 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>
- Øksendal, B. (2003). *Stochastic Differential Equations*. Springer. <https://doi.org/10.1007/978-3-642-14394-6>
- Palmer, J., Huk, A. C., & Shadlen, M. N. (2005). The effect of stimulus strength on the speed and accuracy of a perceptual decision. *Journal of Vision*, *5*(5), 376–404. <https://doi.org/10.1167/5.5.1>
- 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](https://doi.org/10.1016/s1388-2457(99)00141-8)
- 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. (2008). The Diffusion Decision Model: Theory and Data for Two-Choice Decision Tasks. *Neural computation*, 20(4), 873–922. <https://doi.org/10.1162/neco.2008.12-06-420>
- Ratcliff, R., & 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>
- Resulaj, A., Kiani, R., Wolpert, D. M., & Shadlen, M. N. (2009). Changes of mind in decision-making. *Nature*, 461(7261), 263–266. <https://doi.org/10.1038/nature08275>

- 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 Presaccadic 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., 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>
- Shadlen, M. N., & Kiani, R. (2013). Decision making as a window on cognition. *Neuron*, *80*(3), 791–806. <https://doi.org/10.1016/j.neuron.2013.10.047>
- Shadlen, M. N., & Newsome, W. T. (1994). Noise, neural codes and cortical organization. *Current Opinion in Neurobiology*, *4*(4), 569–579. [https://doi.org/10.1016/0959-4388\(94\)90059-0](https://doi.org/10.1016/0959-4388(94)90059-0)
- 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>
- Smith, P. L., & Lilburn, S. D. (2020). Vision for the blind: Visual psychophysics and blinded inference for decision models. *Psychonomic Bulletin & Review*. <https://doi.org/10.3758/s13423-020-01742-7>

- Smith, P. L., & McKenzie, C. R. L. (2011). Diffusive Information Accumulation by Minimal Recurrent Neural Models of Decision Making. *Neural Computation*, *23*(8), 2000–2031. [https://doi.org/10.1162/NECO\\_a\\_00150](https://doi.org/10.1162/NECO_a_00150)
- 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](https://doi.org/10.1016/0001-6918(69)90055-9)
- Stone, C., Mattingley, J. B., & Rangelov, D. (2022). On second thoughts: Changes of mind in decision-making. *Trends in Cognitive Sciences*, *26*(5), 419–431. <https://doi.org/10.1016/j.tics.2022.02.004>
- 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>
- Wang, X.-J. (2002). Probabilistic decision making by slow reverberation in cortical circuits. *Neuron*, *36*(5), 955–968. [https://doi.org/10.1016/s0896-6273\(02\)01092-9](https://doi.org/10.1016/s0896-6273(02)01092-9)
- 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>

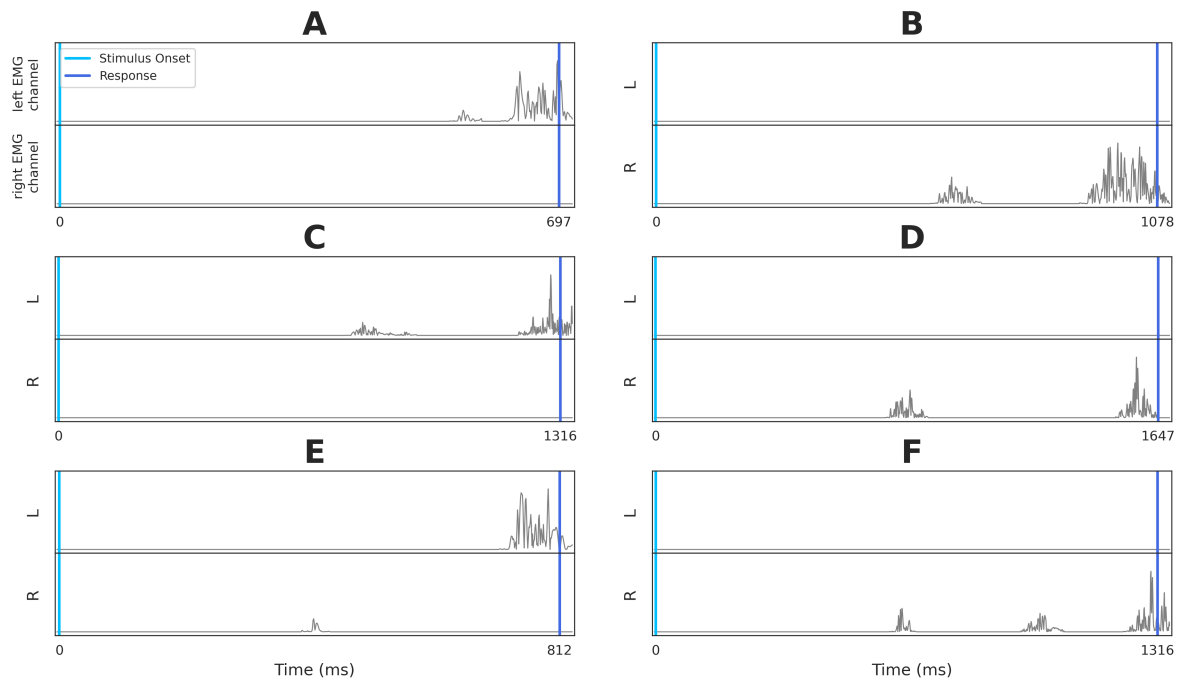
- Wixted, J. T. (2007). Dual-process theory and signal-detection theory of recognition memory. *Psychological Review*, *114*(1), 152–176. <https://doi.org/10.1037/0033-295X.114.1.152>
- Wong, K.-F., & Wang, X.-J. (2006). A recurrent network mechanism of time integration in perceptual decisions. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *26*(4), 1314–1328. <https://doi.org/10.1523/JNEUROSCI.3733-05.2006>

## Appendix A

### Partial EMG bursts

**Figure A1**

*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.

## Appendix B

### Mathematical details

#### Derivation of the Kalman-Bucy filter at the motor preparation level (Equation 6)

Definitions of the decision variable  $x(t)$  and its noisy variant  $\tilde{x}(t)$  received by the motor preparation system are provided in the main text by Equations 1 and 5 respectively. In order to have a tractable mathematical representation of the problem, it is better to write it down in terms of differential equations. This is already the case for  $x(t)$ . However, we cannot differentiate  $\tilde{x}(t)$  as it stands, because we cannot differentiate a white noise. We thus introduce a new process  $q(t)$ , defined as  $q(t) = \int_0^t \tilde{x}(s)ds$ . We have:

$$dq(t) = x(t)dt + \xi dV(t), \quad q(0) = x_0, \quad (\text{B1})$$

where  $V(t)$  is a Brownian motion, independent of  $W(t)$ . Note that processes  $q(t)$  and  $\tilde{x}(t)$  contain the exact same information.

The Kalman-Bucy filtered motor preparation process  $\tilde{y}(t)$  satisfies the following differential equation (Øksendal, 2003):

$$d\tilde{y}(t) = \left( v - \frac{S(t)}{\xi^2} \tilde{y}(t) \right) dt + \frac{S(t)}{\xi^2} dq_t, \quad \tilde{y}(0) = x_0, \quad (\text{B2})$$

where  $S(t) = \mathbb{E}[(x(t) - \tilde{y}(t))^2]$ , and satisfies the Riccati equation:

$$\frac{dS}{dt} = -\frac{S^2}{\xi^2} + \sigma^2, \quad S(0) = 0. \quad (\text{B3})$$

The solution to the Riccati equation is:

$$\begin{aligned} S(t) &= \frac{-\xi\sigma + \xi\sigma e^{\frac{2\sigma}{\xi}t}}{1 + e^{\frac{2\sigma}{\xi}t}} \\ &= \xi\sigma \tanh\left(\frac{\sigma}{\xi}t\right). \end{aligned} \quad (\text{B4})$$

Substituting  $S(t)$  in Equation B2, we obtain:

$$d\tilde{y}(t) = \frac{\sigma}{\xi} \tanh\left(\frac{\sigma}{\xi}t\right) (dq(t) - \tilde{y}(t)dt) + vdt, \quad \tilde{y}(0) = x_0. \quad (\text{B5})$$

Finally, from the definition of  $q(t)$ , we have  $dq(t) = \tilde{x}(t)dt$ , which leads to:

$$d\tilde{y}(t) = \frac{\sigma}{\xi} \tanh\left(\frac{\sigma}{\xi}t\right) (\tilde{x}(t) - \tilde{y}(t))dt + vdt, \quad \tilde{y}(0) = x_0. \quad (\text{B6})$$

### Relationship between Equation 3 and Equation 8

In the main text, we noted that Equation 3 and Equation 8 differ with respect to the input to motor preparation. This difference can be further characterized by solving each equation, and replacing parameter  $\lambda'$  by  $\lambda$  in Equation 8:

$$\begin{aligned} y(t) &= \lambda \int_0^t e^{\lambda(s-t)} x(s) ds \\ \tilde{y}(t) &= \lambda \int_0^t e^{\lambda(s-t)} \tilde{x}(s) ds. \end{aligned} \tag{B7}$$

Since  $\tilde{x}(s)ds = dq(s) = x(s)ds + \xi dV(s)$ , we obtain:

$$\begin{aligned} \tilde{y}(t) &= \lambda \int_0^t e^{\lambda(s-t)} (x(s)ds + \xi dV(s)) \\ &= \lambda \int_0^t e^{\lambda(s-t)} x(s)ds + \lambda\xi \int_0^t e^{\lambda(s-t)} dV(s) \\ &= y(t) + \lambda\xi \int_0^t e^{\lambda(s-t)} dV(s). \end{aligned} \tag{B8}$$

Using Itô isometry, we see that  $\tilde{y}(t)$  is equal to  $y(t)$  plus a Gaussian noise with mean 0 and variance  $\varphi = \lambda\xi^2 \frac{1-e^{-2\lambda t}}{2}$ . This Gaussian noise is independent from  $y$  (because  $V$  and  $W$  are independent). Note that  $\varphi < \frac{\lambda\xi^2}{2}$ , and  $\tilde{y}(t)$  converges towards  $y(t)$  as the amplitude of transmission noise  $\xi$  approaches 0.

## Appendix C

### Fits of a GCF variant that uses variable $\tilde{y}(t)$ to model motor preparation

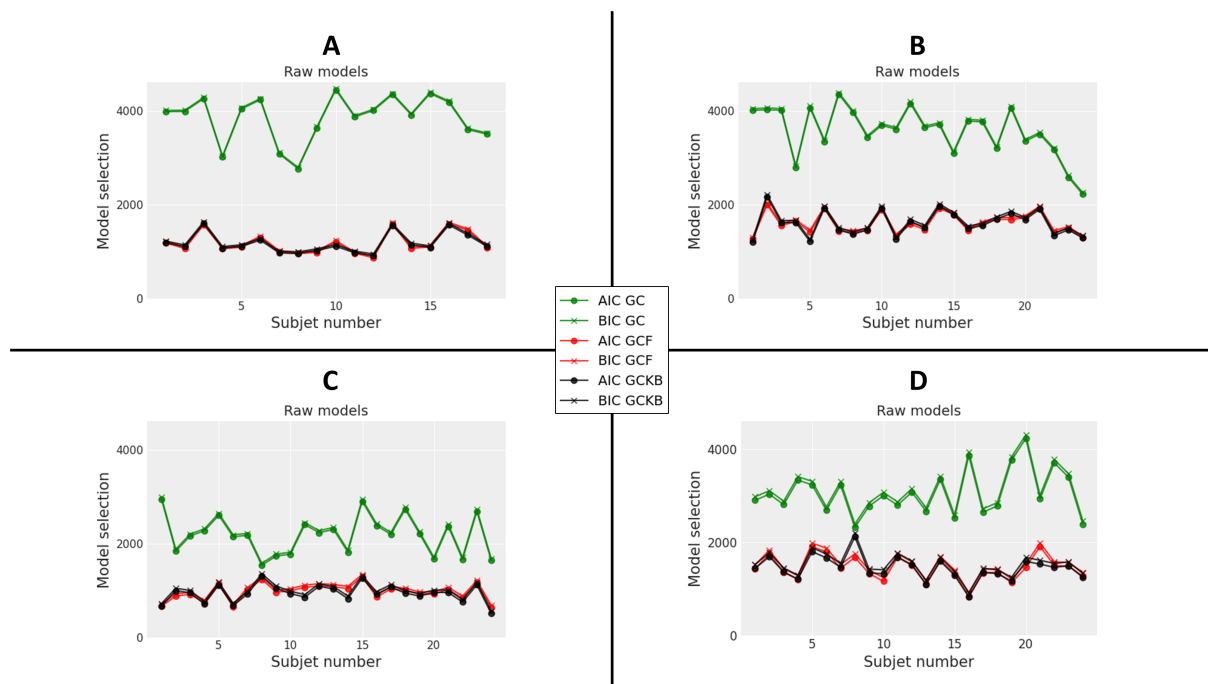
We compared the fit performance of GCF with a model variant that uses  $\tilde{y}(t)$  computed using Equation 7. We refer to this model variant as GCKB, in homage to Rudolf E. Kalman and Richard S. Bucy. Based on our mathematical analysis, we expected a comparable fit performance between the two models. GCKB was fit to behavioral and EMG data from the four choice tasks using a methodology identical to that described in the main text. To avoid complications arising from between-trial variability in processing components, we restricted this analysis to raw models. GCKB has one more free parameter than GCF (the amplitude of transmission noise  $\xi$ ).

Figure C1 displays model comparison statistics (AIC and BIC) for each task. The plot shows a considerable degree of overlap between GCF and GCKB, diagnostic of a similar performance. The AIC (BIC) favors GCKB over GCF for 7 (7) subjects for the RDK task, 12 (10) subjects for the numerosity judgment task, 14 (14) subjects for the recognition memory task, and 11 (10) subjects for the lexical decision task. Each of these differences fails to reach statistical significance (two-sided binomial test).

Best-fitting GCKB parameters averaged across subjects are reported in Table C1. It is worth noting that the prior  $\lambda'$  on the ratio between diffusion noise  $\sigma$  and transmission noise  $\xi$  is significantly smaller than the ratio computed from an ideal observer perspective, resulting in a stronger smoothing of the corrupted decision variable  $\tilde{x}$  than actually needed. This phenomenon could be interpreted as a bias toward response accuracy relative to response speed (Verdonck et al., 2020). This analysis must be taken with caution, as we suspect tradeoffs between GCKB parameters (suggested by the very short residual latency  $Te$  added to predicted PMT in Experiment 1, or the very small prior  $\lambda'$  in Experiment 3). A comprehensive assessment of these tradeoffs is beyond the scope of the present work.

**Figure C1**

*Model selection statistics for Experiments 1-4, including GCKB and focusing on raw models.*



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



Table C1

Parameters from the raw models (GC, GCF, and GCKB) averaged across subjects for Experiments 1-4.

Exp	Model	$v_1$	$v_2$	$v_3$	$v_4$	$v_5$	$k$	$dc$	$\lambda$	$\lambda'$	$\xi$	$g$	$r$	$Te_1$	$Te_2$	$Te_3$	$Te_4$	$Te_5$	$Tr$	$x_0$		
1	GC						1.147				0.085	0.045	0.271							0.079		
1	GCF						0.694		12.519		0.055	0.024	0.193							0.066		
1	GCKB						0.620			7.085	0.0013	0.045	0.019	0.084						0.050		
2	GC	0.020					-0.026				0.058	0.049	0.293							0.089	0.004	
2	GCF	0.015					-0.016	35.937			0.039	0.032	0.267							0.084	0.002	
2	GCKB	0.013					-0.014			20.408	0.0010	0.031	0.028	0.204						0.074	0.002	
3	GC	0.363	0.588	0.628	-0.594						0.259	0.070	0.188							0.067	-0.001	
3	GCF	0.083	0.155	0.223	-0.216				35.690		0.043	0.029	0.367							0.080	0.001	
3	GCKB	0.042	0.105	0.154	-0.148					2.604	0.0028	0.011	0.012	0.225						0.046	0.001	
4	GC	0.438	0.472	0.534	0.583	-0.445					0.225	0.066	0.171	0.167	0.166	0.159	0.180	0.073		0.073	-0.006	
4	GCF	0.111	0.176	0.233	0.282	-0.180			52.543		0.043	0.034	0.337	0.330	0.323	0.305	0.369	0.084		0.084	-0.006	
4	GCKB	0.102	0.169	0.209	0.246	-0.167				21.503	0.0017	0.034	0.028	0.266	0.265	0.250	0.230	0.311	0.072		0.072	-0.005

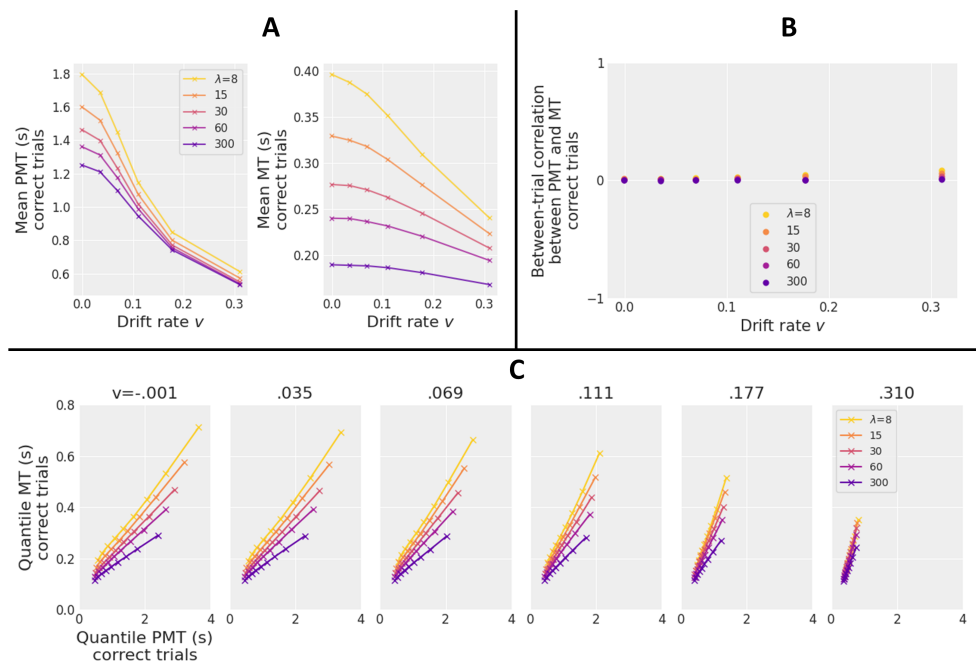
Note. Experiment 3 (recognition memory): parameters  $v_1$  to  $v_4$  correspond to drift rates for conditions old words studied one time, old words studied two times, old words studied four times, and new words respectively. Experiment 4 (lexical knowledge): parameters  $v_1$  to  $v_5$  and  $Te_1$  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.

## Appendix D

## Additional predictions from GCF

Figure D1

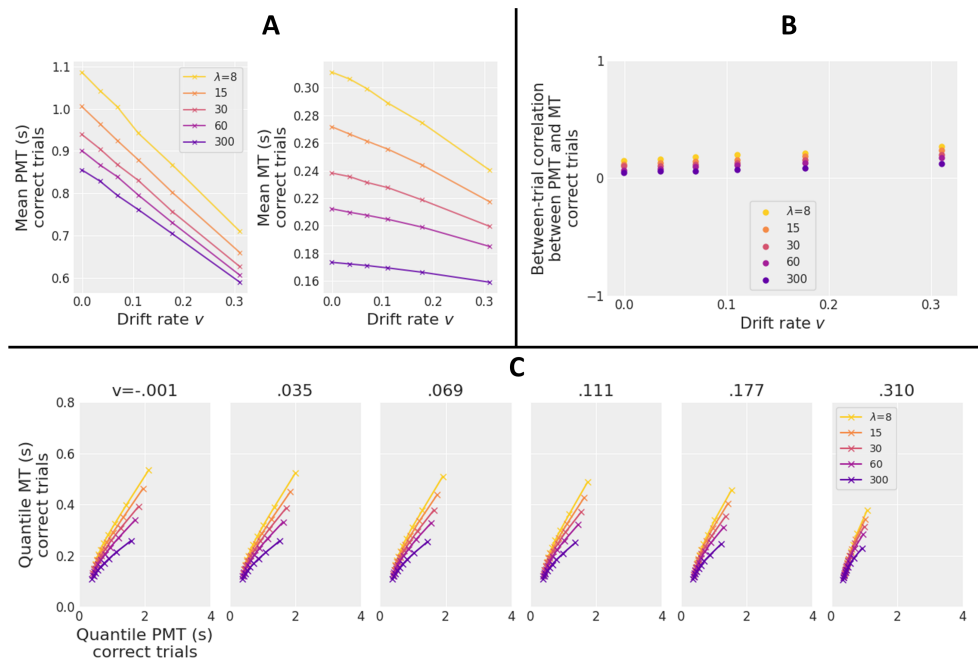
Additional predictions from GCF with varying levels of leak  $\lambda$  and drift rate  $v$ .



*Note.* Apart from the leak parameter, simulations used best-fitting GC 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.

**Figure D2**

*Additional predictions from GCF with varying levels of leak  $\lambda$  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 GC 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.

## Appendix E

### Methods for Experiments 2-4

#### Experiment 2

##### *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 electrophysiological 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). It was not preregistered.

##### *Apparatus*

The experiment took place in a dimly lit room. Subjects sat on a comfortable chair at a distance of 75 cm from a  $34.7 \times 19.5$  cm LCD monitor (resolution:  $1920 \times 1080$ ; 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  $10 \times 10$  virtual grid. Each dot was  $0.24^\circ$  in diameter. The horizontal and vertical separation between two adjacent dots (from center to center) was  $1.15^\circ$ .

### ***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 practice 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 a high level of noise were discarded from analyses (2.7% of trials on average; range 0-11.7%).

### ***Models and fit procedure***

The data were grouped into 8 conditions 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 GC and GCF is a linear function of  $N$  and the criterion (50):

$$v = dc + v_1(N - 50). \quad (\text{E1})$$

Parameter  $v_1$  accounts for interindividual differences in discrimination performance. Parameter  $dc$  (drift criterion) accounts for interindividual differences in the representation of the criterion. Without this parameter (i.e., assuming  $dc = 0$ ), the

representation of the criterion would correspond exactly to 50 dots for every subject, which is not realistic. Between-trial variability in drift rate is assumed to be normally distributed with mean  $v$  and standard deviation  $sv$  defined as follows:

$$sv = \eta_0 + \sigma_1 \sqrt{N^2 + 50^2}. \quad (\text{E2})$$

All parameters were fixed across numerosity conditions. We treated the starting point  $x_0$  of the evidence accumulation process as a free parameter, resulting in seven (12) free parameters for the raw (full) GC and eight (14) free parameters for the raw (full) GCF<sup>8</sup>. All parameters (except  $dc$  and  $x_0$ ) were constrained to be  $\geq 0$ . Between-trial variability parameters  $sx_0$ ,  $sTe$ ,  $sTr$ , and  $s\lambda$  were constrained to not exceed 180% of  $g$ ,  $Te$ ,  $Tr$ , and  $\lambda$  respectively. Parameter  $sx_0$  was further constrained to not exceed  $2 * (g - |x_0|)$ . Model selection statistics (AIC and BIC) were computed using Equations 11 and 12. By a two-sided binomial test, if 18 out of 24 subjects support one model over the other, then the result is significant.

### Experiment 3

#### *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). It was not preregistered.

#### *Apparatus*

The apparatus was identical to Experiment 2.

---

<sup>8</sup> 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 found similar fits for this GC and GCF variant.

### ***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 occurrences 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  $0.76^\circ$ .

### ***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 25 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 studied one time, seven were studied two times, and seven were studied four times. Words were presented in a random order, at a pace of 1 s. The test phase occurred 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 a high level of noise were discarded from analyses (12.4% of trials on average; range 1.6-25%).

### ***Models and fit procedure***

Drift rate  $v$  and between-trial variability in drift rate  $sv$  were the only parameters free to vary between conditions. Similar to Experiment 2, we treated the starting point  $x_0$  of the decision-making process as a free parameter, resulting in nine (16) parameters for the raw (full) GC, and 10 (18) parameters for the raw (full) GCF. All parameters (except drift rates and  $x_0$ ) were constrained to be  $\geq 0$ . Between-trial variability parameters  $sx_0$ ,  $sTe$ ,  $sTr$ , and  $s\lambda$  were constrained to not exceed 180% of  $g$ ,  $Te$ ,  $Tr$ , and  $\lambda$  respectively. Parameter  $sx_0$  was further constrained to not exceed  $2 * (g - |x_0|)$ .

## **Experiment 4**

### ***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). It was not preregistered.

### ***Apparatus***

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 occurrences 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 multilingual 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 Table E1). 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 practice 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 a high level of noise were discarded from analyses (10.4% of trials on average; range 0.06-27.3%).

***Models and 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  $x_0$  of the evidence accumulation process as a free parameter, resulting in 14 (22) parameters for the raw (full) GC, and 15 (24) parameters for the raw (full) GCF. All parameters (except drift rates and  $x_0$ ) were constrained to be  $\geq 0$ . Between-trial variability parameters  $sx_0$ ,  $sTe$ ,  $sTr$ , and  $s\lambda$  were constrained to not exceed 180% of  $g$ ,  $Te$ ,  $Tr$ , and  $\lambda$  respectively. Parameter  $sx_0$  was further constrained to not exceed  $2 * (g - |x_0|)$ .

**Table E1**

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

	Occurrences 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