Assessing Model-Based Inferences in Decision Making With Single-Trial Response Time Decomposition

Gabriel Weindel[1, 2], Royce Anders[1, 3], F.-Xavier Alario[1, 4], and Boris Burle[2]

[1] LPC UMR 7290, Aix Marseille University, CNRS
[2] LNC UMR 7291, Aix Marseille University, CNRS
[3] EMC EA 3082, University Lumière Lyon 2
[4] Department of Neurological Surgery, University of Pittsburgh

The latent psychological mechanisms involved in decision-making are often studied with quantitative models based on evidence accumulation processes. The most prolific example is arguably the drift-diffusion model (DDM). This framework has frequently shown good to very good quantitative fits, which has prompted its wide endorsement. However, fit quality alone does not establish the validity of a model's interpretation. Here, we formally assess the model's validity with a novel cross-validation approach based on the recording of muscular activities, which directly relate to the standard interpretation of various model parameters. Specifically, we recorded electromyographic activity along with response times (RTs), and used it to decompose every RT into 2 components: a premotor time (PMT) and motor time (MT). The latter interval, MT, can be directly linked to motor processes and hence to the nondecision parameter of DDM. In two canonical perceptual decision tasks, we manipulated stimulus strength, speed–accuracy trade-off, and response force and quantified their effects on PMT, MT, and RT. All 3 factors consistently affected MT. The DDM parameter for nondecision processes recovered the MT effects in most situations, with the exception of the fastest responses. The extent of the good fits and the scope of the mis-estimations that we observed allow drawing new limits of the interpretability of model parameters.

Keywords: electromyography, model validity, reaction time fractioning, sequential sampling models

Understanding how decisions are made is an important endeavor at the crossroads of many research programs in cognitive psychology and neuroscience. According to one prevalent theoretical framework, decisions are driven by a cognitive mechanism that sequentially samples goal-relevant information from the environment. Contextual evidence is accumulated in favor of the different alternatives until one of them reaches a threshold that triggers the execution of the corresponding response (Ratcliff et al., 2016; Stone, 1960). Typically, this theoretical framework is implemented quantitatively as a continuous diffusion process. These models are applied empirically to fit response accuracy and latency measures along different alternative-choice tasks and experimental manipulations. Such quantitative modeling procedures yield data-derived process parameters that are used to infer the cognitive dynamics underlying decision-making, which are otherwise not directly observable.

There are multiple variants of these quantitative models, which specify the dynamics of evidence accumulation differently (e.g., Anders et al., 2016; Brown & Heathcote, 2008; Heathcote & Love, 2012; Palmer et al., 2005; Ratcliff, 1978; Ratcliff & McKoon, 2008; Tillman et al., 2020; Usher & McClelland, 2001, among many others). All of these variants share the hypothesis that each response time (RT) is the sum of three major terms (Equation 1; see Luce, 1986): the time needed for stimulus encoding (T_e), the time needed for the accumulated evidence to reach a threshold once accumulation has started ("decision time": DT), and the time needed for response execution (T_r). If the decision-relevant information encoding reaches the same level of completeness on each trial, T_e does not affect the decision-making process of evidence accumulation (DT). It is also usually assumed that the way the response is executed does not depend on how the decision was reached. In other words, T_r is assumed to be independent from DT. For these reasons, presumably, the terms T_e and T_r are most often pooled into a single parameter T_{er} (Equation 2).

$$RT = T_e + DT + T_r \qquad (1)$$

$$T_{er} = T_e + T_r \qquad (2)$$

In the minimal form of the model, the decision time DT is derived from two parameters: the accumulation rate (i.e., amount of evidence accumulated per time unit) and the decision threshold (i.e., the amount of accumulated evidence needed for a response to be triggered; Figure 1). The best fitting parameters are directly estimated from the measured RT and the corresponding response accuracy.

These models provide a framework to determine the cognitive locus of specific experimental observations (e.g., associative vs. categorical priming: Voss et al., 2013; masked vs. unmasked priming: Gomez et al., 2013; among many others). They have also been used to assess populations differences. For example, general slowing associated with aging has been linked to a higher evidence threshold and a longer T_{er}, rather than lower evidence accumulation rate (Ratcliff et al., 2001). The same rationale for group analyses is increasingly used in psychopathology research to characterize information processing alterations in a variety of pathologies such as anxiety (linked to a higher rate of accumulation for threatening stimuli: White et al., 2010), depression (Lawlor et al., 2020; Pe et al., 2013), schizophrenia (Moustafa et al., 2015), Parkinson's disease (Herz et al., 2016), language impairments (Anders et al., 2017), and so on (see Ratcliff et al., 2016, for a review).

This brief overview illustrates the different cognitive applications for which the modeling framework has notably intervened. It reflects the widespread endorsement of the framework by the scientific community as an instrumental cognitive model for characterizing the latent processing dynamics of decision-making. As researchers are faced with the challenge that cognitive dynamics (in decision-making, or elsewhere) are not directly observable, the role of these process models is to provide inferences, or linking functions from currently observable data (e.g., RTs, accuracy) which is an ambitious step beyond the classical statistic models (e.g., regression, mixed models, etc.) that are only descriptive at various depths of analysis (Anders et al., 2018, 2019). Such cognitive process models are a crucial research development to the

quantitative rigor of our domain. Their viability, however, depends not only on their goodness of fit to data, but also on the interpretative (cognitive) validity of the estimated parameters. This indispensable, latter condition depends on a number of key assumptions being met, as follows.

First, in using this framework, one has to assume that the postulated model indeed reflects the process generating the behavior. This assumption is often considered to be supported by the quality of the data fit diagnostics (e.g., by RT quantile residuals Ratcliff & McKoon, 2008), which is generally very satisfactory in most published applications.

However, it has often been reported that a given behavioral data-set can be fitted equally well with different models, that differ substantially in their architectures (e.g., Donkin et al., 2011; Servant et al., 2016). Therefore, although a good fit is a necessary criteria, it might not be sufficient to single out a particular model (Roberts & Pashler, 2000). In addition, although generally robust, model fitting procedures are complex operations that may suffer from problems such as parameter trade-off, sensitivity to trials generated from another generative model (e.g., guessed responses), or other biases that can impact the estimated parameters, and hence notably affect the validity of the inferences made upon them (Ratcliff & Childers, 2015).

Second, one has to assume that the manipulations that modulate the estimated parameters can indeed be attributed to modulations of the presumed cognitive process. The major diagnostic to ascertain this mapping is known as the "Selective Influence Test" (Heathcote et al., 2015). This approach probes whether a given experimental manipulation, presumed to selectively affect a given psychological process, only affects the corresponding parameter of the model. For example, manipulations of stimulus information quality are expected to selectively affect the accumulation rate (Ratcliff & McKoon, 2008; Voss et al., 2004), and manipulations of response execution are expected to selectively affect the T_{er} parameter (Gomez et al., 2015; Voss et al., 2004). Perhaps the paradigmatic example of selective influence has been the manipulation of the participants' response caution, which is implemented by instructions that either emphasize speed or accuracy, in what is known as the *speed–accuracy trade-off* (SAT). This manipulation has originally been shown to selectively affect the threshold parameter, which governs the amount of evidence accumulated by participants before they trigger a response (e.g., Ratcliff & McKoon, 2008).

Unfortunately, selective influence has proven to be an elusive goal. For example, Voss et al. (2004) showed that the SAT manipulation could also affect the nondecision time parameter (see also Palmer et al., 2005; Ratcliff, 2006). Then Rae et al. (2014) observed an effect of SAT manipulation on drift rate, although Starns and Ratcliff (2014) did not find evidence for this effect in multiple data sets. In a multilab collaborative project, Dutilh et al. (2019) asked a number of expert decision-making modelers to map an anonymous experimental manipulation from de-labeled data sets to the appropriate model parameter. Most modelers mapped the effect to the boundary parameter, but also to the drift rate or the nondecision time. The experimental manipulation in the data set was in fact the SAT. In a different dataset, the unknown manipulation was stimulus quality. Some modelers mapped this manipulation to the boundary parameter, in addition to the drift. Dutilh et al. (2019) attributed these uncertainties to an excessive number of

Figure 1

Processing Account of a Single Decision in a Diffusion Model Implementing Evidence Accumulation
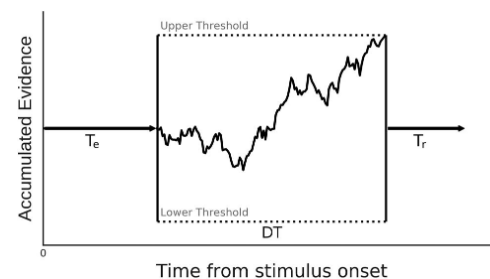
Note. The stochastic path represents the stimulus evidence being accumulated over time through a noisy channel, modeled as a diffusion process with a drift (i.e., rate of accumulation). Accumulation stops once a threshold boundary is reached, and the corresponding alternative is chosen. DT = decision time.

degrees of freedom available to the modelers. Although the authors noted general agreement across modeling approaches in the main parameters at stake, this study illustrates the challenges modelers face in consistently attributing manipulation effects to model parameters. It should be highlighted that the discrepancy might not necessarily be charged onto the modelers and on matters of model fitting (see Smith & Lilburn, 2020, for a recent development). Although the selectivity of the experimental manipulations is assumed, it is by no means mandated.

In short, two crucial assumptions must be scrutinized: the link between the postulated model and the process of interest, and the link between the experimental effects on the parameters and the modulation of the processes of interest. One way to address the aforementioned issues associated with those assumptions is to implement additional modeling constraints through advanced validation measures. These may be derived from data sources coregistered during the experiment, ideally data that could provide a direct correlate or measure of the candidate process. For example, various researchers have linked decision-making variables with neurophysiological activity measured in multiple species, such as rodents (e.g., Brunton et al., 2013), monkeys (e.g., Purcell et al., 2010; Ratcliff et al., 2003; Roitman & Shadlen, 2002), and humans (e.g., Donner et al., 2009; O'Connell et al., 2012; Philiastides et al., 2006). As a result, these approaches have provided essential additional information on latent decision-making dynamics not previously available.

Although this approach, to incorporate other physiological measures that also indicate cognitive processing, is undoubtedly fruitful, it is still suffering from several unresolved challenges that limit its capacity to validate the current modeling framework. First, one major issue is that it is still not clear how to appropriately link physiological activity metrics to psychological processes (Schall, 2004, 2019; Teller, 1984). For example, neuronal discharge frequency increases as a ramping function during decision-making, in a manner very similar to the postulated model dynamic (see Figure 1). But such ramping activity has been observed in several brain areas (Purcell et al., 2010; Roitman & Shadlen, 2002; overview in Lafuente & de Romo, 2006), making it difficult to unambiguously map multiple accumulations into a single cognitive accumulator, or parameter therein. Another limitation comes from the rather low signal-to-noise ratio available in neurophysiological recordings, which results in findings based on averaged data (e.g., peristimulus time histograms, averaged event related potentials, etc.) The information contained in these signals averaged across trials is overly distorted (e.g., Burle et al., 2008; Callaway et al., 1984; Dubarry et al., 2017; Latimer et al., 2015), compared with the distributions afforded by single trial analyses (the resolution at which the RT model in question operates on).

Here, we propose an alternative neurophysiological measure that circumvents some of these limitations, and demonstrate how it can be used to address the previously discussed issues of parameter attribution and inferential validity of the modeling. This neurophysiological measure is the electromyographic (EMG) activity of the effector muscles that perform the responses. In contrast to lateralized readiness potentials computed from electroencephalogram (Osman et al., 2000; Rinkenauer et al., 2004), the high signal-to-noise ratio of EMG allows for a reliable decomposition of the RT of every single trial into two subcomponents: the time from stimulus onset to EMG onset ("pre-motor time" [PMT]),

and the time from EMG onset to the behavioral response ("motor time", MT; Botwinick & Thompson, 1966; Burle et al., 2002; Figure 2). Although the PMT certainly contains many processes, linking the recorded MT with a psychological process is more straightforward. As (Luce, 1986, p. 97) states, "the time from that event [EMG onset] to the response is a proportion of the entire motor time and that, in turn, is a proportion of the residual $\mathbf{R}$ [in our notation, T_{er}]." In other words, the time postulated in the model to execute the response (T_r in Equation 2) should be strongly related to the recorded MT.

The relationship between MT and T_r might be complicated by the fact that typical fitting procedures provide values for T_{er}, not T_r. However, as proposed by (Luce, 1986, p. 118):

The assumption is that if $\mathbf{R}'$ [here, MT] exhibits a dependence upon signal intensity [or other factors], then the chances are that $\mathbf{R}$ [here, T_{er}] will be affected also, which it will be unless the effects on $\mathbf{R}'$ and $\mathbf{R} - \mathbf{R}'$ are equal and opposite—an unlikely possibility.

Following this rationale, if the covariance *across trials* between MT and T_{er} turned out to be weak or nonexistent, then one would have to conclude that the measured MT is not contained in the estimated T_{er}. Consequently, if it is hence the case that T_{er} does not incorporate response execution processes, this would severely put into question the current cognitive interpretative validity of this parameter, and hence of the other parameters too.

Previous research has already shown that the measured MT is not a mere constant value added to the PMT. Its variability contributes to the overall performance one (i.e., of the RTs). Furthermore, MTs have been found to be modulated by certain experimental manipulations. Early on, Grayson (1983) suggested that stimulus intensity affects the estimated MT in a simple reaction task, and Servant et al. (2016) reports a recent confirmation. It has also been shown that MTs are shortened if advance information is provided about the timing (Tandonnet et al., 2003, 2006) or about the nature (Possamaï et al., 2002) of the forthcoming stimulus or response.

Even if the functional interpretation of such perceptual and decisional effects on late motor processes is still under discussion, the observations suggest that certain effects detected in the RTs, even among those assumed to be "cognitive/decisional," could be

Figure 2

Single-Trial Response Time (RT) Decomposition Using Electromyograph (EMG)
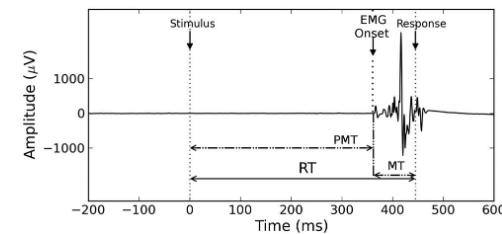
Note. The premotor time (PMT) is the time between stimulus and EMG onsets; the motor time (MT) Is the time between the EMG onset and the mechanical response. See Figure A1 for additional examples of EMG recordings.

driven in part by modulations in motor processes. Inasmuch as motor time can be mapped to the parameter capturing response time execution T_r (see above), this conclusion contradicts one of the assumptions often made about nondecision processes, which was summarized by (Turner et al., 2015, p. 316): "The non-decision time parameter captures effects that are not cognitively interesting." The alternative possibility that nondecision time *is* cognitively relevant has important implications for models of decision making.

The Present Study

The present study is based on two experiments, each involving a standard perceptual decision task. Our main hypothesis in this study is that decomposing RTs through EMG techniques into PMT and MT will help (a) better establish the locus of certain experimental manipulations, (b) test some central assumptions of the modeling framework regarding the postulated cognitive processes, and (c) assess to what extent EMG-based and model-based decompositions of RTs provide (di)similar information about the nondecision time parameter.

Visual contrast was manipulated to adjust task difficulty, and verbal instructions emphasizing speed or accuracy were used to implement an SAT setting. Across the two experiments, we also manipulated the force required for the response to be produced. We recorded the EMG activation and the latency of every manual response, hence deriving single trial distributions not only for RT, but also for PMT and MT, and assessed the impact of the experimental factors on each of these variables. Based on previous studies (Spieser et al., 2017; Steinemann et al., 2018, see also Osman et al., 2000; Rinkenauer et al., 2004, who suggested the existence of a SAT effect on motor processes using LRP), we expected that the measured MT would be affected by SAT, and possibly by stimulus strength (Grayson, 1983; Servant et al., 2016). In addition, by having a reliable EMG measure for every single trial, we were able to assess the stochastic dependency between PMT and MT. In other words, this provides a test for the subsidiary assumption of independence between decision and nondecision processes, a test that is not possible in most regular parameter estimation procedures.

Following this empirical exploration, we estimated the parameters of the Drift Diffusion Model (DDM; Ratcliff, 1978; Ratcliff & McKoon, 2008) from the data using a hierarchical Bayesian fitting method (Wiecki et al., 2013, HDDM package). First, we applied a model selection procedure to find the best fitting model. Second, we evaluated the correlation between the estimated T_{er} parameter and the measured MT, across participants. Finally, we fitted a joint DDM that takes the MT as a by-trial regressor, to assess the link between the T_{er} parameter and the measured MT.

Experiment 1

Method

Participants

Sixteen participants (eight men and eight women, mean age = 23.5, 2 left-handed) that were students at Aix-Marseille University, were recruited for this study. They were compensated at a rate of 15 € per hour. All participants reported having normal or corrected vision, and no neurological disorders. The experiment was approved by the ethical experimental committee of Aix-Marseille University, and by the "Comité de Protection des Personnes Sud Méditerrannée 1" (Approval No. 1041). Participants gave their informed written consent, according to the declaration of Helsinki.

Apparatus

Participants performed the experiment in a dark and sound-shielded Faraday cage. They were seated in a comfortable chair in front of a 15-in. CRT monitor placed 100 cm away, which had a refresh rate of 75 Hz. Responses were given by pressing either a left or a right button with the corresponding thumb. Buttons were fixed on the top of two cylinders (3 cm in diameter, 7.5 cm in height) separated by a distance of 20 cm. They were mounted on force sensors allowing to continuously measure the force produced (A/D rate 2048 Hz), and to set the force threshold needed for the response to be recorded. In Experiment 1, this response threshold was set to 6N (600g). Response signals (threshold crossing) were transmitted to the parallel port of the recording computer with high temporal accuracy (<1 ms). At button press, participants heard a 3-ms sound feedback at 1,000 Hz (resembling a small click). The forearms and hypothenar muscles of the participants rested comfortably on the table to minimize tonic muscular activity compromising the detection of voluntary EMG bursts. We measured the EMG activation of the flexor pollicis brevis of both hands with two electrodes placed 2 cm apart on the thenar eminences. This activity was recorded using a BioSemi Active II system (BioSemi Instrumentation, Amsterdam, the Netherlands). The sampling rate was 2,048 Hz.

Stimuli

Stimuli presentation was controlled by the software PsychoPy (Peirce, 2007). Each stimulus was composed of two Gabor patches, presented to the left and the right of a fixation cross. The Gabor patches had a spatial frequency of 1.2 cycles/visual angle degree and had a size of 2.5 visual angle degrees each. The standard Gabor patch contrast was set to 0.5, on a scale between 1 (maximum contrast) and 0 (uniform gray). Five levels of stimulus contrast were used (0.01, 0.025, 0.07, 0.15, 0.30). These contrast values were added to the target Gabor patch and subtracted from the distractor Gabor patch. These contrast levels were decided based on performance from a pilot study where they were found to yield a full range of performance quality: from near perfect to almost chance level accuracy. The task of the participants was to press the button (left or right) ipsilateral to the patch with the highest contrast.

Procedure

All participants performed one single session with 24 blocks of 100 trials each. Session duration was approximately 1h30, including a training session of 15 min and self-paced breaks between each block. During the training session, participants were instructed that "Speed" instructions required a mean RT near 400 ms and that "Accuracy" instructions required a minimal response accuracy near 90% while maintaining RTs below 800 ms. Participants were also informed to keep their gaze on the central fixation cross during the blocks.

The beginning of each speed or accuracy block was preceded by the corresponding visual instruction (the French word "Vitesse" or "Précision" for Speed and Accuracy, respectively). The end of each block was followed by the presentation of the recorded mean RT and response accuracy performance of the block, along with oral feedbacks from the experimenter in those cases were the participant did not satisfy the condition goals. The training session included 40 trials without performance instructions followed by two blocks of 10 trials in the Speed condition, followed by two blocks in the Accuracy condition, and ended with four blocks of 10 trials with alternating instructions. For the experimental session, speed instructions alternated every three consecutive blocks. The order of the instructions was counterbalanced across participants. The contrast conditions of the stimuli were fully randomized across the five levels within each block. No response deadline was applied, the stimulus disappeared when participant produced a button press, the response-stimulus interval was fixed to 1,000 ms.

EMG Processing

The EMG recordings were read in Python using the MNE module (Gramfort et al., 2013) and filtered using a Butterworth 3rd order high-pass filter at 10Hz from the scipy Python module (Oliphant, 2007). The by-trial EMG signal was then processed in a window between 150 ms before and 1,500 ms after stimulus onset. A variance-based method was used to detect whether EMG activation was significantly present in either hand's channel. The precise onset was then identified with an algorithm based on the "Integrated Profile" of the EMG burst. This method takes the cumulative sum of the rectified EMG signal on each epoch and subtracts it from the straight line joining the first and the last data-points (corresponding to the cumulative sum of an uniform distribution). The onset of the EMG burst corresponds to the minimum of this difference (see Liu & Liu, 2016; Santello & Mcdonagh, 1998 for more details[1]). The EMG onsets determined by the algorithm were examined by the experimenter who could perform manual corrections as needed (18.3% of the trials). For this processing stage, the experimenter was unaware of the trial type he was examining, to avoid any correction bias. Every muscular event (rapid change in the signal followed by a return to the baseline) in the trial was marked, thus quantifying how many times the muscle was triggered, even when the identified activation did not lead to an overt response (see Figure A1 in Appendix A for examples of recorded EMG activity).

MT was defined as the time between the onset of the last EMG activation preceding the responding hand button press. PMT was defined as the time between stimulus onset and this last EMG onset. In this way, any intervening EMG activations were discarded. For the purpose of this study, we treated the trials with multiple activities (25.90% of the total number of trials, see e.g., participants S4 and S5 EMG plot in Figure A1) as trials with only the last EMG activation. Notice that if only RT was measured these trials would not have had any special status. It has not escaped our attention that these trials can represent a challenge for evidence accumulation models (Servant et al., 2015, 2016), and

that they will have to be investigated more thoroughly in future research.

Statistical Procedure

Bayesian Statistics. Apart from a few exceptions (see below), the analyses were performed within the Bayesian framework. Bayesian methods aim to estimate an unknown parameter (or set of parameters) and the uncertainty around it. More explicitly, Bayesian methods implement Bayes' rule to generate a posterior distribution for each parameter based on a combination of prior information and the likelihood of the data given the parameters. This posterior distribution can then be naturally interpreted as the probability of any given parameter value given the data, the priors and the tested model. In our study, we summarize the posterior distribution using the mean, standard deviation and the 95% Bayesian credible interval (CrI). Our criterion to assess the presence of an effect was that the null value lied outside the CrI. Although this method does not quantify the evidence in favor of the null hypothesis, it does provide an estimation of the effect size and its uncertainty. All the priors used in the article are detailed in Appendix B.

Bayesian Mixed Models. To test our hypotheses on the behavioral and EMG variables we used linear mixed models (LMMs). These models estimate fixed effects (e.g., the effect of SAT on RT) while accounting for random effects (e.g., the interindividual differences in the effect of SAT on RT), making them particularly useful in repeated measure design such as the one used in this study. Estimating interindividual differences as random effects shares the information gathered from each participant while providing separate (but not independent) estimates for each one of them. Given our analysis approach, we derived one generic LMM fitted independently for all chronometric dependent variables: RT, PMT and MT. In these LMMs, the log transformation of the chronometric variables on the ith trial for the jth participant (y_{ij}) was assumed to be drawn from a normal distribution with mean μ_j and residual standard deviation σ_r:

$$y_{ji} \sim \mathcal{N}(\mu_j, \sigma_r) \tag{3}$$

where $\sim$ stands for "distributed as." The mean of each participant μ_j is then defined by an intercept (α_j) and slope coefficient (β_j) for each experimental factor and their interactions.

$$\begin{aligned}
\mu_j = \alpha_j &+ \beta_{1j}\,SAT + \beta_{2j}\,Cont. + \beta_{3j}\,Corr. + \beta_{4j}\,RS \\
&+ \beta_{12j}SAT \times Cont. + \beta_{13j}\,SAT \times Corr. \\
&+ \beta_{23j}\,Cont. \times Corr. + \beta_{123j}\,SAT \times Cont. \times Corr.
\end{aligned} \tag{4}$$

where $Cont.$ stands for "Contrast," $Corr.$ stands for "Correctness," and RS stands for "Response side."[2] The individual intercepts (α_j) and slopes of each predictor x (β_{xj}) are modeled as drawn from a normal distribution:

$$\alpha_j \sim \mathcal{N}(\mu_\alpha, \sigma_\alpha) \tag{5}$$

---

[1] Software implementing this two-step procedure will soon be released with an open-source license and is already accessible upon request.
[2] The common R syntax for these LMMs would be: y ~ SAT * Cont. * Corr. + RS + (SAT * Cont. * Corr. + RS | participant).

$$\beta_{xj} \sim \mathcal{N}(\mu_{\beta_x}, \sigma_{\beta_x}) \qquad (6)$$

where μ_α and μ_{β_x} are the population estimated intercept and slope while σ_α and σ_{β_x} the estimated population variance of the intercept and the slope (i.e., the random effect). To test our hypothesis, we report for each LMM the posterior distribution of the population-estimated intercept and regression coefficients.

In Equation 4, correctness of the response was included as a predictor because the relationship between the distributions of RT on correct and incorrect trials is known to change under speed pressure (Grice & Spiker, 1979) and because MT has been previously reported to be affected by this factor (e.g., Allain et al., 2004; Rochet et al., 2014; Śmigasiewicz et al., 2020). Response side was included as an additive predictor because left and right RTs often differ. Two remarks are in order concerning this last point. First, we did not expect any interaction with the other predictors. Second, motoneurons synchronization has been shown to depend on handedness (Schmied et al., 1994). As a consequence, we can expect Response side to affect MT and the effect to be, at least substantially, of motor origin.

We also tested the effects of these factors on the proportion of correct responses using a generalized linear mixed model assuming that each response (correct or incorrect) was drawn from a Bernoulli distribution whose parameter depends on the same predictors as the LMM (except the correctness factor and its interactions).

For each LMM and generalized LMM, six Markov Chain Monte Carlo (MCMC) sampling processes were run in parallel, each composed of 2,000 iterations among which the first 1,000 samples were discarded as warm-up samples. We assessed convergence of the MCMC chains both by computing the potential scale reduction factor ($\hat{R}$, see Gelman & Rubin, 1992) and by means of visual inspection of the MCMC chains. We also visually checked the assumptions of the linear regression by inspecting the normality of the residuals through QQ-plots and assessment of homeoscedasticity. The LMM and generalized LMM were fitted with a custom Stan code, available in the online repository, inspired from the code provided by Nicenboim et al. (2018) and using the `pystan` package (Stan Development Team, n.d.). Summary statistics and plots of the parameters were created with the `arviz` python package (Version 0.4.1, Kumar et al., 2019).

We also performed a frequentist replication of each G/LMM using the `lme4` R package as a means to check for prior sensitivity. Any discrepancy with the main (Bayesian) analysis is reported in the Results.

Factor Coding and LMM Parameter Interpretation. For all LMMs, sum-contrasts were used for SAT (-0.5 for speed and 0.5 for accuracy) and for response side (-0.5 for right responses and 0.5 for left responses). Treatment-contrast was used for correctness (0 for correct and 1 for incorrect responses). The stimulus strength factor was centered on its middle value and transformed such that $-.5$ represented the lowest possible contrast and $.5$ the highest possible contrast. These coding features were chosen to ease the interpretation of the resulting coefficients. When the binary predictor is sum-contrasted (-0.5 and 0.5), the estimated β value can be read as the difference between both conditions. When the binary predictor is treatment-contrasted (0 and 1), the estimated β can be read as the difference to add to the intercept (predictor at 0) to obtain the mean of the condition where the predictor is at

value 1. Hence, in our analysis, the intercept can be read as the predicted time for the reference condition where the response is correct, and at an intermediate value for the predictor SAT. The main effects and interactions can be read according to the coding scheme used; for example, the correctness slope represents the benefit or cost of an error when contrast is at midlevel. The contrast slope represents the benefit or cost of a higher contrast when the response is correct. The interaction between contrast and correctness represents how the correctness/contrast effect changes when contrast is higher than midlevel/an error was made.

RT, PMT and MT Specific Adjustments. By-trial RT, PMT and MT were log transformed prior to the analysis because these variables have heavily skewed distributions that would violate the normality assumption of residuals of the LMM. To ease the interpretation of the estimated LMM parameters, we back-transformed the intercepts by taking their exponential, and each slope by subtracting the exponential of the intercept from the exponential of the sum of the intercept and slope. We applied this back-transformation at each iteration of the MCMC procedure, hence computing the uncertainty around the parameter values on their natural scale.

Fast Guess Detection. Before applying any analysis we performed the Exponentially Weighted Moving Average (EWMA) filter developed by Vandekerckhove and Tuerlinckx (2007). This method iteratively computes a weighted accuracy measure (amount of correct responses relative to errors) on the sorted RT distribution, from the fastest to the longest RT. Participants are considered as being in a fast guess state until the weighted accuracy is higher than a defined threshold. The RT at which this change of state occurs is identified, and all trials faster than this RT are censored. The user defined parameters in this method are the initial starting point of the weighted accuracy, the accuracy threshold for defining nonguess trials, and the amount of preceding trials (weight) retained in the accuracy computation. The starting point was defined at 0.50 based on the assumption that a guessing strategy yields a 50% chance of correct response. The threshold was fixed at 0.60 based on a reasonable assumption that participants were not guessing when accuracy was superior to 0.60. The weight was heuristically fixed at 0.02 (bounded from 0 to 1, with 0 being all preceding trials used), after visual inspection of the rejection plots with different weights. This method was applied for each participant's RT distribution separately in the speed and accuracy conditions, as fast-guesses can have different latencies across both conditions. The figures illustrating these rejection procedures can be found in the online repository. We thank Michael Nunez for kindly providing the code used for this method (https://github.com/mdnunez/bayesutils/blob/master/wienerutils.py).

Results

Two of the 16 participants were excluded because of high tonic activity in the electromyogram, which otherwise would have made the detection of their EMG onsets too unreliable. This rejection was decided before performing any analyses. Because of technical constraints on the methods for detecting EMG onsets, we applied an upper limit of 1,500 ms to the RTs, resulting in the loss of 0.50% of the trials. Trials with low signal-to-noise ratio, and trials with high spontaneous tonic activity, making appropriate EMG onset detection difficult, were also removed. In total, these accept-

ability conditions led to the exclusion of 4.08% trials for all analyses. Additionally, the EWMA method removed 5.80% trials of the data.

The descriptive statistics discussed for the chronometric variables are summarized in Figure 3. The parameters of the Bayesian LMM are represented on the same scale (in milliseconds) in Table 1 and in Figure 4. Additional analyses of response accuracy are provided in Appendix C.

Reaction Time and Premotor Time

PMTs and RTs become shorter as contrast increases and when speed is stressed. Although the credible intervals contained the null value, we observe a weak positive main effect of correctness. However, an interaction with SAT instructions showed that when speed is emphasized, errors are faster, and when accuracy is emphasized, errors are slower. A strong three-way interaction furthermore specified that this correctness effect according to SAT conditions is even stronger for easier contrasts. Overall, the results for PMTs and RTs mirrored each other except for one difference: response side (laterality) had a significant effect on the RTs but not on the PMTs (not shown in Figure 3).[3]

Motor Time

Replicating previous results, MTs were also faster under speed emphasis. However, contrary to RT and PMT, the effect of correctness was in the same direction in both SAT conditions (slower MTs during errors). MTs were faster when participants responded with their right hand, which explains that the previous laterality effect observed on RTs (but not on PMTs) was attributable to MT differences. Finally, there was a small but consistent effect of contrast on MT. This effect interacted significantly with response correctness indicating that the lengthening of response execution for errors (see Allain et al., 2004; Śmigasiewicz et al., 2020, for interpretation) was larger in the easier contrast conditions. No additional interactions were observed (see Table 1).

Correlations Between PMT and MT

In most implementations of the evidence accumulation framework, the decision and nondecision stages are assumed to be independent from one another. We tested this assumption by examining, for each participant, the Spearman correlation between by-trial PMTs and MTs, in the speed and the accuracy conditions separately. It is important to note that certain trial features may bias the correlation estimates. For example, fast-guess trials defined on the basis of their RT value could bias the computed correlation coefficients toward negative values. Indeed, as the total RT is the sum of PMT and MT, fast RTs most likely result from both fast MT and PMT. Trimming lower RT values would remove trials in which PMT and MT likely show a positive covariation, thus biasing the correlation estimates toward negative values. Additionally, as suggested by Stone (1960), trials with long PMT might be subject to a trade-off between PMT and MT based on an implicit deadline (a process that is not implemented in the DDM), hence generating a negative correlation.

To address these concerns, the correlation between these RT subcomponents was assessed separately for trials identified as fast-guesses and, the remaining trials, for which the correlation was computed across five different quantiles (namely, .1, .3, .5, .7,

and .9). With respect to fast guesses, these were identified as before, with the exponentially weighted moving average (Vandekerckhove & Tuerlinckx, 2007) now based on the sorted distribution of PMT rather than the distribution of RT.[4] There was an average of 79.14 fast-guess trials per participant and SAT condition (range: 17–249). There was an average of 1065.93 remaining trials per participant and SAT condition (range: 786–1165), to be divided in five quantiles.

For the trials identified as fast guesses, we computed Spearman correlations between PMT and MT within each participant and SAT condition and submitted these values to an LMM with SAT emphasis as a predictor.[5] The Accuracy condition was coded as 0 and Speed as 1, thus allowing us to interpret the intercept of the LMM as the mean correlation value when accuracy is emphasized and the slope (effect) as the change in this mean correlation when speed is emphasized. Fast guesses presented a negative mean correlation in the accuracy condition, as shown by the intercept of the LMM (m = −0.22, 2.5% = −0.33, 97.5% = 0.10). The slope of the LMM (in other words, the change in this mean correlation when speed is emphasized) did not suggest that the correlation differed between conditions (m = −0.03, 2.5% = −0.15, 97.5% = 0.09; see Figure 5).

For the correlation analysis along quantiles, the likelihood of the mean correlation coefficient at each quantile was assessed with a Monte-Carlo procedure. We computed the Spearman correlation between draws from two random variables following a normal distribution (with $M = 0$ and $SD = 1$) for 14 simulated participants divided into two conditions, and five bins of data with the same amount of trials as the real data; this procedure was repeated 1,000 times. This nonparametric analysis was motivated by the fact that we did not have a specific hypothesis (e.g., a linear trend) for the effect of the quantiles on the mean correlation value. The first quantile in the speed condition, was outside the range of expected values for the normal variables (see Figure 5). Likewise, the first quantile in the accuracy condition was also outside the expected range, even if less negative than in the speed condition. The other correlation values do not differ from random levels. To illustrate the relationship between PMT and MT at the participant level, in Appendix D we provide the scatter plots for the first five participants across the five quantiles.

Discussion

With respect to the RTs, the different manipulated factors led to clear results that were congruent with previous works discussed in the Introduction. Because theories for the potential effects on PMT and MT are less developed, these results will be discussed in more detail in this section.

First, consistent with previous studies, the SAT manipulation significantly affected RT and response accuracy (the latter is reported in Appendix C). In agreement with a decisional locus of SAT, this factor also had a large impact on PMT. However, the

---

[3] A frequentist replication of these tests provided the same results, except it included an additional significant interaction: between contrast and correctness for PMT ($log[\beta] = -0.07$, $t = 2.97$).

[4] Both applications of the method revealed a strong but not perfect correlation on the amount of censored trials, $r(28) = 0.76$, $p < .001$.

[5] Note that, given low number of points ($N = 14$), this LMM only included the by-participant intercept as a random effect.

Figure 3

Observed Results in Experiment 1, Effect of Stimulus Strength Level (Contrast) on the Mean Response Time (RT; Left Column), Premotor Time (PMT; Center Column) and Motor Time (MT; Right Column)
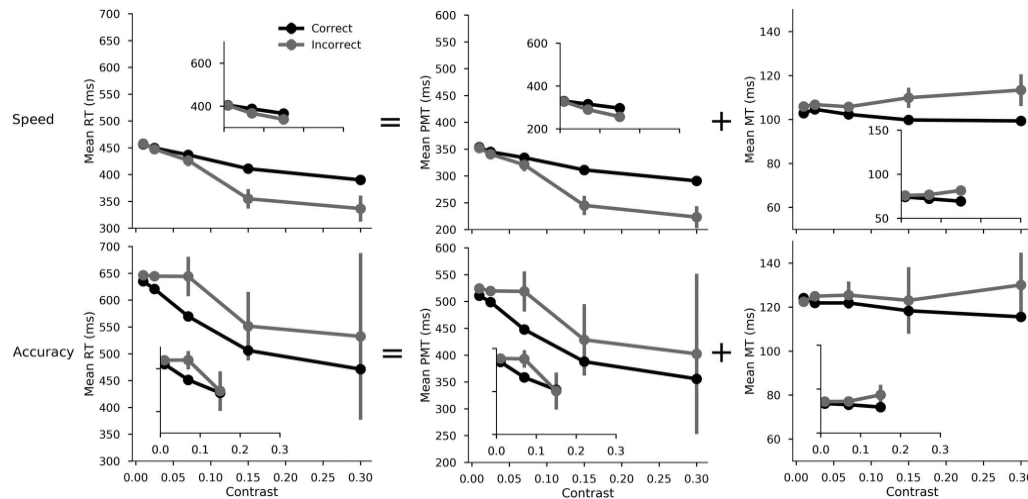
Note. This plot illustrates the interaction between contrast (*x* axis), speed–accuracy trade-off (SAT) conditions (top vs. bottom rows), and correctness of the response (black – correct vs. gray – incorrect). Bars around the mean represent 95% confidence intervals corrected for within-subject design using the method developed in Cousineau (2005). To assess replication, the small insets provide the results obtained in Experiment 2 for a subset of the contrast levels (0.01, 0.07, 0.15). These figures analyze the means in millisecond units, whereas the linear mixed model (LMM) analysis presented in Table 1 model the means in log-transformed units.

effect of SAT was not restricted to PMT, because this factor also affected MT. This latter observation replicates recently reported results (Spieser et al., 2017; Steinemann et al., 2018) and is thus taken to be robust. Furthermore, the effect of SAT on MT is large, accounting for 11% of the whole SAT effect measured on RT (it was up to 20% in Spieser et al., 2017).

The issue of how to properly account for observed RT variation between correct and error responses has a long history in the modeling of decision making. Because early decision-making

models could not account for errors resulting in different latencies, additional parameters were added (e.g., to the DDM: Ratcliff & Tuerlinckx, 2002). In our EMG decomposition, first we see that PMT was shorter for errors than correct responses when speed was stressed. This suggests that errors tend to be made based on either shorter encoding times and/or shorter decision times. Second, replicating previous reports in other task-settings (e.g., Allain et al., 2004; Rochet et al., 2014; Roger et al., 2014), MT was longer for errors than for correct responses. Such longer MTs on errors

Table 1

Results of the LMMs in Experiment 1 for Each Latency Measure

Predictor	RT				PMT				MT			
	Coeff.	SE	2.5%	97.5%	Coeff.	SE	2.5	97.5%	Coeff.	SE	2.5%	97.5%
Intercept	469	12	446	494	354	8	338	370	105	6	93	118
SAT	135	19	101	173	118	17	85	150	15	3	9	22
Contrast	−94	10	−114	−74	−81	9	−99	−62	−6	1	−9	−3
Correctness	11	6	0	22	0	6	−11	12	6	1	4	8
Resp. side	18	6	6	31	0	8	−16	17	17	4	10	24
SAT × Contrast	−65	9	−82	−47	−62	8	−78	−44	−1	1	−3	1
SAT × Correctness	71	12	48	96	75	14	49	105	−1	1	−4	2
Contrast × Correctness	26	15	−3	55	3	16	−28	35	13	3	8	18
SAT × Contr. × Corr.	153	32	90	216	164	33	95	228	1	3	−5	7

Note. RT = reaction time; PMT = premotor time; M = motor time; Coeff. = estimated coefficients of the linear mixed model (LMM) fitted on the log scale and back-transformed to the millisecond-scale; SE = standard error for the coeff; SAT = speed–accuracy trade-off; 2.5 and 97.5% = the lower and upper Bayesian credible interval (CrI) around the coeff. For further details, see the Statistical Procedure section.

Figure 4

Posterior Distributions for the Regression Coefficients From the Linear Mixed Model (LMM) Fitted on the Response Time (RT), Pre-motor Time (PMT), and Motor Time (MT) of Experiment 1 (Black) and Experiment 2 (Gray)
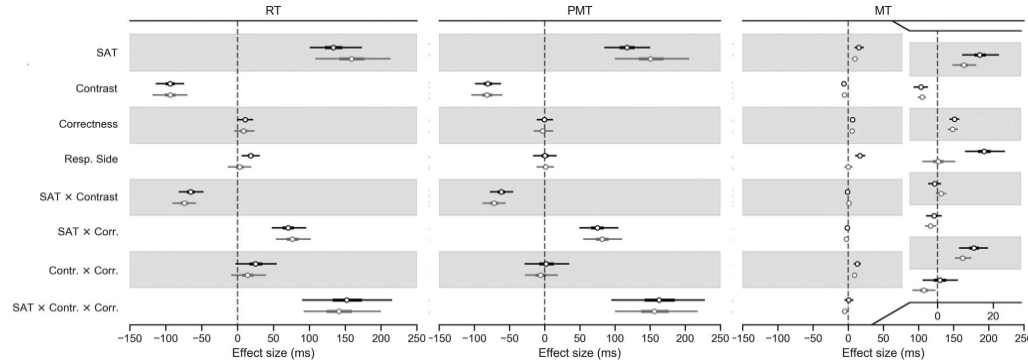
Note. The segments around the estimates represent 95% Bayesian credible interval. Because the coefficients for MT are of smaller magnitude, we provide a zoomed inset on the far right of the figure to ease their visualization.

are associated with a modulation of the EMG burst leading to the response (not explored here, but see: Allain et al., 2004; Rochet et al., 2014; Śmigasiewicz et al., 2020), which has prompted the interpretation that they could reflect a desperate attempt to stop the incorrect response, hence revealing an online process of cognitive control. To date, no decision making models incorporate such a process. The effect magnitude estimated for the reference condition (intermediate contrast) was a modest 6 ms, which may suggest that it is a reasonable simplification to not aim to account for this effect in decision making models. However for easy stimuli, the difference between correct and incorrect trials was threefold (mean MT difference equal to 18ms in the condition emphasizing accu-

racy, see Figure 3), which highlights the importance of considering this effect, depending on the experimental conditions. Most importantly, this observation suggests that motor processes can indeed be affected by experimental manipulations that have been considered to be purely "decisional" in previous research.

Stimulus contrast has been classically considered to affect evidence accumulation processes (e.g., Palmer et al., 2005, Experiment 5). In agreement with this view, its effect on PMT was clear and very similar in magnitude to that observed on total RT. More surprisingly, a small but highly reliable effect of stimulus contrast was observed on MT, in the same direction as on PMT. The presence of perceptual effects on motor processes has been previ-

Figure 5

Mean Spearman Correlations Between Premotor Time (PMT) and Motor Time (MT) for the Trials Identified as Fast-Guesses (Triangles), and the PMT Quantiles (Circles) Across Both Speed (Gray) and Accuracy (Black) Conditions for Experiment 1 (Left) and Experiment 2 (Right)
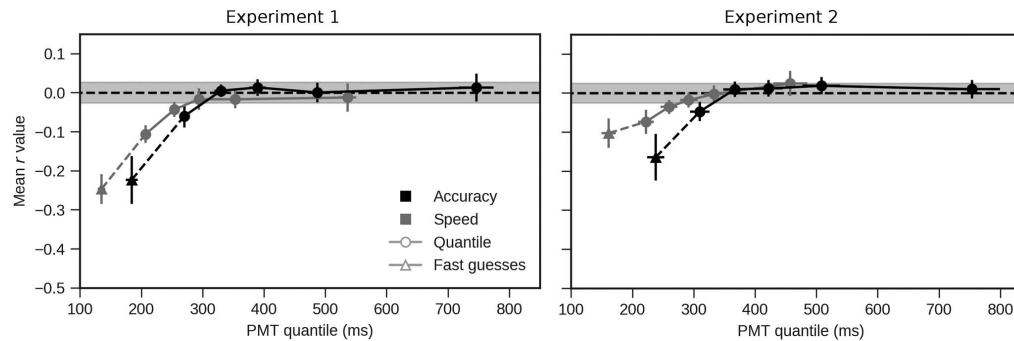
Note. Error bars represent one standard error of the mean. Shaded intervals represent 95% confidence intervals of the correlation coefficient based on 1,000 draws of the simulated random, normally distributed variables.

ously debated. For instance, in using a double response paradigm, Ulrich and Stapf (1984) showed that increasing stimulus duration shortened RT but also increased output response force. In three separate (unpublished) experiments, Grayson (1983) also reported evidence that higher signal intensities shorten MT in a simple RT task. More recently, Servant et al. (2015) reported color saturation effects on MT in a color discrimination conflict task. In contrast, other studies have reported that MT is unaffected by stimulus intensity (Bartlett, 1963; Miller et al., 1999; Smith, 1995). For example, Smith et al. (1992, unpublished article, cited in Smith, 1995) reported the invariance of mean motor time across stimulus conditions. Miller et al. (1999) also reported no effect of stimulus intensity on the lateralized readiness potential nor on the EMG-based MT in two independent experiments using a forced choice task.[6] In the present study, the statistical robustness, along with the linear trend observed across contrast levels, leave no doubt that this effect exists in the data sets acquired. The question remains open, however, as to whether this effect reflects a "cognitive" or "energetic" process (see Sanders, 1983). For our current purpose, and irrespective of the origin of the effect, the important aspect is that the factor contrast affects PMT and MT in a similar direction.

Response side affected the RTs, with longer RTs for the left than the right hand. This effect was selectively localized to MT rather than PMT. The effect is likely attributable to a left–right difference in the innervation of the motor units (Schmied et al., 1994). Such a pure motor laterality effect (no effect on the PMT) indicates that, contrary to what the RT data may have suggested, decision latencies are independent of the hand with which the response is given.

Finally, PMT and MT were not significantly correlated in most PMT quantiles. It is important to dissociate the correlations between-trials, reported in the results, and the correlations between-participants s, which would be computed on average measures per participant. The former were close to zero, which is a necessary, if not sufficient, condition for stochastic independence between the two measures. The latter, between participants, appears to be positive, indicating that slow participants tend to be slow on both decisional and motor components. To the best of our knowledge, the absence of a significant between-trial correlation is formally reported here for the first time, but it was already observed in previous data sets (unpublished observations made on published data, e.g., Burle et al., 2002) instilling confidence on its reliability. A more detailed analysis showed that there is a negative correlation between PMT and MT on the early quantile of the PMT distribution, irrespective of the SAT instructions (see Figure 5). Such modulation of the correlation pattern could indicate that the trials are not generated from the same architecture across the quantiles of either condition. This interpretation is in line with the observation that trials identified as fast-guesses also present a negative correlation. We come back to this issue in the General Discussion.

In summary, MT appears to be substantially affected by various experimental factors. These results were robust, and most of them were consistent with previously reported, or unreported, findings. Before interpreting these observations any further, however, we take up the issue that the force threshold for triggering a response was rather high in this experiment, which motivated Experiment 2. A high force might lengthen MT in such a way that it becomes modulated by parameters that do not affect it in more canonical decision settings. The other important limitation of the experiment,

potentially connected to the high force setting, was there being a high rate of trials exhibiting multiple EMG activations. Repeated muscle triggering during very short intervals could modify the activation dynamics of the cortical and spinal neurons, as well as the excitability of the neuromuscular junction. This could result in a mis-estimation of the motor time for these trials, compared with trials showing a single EMG activation. To address these concerns, we hence sought to replicate our findings in a second experiment with lower response force requirements. By reducing the force required, we expected to, hopefully selectively, affect the motor components (MTs and T_r), and reduce the rate to which trials with multiple EMG activations occur (Burle et al., 2002).

Experiment 2

The main goal of Experiment 2 was to replicate and extend the previous results by refining the design. Specifically, this experiment differs from the first one based on the following two adjustments. First, the force threshold needed to respond was divided by 3 (from 6 to 2 N). Second, to increase the total number of trials per design cell (Participant $\times$ SAT $\times$ Contrast level) from 240 to 432, the number of contrast levels was lowered (from 5 to 3). Higher trial counts would allow us to reject trials with multiple EMG activity while keeping a sizable amount of trials.

Method

Participants

Sixteen participants (eight men and eight women, mean age = 23.6, 1 left-handed) were recruited. None had participated in Experiment 1. They were all students from Aix-Marseille University. All reported having normal or corrected vision and no neurological disorder. Participants gave their informed written consent according to the declaration of Helsinki and were compensated at a rate of 15 € per hour.

Procedure

All participants completed a single experimental session comprised of 24 blocks with 108 trials each (2,592 trials per participant). Session duration was similar to Experiment 1 (~1h30), including an initial training component of 15 min and self-paced breaks between each block. The duration of the training session was shortened compared with Experiment 1, because asymptotic performance was reached quickly. Contrast levels were chosen from Experiment 1, targeting a full range of performance from almost-chance level to near-perfect (i.e., 0.01, 0.07, 0.15). The statistical procedures, EMG recordings, and processing techniques were the same as in Experiment 1.

Statistical Analysis

As a part of data analysis, no participants presented conditions for exclusion. However, during data collection, two participants

---

[6] There is potentially a very serious flaw in the EMG recording of this study. The signal was low-pass filtered at 500 Hz before being sampled at 250 Hz. According to the Shannon-Nyquist theorem, the minimal sampling frequency given the filtering should have been at 1,000 Hz. As a result, strong aliasing of the signal may have occurred, which could jeopardize the validity of its conclusions.

were stopped for excessively high tonic activity in the EMG that could not be reduced. As for the identification of fast guesses and value transformations (log transform for RT, PMT and MT), these were performed as in Experiment 1. The same factor coding features and priors were applied in the LMM models.

Results

The upper limit of 1,500 ms for the RTs resulted in the removal of less than 1% of trials. Next, trials with low signal-to-noise ratio, high spontaneous tonic activity, or multiple activities led to the exclusion of 14.18% of trials. With regard to the effectiveness of lowering the force needed to respond, the rate of occurrence of trials with multiple EMG activations was successfully reduced in this experiment to 12.73% (from 25.90% in Experiment 1). After exclusion of the multiple activity trials, EMG onsets detected by the algorithm had to be visually corrected on 5.9% of the remaining trials. Additionally, the EWMA method removed 7% of trials taken to be fast guesses. As was done for Experiment 1, the analysis of the error rates is provided in Appendix C, and the overall pattern of variation on RT, PMT, and MT is represented in Figure 3 (insets). All discussed LMM parameters can be found in Table 2 and are represented in Figure 4.

Reaction Time and Premotor Time

As in Experiment 1, PMT and RT followed the same trends, as reflected by effect estimates with the same sign and comparable magnitudes (Table 2, Figure 4). Both PMT and RT were faster with increases in contrast or when speed is stressed, and all CrIs of the interactions excluded 0 as a plausible effect, with the exception of the interaction between correctness and contrast. The only notable difference with Experiment 1 was the result of response side not significantly affecting RT, which was localized to an MT effect.[7]

Motor Time

All of the effects observed in Experiment 1 were replicated, except the effect of response side (see Table 2). This is shown As Figure 4, where all estimates are of close magnitude across experiments, and the corresponding CrIs largely overlap, except for the response side factor. We hence successfully replicated the finding that MT is sensitive to SAT, correctness, and contrast; and that the correctness and contrast factors interact.

Correlations Between PMT and MT

As in Experiment 1, we again computed the Spearman correlations on the fast-guess trials, as identified using the EWMA method on the PMT distribution, as well as on the quantiles of the PMT distribution that excludes fast guesses. As before, fast guess trials are associated with a negative correlation between PMT and MT in accuracy ($m = -0.17$, $2.5\% = -0.26$, $97.5\% = -0.06$); emphasizing speed over accuracy did not change the negative correlation ($m = 0.07$, $2.5\% = -0.08$, $97.5\% = 0.20$). The by-quantile analysis revealed the same pattern as in Experiment 1, where early quantiles are associated with a negative correlation (see Figure 5). Similarly as before, two quantiles also fell below the random simulation results in the speed condition, whereas only the first quantile did so in the accuracy condition.

Discussion

This experiment successfully replicated the principal results observed in Experiment 1. With the response force settings being lower, shorter MT durations were obtained, and these force conditions are more resembling to the canonical publications in the field of decision making. Under these circumstances, the fact that we observed a sensitivity of motor processes (indexed on the basis EMG activations) to SAT instructions, stimulus contrast, and response correctness, indicates that these effects are not an artifact induced by the requirement of a large response force. Moreover, the exclusion of the multiple activity trials in Experiment 2 did not lead to a pattern of results that were different from those reported in Experiment 1. The only noticeable difference was the disappearance of the response-side effect on MT. This could indicate that, under low force requirements, the left-right difference in the innervation of motor units (Schmied et al., 1994) can be functionally compensated for.

The negative correlation between PMT and MT on the early quantiles of the PMT distribution was also replicated, confirming that the temporal relationship between "decisional" and "motor" processes changes with the duration of the processes contained in the PMT.

Overall, these results consolidate the interpretation that motor processes in decision making are not fixed ballistic processes, and that the factors thought to affect decision processes can also impact motor-related components. We will come back to the size of these effects in the General Discussion.

Having reliably established, from two experiments, how these experimental conditions affect decisional and motor components of RTs, it is worthwhile to explore the extent to which the parameters of a formal decision making model may covary with these EMG-based decompositions. We carry out this analysis with the DDM, which also aims to decompose RTs into decision and nondecision processes.

Modeling

Method

The DDM (thoroughly reviewed by Ratcliff & McKoon, 2008) was fitted to the RTs obtained in Experiments 1 and 2. We first determined the model that best fitted the behavioral data of both experiments. Then, to assess whether the EMG and model-based decompositions may lead to the same conclusions, we evaluated how the variation of T_{er} across participants covaried with the variation of MT across participants. If such covariation is present, modelers can presumably get some information about the motor system in the absence of EMG activity recordings by comparing T_{er} parameters between (groups of) participants. Finally, to formally probe the relationship between nondecision time (DDM) and motor processes (EMG), we estimated a linear dependency between T_{er} and MT within the best-fitting model.

As for the behavioral analysis in Experiment 1 and Experiment 2, the data used for the following modeling section was filtered for

---

[7] As for Experiment 1, the frequentist replication shows the same results except the significant interaction between contrast and correctness for PMT ($\beta = -0.06$, $t = 2.45$).

Table 2
Results of the LMM Models Performed on Each Latency Measure From Experiment 2

	RT				PMT				MT			
Predictor	Coeff.	SE	2.5%	97.5%	Coeff.	SE	2.5	97.5%	Coeff.	SE	2.5%	97.5%
Intercept	449	29	393	511	368	26	319	417	72	5	63	83
SAT	160	27	108	213	152	27	100	206	10	2	5	14
Contrast	−94	12	−118	−70	−82	11	−104	−60	−6	1	−7	−4
Correctness	9	7	−5	24	−3	7	−16	12	5	1	4	7
Resp. side	3	8	−14	19	1	6	−11	13	0	3	−5	6
SAT × Contrast	−74	8	−91	−58	−72	8	−88	−55	1	1	−1	3
SAT × Correctness	77	13	53	102	82	14	55	110	−3	1	−5	0
Contrast × Correctness	14	13	−9	40	−5	12	−28	19	9	1	6	12
SAT × Contr. × Corr.	142	27	92	200	158	30	100	218	−5	2	−9	−1

Note. RT = reaction time; PMT = premotor time; M = motor time; Coeff. = estimated coefficient of the linear mixed model (LMM) fitted on the log scale and back-transformed to the millisecond scale; *SE* represent standard error for the estimate; SAT = speed–accuracy trade-off; 2.5% and 97.5% = respectively, the lower and upper Bayesian credible interval (CrI) around the estimate. The intercepts correspond to the mean value of the latency when all the predictors are kept at 0. Main effects can be interpreted as the change in the mean value when the other predictors are kept at null value. Interactions can be read as the change in the effect of the main effects when the other variable is added (cf. section on Statistical Analysis of Experiment 1).

fast-guesses using the EWMA method and trials with RT longer than 1,500 ms or with ambiguous EMG onset were discarded. No other filtering was applied.

Model Estimation

We used a hierarchical Bayesian estimation method for the model fit. As discussed in the Methods of Experiment 1 (section on the behavioral analysis), a hierarchical Bayesian estimation of a cognitive model assumes that each parameter of a participant is drawn from a population distribution described by hyper-parameters, often the mean and variance of a normal distribution. This method preserves the uncertainties associated with the parameter values (because of its Bayesian procedure) while sharing the information between participants to estimate individual parameters (because of its hierarchical nature). As was done for the LME in the behavioral analysis, we only report the hyper-parameter of the estimated population mean for each parameter.

We used the implementation of a hierarchical Bayesian DDM provided in the *HDDM* python package (Wiecki et al., 2013). For each model, both in the Model Selection section below and the model including MT as a covariate, we ran 18,500 burn-in samples and 1,500 actual recorded samples across four Markov chains Monte-Carlo (MCMC). We inspected each parameter of each chain visually to assess whether they reached their stationary distribution, and whether the $\hat{R}$ (Gelman & Rubin, 1992) was under the conventional threshold of 1.01. Additionally, we examined the autocorrelation of each chain to ensure that samples were drawn independently. For the priors, because our design is canonical and to ease convergence, we used the default informative priors used in HDDM based on the work of Matzke and Wagenmakers (2009).

Model Selection

We designed a base model and added parameters according to our hypothesis. The base model was chosen based on previous studies and on the data reported in the previous section. For this base model, the boundary parameter was free to vary with SAT instructions, as it is thought to capture SAT changes. The drift rate was free to vary with the contrast, because this parameter has been

shown to be associated with stimulus strength. The nondecision time was free to vary with SAT, because it has been observed (including in the present report) that this parameter also varies with SAT conditions (Palmer et al., 2005; Ratcliff, 2006; Voss et al., 2004). The accumulation starting point was assumed to be constant because the boundaries were accuracy-coded (correct and incorrect). We also added intertrial variability of the drift rate and the nondecision time, because of their ability to reduce the influence of contaminant fast-trials (Lerche et al., 2017). Finally, we added the intertrial variability of the starting point parameter which was free to vary with SAT instructions, because the data analysis clearly shows that the speed of errors compared with the speed of correct responses does change according to the SAT condition. Almost all parameters were estimated individually with the constrain of being drawn from a common normal distribution (or half-normal depending on the boundaries, e.g., variability parameters cannot have a negative value). Only the intertrial variability parameters of the drift rate, of the starting point, and of the nondecision time were estimated at the group-level because they are notoriously difficult to estimate (Boehm et al., 2018; Wiecki et al., 2016).

In addition to the base model, we also tested the following hypothesis, and combinations thereof: whether the drift rate also varies with SAT, as in Rae et al. (2014), whether the T_{er} varies with the response side, as was observed for MT, and whether the model needs to account for potential bias in the starting point of accumulation.[8] The various possible combinations of these hypothesis is summarized in Table F1.

We used the deviance information criterion (DIC) to select among competing models. The DIC is an analog to the Akaiake

---

[8] Estimating the starting point required a change in the coding of the boundaries, from "correct" and "incorrect" to "left" and "right" responses. This change in coding does also change the meaning of the drift rate as it will represent evidence in favor of left/right instead of evidence for correct/incorrect. To keep the same meaning, and to avoid estimating a drift rate for each side (times the number of stimulus strength levels), we simply took the positive or negative sign according to the side of the correct response. Note that this last modification does not allow to recover a left/right bias in the drift rate but still allows to estimate a starting point bias between left and right responses.

information criterion (AIC) generalized to the hierarchical Bayesian estimation method, in which the improvement of the log-likelihood is weighted against the cost of additional parameters. Because, it has repeatedly been shown that DIC tends to select overfitted models, we also report for each model the Bayesian predictive information criterion (Ando, 2007, BPIC). BPIC is intended to correct DIC's bias in favor of overfitted models by increasing the penalty term for the number of parameters. For all these measures, a lower value of DIC or BPIC indicates a preferred model.

Assessing Covariance

There are two reasons why we cannot use standard correlation coefficients to evaluate between-participants covariance of T_{er} and MT. First, by taking a point estimate, a correlation coefficient ignores the uncertainty associated with the parameter estimation procedure. Second, point estimates are assumed to be independent from one another, which is not the case when using a hierarchical estimation method. To overcome these two issues, we used the plausible values method developed in Ly et al. (2017). This method consists in drawing participants' parameters (i.e., plausible values) from the posterior distribution and correlating them with the variable of interest at each draw. To generalize this sample plausible correlation distribution to the population, we then used an analytic posterior method (see Ly et al., 2017) using an R code provided with the evidence accumulation model fitting package Dynamic Model of Choice (Heathcote et al., 2019).

To evaluate the by-trial covariance between T_{er} and MT we use the implementation of a regression model provided by the HDDM package. This implementation allows to specify a by-trial linear relationship between a model parameter and an external covariate

$$T_{er} = \alpha + \beta MT \qquad (7)$$

where α is the intercept of the linear model, that is, the mean T_{er} when the (centered) MT is at 0, and β is the slope of the predictor MT, hence the linear relationship between T_{er} and MT. This by-trial covariance was estimated for each participant individually, while keeping the hierarchical constrain.

Results and Discussion

Analysis of the Model

The model selection analysis shows that including a modulation of the starting point always leads to lower criteria and that allowing the drift rate to vary with SAT conditions leads to lower DIC and lower BPIC. We therefore chose to use the model with starting point estimation and drift free to vary between stimulus strength level and SAT conditions for both experiments (see Appendix F for the complete analysis of the model selection procedure). A summary of the estimated values of the parameters from the winning model is presented in Table 3, and an average representation of the goodness-of-fit for the selected model is provided in Figure 6. We provide a representation of the posterior distributions for each parameter that was free to vary in the different conditions in Figure 7. Table 3 and Figure 7 reveal a remarkable stability in the estimation of parameters across the two experiments, despite them involving two different samples of participants. We will not provide here a precise description of how the different manipula-

Table 3

Posterior Distributions for the Estimated Parameters of the Selected DDM

Parameter	Experiment 1			Experiment 2		
	M	2.5%	97.5%	M	2.5%	97.5%
a_{Acc}	1.18	1.10	1.26	1.23	1.12	1.34
a_{Spd}	0.93	0.86	1.01	0.88	0.78	1.00
v_{Acc1}	0.22	−0.09	0.55	0.24	−0.16	0.64
v_{Acc2}	0.70	0.37	1.02			
v_{Acc3}	2.04	1.70	2.38	2.28	1.89	2.69
v_{Acc4}	3.52	3.17	3.89	3.94	3.53	4.36
v_{Acc5}	4.52	4.14	4.90			
v_{Spd1}	0.21	−0.12	0.54	0.27	−0.13	0.68
v_{Spd2}	0.70	0.34	1.03			
v_{Spd3}	2.04	1.70	2.39	2.43	2.02	2.84
v_{Spd4}	3.61	3.25	3.98	4.21	3.79	4.63
v_{Spd5}	4.79	4.41	5.17			
T_{erAcc}	0.33	0.31	0.35	0.33	0.30	0.36
T_{erSpd}	0.29	0.28	0.31	0.28	0.25	0.32
sv	0.90	0.78	1.02	0.72	0.60	0.84
sz_{Acc}	0.17	0.02	0.29	0.25	0.15	0.33
sz_{Spd}	0.65	0.60	0.68	0.67	0.64	0.71
st	0.09	0.08	0.09	0.07	0.06	0.07
z	0.54	0.52	0.56	0.50	0.48	0.52

Note. DDM = drift-diffusion model; 2.5% and 97.5% represent boundaries of Bayesian credible interval (CrI) of the mean. Note that T_{er} is on the second, rather than millisecond, scale.

tions map onto DDM parameters, because the manipulations and the results are congruent with those reported in the literature.

We started with a comparison between the SAT effect size quantified in the winning model and in the EMG results, bearing in mind that both estimates stem from different methods: subtraction of both SAT levels for T_{er} versus LMM estimate for MT. The mean estimated SAT effect on the T_{er}, computed on the subtraction of estimated T_{er} in speed from the estimated T_{er} in accuracy at each MCMC iteration, was of 38 ms (2.5% = 11 ms, 97.5% = 64 ms) in Experiment 1 and 42 ms in Experiment 2 (2.5% = 0 ms, 97.5% = 84 ms). For MT, the observed effects were 15 ms (2.5% = 9 ms, 97.5% = 22 ms) and 10 ms (2.5% = 5 ms, 97.5% = 14 ms), respectively, in Experiments 1 and 2. If we were to assume a direct link between T_{er} and MT, we would need to account for the observation that the effect on T_{er} was approximately twice the effect observed on MT. One ad hoc assumption would be that encoding processes are also affected by the SAT instructions, to a comparable extent, however we cannot yet exclude the possibility that the SAT effect could also/partially be contained in motor processes that precede the EMG activity onset.

Between-Participants Dependence of MT and T_{er}

Having estimated a T_{er} parameter for each SAT condition, we can perform the plausible population correlation separately in each condition and for each experiment. In Experiment 1 the plausible population correlation was positive and high, but with high uncertainty in both the accuracy (m = 0.60, 2.5% = 0.17, 97.5% = 0.85) and the speed conditions (m = 0.56, 2.5% = 0.11, 97.5% = 0.83). In Experiment 2, in contrast, the plausible population correlation was lower and its CrI contained 0 in both accuracy (m = 0.24, 2.5% = −0.24, 97.5% = 0.63) and speed conditions (m = 0.20, 2.5% = −0.28, 97.5% = 0.60). Although posterior distri-

Figure 6

Quantile-Probability Plots (Ratcliff & McKoon, 2008) for Experiment 1 (Left Column) and Experiment 2 (Right Column), in the Accuracy (Upper Row) and Speed (Lower Row) Conditions, Computed From the Winning Model
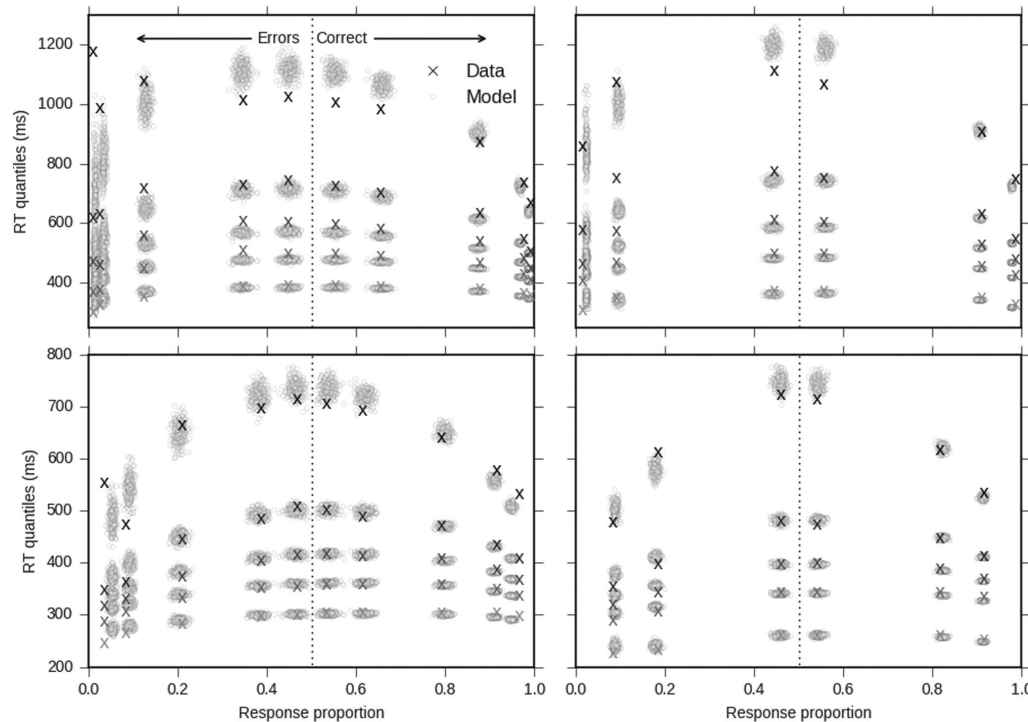
Note. The *x* axis displays obtained response proportion across contrast levels, symmetrically for errors (left side) and correct responses (right side). The *y* axis displays the fitted (dot) and observed (cross) RT binned in 5 quantiles (.1, .3, .5, .7, and .9 quantiles, from bottom to top). Observed response proportion and RT quantiles were computed from values pooled across participants. Model predictions were obtained by drawing 500 parameter values from the joint posterior distribution and computing their associated predicted performance. The substantially larger misfits observed on the far left of the QP-plot for Experiment 1 in the accuracy condition are likely attributable to the lower amount of errors in these conditions.

butions of plausible correlation values clearly overlap between Experiment 1 and 2 (see Figure 8), hindering any strong conclusion, we speculate that the higher values in Experiment 1 could be linked to the higher force requirement therein. We also observed that the pattern is consistent within the experiments where values are always higher for accuracy (which yields longer MT) than speed conditions. Hence, when a modeler is searching to interpret the T_{er}, it is highly likely that the amount of information about the motor system one can get by comparing T_{er} parameter between participants depends on the amount of time spent on executing the response.

Within-Participant Link Between MT and T_{er}

We added the MT as a covariate of the T_{er} to the models selected for Experiments 1 and 2. The estimated slope linking T_{er} and MT indicates that MT weakly but consistently covaries with T_{er} in

Experiment 1 (β = 0.22, 2.5q = 0.10, 97.5q = 0.34) and in Experiment 2 (β = 0.27, 2.5q = 0.12, 97.5q = 0.41). Although the population estimate is rather concentrated, at the individual level the mean of the posterior distribution on the β coefficient varies from -0.10 to.78 across both experiments. This would indicate that, for some participants at least, measured MT does not covary with estimated T_{er}, although it should be noted that this could also be due to measurement error.

Source of Between Participant Variation in the By-Trial Covariance

The previous section showed that there is a high variability, between participants, in the degree to which T_{er} and by-trial MT covary. We thought to link this variability with the previous finding from the EMG section which shows that, for fast decision,
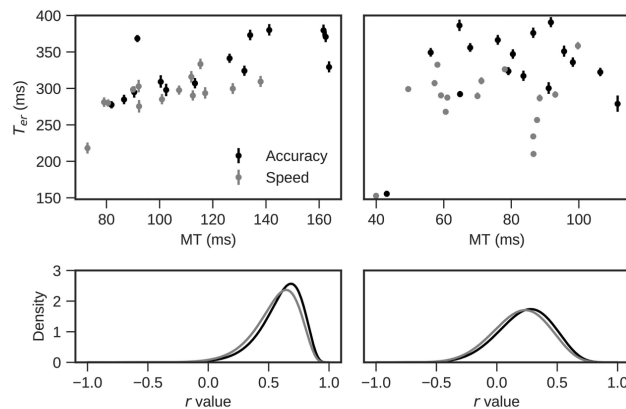
Figure 7

Posterior Distributions of the Parameters Left Free to Vary Across Conditions
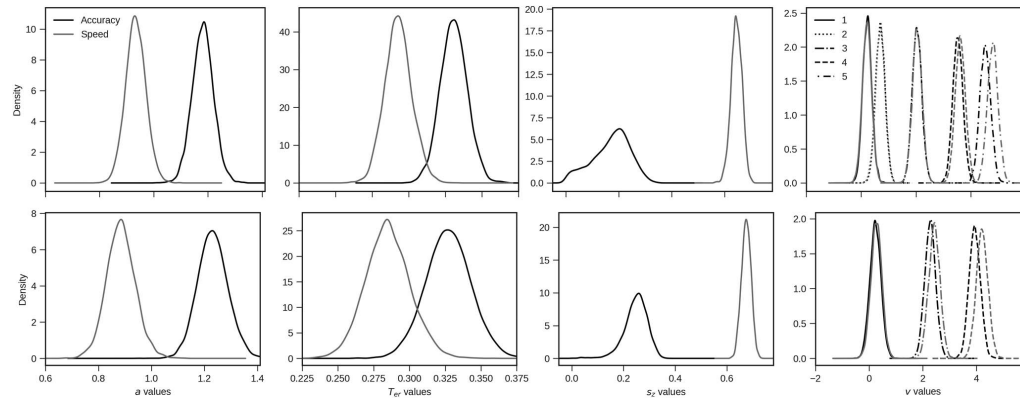
Note. Top row is for Experiment 1, bottom row for Experiment 2. a, T_{er}, s_z, v stand for boundary separation, nondecision time, intertrial variability of the starting point, and drift rate. For the drift rate (last panel), numbers in the legend represent contrast levels.

PMT and MT are negatively correlated. Such correlation violates the independence assumption of the model. One can hence assume that the larger this violation, the lower the capacity of the model to fit the data and to correctly recover the parameters. As a consequence, the relationship between T_{er} and MT should be all the weaker that the negative correlation is high. To test this hypothesis we took the average Spearman correlation value between PMT and MT across all trials in both SAT conditions for each participant.

We then test the plausible population correlation between the estimated β of the linear regression between T_{er} and MT and the correlation between PMT-MT.

From this analysis (see Figure 9) we see that there is a high association between the regression $β_{MT}$ and the value of the correlation between PMT and MT, both in Experiment 1 (m = 0.50, 2.5% = 0.03, 97.5% = 0.81) and even stronger in Experiment 2 (m = 0.70, 2.5% = 0.36, 97.5% = 0.89). Hence this

Figure 8

Correlation Between the Estimated T_{er} and the Measured Motor Time (MT), Computed by Participant

Note. Upper row, Scatterplots of the *MT* and T_{er} pairs for each participant. Lower row, Plausible population correlation density. Left column, Experiment 1; Right column, Experiment 2.

Figure 9

Representation of the Correlation Between the Regression Coefficient of Motor Time (MT) Over T_{er} Plotted Against the Correlation Between Premotor Time (PMT) and MT
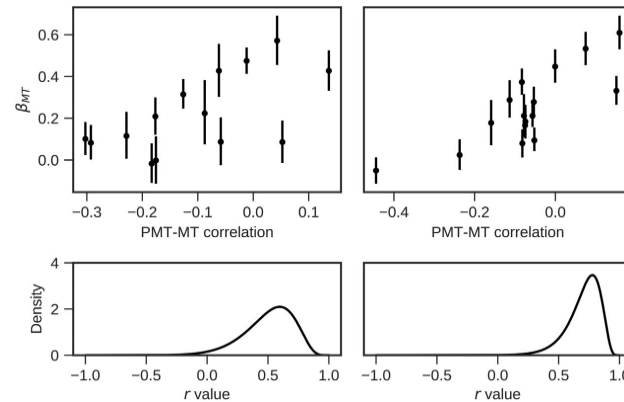
Note. Upper panels depicts scatterplots of both variables, whereas lower panels represent plausible population correlation density for Experiment 1 (left) and Experiment 2 (right).

analysis suggest that the less negative the correlation between PMT and MT is, the better T_{er} recovers MT.

General Discussion

Our understanding of mental functions is largely owed to chronometric analyses of performance (Luce, 1986). Both experimental and mathematical psychologists assume that response times (RTs) result from different subprocesses, although the nature and the temporal relationship between these processes has long been, and still is, largely debated (e.g., Dubarry et al., 2017; McClelland, 1979; Miller, 1988; Requin et al., 1988; Sternberg, 1969). A minimal, commonly accepted form of RT decomposition assumes that response latencies result from (at least) two components: a decision time (DT) and a residual nondecision time (T_{er}), where the T_{er} itself can be decomposed into an encoding time (T_e) and a motor execution time (T_r). In current and widely used practice, these components and subcomponents are assumed to be additive and independent (see Eq. 1 and 2), two subsidiary assumptions embraced by many formal models of decision making since their original formulation (Stone, 1960).

While experimental/mathematical psychologists and neuroscientists aim to understand the processes leading to behavior, the main dependent variable, the RT, only gives access to the overall outcome of all those processes, and does not reveal, by itself, the individual underlying processes. Although analyzing adequate experimental manipulations, and their interaction patterns, has proved very powerful in RT research (see, e.g., Sternberg, 1969), the validity of the inferences one can do on the individual processes is difficult to assess (see, e.g., Pieters, 1983). To better study separately the underlying cognitive processes, two main approaches have been followed. The first consists in using physiological variables (e.g., Coles, 1989; Requin et al., 1988) which

are thought to better reflect the intermediate processes. Such approach, however, requires to explicitly define "linking functions" (Schall, 2004; Teller, 1984) mapping physiological measures into cognitive processes, which is not as easy as it may seem (Schall, 2019).

The second approach is to use process models that are fitted to empirical data (RTs and error rates) to derive inferences about latent cognitive processes, which may be reflected in the estimated parameters. The validity of the inferences made hinges on the validity of the process model assumptions themselves. Despite the importance of assessing such validity, however, few empirical studies have directly put these assumptions to the test.

In recent years, several studies have yoked behavioral and physiological markers to constrain, and hopefully improve, the recovery of underlying cognitive processes (Ravenzwaaij et al., 2017; Turner et al., 2015). The perspective followed in the present work is (partially) different and motivated by the vast majority of studies where formal decision-making models are fitted to data solely on the basis of behavioral measures. We aimed at estimating the information provided by the nondecision parameter of a decision making model (namely, the DDM fitted on response times) about the motor processes captured through EMG decomposition. In other words, we used physiological data as an external validator, to assess whether the inferences drawn from model fits are compatible with the independent information provided by the EMG decomposition.

As argued in the Introduction, for such validation to be effective, there must be a strong link between the motor execution part of the T_{er} and the MT. Any discrepancy will either be generated by an incorrect linking function between MT and T_{er} (which is unlikely) or by an inappropriately specified model. Importantly, identifying a mis-estimation of the T_{er} signals the risk of a mis-

specification of the other parameters, because the correct estimation of any parameter depends on the correct estimation of the others.

In the present study, the two RT decompositions, one physiological and the other model-based, were first performed separately, and then combined. We will hence first discuss the conclusions one may reach based on each approach individually, and then evaluate the added value of combining them.

Relationship Between Decisional and Motor Processes Based on Physiological Fractionation of RT

Recent studies analyzing movement trajectories in reaching tasks have revealed that motor execution can be affected by decision processes (Buc Calderon et al., 2015; Dotan et al., 2018; Resulaj et al., 2009), thus questioning the assumption of independence between these two levels. Despite their theoretical importance, these prior results were obtained in experimental settings (reaching behaviors) that depart from the vast majority of studies (simple key-pressing) in which formal decision making models are fitted, and on which many cognitive inferences have been made. Such experimental settings can afford strategic adjustments like optimizing the distance to the targets according to the online evolution of the decision parameter.

In the present study, we tested the relationship between decisional and motoric processes using conventional setups but added a neurophysiological measure on each trial.

In the two experiments, we manipulated perceptual difficulty and SAT conditions, and, between experiments, the required response force was adjusted to further verify our findings. All manipulations were successful and produced clear and expected effects on the RTs. With the noticeable exception of response side (affecting performance in Experiment 1 but not in Experiment 2), all factors affecting the RTs did so by affecting PMTs. This was not surprising. More notably, MT was also affected by several experimental manipulations that we now detail.

Stimulus contrast and SAT impacted not only PMT, but also MT; the effects on the two variables were in the same direction (i.e., shorter PMTs go with shorter MTs). This replicates the observation that SAT has an impact (approximately 10% of the SAT effect on RT and 14% of the mean MT) on motor execution (Spieser et al., 2017; Steinemann et al., 2018). Whereas effects of SAT on T_{er} estimates are regularly observed, to date the genuine nature of the SAT effect on T_{er} (specifically motor processes) is still debated and yet to be validated (Lerche & Voss, 2018; Smith & Lilburn, 2020). Our work, which assessed this phenomenon in a canonical, perceptual decision-making task with an isometric responding format, has provided a proof, or validated evidence by measures of muscular activity, that the motor processes assumed to be contained in the T_{er} are indeed sensitive to SAT.

As already discussed above, the effect of stimulus contrast on MT (approximately 7% of the effect of contrast on RT and 8% of the mean of MT) is somehow more surprising, but appears very reliable despite a small effect size. While the origin of this effect remains to be deciphered (as a "cognitive" or "energetic" process, see Sanders, 1983), one will retain here that stimulus contrast also modulates the two variables in the same direction. In contrast, replicating previous observations made in the context of conflict tasks (Allain et al., 2004; Rochet et al., 2014; Śmigasiewicz et al.,

2020), correctness affected the two intervals in opposite directions: on errors, PMT was shorter while MT was longer, an effect that was enhanced by stimulus contrast. Comparing the two experiments, revealed that changing response force had a large effect on MT, but not on PMT (if any, there would be a trend in the opposite direction). Finally, in Experiment 1 only, response side selectively affected MT, but spared PMT.

Two main conclusions emerge from this brief summary. First, the recorded MT is not merely a constant factor, because it can be modulated by several experimental factors, including cognitive processes. Second, some factors selectively influence one of the two intervals, some affect both in a similar way, and others do so in an opposite way. Altogether, these results indicate that PMT and MT can be manipulated independently. In the literature, it has been repeatedly shown that certain experimental manipulations, such as stimulus-response compatibility, selectively affect PMT and leave MT unchanged (Burle et al., 2002; Hasbroucq et al., 1999; Spieser et al., 2017). A mirror selective influence was recently observed: Dopaminergic-agonist treatment on Parkinson's disease patients shortens RT by selectively affecting MT, while leaving PMT unchanged (Fluchère et al., 2018). Overall, these findings reveal a form of independence between the two intervals, and hence of the underlying processes they index. Stating such independence does not imply that PMT and MT are solely affected by disjoint groups of factors. As a matter of fact, contrast (i.e., stimulus quality) and SAT affect both PMT and MT.

The independence between PMT and MT is further strengthened by the absence of correlation between the two intervals for most of the trials, as observed consistently in the two experiments. Crucially, however, this does not hold for the trials with very short PMTs, in both SAT conditions, where a negative correlation between the two intervals is observed. From a functional point of view, this variation in stochastic dependency suggests that, even within a condition, the temporal relationship between the underlying processes is variable, and hence that fast trials might not arise from the same generative model as slower ones. We will come back to this issue later.

Model-Based Decomposition and Comparison With Physiological Fractionation

We first discuss the results from the model fit by themselves, irrespective of the information gained from EMG. As expected, contrast had a clear effect on the drift rate, and SAT on response threshold. Besides those canonical effects, SAT also affected the drift rate and T_{er}, confirming that SAT is not purely a threshold effect. Interestingly, the force manipulation between the two experiments was not captured by any difference in T_{er}, which, absent any RT fractionation, would have lead to the conclusion that response force does not affect response execution, but rather upstream processes.[9]

The herein observed presence of systematic effects on MT puts into question the widely practiced assumption that the T_{er} parameter is insensitive to experimental manipulations which affect decisional processes, such as stimulus difficulty. Many modelers, out of convenience, implement the DDM with the assumption that

---

[9] Unless one assumes that the effect in T_r is perfectly compensated by an opposite effect on T_e which, as argued in the introduction, is very unlikely.

T_{er} does not vary across experimental conditions (i.e., this reduces the number of free parameters to estimate). The present results challenge this strategy. It is highly likely that when true effects are ignored by fixing nondecision time (e.g., Ratcliff & McKoon, 2008; Ratcliff & Rouder, 1998), the estimation of the other model parameters will be biased. While this is probably true for factors with moderate effect sizes such as SAT or response side on MT in our experiment, it should be of lesser importance for factors with smaller effect sizes, such as contrast and correctness (although the two interact, leading to larger effect of correctness for the highest contrasts).

In our model selection procedure, T_{er} was left free to vary as a function of SAT and response side for some models, but was always fixed for contrast and correctness. We note that current implementations of the DDM do not allow any parameter to vary between correct and error responses, because correctness is the product of decision related parameters. As a consequence, the model in its current state is not equipped to capture correctness effects on MT. In a similar vein, despite an effect of stimulus contrast on MT, T_{er} was not allowed to vary as a function of contrast because (a) this effect on MT is small and the cost of 5 additional parameters was judged too high and (b) contrast interacts with correctness and this interaction cannot be recovered. Hence, owing to the model fitting method, we were limited to models with the nondecision parameter free to vary with SAT, response side and their interaction. We also tested for an effect of SAT on drift rate as well as the presence of a bias in the starting point. Although we do believe that our modeling strategy was the best possible, the empirical data indicates that it is not perfect.

After the initial model selection procedure suggested that boundaries, drift rate and nondecision time vary with the speed emphasis in our experiment, we showed that, the by-participant mean MT and the by-participant T_{er} estimated on behavior alone share a close relationship (see Figure 8) in Experiment 1 but not in Experiment 2. At first sight, this could provide an argument against the idea that MT is indeed contained in the T_{er}, which would be problematic for the model assumptions. However, a closer look leads to a more nuanced interpretation. Indeed, their actual covariation depends on how much of T_{er} represents the motor processes. If one reasonably assumes that T_e (encoding time) is approximately the same for the two experiments because the same stimuli are used, then the proportion of T_r (execution time) into T_{er} is lower in Experiment 2 (shorter MT due to lower force) than in Experiment 1 (longer MT). The contribution of T_e (that should not correlate with MT) is hence larger for Experiment 2 than Experiment 1.

It follows from these considerations that the pattern of substantial correlation for high force and much reduced for lower force is actually predicted by, or consistent with, the model. A potentially more problematic result is the observation that, despite a substantial reduction in the force needed to execute the response, captured by MT, as shown by the lower intercept for the LMM between Table 2 and Table 1, the mean nondecision time remains almost exactly the same in both experiments (see Table 3). This results implies that participants in Experiment 2, that is, when the force needed to execute the response is lower, have a process inside the PMT and captured by the T_{er} parameter that is slower than it was for participants in Experiment 1. However in this study we cannot adjudicate between a genuine cognitive effect, a participant sam-

pling bias, a consequence of the design modification between both experiments, or a misattribution of a motoric effect by the DDM. Hence, another study containing a within-participant manipulation of the force is needed to test these competing hypothesis. It is also important to note that the estimated SAT effect on T_{er} was twice as big as the effect observed on MT. There are two possible explanations of this result: either the SAT effect could be contained in the remaining motor processes preceding the onset of EMG, or SAT could also affect early nondecision processes such as the T_e.

We do not have direct (physiological) measure of T_e. However, if we look at the value of T_{er} and the mean values of MT we see that subtracting MT from T_{er} would result in a residual time of approximately 200 ms, although the estimated value varies across experiments. This is fairly close to previous estimations of visual encoding times ($\simeq$ 200 msec) made by Roitman and Shadlen (2002) using intracerebral recordings in monkeys or Nunez et al. (2019) using EEG in humans. Although this result has to be taken with caution, its consistency with previous estimates is noticeable.

We finally tested for a linear dependency between the EMG measured MT and the T_{er} by allowing a by-trial dependency between both measure in the estimation procedure. The corresponding fits show that the by-trial recorded MT predicts the T_{er}. We do however observe that this linear dependency is variable across participants, with some participants presenting a low-to-null dependency. We linked this interindividual variability in the dependency between MT and T_{er} with the interindividual difference in correlation between PMT and MT, showing that the more negatively the two intervals were correlated, the less there was a link between MT and T_{er}. This is important because it shows that when we depart from the independence assumption of the DDM, the MT no longer shows a relationship with the modeling estimation of nondecision processes.

Taken together, our findings indicate a fair concordance between decision-making models (e.g., the DDM and its parameters) and EMG-derived chronometry measures. There are cases, however, in which the model fitting completely missed, or misattributed, some of the empirically observed effects. Although some coherent patterns can already be suspected, it is certainly too early to try to establish which types of effects can be appropriately recovered and which cannot. The present methodology, used systematically to probe other experimental manipulations may help to better delineate when the inferences based on model fitting are likely to be valid, and when one can doubt about them. In turn, this will certainly help to asses the strengths and weaknesses of the current models, and allow to improve them.

A Single Model With Parametric Modulations?

A last comment is in order. The core architecture of the DDM is intended to be generic, geared to account for data obtained in large variety of situations. These situations will induce changes in the parameters that capture different processes, always within the very same core architecture. The correlations observed between PMT and MT, however, challenge this view. Indeed, while no correlation was observed between these two components in most of the trials, in agreement with the additive assumption (see Equation 1), a negative correlation was ob-

served for short PMTs.[10] This has several consequences. These negatively correlated trials cannot easily be mapped into any parameter adjustment. Accordingly, participants' speed-up does not only reflect the modulation of one or another subprocess (speed of accumulation, response caution, etc.), but also partly reflects a change in the generative model of the decisions (see Dutilh et al., 2011; Heitz, 2014, for discussions). The hypothesis of a different generative model is strengthened by the fact that the more negative the correlation between the two EMG intervals is, the less correspondence we observe between estimated nondecision time and motor time (see Figure 9).

One approach that could accommodate these considerations might be a model in which decisions performed under SAT are generated from a mixture between fast-guesses and slow controlled decisions (an hypothesis formulated by Ollman, 1966; for a review see Heitz, 2014). The fact that trials identified as fast-guesses by the EWMA method (Vandekerckhove & Tuerlinckx, 2007) present a high negative correlation between PMT and MT, whereas the other trials do not, supports this hypothesis. However additional analysis are needed to explore the nature of the generative model on these trials.

Even more than the difficulties the DDM had in accounting for some effects obtained on MT, such as the difference between correct and incorrect trials, a change in the architecture would strongly challenge the idea that DDM is a generic model that can account for the dynamics of information processing across response speeds. It must be noted, however, that despite this conceptual mismatch between the model's single process architecture and the presumed mixture of strategies the data provides evidence for, the model nonetheless efficiently fitted the data in the speed condition (see lower panels of Figure 6). This confirms that a good fit, although necessary, is not a sufficient argument for an agreement between the empirical observations and interpretative cognitive validity.

Conclusion

Through two experiments, we reliably showed that response execution time measured with EMG is far from being a ballistic movement subsequent to the decision. This conclusion is in contrast to what is assumed in major decision-making models and their fitting applications. Rather, response execution time is sensitive to the Speed-Accuracy Trade-off, stimulus strength, and correctness of the response. Importantly, EMG recording allows for the measurement of effect sizes, in turn allowing researchers to assess the impact of ignoring such variations. Additionally, we showed that a subsidiary assumption of most models of RT, the independence between decision and nondecision time, while potentially valid in most of the trials, is violated on the early quantiles (i.e., the fastest decision) of the PMT distribution.

Despite such discrepancies between our observations and model assumptions, we show that although relying on only behavioral measures to estimate the nondecisional latencies, the DDM nondecision time parameter provides, under some conditions, an accurate representation of the EMG-observed response execution time. This observation is true both between participants, when the amount of motor processes in the RT is fairly high, and within participants, as is shown with the linear dependency between the by-trial EMG MT latency and the estimated T_{er}. However, the

by-trial negative correlation between premotor and motor time on faster trials, points to possible trade-offs between latency components that are dependent on task requirements. This could imply that parameter interpretability (or attribution) can change according to the validity of the independence assumption.

---

[10] Importantly this is true even when considering only correct responses or computing correlations within each stimulus contrast condition.

References

Allain, S., Carbonnell, L., Burle, B., Hasbroucq, T., & Vidal, F. (2004). On-line executive control: An electromyographic study. *Psychophysiology*, *41*(1), 113–116. https://doi.org/10.1111/j.1469-8986.2003.00136.x

Anders, R., Alario, F. X., & Maanen, L. van. (2016). The shifted wald distribution for response time data analysis. *Psychological Methods*, *21*(3), 309–327. https://doi.org/10.1037/met0000066

Anders, R., Oravecz, Z., & Alario, F.-X. (2018). Improved information pooling for hierarchical cognitive models through multiple and covaried regression. *Behavior Research Methods*, *50*(3), 989–1010.

Anders, R., Riès, S., van Maanen, L., & Alario, F. X. (2017). Lesions to the left lateral prefrontal cortex impair decision threshold adjustment for lexical selection. *Cognitive Neuropsychology*, *34*(1–2), 1–20. https://doi.org/10.1080/02643294.2017.1282447

Anders, R., Van Maanen, L., & Alario, F.-X. (2019). Multi-factor analysis in language production: Sequential sampling models mimic and extend regression results. *Cognitive Neuropsychology*, *36*(5–6), 234–264. https://doi.org/10.1080/02643294.2019.1610371

Ando, T. (2007). Bayesian predictive information criterion for the evaluation of hierarchical Bayesian and empirical Bayes models. *Biometrika*, *94*(2), 443–458.

Barr, D. J. (2013). Random effects structure for testing interactions in linear mixed-effects models. *Frontiers in Psychology*, *4*, 328. https://doi.org/10.3389/fpsyg.2013.00328

Bartlett, N. R. (1963). A comparison of manual reaction times as measured by three sensitive indices. *Psychological Record*, *13*, 51–56.

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., Wiecki0, 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.

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

Brunton, B. W., Botvinick, M. M., & Brody, C. D. (2013). Rats and humans can optimally accumulate evidence for decision-making. *Science*, *340*(6128), 95–98. https://doi.org/10.1126/science.1233912

Buc Calderon, C., Verguts, T., & Gevers, W. (2015). Losing the boundary: Cognition biases action well after action selection. *Journal of Experimental Psychology: General*, *144*(4), 737.

Burle, B., Possamaï, C.-A., Vidal, F., Bonnet, M., & Hasbroucq, T. (2002). Executive control in the Simon effect: An electromyographic and distributional analysis. *Psychological Research*, *66*(4), 324–36. https://doi.org/10.1007/s00426-002-0105-6

Burle, B., Roger, C., Vidal, F., & Hasbroucq, T. (2008). Spatio-temporal dynamics of information processing in the brain: Recent advances, current limitations and future challenges. *International Journal of Bioelectromagnetism*, *10*(1), 17–21.

Callaway, E., Halliday, R., Naylor, H., & Thouvenin, D. (1984). The latency of the average is not the average of the latencies. *Psychophysiology*, *21*, 571.

Coles, M. G. (1989). Modern mind-brain reading: Psychophysiology, physiology, and cognition. *Psychophysiology*, *26*(3), 251–269.

Cousineau, D. (2005). Confidence intervals in within-subject designs: A simpler solution to Loftus and Masson's method. *Tutorials in Quantitative Methods for Psychology*, *1*(1), 42–45. https://doi.org/10.20982/tqmp.01.1.p042

Donkin, C., Brown, S., Heathcote, A., & Wagenmakers, E. J. (2011). Diffusion versus linear ballistic accumulation: Different models but the same conclusions about psychological processes? *Psychonomic Bulletin & Review*, *18*(1), 61–69. https://doi.org/10.3758/s13423-010-0022-4

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*, *19*(18), 1581–1585. https://doi.org/10.1016/j.cub.2009.07.066

Dotan, D., Meyniel, F., & Dehaene, S. (2018). On-line confidence monitoring during decision making. *Cognition*, *171*, 112–121.

Dubarry, A.-S., Llorens, A., Trébuchon, A., Carron, R., Liégeois-Chauvel, C., Bénar, C.-G., & Alario, F.-X. (2017). Estimating parallel processing in a language task using single-trial intracerebral electroencephalography. *Psychological Science*, *28*(4), 414–426. https://doi.org/10.1177/0956797616681296

Dutilh, G., Jeffrey, A., Brown, S. D., Cassey, P., Evans, N. J., Grasman, R. P. P. P., Hawkins, G. E., Heathcote, A., Holmes, W. R., Krypotos, A.-M., Kupitz, C. N., Leite, F. P., Lerche, V., Lin, Y.-S., Logan, G. D., Palmeri, T. J., Starns, J. J., Trueblood, J. S., van Maanen, L., . . . Donkin, C. (2019). The quality of response time data inference: a blinded, collaborative assessment of the validity of cognitive models. *Psychonomic Bulletin & Review*, *26*(4), 1051–1069. https://doi.org/10.3758/s13423-017-1417-2

Dutilh, G., Wagenmakers, E.-J., Visser, I., & van der Maas, H. L. J. (2011). A phase transition model for the speed-accuracy trade-off in response time experiments. *Cognitive Science*, *35*, 211–250. https://doi.org/10.1111/j.1551-6709.2010.01147.x

Fluchère, F., Burle, B., Vidal, F., van den Wildenberg, W. P. M., Witjas, T., Eusebio, A., Azulay, J. P., & Hasbroucq, T. (2018). Subthalamic nucleus stimulation, dopaminergic treatment and impulsivity in Parkinson's disease. *Neuropsychologia*, *117*, 167–177.

Gelman, A., & Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science*, *7*(4), 457–472.

Gomez, P., Perea, M., & Ratcliff, R. (2013). A diffusion model account of masked versus unmasked priming: Are they qualitatively different? *Journal of Experimental Psychology: Human Perception and Performance*, *39*(6), 1731–1740. https://doi.org/10.1037/a0032333

Gomez, P., Ratcliff, R., & Childers, R. (2015). Pointing, looking at, and pressing keys: A diffusion model account of response modality. *Journal of Experimental Psychology: Human Perception and Performance*, *41*(6), 1515–1523. https://doi.org/10.1037/a0039653

Gramfort, A., Luessi, M., Larson, E., Engemann, D. A., Strohmeier, D., Brodbeck, C., Goj, R., Jas, M., Brooks, T., Parkkonen, L., & Hämäläinen, M. (2013). MEG and EEG data analysis with MNE-Python. *Frontiers in Neuroscience*, *7*, 267. https://doi.org/10.3389/fnins.2013.00267

Grayson, D. (1983). *The role of the response stage in stochastic models of simple reaction time* [Unpublished doctoral dissertation]. University of Sydney.

Grice, G., & Spiker, V. (1979). Speed-accuracy tradeoff in choice reaction time: Within conditions, between conditions, and between subjects. *Attention, Perception & Psychophysics*, *26*(2), 118–126. https://doi.org/10.3758/bf03208305

Hasbroucq, T., Possamaï, C.-A., Bonnet, M., & Vidal, F. (1999). Effect of the irrelevant location of the response signal on choice reaction time: An electromyographic study in humans. *Psychophysiology*, *36*(4), 522–526.

Heathcote, A., Brown, S. D., & Wagenmakers, E.-J. (2015). An introduction to good practices in cognitive modeling. In B. U. Forstmann, & E.-J. Wagenmakers (Eds.), *An introduction to model-based cognitive neuroscience* (pp. 25–48). Springer.

Heathcote, A., Lin, Y.-S., Reynolds, A., Strickland, L., Gretton, M., & Matzke, D. (2019). Dynamic models of choice. *Behavior Research Methods*, *51*(2), 961–985.

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

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

Herz, D. M., Bogacz, R., & Brown, P. (2016). Neuroscience: Impaired decision-making in Parkinson's disease. *Current Biology*, *26*, R671–R673. https://doi.org/10.1016/j.cub.2016.05.075

Kumar, R., Carroll, C., Hartikainen, A., & Martin, O. A. (2019). ArviZ a unified library for exploratory analysis of Bayesian models in Python. *The Journal of Open Source Software*. Advance online publication. https://doi.org/10.21105/joss.01143

Lafuente, V., & de Romo, R. (2006). Neural correlate of subjective sensory experience gradually builds up across cortical areas. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(39), 14266–14271. https://doi.org/10.1073/pnas.0605826103

Latimer, K. W., Yates, J. L., Meister, M. L. R., Huk, A. C., & Pillow, J. W. (2015). Single-trial spike trains in parietal cortex reveal discrete steps during decision-making. *Science*, *349*(6244), 184–187. https://doi.org/10.1126/science.aaa4056

Lawlor, V., Webb, C., Wiecki, T., Frank, M., Trivedi, M., Pizzagalli, D., & Dillon, D. (2020). Dissecting the impact of depression on decision-making. *Psychological Medicine*, *50*(10), 1613–1622.

Lerche, V., & Voss, A. (2018). Speed–accuracy manipulations and diffusion modeling: Lack of discriminant validity of the manipulation or of the parameter estimates? *Behavior Research Methods*, *50*(6), 2568–2585.

Lerche, V., Voss, A., & Nagler, M. (2017). How many trials are required for parameter estimation in diffusion modeling? A comparison of different optimization criteria. *Behavior Research Methods*, *49*(2), 513–537. https://doi.org/10.3758/s13428-016-0740-2

Lewandowski, D., Kurowicka, D., & Joe, H. (2009). Generating random correlation matrices based on vines and extended onion method. *Journal of Multivariate Analysis*, *100*(9), 1989–2001.

Liu, J., & Liu, Q. (2016, feb). Use of the integrated profile for voluntary muscle activity detection using EMG signals with spurious background spikes: A study with incomplete spinal cord injury. *Biomedical Signal Processing and Control*, *24*, 19–24. https://doi.org/10.1016/j.bspc.2015.09.004

Luce, R. (1986). *Response times: Their role in inferring elementary mental organization*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195070019.001.0001

Ly, A., Boehm, U., Heathcote, A., Turner, B. M., Forstmann, B., Marsman, M., & Matzke, D. (2017). A flexible and efficient hierarchical bayesian approach to the exploration of individual differences in cognitive-model-based neuroscience. In A. A. Moustafa (Ed.), *Computational models of brain and behavior* (pp. 467–480). Wiley.

Matzke, D., & Wagenmakers, E.-J. (2009). Psychological interpretation of the ex-gaussian and shifted wald parameters: A diffusion model analysis. *Psychonomic Bulletin & Review*, *16*(5), 798–817.

McClelland, J. L. (1979). On the time relations of mental processes: An examination of systems of processes in cascade. *Psychological Review*, *86*(4), 287–330.

Miller, J. (1988). Discrete and continuous models of human information processing: Theoretical distinctions and empirical results. *Acta Psychologica*, *67*(3), 191–257.

Miller, J., Ulrich, R., & Rinkenauer, G. (1999). Effects of stimulus intensity on the lateralized readiness potential. *Journal of Experimental Psychology: Human Perception and Performance*, *25*(5), 1454–1471. https://doi.org/10.1037/0096-1523.25.5.1454

Moustafa, A. A., Kéri, S., Somlai, Z., Balsdon, T., Frydecka, D., Misiak, B., & White, C. (2015). Drift diffusion model of reward and punishment learning in schizophrenia: Modeling and experimental data. *Behavioural Brain Research*, *291*, 147–154. https://doi.org/10.1016/j.bbr.2015.05.024

Nicenboim, B., Vasishth, S., Engelmann, F., & Suckow, K. (2018). Exploratory and confirmatory analyses in sentence processing: A case study of number interference in German. *Cognitive Science*, *42*(S4), 1075–1100.

Nunez, M. D., Gosai, A., Vandekerckhove, J., & Srinivasan, R. (2019). The latency of a visual evoked potential tracks the onset of decision making. *Neuroimage*, *197*, 93–108.

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

Oliphant, T. E. (2007). SciPy: Open source scientific tools for Python. *Computing in Science and Engineering*, *9*, 10–20. https://doi.org/10.1109/MCSE.2007.58

Ollman, R. (1966). Fast guesses in choice reaction time. *Psychonomic Science*, *6*(4), 155–156.

Osman, A., Lou, L., Muller-Gethmann, H., Rinkenauer, G., Mattes, S., & Ulrich, R. (2000). Mechanisms of speed-accuracy tradeoff: Evidence from covert motor processes. *Biological Psychology*, *51*(2–3), 173–199. https://doi.org/10.1016/S0301-0511(99)00045-9

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), 1. https://doi.org/10.1167/5.5.1

Pe, M. L., Vandekerckhove, J., & Kuppens, P. (2013). A diffusion model account of the relationship between the emotional flanker task and rumination and depression. *Emotion*, *13*(4), 739–747. https://doi.org/10.1037/a0031628

Peirce, J. W. (2007). PsychoPy-Psychophysics software in Python. *Journal of Neuroscience Methods*, *162*(1–2), 8–13. https://doi.org/10.1016/j.jneumeth.2006.11.017

Philiastides, M. G., Ratcliff, R., & Sajda, P. (2006). Neural representation of task difficulty and decision making during perceptual categorization: A timing diagram. *Journal of Neuroscience*, *26*(35), 8965–8975.

Pieters, J. P. (1983). Sternberg's additive factor method and underlying psychological processes: Some theoretical considerations. *Psychological Bulletin*, *93*(3), 411–426.

Possamaï, C.-A., Burle, B., Osman, A., & Hasbroucq, T. (2002). Partial advance information, number of alternatives, and motor processes: An electromyographic study. *Acta Psychologica*, *111*(1), 125–139.

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. *Psychological Review*, *117*(4), 1113–43. https://doi.org/10.1037/a0020311

Rae, B., Heathcote, A., Donkin, C., Averell, L., & Brown, S. (2014). The Hare and the Tortoise: Emphasizing speed can change the evidence used to make decisions. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *40*(5), 1226–1243. https://doi.org/10.1037/a0036801

Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review*, *85*(2), 59–108. https://doi.org/10.1037/0033-295X.85.2.59

Ratcliff, R. (2006). Modeling response signal and response time data. *Cognitive Psychology*, *53*(3), 195–237.

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.

Ratcliff, R., & Childers, R. (2015). Individual differences and fitting methods for the two-choice diffusion model of decision making. *Decision*, *2*(4), 237–279. 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., & Rouder, J. N. (1998). Modeling response times for two-choice decisions. *Psychological Science*, *9*(5), 347–356.

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. (2001). The effects of aging on reaction time in a signal detection task. *Psychology and Aging*, *16*(2), 323–341. https://doi.org/10.1037//0882-7974.16.2.323

Ratcliff, R., & Tuerlinckx, F. (2002). Estimating parameters of the diffusion model: Approaches to dealing with contaminant reaction times and parameter variability. *Psychonomic Bulletin & Review*, *9*(3), 438–481. https://doi.org/10.3758/BF03196302

Ravenzwaaij, D., van Provost, A., & Brown, S. D. (2017). A confirmatory approach for integrating neural and behavioral data into a single model. *Journal of Mathematical Psychology*, *76*, 131–141.

Requin, J., Riehle, A., & Seal, J. (1988). Neuronal activity and information processing in motor control: From stages to continuous flow. *Biological Psychology*, *26*, 179–198.

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

Rinkenauer, G., Osman, A., Ulrich, R., Muller-Gethmann, H., & Mattes, S. (2004). On the locus of speed-accuracy trade-off in reaction time: Inferences from the lateralized readiness potential. *Journal of Experimental Psychology: General*, *133*(2), 261–82. https://doi.org/10.1037/0096-3445.133.2.261

Roberts, S., & Pashler, H. (2000). How persuasive is a good fit? A comment on theory testing. *Psychological Review*, *107*(2), 358–367. https://doi.org/10.1037//0033-295X.107.2.358

Rochet, N., Spieser, L., Casini, L., Hasbroucq, T., & Burle, B. (2014). Detecting and correcting partial errors: Evidence for efficient control without conscious access. *Cognitive, Affective & Behavioral Neuroscience*, *14*(3), 970–982. https://doi.org/10.3758/s13415-013-0232-0

Roger, C., Núñez Castellar, E., Pourtois, G., & Fias, W. (2014). Changing your mind before it is too late: The electrophysiological correlates of online error correction during response selection. *Psychophysiology*, *51*, 746–760. https://doi.org/10.1111/psyp.12224

Roitman, J. D., & Shadlen, M. N. (2002). Response of neurons in the lateral intraparietal area during a combined visual discrimination reaction time task. *The Journal of Neuroscience*, *22*(21), 9475–9489. https://doi.org/10.1016/S0377-2217(02)00363-6

Sanders, A. (1983). Towards a model of stress and human performance. *Acta Psychologica*, *53*(1), 61–97. https://doi.org/10.1016/0001-6918(83)90016-1

Santello, M., & Mcdonagh, M. J. (1998). The control of timing and amplitude of EMG activity in landing movements in humans. *Experimental Physiology*, *83*(6), 857–874. https://doi.org/10.1113/expphysiol.1998.sp004165

Schad, D. J., Betancourt, M., & Vasishth, S. (in press). Toward a principled bayesian workflow in cognitive science. *Psychological Methods*.

Schall, J. D. (2004). On Building a Bridge Between Brain and Behavior. *Annual Review of Psychology*, *55*(1), 23–50. https://doi.org/10.1146/annurev.psych.55.090902.141907

Schall, J. D. (2019). Accumulators, neurons, and response time. *Trends in Neurosciences*, *42*(12), 848–860.

Schmied, A., Vedel, J. P., & Pagni, S. (1994). Human spinal lateralization assessed from motoneurone synchronization: Dependence on handedness and motor unit type. *The Journal of Physiology*, *480*, 369–387. https://doi.org/10.1113/jphysiol.1994.sp020367

Seabold, S., & Perktold, J. (2010). *Statsmodels: Econometric and statistical modeling with python*. Proceedings of the 9th python in science conference (Vol. 57, p. 61).

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*, *35*(28), 10371–10385. https://doi.org/10.1523/JNEUROSCI.0078-15.2015

Servant, M., White, C., Montagnini, A., & Burle, B. (2016). Linking theoretical decision-making mechanisms in the Simon task with electrophysiological data: A model-based neuroscience study in humans. *Journal of Cognitive Neuroscience*, *28*(10), 1501–1521. https://doi.org/10.1162/jocn_a_00989

Śmigasiewicz, K., Ambrosi, S., Blaye, A., & Burle, B. (2020). Inhibiting errors while they are produced: Direct evidence for error monitoring and inhibitory control in children. *Developmental Cognitive Neuroscience*, *41*, 100742.

Smith, P. L. (1995). Psychophysically principled models of visual simple reaction time. *Psychological Review*, *102*(3), 567–593. https://doi.org/10.1037/0033-295X.102.3.567

Smith, P. L., & Lilburn, S. D. (2020). Vision for the blind: Visual psychophysics and blinded inference for decision models. *Psychonomic Bulletin & Review*, *27*, 882–910.

Spieser, L., Servant, M., Hasbroucq, T., & Burle, B. (2017). Beyond decision! Motor contribution to speed-accuracy trade-off in decision-making. *Psychonomic Bulletin & Review*, *24*(3), 950–956. https://doi.org/10.3758/s13423-016-1172-9

Stan Development Team. (n.d.). *Pystan: The python interface to stan*. Retrieved from http://mc-stan.org

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.

Starns, J. J., Ratcliff, R., & McKoon, G. (2012). Evaluating the unequal-variance and dual-process explanations of zROC slopes with response time data and the diffusion model. *Cognitive Psychology*, *64*(1–2), 1–34.

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.

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

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

Tandonnet, C., Burle, B., Vidal, F., & Hasbroucq, T. (2003). The influence of time preparation on motor processes assessed by surface laplacian estimation. *Clinical Neurophysiology*, *114*(12), 2376–2384.

Tandonnet, C., Burle, B., Vidal, F., & Hasbroucq, T. (2006). Knowing when to respond and the efficiency of the cortical motor command: A laplacian erp study. *Brain Research*, *1109*(1), 158–163. https://doi.org/10.1016/j.brainres.2006.06.052

Teller, D. Y. (1984). Linking propositions. *Vision Research*, *24*(10), 1233–1246. https://doi.org/10.1016/0042-6989(84)90178-0

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*. https://doi.org/10.3758/s13423-020-01719-6

Turner, B. M., Van Maanen, L., & Forstmann, B. U. (2015). Informing cognitive abstractions through neuroimaging: The neural drift diffusion model. *Psychological Review*, *122*(2), 312.

Ulrich, R., & Stapf, K. H. (1984). A double-response paradigm to study stimulus intensity effects upon the motor system in simple reaction time experiments. *Perception & Psychophysics*, *36*(6), 545–558. https://doi.org/10.3758/BF03207515

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

Vandekerckhove, J., & Tuerlinckx, F. (2007). Fitting the Ratcliff diffusion model to experimental data. *Psychonomic Bulletin & Review*, *14*(6), 1011–1026.

Voss, A., Rothermund, K., Gast, A., & Wentura, D. (2013). Cognitive processes in associative and categorical priming: A diffusion model analysis. *Journal of Experimental Psychology: General*, *142*(2), 536–559. https://doi.org/10.1037/a0029459

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

White, C. N., Ratcliff, R., Vasey, M. W., & McKoon, G. (2010). Anxiety enhances threat processing without competition among multiple inputs: A diffusion model analysis. *Emotion*, *10*(5), 662–677. https://doi.org/10.1037/a0019474

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.

Wiecki, T. V., Sofer, I., & Frank, M. J. (2016). *Hddm 0.6. 0 documentation*. http://ski.clps.brown.edu/hddm_docs/index.html
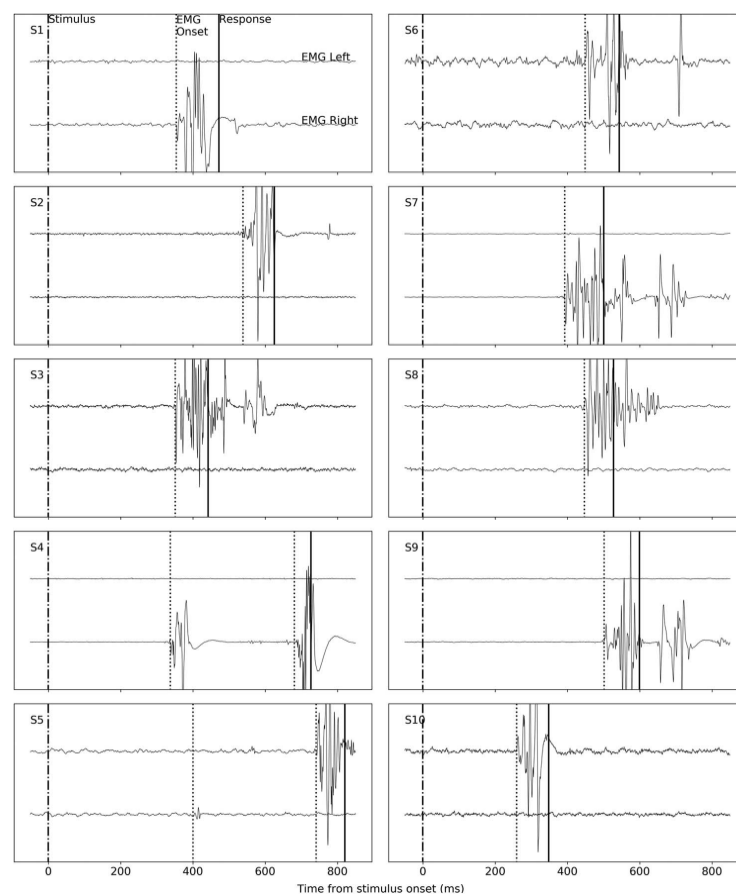
(*Appendices follow*)

Appendix A

EMG Traces

Figure A1 provides examples of the analyzed EMG signal for the first trial of 10 participants.

Figure A1

Graphic Representation of the Recorded Electromyograph (EMG) Signal for the First Trial of 10 Participants in Experiment 1

Note. Time is centered at stimulus onset (dot-dashed line); EMG onset(s) and recorded response are shown respectively with dotted line(s) and plain line. Left and right EMG signals are displayed respectively on the upper and lower half of each panel. For graphical purpose values are z scores of the amplitude (usually displayed in µV); the data were read in Python using the MNE module (Gramfort et al., 2013).

(*Appendices continue*)

Appendix B

Choice of Priors

Linear Mixed Models on Chronometric Data

In agreement with the suggestion made by Schad, Betancourt, and Vasishth (in press) for each applied LMM we selected reasonable priors and estimated the predicted values from the specified priors and whether these values are in the range of expected values (*e.g.*, most likely values for mean RT are sub second given our design, hence a factor effect of 100 ms on RT is more likely than a factor effect of 1000 ms). It should be noted that given the large number of observations in both experiments, the chosen priors have little influence as long as they do not exclude plausible values. All priors for the chronometric variables are expressed on the log scale.

We defined the following priors for the populations intercepts (μ_α) of the LMMs on RT, PMT and MT:

$$\mu_{\alpha_{RT}} \sim \mathcal{N}(\mu = 6.15, \sigma = 0.3)$$

$$\mu_{\alpha_{PMT}} \sim \mathcal{N}(\mu = 5.87, \sigma = 0.3)$$

$$\mu_{\alpha_{MT}} \sim \mathcal{N}(\mu = 4.67, \sigma = 0.3)$$

The population slope parameters (μ_{β_x}), as well as the random effects on the intercept (σ_α) and the slopes (σ_{β_x}), and the residual standard deviation (σ_r), were all given the same prior irrespective of the chronometric data fitted:

$$\mu_{\beta_x}, \sigma_\alpha, \sigma_{\beta_x}, \sigma_r \sim \mathcal{N}(\mu = 0, \sigma = 0.4)$$

Allowing for a wide range of effect sizes, random effects and residual standard deviation with values close to 0 more likely than large values.

Finally, as in Schad et al. (in press) we chose an LKJ prior, so-called because it was first described by Lewandowski et al. (2009), with a value of 2 for the correlation matrix of the by-participant parameter adjustments, firstly because we do not expect strong correlations among participant specific parameter adjustments and because a weakly informative priors on the correlation matrix helps estimating full random-effect structure as advocated by Barr (2013).

Linear Mixed Models on PMT-MT Correlation Values

The parameters of the LMM on the correlation values between PMT and MT were all given the same prior:

$$\mu_\alpha, \sigma_\alpha, \mu_{\beta_{SAT}}, \sigma_{\beta_{SAT}}, \sigma_r \sim \mathcal{N}(\mu = 0, \sigma = 0.3)$$

Therefore, allowing all possible values of intercept (μ_α), slopes of SAT ($\mu_{\beta_{SAT}}$), random effect of the intercept (σ_α) and the slope ($\sigma_{\beta_{SAT}}$), and residual standard deviation (σ_r) given that correlation values are contained between -1 and 1. In agreement with the defined normal distribution, values close to 0 are more likely than extreme values coherent with what we expect on the correlation value between PMT and MT.

HDDM

The informative priors used for the fit of the Hierarchical drift diffusion model are given by Wiecki et al. (2013) based on the analysis of range of plausible values done by Matzke and Wagenmakers (2009).

(Appendices continue)

Appendix C

Accuracy Analysis

To analyze accuracy we fitted a General LMM with contrast, SAT, response side and the interaction between contrast and SAT as fixed effects. Random effects included the random intercept by participants and random slopes for all predictors. In this analysis for Experiment 1, we see a main effect of contrast, SAT instructions and response side. The interaction between SAT and contrast was also significant. Experiment 2 replicates these effects except the response side effect (see Table C1). The presence of an effect for response side only in Experiment 1 is surprising. This could be linked to the observed effect on MT in Table 1, if right responses are faster they might also be less costly to execute, hence inducing a small bias in these responses.

Table C1
Results of the Generalized LMM Model on Accuracy in Experiment 1 (Left Column) and in Experiment 2 (Right Column)

Model	Exp. 1				Exp. 2			
	M	*SD*	2.5%	97.5%	*M*	*SD*	2.5%	97.5%
Intercept	0.84	0.02	0.80	0.87	0.75	0.06	0.63	0.86
SAT	0.06	0.01	0.03	0.08	0.12	0.03	0.07	0.18
Contrast	0.15	0.02	0.11	0.19	0.18	0.03	0.13	0.23
Resp. side	0.03	0.01	0.01	0.06	−0.01	0.03	−0.07	0.05
SAT × Contrast	0.08	0.02	0.04	0.12	0.17	0.04	0.10	0.25

Note. As for the chronometric linear mixed models (LMM)s, the fitting was performed on a transformed scale (logit) but parameters where back-transformed at each Markov Chain Monte Carlo iteration (by taking the inverse logit). *M*, *SD*, and 2.5 and 97.5% summarize the posterior distribution of the parameters.

(Appendices continue)

Appendix D

Scatter Plot of the Quantile Correlation

To convey a sense of the correlation values contained in Figure 5, we present scatter plots for the first five participants of Experiment 1 in Figure D1.

Figure D1

Scatter Plots of Premotor Time (PMT) Versus Motor Time (MT) for the Five First Participants in Experiment 1 (Participant 1 to 5 From Top to Bottom) Across the Five Quantiles (.1, .3, .5, .7 and .9 Quantiles, From Left to Right) Either With the Speed (Gray) or Accuracy (Black) Instructions
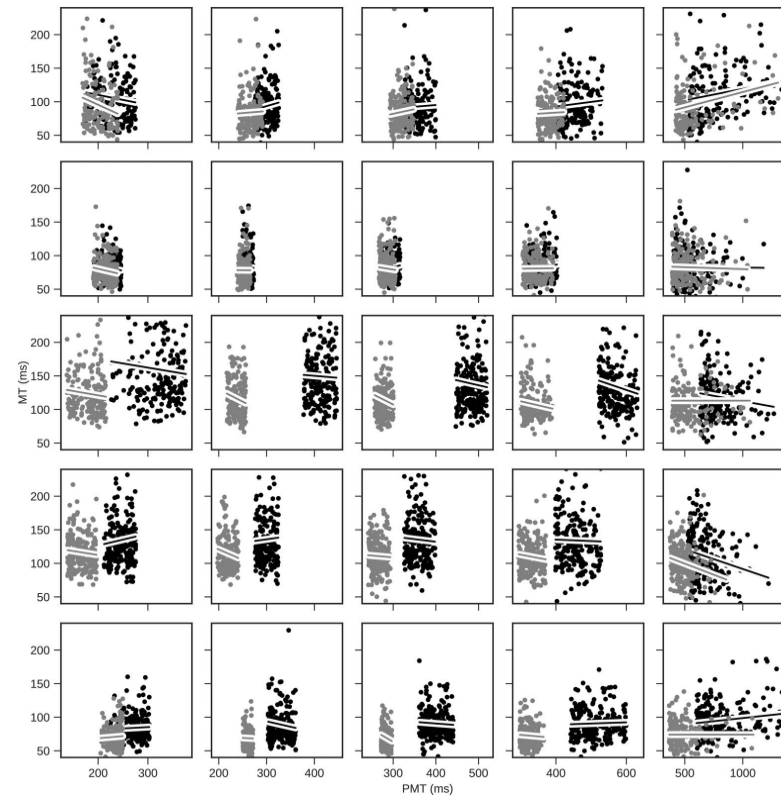
Note. To give a sense of the linear relationship between the pairs of PMT and MT, each scatter plot is associated with a line drawn from the parameters of a linear regression of MT over PMT estimated using the ordinary least squares method as implemented in the `statsmodel` python package (Seabold & Perktold, 2010).

(Appendices continue)

Appendix E

Mean Table of the Chronometric Variables

In Table E1 we report the mean values for correct response, for each chronometric variable, in the experimental cells combining SAT and contrast levels.

Table E1
Means (in ms) by Experiment, Speed-Accuracy Trade-Off, and Contrast Condition for Each Chronometric Data for Correct Responses Only

Exp.	Data	SAT	1	2	3	4	5
1	RT	Acc.	636	621	571	505	470
		Speed	457	451	437	410	388
	PMT	Acc.	512	500	448	387	355
		Speed	355	347	335	311	290
	MT	Acc.	124	121	122	118	116
		Speed	102	103	101	99	99
2	RT	Acc.	628		551	487	
		Speed	407		393	366	
	PMT	Acc.	545		470	408	
		Speed	334		322	299	
	MT	Acc.	82		81	79	
		Speed	73		71	68	

Note.   RT = reaction time; PMT = premotor time; MT = motor time; SAT = speed–accuracy trade-off.

Appendix F

Model Selection Results

The models where a modulation of the starting point and a response side effect on T_{er} were simultaneously estimated did not reach convergence, despite a high amount of MCMC iterations. This can be due to the difficulty to separate a "cognitive" bias effect from a motor bias (i.e. right hand is preferred or right stimulus is preferred) when response side and stimulus location are not counterbalanced.

For the remaining models, including a modulation of the starting point always lead to lower criteria, whichever was chosen, for both experiments. The models allowing the drift rate to vary with SAT conditions were systematically selected when considering the DIC criterion, coherent with Rae et al. (2014) (but see Starns et al., 2012 for an interpretation of SAT effects on drift rate). However, when considering the BPIC and the fit on data from Experiment 1, the models estimating only one drift rate across the SAT condi-tions were preferred. We also note that, when considering the models with a fixed starting point only (models 1 to 4 in Table F1), the fit always improved when T_{er} was free to vary between left and right responses, for both experiments. This result is surprising because we do not observe an effect of response side on MT in the second experiment. Given the issue raised above concerning the possible confound between cognitive and motor bias it could be that this result is simply caused by not allowing the starting point to vary, hence forcing the cognitive bias to be captured by the non-decision time.

Overall, only the BPIC suggests a different model for Experiment 1 than for Experiment 2. For the sake of simplicity, we chose to use the same model for both experiments, namely model 6 in Table F1, with starting point estimation and drift free to vary between stimulus strength level and SAT conditions.

(Appendices continue)

Table F1

Model Comparison for Models Fitted to Data From Experiment 1 and Experiment 2

Model	a	v	T_{er}	z	s_z	s_v	s_t	$DIC_{Exp.\ 1}$	$DIC_{Exp.\ 2}$	$BPIC_{Exp.\ 1}$	$BPIC_{Exp.\ 2}$
1	SAT	Contrast	SAT	0	SAT	1	1	−21927	−31552	−21802	−31438
2	SAT	SAT × Contrast	SAT	0	SAT	1	1	−21961	−31722	−21782	−31568
3	SAT	Contrast	SAT × Response	0	SAT	1	1	−22295	−31864	−22143	−31721
4	SAT	SAT × Contrast	SAT × Response	0	SAT	1	1	−22329	−32031	−22116	−31842
5	SAT	Contrast	SAT	1	SAT	1	1	−22898	−32664	**−22761**	−32538
6	SAT	SAT × Contrast	SAT	1	SAT	1	1	**−22931**	**−32844**	−22732	**−32671**
7	SAT	Contrast	SAT × Response	1	SAT	1	1	—	—	—	—
8	SAT	SAT × Contrast	SAT × Response	1	SAT	1	1	—	—	—	—

Note. a = boundary parameter; v = drift rate; T_{er} = nondecision time; z = starting point of the accumulation. s_z, s_v, and s_t are, respectively, the intertrial variability of the starting point, of the drift rate and of the nondecision time. 0 and 1 indicate whether a parameter was estimated or not. SAT = speed–accuracy trade-off; DIC = deviance information criterion; BPIC = Bayesian predictive information criterion. The bold values correspond to the lowest DIC/BPIC values, and hence indicate the retained model.