The Decisive Role of Non-Decision Time for Interpreting the Parameters of Decision Making Models

Gabriel Weindel, Thibault Gajdos, Boris Burle, F.-Xavier Alario

The Decisive Role of Non-Decision Time for Interpreting Decision Making Models

Gabriel Weindel[1,2], Thibault Gajdos[1], Boris Burle[2], F.-Xavier Alario[1]
[1]Laboratoire de Psychologie Cognitive, Aix Marseille Univ, CNRS, LPC, Marseille, France
[2]Laboratoire de Neurosciences Cognitives, Aix Marseille Univ, CNRS, LNC, Marseille, France

Computational models of decision making are becoming increasingly popular to interpret reaction time and choice data in terms of decision and non-decision related processes. But current evidence remains scarce as to whether parameters of a mathematical model such as the Drift Diffusion Model can recover genuine latent psychological processes. In this study, we combine an experimental approach using a decision making task with a physiological decomposition of each reaction time into a motor and pre-motor time using electro-myography. The aim is to test whether the non-decision time parameter of a DDM, assumed to contain encoding and motor processes, varies according to both psychophysical predictions of stimulus encoding and the physiological measurement of motor processes. Our results show that 1) the encoding time is accounted by a DDM only in the case of instructions emphasizing speed over accuracy and 2) that the onset of muscular activity does not sign the end of the accumulation of evidence. This questions the ability of DDM to account for how participants achieve speed-accuracy tradeoff as well as the interpretability of its parameters in terms of decision and non-decision processes.

Introduction

Making a decision takes time. In the paradigmatic example of perceptual decisions, where a visual stimulus calls for a motor response, this time is taken by the neural conduction delays from the retina to the brain and from the brain to the muscles (von Helmholtz, 1850, cited by Schmidgen, 2002), and, in between, by the mental action of deciding between alternatives (Donders, 1868). The primary measure used to explore how these decisions are made has been the time elapsing between the stimulation and the participant's overt reaction, or "reaction time" (*RT*). While *RT* measures collapse conduction and mental action delays, quantitative processing models have explored the possibility of breaking down *RT*s into more elementary components.

There is a general agreement that *RT*s can be broken down into at least two parts: a "decision time" (T_D) and a time outside the decision, or "non-decision time" (T_0). The two time periods T_D and T_0 are generally thought to reflect sequential events with additive duration (Ratcliff, 1978; Ratcliff, Smith, Brown, & McKoon, 2016).

In most decision making models, T_D is defined within the evidence accumulation framework, as the time elapsing from the start of the accumulation of evidence to the time at which the accumulated evidence reaches a decision threshold (Brown & Heathcote, 2005; Heathcote & Love, 2012; Ratcliff, 1978; Ratcliff & Rouder, 1998; Stone, 1960; Usher & McClelland, 2001, see Stine, Zylberberg, Ditterich, & Shadlen, 2020, for recent alternatives). T_D is computed on the basis of at least three parameters represented in Figure 1: the *boundaries* (a), usually interpreted as the response caution of the participant, the *drift* (v), interpreted as the speed of evidence accumulation, and the *bias* (z), reflecting the starting point of the evidence accumulation process.

T_0, in turn, is defined as the time needed to form an internal representation of the stimulus ($T_{encoding}$) summed to the time needed to execute the response ($T_{response}$). T_0 is often referred to as the "residual" time, a term that reveals the secondary importance given to these aspects of the *RT* in decision making. Perhaps because of the diagnostic power of T_0 (*e.g.,* see Smith & Lilburn, 2020), considerations of this time are generally technical and focused around fitting quality issues rather than functional interpretations. In this paper, we focus on those so-called residual components and show how they shape our understanding of decision making

The authors thank Iván Ballasch for his help in data collection, Anna Montagnini, Mathieu Servant, and Thierry Hasbroucq for fruitful discussions, Frédéric Chavane and Alexandre Reynaud for sharing and explaining the raw data from their article.This research was supported by grant ANR-16-CONV-0002 (ILCB) and the Programme "Investissements d'Avenir", Initiative d'Excellence d'Aix-Marseille Université via A*Midex funding (AMX-19-IET-004), and ANR (ANR-17-EURE-0029).

Correspondance should be adressed to Gabriel Weindel, Laboratoire de Neurosciences Cognitives, Case C, Aix-Marseille Université, 3 Place Victor Hugo, 13331 Marseille, cedex 3. E-mail: gabriel.weindel@gmail.com

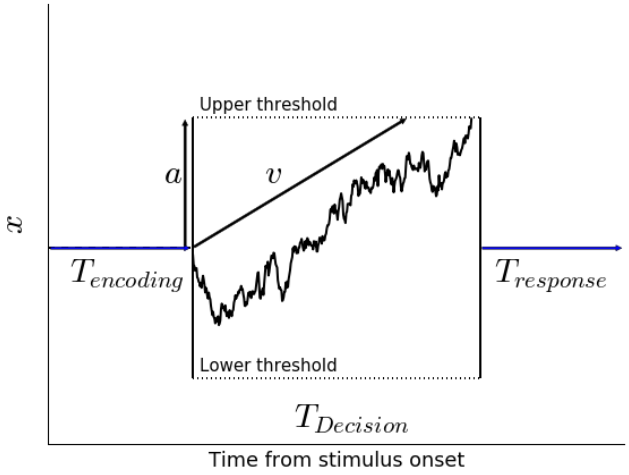

Figure 1. Processing account of a single decision in a diffusion model implementing evidence accumulation.
The stochastic path represents the stimulus evidence being accumulated over time through a noisy channel, modeled as a diffusion process with a drift v (i.e., , rate of accumulation). Accumulation starts at a given amount of evidence (referred to as the bias, figure represents an unbiased decision) and stops once a threshold boundary, a is reached, and the corresponding alternative is chosen. The time between the onset of the accumulation to its termination is referred to as the T_D, the time before and after, in blue, the T_0.

processes.

Formal quantitative models of decision making allow using RTs to test theoretically grounded predictions about cognitive processes. So, in theory, manipulating the difficulty of the decision should specifically modulate T_D and be reflected in a change of v, the drift parameter. This result is indeed observed in several studies (Gomez, Ratcliff, & Childers, 2015; Palmer, Huk, & Shadlen, 2005; Voss, Rothermund, & Voss, 2004). Another example is the manipulation of the speed-accuracy trade-off (SAT), which has been hypothesized to affect response caution, and thus, presumably, to lengthen selectively T_D through an adjustment of a, the decision threshold. Such is indeed the case (Ratcliff & McKoon, 2008; Ratcliff & Tuerlinckx, 2002a), but the SAT manipulation is not always selective, it can also affect the estimation of T_0 (Dutilh et al., 2016; Palmer et al., 2005; Ratcliff, 2006; Voss et al., 2004). Finally, manipulating the brightness of a visual stimulus or the force needed to produce the response should affect non-decisional input and output processes, respectively. Both manipulations indeed result in a modulation of T_0 (Gomez et al., 2015; Ho, Brown, & Serences, 2009; Servant, White, Montagnini, & Burle, 2016; Voss et al., 2004); however, they also impact the parameters determining T_D: brightness manipulations affect v (Servant et al., 2016), force manipulations affect z (Voss et al., 2004) and a (Gomez et al., 2015; Ho et al., 2009).

In other words, experimental manipulations are not always reflected in modulations of the parameters predicted by the theory. This discrepancy may receive two alternative explanations: either, the model does not appropriately implement the theory, or the theory itself (*i.e.,* the psychological interpretation of the experimental manipulations) is incorrect. A clarification of this issue is required to better ground the theoretical interpretation of the parameters, especially in the context of an increasing reliance on the DDM to address both basic and clinical research.

The present study

Our research on the links between parameters and processes in decision making models focused on the "residual" non-decision time T_0 estimated in the Drift Diffusion Model (DDM, Ratcliff & Tuerlinckx, 2002b). To test how faithfully T_0 reflects the cognitive delays it is intended to capture (*i.e.,* $T_{encoding}$ plus $T_{response}$), we used targeted experimental manipulations coupled with empirical response time decomposition in a perceptual decision task.

The RTs of every trial were decomposed into pre-motor and motor times (PMT and MT), based on muscle activity onsets revealed by the electromyogram (EMG) of the responding hand (Burle, Possamaï, Vidal, Bonnet, & Hasbroucq, 2002). Following (Luce, 1986, p. 97) and our own previous work (Weindel, Anders, Alario, & Burle, 2021), we assumed that MT starts after accumulation has ended and hence allows $T_{response}$ to be measured rather than estimated. As a result, the remainder PMT was thought to correspond to the sum of $T_{encoding}$ and $T_{decision}$. The DDM could then be fitted either to RTs, to estimate T_0 as is usually done, or be fitted to PMTs, to provide a reasonable estimate of $T_{encoding}$.

There were three experimental manipulations. The two first ones were intended to primarily affect T_0: a manipulation based on stimulus contrast was predicted to affect the process of encoding, and a manipulation of the response force requirements was predicted to affect the process of responding. Finally, the participants' strategy was manipulated with a classic Speed-Accuracy trade-off, assumed to affect the decision process. The details for these three manipulations are as follows.

Affecting encoding processes. Participants had to choose which of two sinusoidal gratings (Gabor patches) had the highest contrast. We manipulated the mean contrast of the gratings, while keeping their absolute difference constant. It has been shown that stimuli become more accessible, and are processed faster, when their contrast increases (*e.g.,* Harwerth & Levi, 1978, in a simple reaction time task). We thus expected shorter $T_{encoding}$ for higher contrasts. Conversely, Weber-Fechner's law (Stevens, 1961) states that the just noticeable difference between a pair of stimulus increases linearly with stimulus intensity. We thus expected that discrim-

inating between the two Gabor patches will be more difficult for higher contrasts, thereby inducing longer T_D through a decrease in the *drift* parameter (for a detailed account on how such manipulation can affect the T_D see Ratcliff & McKoon, 2018). In sum, the contrast manipulation was expected to have opposite effects on $T_{encoding}$ and on T_D.

Affecting motor processes. We manipulated the force required to produce the responses, a manipulation known to affect processes related to motor execution (Burle et al., 2002). We expected that an increase in the required response force will result in an increase of $T_{response}$. This would be consistent with studies where similar response output manipulations resulted in changes in the estimated T_0 parameter of the drift diffusion model (Gomez et al., 2015; Ho et al., 2009; Voss et al., 2004). Conversely, the T_0 fitted on PMT (*i.e.,* putatively $T_{encoding}$) was not expected to be affected by the Force manipulation. It has occasionally been reported that force requirements can also influence the decision related parameters bias and threshold (Gomez et al., 2015; Voss et al., 2004), but no explicit interpretation has been offered. Our EMG decomposition and the hypothesized absence of decision-related latency variations in MT allows the following subsidiary expectations. If the force manipulation genuinely affect the decision processes, the effect should be invariant whether the fit is performed on RT or on PMT. Conversely, observing that the effects vary between both fits would suggest an incorrect separation of T_D and T_0.

Speed accuracy trade-off (SAT). Finally, we manipulated the SAT level required from participants through verbal instructions. This manipulation is classicallydescribed as an adjustment of the level of evidence needed before a decision is made, therefore linked to a change in the *boundary* parameter of the DDM. However, the manipulation of SAT has also been shown to modulate the speed of encoding processes (Steinemann, O'Connell, & Kelly, 2018) and of motor execution (Spieser, Servant, Hasbroucq, & Burle, 2017; Steinemann et al., 2018; Weindel et al., 2021). There is a debate between accounts of SAT, either in terms of parameter (i.e. process modulations) within DDM or in terms of changes in the nature of the generative model (e.g. Cisek, Puskas, & El-Murr, 2009; Dutilh, Wagenmakers, Visser, & van der Maas, 2011; Ollman, 1966; Verdonck, Loossens, & Philiastides, 2020)a. Here, decision instructions were manipulated to test whether the above derived predictions on encoding and motor processes hold across all SAT spectrum.

Methods

Participants. Sixteen participants (6 men and 10 women, mean age = 24.5 years, 2 left-handed) that were students at Aix-Marseille University, were recruited for this experiment. 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 n° 1041). Participants gave their informed written consent, according to the declaration of Helsinki. They received a compensation at a rate of €15 per hour.

Apparatus. Participants performed the experiment in a dark and sound-shielded Faraday cage. They were seated in a comfortable chair about 100 cm away from a 15 inch CRT monitor that had a refresh rate of 75 Hz. The CRT monitor was gamma corrected by a psychophysical procedure provided by the software PsychoPy (Peirce, 2007). Responses were given by pressing either a left or a right button with the corresponding thumb. The buttons were fixed on top of two cylinders (3 cm in diameter, 7.5 cm in height). The cylinders were fixed on a tablet and separated by a distance of 20 cm. The buttons were mounted on force sensors that recorded a continuous measure of the force produced at a sampling rate of 2048 Hz. The behavioral response was recorded when a force threshold was exceeded. The device allowed adjusting the force threshold needed for a response to be received. The threshold was manipulated across conditions, as described below. Response signals were transmitted to the parallel port of the recording computer. At button press, participants received a 3ms sound feedback (1000 Hz pure tone).

The participants' forearms and hypothenar muscles rested comfortably on the table, to minimize muscle recruitment during response execution. We measured the EMG activity 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 2048 Hz.

Stimuli. Stimulus presentation was controlled by the software PsychoPy (Peirce, 2007). Each stimulus was composed of two vertical oriented Gabor patches, on the left and right of a fixation cross separated by 1.4 visual angle degrees. The Gabor patches had a spatial frequency of 1.2 cycles per visual angle degree and a size of 2.5 visual angle degrees each. At each trial, the same amount of contrast (7%) was subtracted to the randomly assigned incorrect Gabor patch and added to the correct one, resulting in a 14% contrast difference. The task of the participant was to press the button ipsilateral to the highest contrast.

Experimental manipulations.

Contrast. We choose to manipulate the mean contrast of both Gabor patches while keeping a constant difference of 14% on a scale between 0 and 100% (where 0% is uniform grey) between them. Six levels of mean contrast (23%, 37%, 51%, 65%, 79%, 83%) were selected based on a pilot study, targeting a performance that would typically span from near-perfect accuracy to near chance level. The mean contrast across both patches was randomly chosen at each trial with a fixed rate of occurrence (1/6) within every block.

Force. The Force factor had two levels: strong and weak. These levels were tailored to each participant before the experiment started. Participants were asked to press twice the right and then the left button, with the maximum force they could apply. The maximum voluntary force was defined as the maximum between the two trials from the weakest of the two hands. Defining maximum voluntary force this way was chosen to avoid muscular fatigue from the weakest hand. The actual force levels for the strong and weak conditions were then defined as, respectively, 2% and 20% of this maximum voluntary force level (generating force levels around 1.20 and 12N respectively).

SAT. The speed-accuracy trade-off (SAT) instruction was manipulated between blocks. Participants were instructed that "Speed" instructions required a mean reaction time near 400 ms and that "Accuracy" instructions required a percentage of correct responses near 90% while maintaining RTs below 800 ms. Each block started with the presentation on the center of the screen of its corresponding instruction: the French word for Speed ("Vitesse") or Accuracy ("Précision"). The end of each block was followed by feedback about mean reaction time and mean accuracy, along with oral feedback from the experimenter, if the participant had not satisfied the condition goals of the block.

Procedure. All participants performed a single experimental session with 24 blocks of 100 trials each. Session duration was around 1h30 including a training session of 15 minutes and self-paced breaks between each block. Participants were asked to keep their gaze on the central fixation cross throughout each block, and to respond to the visual stimuli according to the corresponding SAT instruction.

The training session started with 40 trials without specific SAT instructions, followed by 2 blocks of 10 trials in the Speed condition, followed by 2 blocks in Accuracy condition, and ended with 2 blocks of 10 trials with alternating instructions. During the experimental session, SAT instructions alternated every three consecutive blocks. The force settings varied every six blocks, with an on-screen message to inform the participant beforehand. The order of the SAT instructions and the force requirement was counterbalanced across participants so that every possible order combination was presented to 4 participants. Within each block, the 6 levels of mean contrast value were fully randomized across trials. No response deadline was applied, and the inter-trial interval was fixed to 1000 ms from button press to next stimulus onset.

EMG processing. The EMG recordings were read in Python using the MNE module (Gramfort et al., 2013). The signal was filtered using a Butterworth 3rd order high pass filter at 10Hz from the scipy Python module (Oliphant, 2007), then segmented by-trial in windows between 150 ms before and 1500 ms after stimulus onset. We used a variance-based method to detect whether EMG activity was significantly above threshold in either hands' channels. The precise burst onset was then identified with an algorithm based on the "Integrated Profile" of the EMG burst (see Liu & Liu, 2016; Santello & Mcdonagh, 1998, for details). If the algorithm failed to locate or detect the EMG burst onset, the experimenter corrected or added them manually. At this stage of signal processing, the experimenter was unaware of the trial type he was annotating to avoid any bias. Every muscular event (above-threshold change in the signal followed by a return to the baseline) in the trial was marked, even when the activation was not immediately followed by an overt response.

In trials where a single EMG burst was detected, motor time (MT) was defined as the time between the onset of EMG burst and the force threshold crossing recorded. Pre-motor time (PMT) was defined as the time between stimulus onset and the EMG burst onset. Multiple EMGs were observed in 21% of trials. Such observations are not new (e.g. Weindel et al., 2021, and others), but a precise account of these multiple activities is still lacking (although see Servant, Logan, Gajdos, & Evans, 2021, for a tentative account of such trials). Minimally, they show that participants where not always engaged in a pure sequential encoding-decision-execution process. Therefore, we removed these trials from all the analysis in the study.

Statistical procedure

Bayesian Statistics. All analysis were performed in a Bayesian framework. Bayesian methods try to estimate an unknown parameter (or set of parameters) and the uncertainty around it. More explicitly, Bayesian methods combine prior information and Bayes' rule to quantify the likelihood of the parameters by generating a posterior distribution for each of them. This posterior distribution can be naturally interpreted as the probability of any given parameter value given the priors, the data, and the tested model. In our study we summarize the posterior distribution using the mean and the Credible Interval (CrI), the 95% Highest Probability Density interval (HPD; Kruschke, 2010).

Hierarchical Modelling. All models, including linear models, were constrained to follow a hierarchical structure with parameters from each participant as units assumed to be drawn from a population distribution. This parametrization allows to estimate population parameters (*e.g.,* the slope of the effect of stimulus contrast on RTs) along with individual parameters (*e.g.,* the inter-individual differences in the slope of contrast with RT), often referred to respectively as fixed and random effects in the case of linear models. Hierarchical modelling remains Bayesian thus preserving the uncertainties associated with parameter values. Such approach allowed directly testing our hypotheses, by comparing the posterior distributions for the population effects across conditions.

Linear Mixed Models. We used linear mixed models (LMM) on the log transformed RT, PMT, and MT, and generalized linear mixed models (gLMM) for the proportion of correct responses. Formally, LMM model the dependent variable as drawn from a normal distribution who's parameters are constrained by the experimental design (e.g. the mean of the normal distribution changes with SAT instructions).

Given our analysis plan, we derived generic LMMs for RT, PMT and MT where all fixed effects and all random effects were estimated. The effects of the experimental factors were modeled on the mean parameter for normally distributed dependent variables (DV), assuming equal variance across conditions.

$$y_{ji} \sim \mathcal{N}(\mu_j, \sigma^2) \tag{1}$$

$$\begin{aligned} \mu_j = \alpha_j + \beta_{1j}\, SAT + \beta_{2j}\, FC + \beta_{3j}\, Con. \\ + \beta_{12j} SAT \times FC + \beta_{23j}\, SAT \times Con. \\ + \beta_{13j}\, FC \times Con. + \beta_{123j}\, FC \times SAT \times Con. \end{aligned} \tag{2}$$

Where y_{ji} represents the modeled DV (RT, PMT or MT) on the ith trial for the jth participant and is assumed to be normally distributed with mean μ_j and standard deviation σ^2. As seen in Equation 2, μ is dependent of the experimental factors (Con. for mean contrast, FC for Force Condition and SAT for speed accuracy trade-off instructions). The LMM where only fitted on correct responses as Weindel et al. (2021) already reported the effect on these variables on errors and, for the sake of simplicity, to limit the analysis to three factors.

Response correctness was modelled with a gLMM, whereby proportion of correct responses is hypothesized to follow a Bernoulli distribution, modulated by the same factors than in Equation 2:

$$p(response_{ij} = 1) \sim Bernoulli(logit(\mu_j)) \tag{3}$$

To reiterate, in the LMMs and the gLMM, the intercept and all factors and interactions are modeled as random effects :

$$\alpha_j \sim \mathcal{N}(\mu_\alpha, \sigma_\alpha^2) \tag{4}$$

$$\beta_{xj} \sim \mathcal{N}(\mu_{\beta_x}, \sigma_{\beta_x}^2) \tag{5}$$

where μ_α and μ_{β_x} are the population estimated intercept and regression coefficient and σ_α^2 and $\sigma_{\beta_x}^2$ the estimated random effect of the population sampling.

Fitting procedure for the g/LMMs. For each LMM and gLMM, six Markov Chain Monte Carlo (MCMC) sampling processes were run in parallel, each composed of 2000 iterations among which the first 1000 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, et al., 1992) and by means of visual inspection of the MCMC chains. We 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, Vasishth, Engelmann, and Suckow (2018) and using the `pystan` package (Stan Development Team, n.d.). The summary statistics and plots of the parameters were created using `arviz` python package (version 0.4.1, Kumar, Carroll, Hartikainen, & Martin, 2019).

Priors for the g/LMMs. The priors for the LMM and gLMM are intended to ease the fitting procedure, we chose to use the relatively broad informative priors described as described in Weindel et al. (2021).

Estimated difference between condition levels $\hat{d}$. In order to estimate the magnitude of the difference ($\hat{d}$) between the levels of the experimental factors Contrast, Force and SAT, we chose to use the predictions of the fitted linear models. For each dependent variable, we first computed the predicted difference between both SAT level with all other predictors set at 0 (see Appendix A). We then computed the predicted differences between the lowest and highest contrast level as well as the weak and high force condition for each SAT instruction separately. The results are thus composed of the effect of SAT, the effects of force and contrast in each SAT condition and the difference of these effects across SAT conditions. Due to the Bayesian nature of the analysis, the uncertainties associated with the regression parameters are preserved in these estimated differences. Thanks to the hierarchical nature of the regression models, we directly estimate a population difference. Both the Bayesian and hierarchical nature of the method therefore allow to directly infer the population level effect size with its uncertainty using the posterior distribution of the predicted difference. The strength of evidence for the presence or absence of an effect was determined based on the credible values of the differences as provided by the mean and the 95% CrI of the posterior distribution.

Drift Diffusion Modelling.

Model fitting procedure. We used the implementation of a hierarchical Bayesian DDM provided in the *HDDM* python package (T. V. Wiecki, Sofer, & Frank, 2013). Note that *HDDM* uses the diffusion coefficient (See Ratcliff & McKoon, 2008, for a review of the DDM parameters) as a scaling parameter by fixing it to a value of 1 (contrary to a value of 0.1 in some applications of the DDM). For each model on RT and on PMT, both in the "Model selection" section below and for the model including co-variates, we ran 32500 burn-in samples and 2500 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 et al., 1992) was under the con-

ventional 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 in order to ease convergence, we used the default informative priors used in HDDM based on the work of Matzke and Wagenmakers (2009). 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 inter-trial variability parameters of the drift rate, of the bias, and of the non-decision time were estimated at the group-level because they are notoriously difficult to estimate (Boehm, Annis, et al., 2018; T. Wiecki, Sofer, & Frank, 2016). The informative piors used for the fit of the Hierarchical drift diffusion model are given by T. V. Wiecki et al. (2013) based on the analysis of range of plausible values done by Matzke and Wagenmakers (2009).

Model selection. We designed a base model and added parameters according to our hypotheses. The base model was chosen based on previous studies. For this base model, the boundary parameter was free to vary with SAT instructions. The drift rate was free to vary with the contrast[1], as this parameter has been shown to be associated with stimulus strength. The T_0 was free to vary with SAT, as it has been observed that this parameter also varies with SAT conditions (Palmer et al., 2005; Ratcliff, 2006; Voss et al., 2004), with the Force Condition and with the contrast factor, as all three are the factors of interest in the study of T_0 [2]. The accumulation bias was free to vary for each participant. We also added inter-trial variability of the drift rate and the non-decision time, because of their ability to reduce the influence of contaminant fast-trials (Lerche, Voss, & Nagler, 2017). Finally, we added the inter-trial variability of the starting point parameter which was free to vary with SAT instructions, because it is often reported that the latency contrast between errors and correct response does change according to the SAT condition and that this pattern is captured by a different bias variability.

In addition to the base model, we tested the following hypothesis, and combinations thereof: whether the drift rate also varies with SAT (Rae, Heathcote, Donkin, Averell, & Brown, 2014), or with the Force Condition (Voss et al., 2004); and whether the bias and the boundaries are variable between Force conditions (respectively Voss et al., 2004 and Gomez et al., 2015; Ho et al., 2009). The 16 possible combinations of hypotheses are summarized in Table B1.

We used the deviance information criterion (DIC) to select among competing models. The DIC is an analog to the Akaiake 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. But because DIC tends to select over-fitted models (Ando, 2007)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 over-fitted 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.

DDM regression analysis. Once the best-fitting model was identified and selected, the effects of the experimental factors on the parameters were assessed by further embedding a hierarchical regression in the model fitting procedure (Boehm, Marsman, Matzke, & Wagenmakers, 2018). The three experimental factors and their interactions were included as predictors in the regression, on the condition that had been left free to vary across conditions in the model selection procedure. Each parameter that is free to vary with one or more factors was estimated with one intercept and one slope for each factor and interactions. This allowed to use the posterior distribution of intercept and slopes to test directly for the presence and the direction of an effect by inspecting whether 0 is included in the posterior distribution. We compared the results of joint DDM-regression fits on the PMT and RT.

The hierarchical nature of the data is preserved in these models because each intercept and slope parameters is estimated as being drawn from a population distribution. The parameters that do not vary with experimental factors (*i.e.,* inter-trial variability of the drift rate and the non-decision time) are estimated as described in the model selection section. The inter-trial variability of the bias was free to vary between SAT instructions but the corresponding effect size was not estimated with a regression. This is because, first, we only have one estimate for the population due to the difficulty to estimate it and, second, we do not have specific hypothesis about this parameter. As in the model selection procedure, the models were fitted using the *HDDM* python package (T. V. Wiecki et al., 2013).

Fast-guess detection and removal. Fast guess trials can be problematic when studying decision making in the context of evidence accumulation models. Before performing any statistical analysis, we applied an exponentially weighted moving average filter (EWMA; Vandekerckhove & Tuerlinckx, 2007). This method iteratively computes a weighted accuracy measure (amount of correct responses relative to errors) from the fastest to the slowest response time. The

[1] As boundaries were coded as right and left responses (respectively upper and lower threshold), and in order to avoid estimating one drift for each combination of stimulus side and contrast, the model was coded to take negative drift value when the correct stimulus was on the left

[2] We started by including models that do not allow the parameter T_0 to take different values for the different contrast levels but these models failed to converge probably because of the effects reported in the behavioral results section

method is usually performed on the sorted *RT* distribution, but was applied here to the sorted *PMT* distribution. Participants are considered to be in a fast guess state until the weighted accuracy is higher than a defined threshold. The *PMT* at which this change of state occurs is identified, and faster trials are censored. EWMA involves the following user-defined parameters: the initial starting point of the weighted accuracy, the amount of preceding trials (weight) retained in the accuracy computation, and the accuracy threshold for defining non-guess trials. The starting point was defined at 0.50 based on the assumption that a guessing strategy yields a 50% chance of correct response. The weight (bounded from 0 to 1, with 0 being all preceding trials used) was fixed at 0.01 as in the description of the method by Vandekerckhove and Tuerlinckx (2007). The threshold was fixed at 0.55 based on a reasonable assumption that participants could not reach an accuracy superior to 0.55 on the basis of guessing.

The EWMA filter was applied for each participant's *PMT* distribution, separately in the speed and accuracy conditions; fast-guesses can have different latencies across both conditions. *PMT*s rather that *RT*s were used for the EWMA, first because Weindel et al. (2021) showed a high reliability when the method was applied separately to *PMT* and *RT*, and second because trials that do not appear very fast on *RT* can sometimes be fast on the *PMT* and therefore be problematic when fitting a DDM on the *PMT* as done here. 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[3].

Results

The method implemented for detecting EMG onsets imposed an *RT* upper limit of 1500 ms, whereby 1% of the trials were excluded. Trials with low signal-to-noise ratio or with high spontaneous tonic activity that resulted in uncertain EMG onset detection were excluded (7%). Trials that presented more than one EMG activity (see Method section on EMG analysis) were also excluded (21%)[4]. Finally, the trim criterion derived from the fast-guess detection method lead to the exclusion of 8% of the data. Thus, the combined EMG and statistical criteria resulted in the exclusion of 37% of the trials. Censoring errors, for the LMM analysis of RT, PMT, and MT, removed 13% of the remaining data. On average, 1513 trials (SD = 310) were available per participant. All estimated differences ($\hat{d}$) are presented on the data scale, milliseconds for chronometric variables and proportion correct for accuracy.

Behavioral results (*RT* and error rates)

Linear mixed models. The following observations are illustrated in Figure 2 and in Figure 3 (two left columns). When speed is emphasized, *RT* decreases, $\hat{d}_{SAT} = -165$, $CrI = [-197, -134]$, and so does the accuracy rate, $\hat{d}_{SAT} = -0.10$, $CrI = [-0.13, -0.07]$. The effect of contrast on *RT* proved to be different between SAT instructions $\hat{d}_{Contrast:Speed-Acc.} = -55$, $CrI = [-70, -39]$. When participants are asked the emphasize accuracy, an increase in contrast lengthens *RT*s, $\hat{d}_{Contrast:Acc.} = 60$, $CrI = [35, 83]$, and reduces response accuracy, $\hat{d}_{Contrast:Acc.} = -0.22$, $CrI = [-0.27, -0.18]$. The contrast effect is essentially canceled when participants are asked to speed their responses, $\hat{d}_{Contrast:Speed} = 5$, $CrI = [-13, 22]$. The proportion of correct responses however displayed similar effects of Contrast in both the Speed and Accuracy conditions, $\hat{d}_{Contrast:Speed} = -0.22$, $CrI = [-0.27, -0.16]$, $\hat{d}_{Contrast:Speed-Acc.} = 0.02$, $CrI = [-0.01, -0.04]$.

When force requirements are higher, *RT* increases both in the Accuracy, $\hat{d}_{Force:Acc.} = 48$, $CrI = [22, 75]$ and the Speed conditions, $\hat{d}_{Force:Speed} = 35$, $CrI = [16, 53]$. The proportion of correct response is not affected by the Force factor neither in accuracy $\hat{d}_{Force:Acc} = 0.01$, $CrI = [-0.01, 0.02]$ nor in the speed condition $\hat{d}_{Force:Speed} = 0.03$, $CrI = [0.00, 0.05]$ although the CrI barely included 0.

Unexpectedly, the interaction between Force and Contrast had an effect on *RT* selectively in the Accuracy condition, $\hat{d}_{Force\times Contrast:Acc.} = -47$, $CrI = [-68, -27]$ but not in the Speed condition $\hat{d}_{Force\times Contrast:Speed} = 7$, $CrI = [-11, 23]$. The proportion of correct responses was not sensitive to the interaction between force and contrast neither in Accuracy $\hat{d}_{Force\times Contrast:Acc} = -0.03$, $CrI = [-0.09, 0.03]$) nor in Speed, $\hat{d}_{Force\times Contrast:Speed} = -0.04$, $CrI = [-0.08, 0.07]$.

Drift Diffusion Model selection on RT. The model selection procedure is fully described in Appendix B, with DIC and BPIC estimates summarized in Table B1. In the model that was ultimately selected, one boundary parameter was estimated for each combination of SAT and force condition levels, one drift for each level of contrast, one starting point for each force level, and one non-decision time for each experimental cell of the three factors SAT × Force × Contrast (see Table B1). The effects of the experimental factors on these model parameters are summarized in Table 1 and spelled out below.

Effects on T_0 when fitted on RT. Estimated T_0 was longer when accuracy was emphasizedon T_0. SAT and force interacted: increasing force had a strong effect on T_0 in the accuracy condition, and the interaction term indicated

[3] https://github.com/mdnunez/bayesutils/blob/master/wienerutils.py

[4] Those trials display a longer mean RT. While usual behavioral experiments would analyse these data, EMG allows to filter them out as they could represent alternative response modes (e.g. mind wandering) or trials in which decision threshold were crossed multiple times (Servant et al., 2021), thus departing from decision making as implemented by a DDM

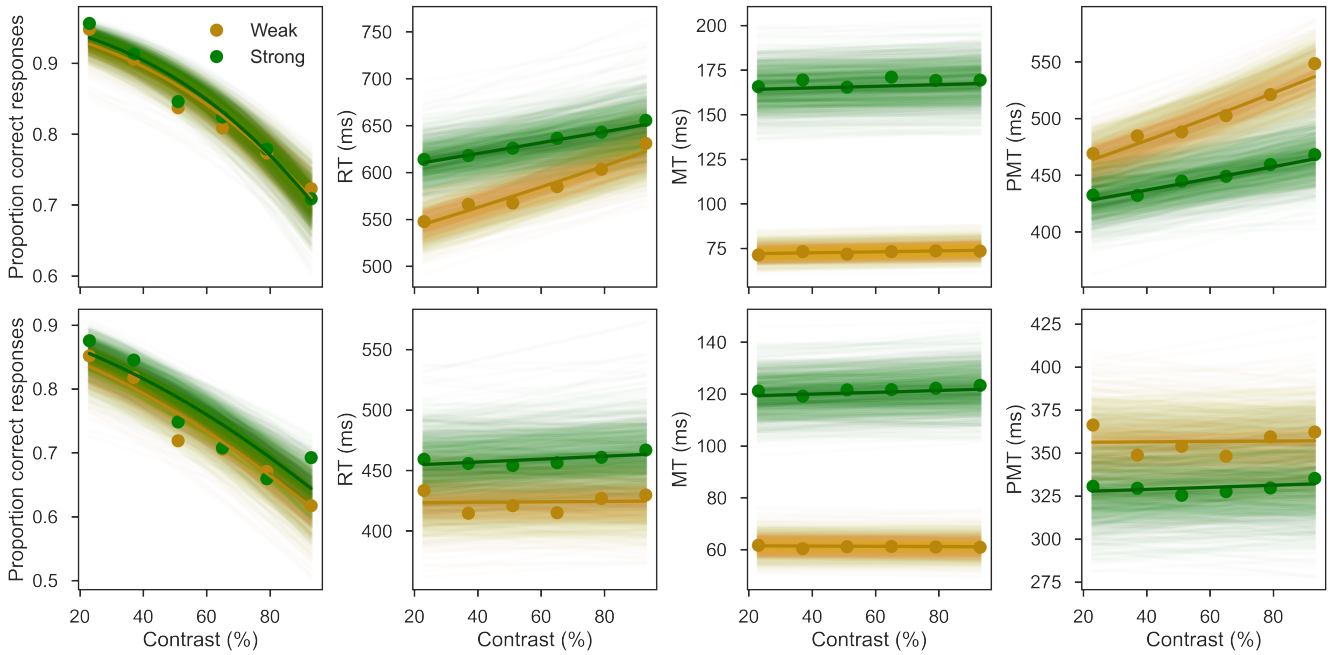

Figure 2. Average values for proportion correct, *RT*, *MT*, and *PMT* (columns from left to right) plotted for each SAT condition (accuracy on the top row and speed on the bottom row), broken down by contrast levels and by force condition. The lines represent 1000 random draws from the joint posteriors of the combined MCMC chains of the corresponding G/LMM fits. The thick line represent the predicted regression line with all parameters set at their maximum a posteriori value.

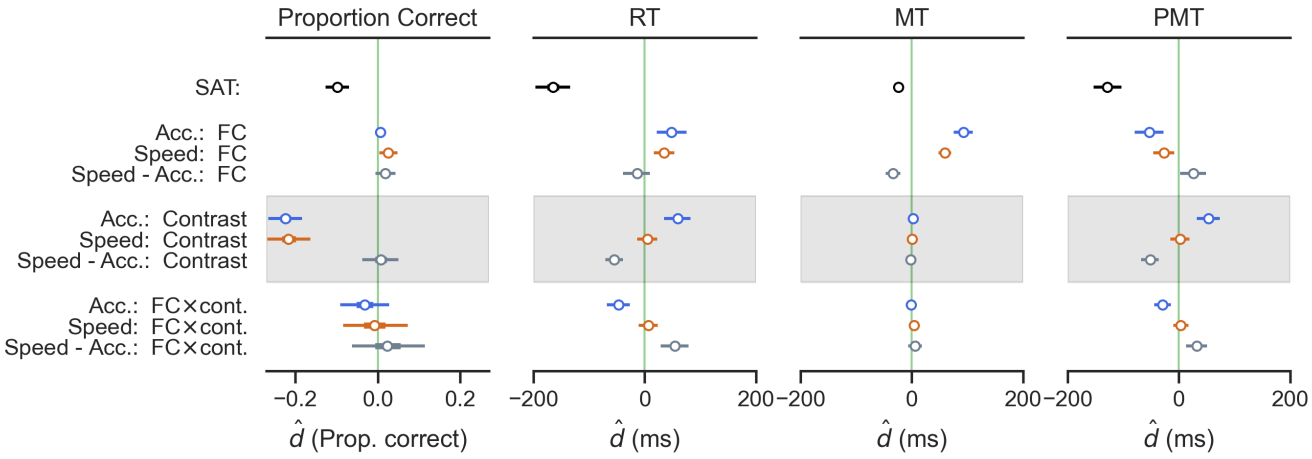

Figure 3. Estimated differences between condition levels ($\hat{d}$) for SAT, Force (FC), Contrast and their interactions on the millisecond scale of the data in the Accuracy condition (blue), the Speed condition (chocolate), and the difference between both conditions (i.e. interaction; grey). Dots represent the maximum a posteriori, and bars the 2.5% and 97.5 % HPD of the corresponding marginal posterior distributions.

a smaller yet reliable effect in the speed condition (see left column of Figure 4). SAT and Contrast also interacted (see left column, middle and bottom panels, of Figure 5): there was no evidence for an effect of contrast on T_0 in the accuracy condition, but the interaction term indicated a negative effect when speed was emphasized.. Finally, there was no evidence for an interaction between Force and Contrast, nor for the three way interaction (although in this latter case the CrI barely included 0). These estimates are summarized in Table 1.

Effects on decision related parameters when fitted on RT. The model selection procedure revealed that the drift rate was affected by Contrast (see left column top panel of Figure 5 and Table 1), but not by SAT nor Force (for de-

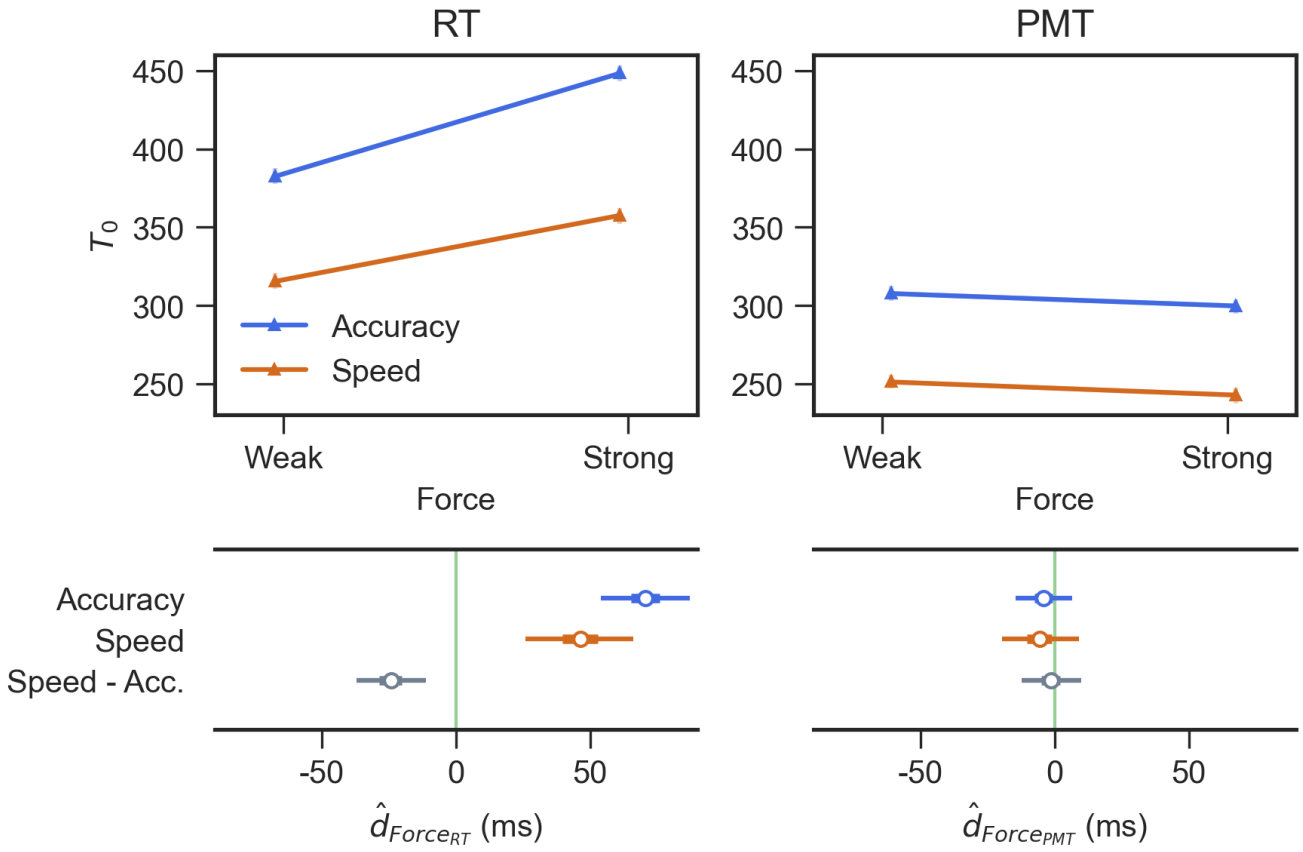

Figure 4. Effect of Force on the parameter T_0 estimated on RTs (left panel) vs. on PMTs (right panel).
Upper row: estimated mean values and one SD as shaded lines (barely visible).
Bottom row: posterior distribution of the DDM parameter for the effect of Force on T_0 at each level of SAT, and of its interaction with SAT (see T_0 SAT × FC in Table 1).

tails, see model selection results in Table B1). The boundary parameter was affected by SAT, being smaller for the speed condition, and it was not reliably affected by Force in either SAT condition (see left panel in Figure 6). The starting point parameter revealed no evidence for an effect of Force (with the provision that the corresponding CrI barely included 0; Table 1).

Summary and discussion of behavioral observations and the DDM fit on RT. Regarding the targeted encoding processes, performance decreased with increasing contrast. This variation was captured by the model as the predicted negative relation between contrast and T_0, as well as the predicted negative relation between contrast and drift – the latter resulting in a positive relation between contrast and T_D. However, the effect of Contrast on T_0 was only present when participants were asked to emphasize speed over accuracy. Because T_0 aggregates $T_{encoding}$ and $T_{response}$, which cannot be estimated separately, a *post-hoc* account of the unexpected absence of Contrast effect in the Accuracy condition would be to hypothesize opposite effects on the two components of T_0. We show in the next sections on *MT* and *PMT* measures that this account is very unlikely.

Regarding the targeted motor processes, *RT* increased with higher Force demands, while the rate of correct responses was not affected by Force. The variation in *RT*s was captured by the DDM as a rather selective increase in T_0 that left all other parameters unaffected (with the possible exception of bias). However, there was a discrepancy between the effects of Force estimated on *RT* vs. on the T_0 fitted values. In the *RT* analysis, the effect size of Force was 48 ms in the accuracy condition and 35 ms in the speed condition, and Force did not interact with SAT; in the T_0 fitted values, Force interacted with SAT and its values for the Accuracy and Speed conditions were 71 ms and 46 ms, respectively. In the next section, we quantify factor effects on *MT* and we build on our assumption linking *MT* to $T_{response}$ to assess whether the more faithful capture of motor processes comes from linear regression on *RT* or from DDM fitting.

Regarding the parameters affecting the decision time T_D, the observed effects were exactly as predicted: the drift was

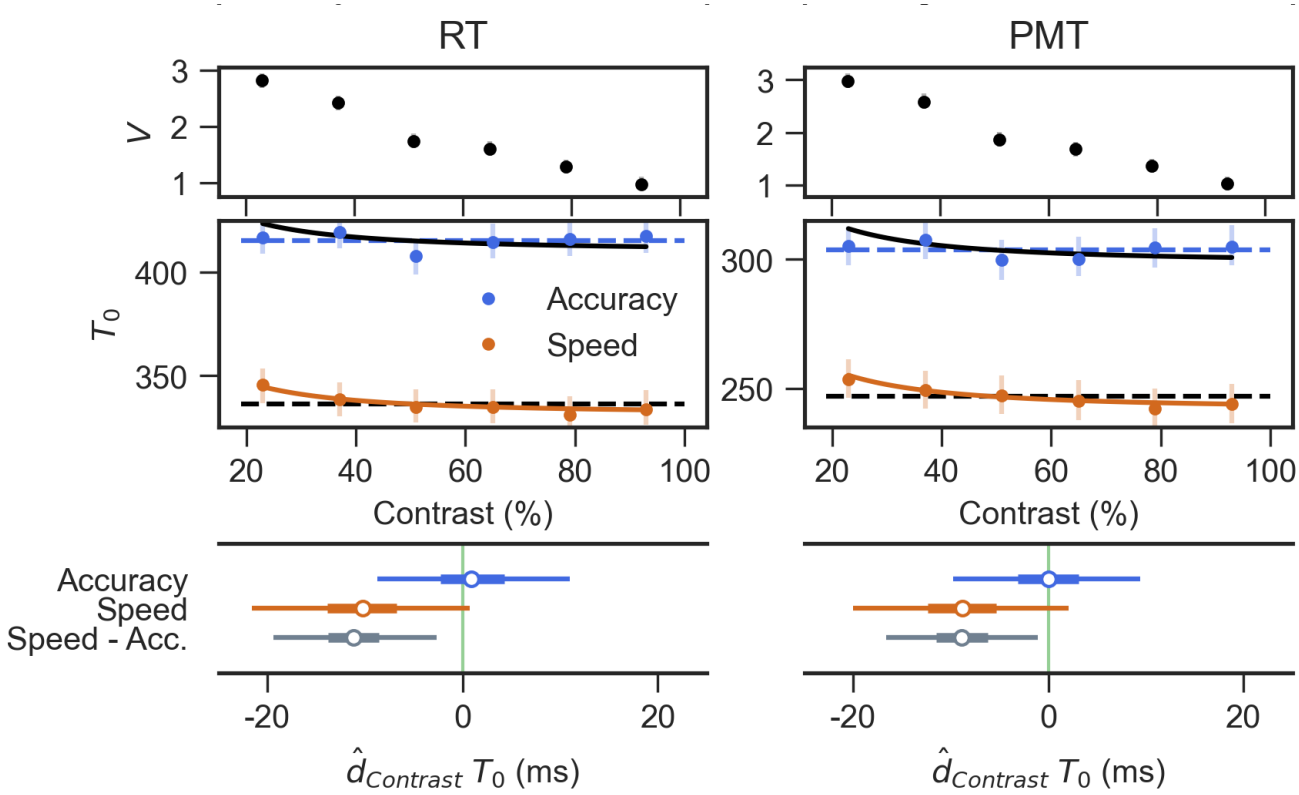

Figure 5. Effect of Contrast on the parameters drift rate (V) and T_0 estimated with DDM on RT (right column) and PMT (left column) in the speed and accuracy conditions.
Panels in the top row show the estimated drift rate (V) mean values and one SD as shaded lines along with the contrast levels. Panels in the middle show the estimated T_0 mean values and one SD as shaded lines. In the General Discussion section we describe external data whose fit is represented by the plain line and compared to the mean T_0 of each condition represented by the dashed lines. The best fit, as assessed using an R^2, between both lines is colored according to the color code of the condition.
Panels in the bottom row show the posterior distribution of the effect of contrast in accuracy and speed along with the difference between both conditions (T_0 SAT × FC in Table 1).

selectively affected by the Contrast manipulation and the boundary was selectively affected by the SAT manipulation.

Motor times (*MT*)

Linear mixed model. The following observations are illustrated in Figure 2 and in Figure 3 (third column). The SAT condition Speed reduced MT, $\hat{d}_{SAT} = -24$, $CrI = [-30, -17]$. A higher force requirement increased MT, both in the accuracy, $\hat{d}_{FC:Acc.} = 93$, $CrI = [75, 110]$, and the speed conditions, $\hat{d}_{FC:Speed} = 59$, $CrI = [48, 71]$. The interaction term confirmed that the effect of Force was indeed smaller when speed was emphasized $\hat{d}_{FC:Speed-Acc.} = -34$, $CrI = [-47, -20]$.

Contrast and its interaction with Force had no reliable effect on MT across any of the SAT conditions, as indicated by coefficients restricted to low effect sizes and CrIs containing 0.

Summary and discussion of *MT* results. We found no evidence for an effect of contrast on MT nor of its interaction with SAT. In the section on behavioral results (RT and error rates), we reported an interaction between Contrast and SAT, and speculated that Contrast may have opposite effects on $T_{encoding}$ and $T_{response}$, the two components of T_0, in the Accuracy condition. The current analysis of MT shows that this hypothesis is implausible. One alternative possibility is that response mechanisms in the Speed and Accuracy conditions are different enough that they are differently sensitive to the encoding of contrast. We come back to this issue in the General Discussion.

The expected effect of Force on MT was clear, but its size (93 ms and 60 ms in the Accuracy and Speed conditions, respectively) was much larger than that observed on RT (48 ms and 35 ms). Because each RT is the sum of its corresponding MT and PMT, this discrepancy can only be explained by an

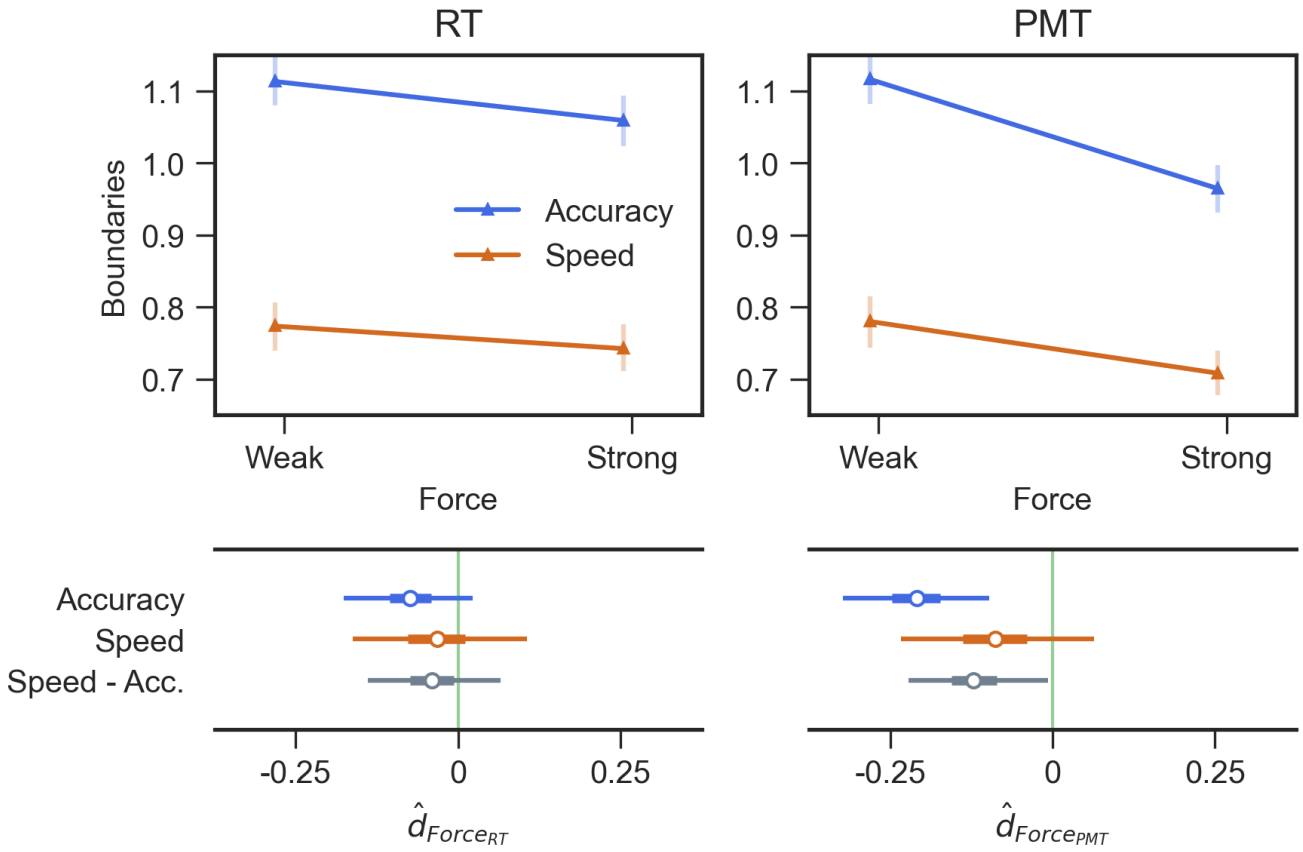

Figure 6. *Effect of Force on the boundary parameter* a *when estimated on RTs (left panel) vs. on PMTs (right panel).* Upper row: estimated mean values and one SD as shaded lines.
Bottom row: posterior distribution of DDM regression coefficients for the effect on a of Force in accuracy, speed and their difference on either chronometric variable (see *Boundaries* in Table 1).

opposite effect of Force on PMT. We pursue this issue in the next section on PMT.

The effect size of Force on MT was also remarkably higher than the effect size estimated in the previous section on T_0 (71 and 47 ms). Our assumption linking MT to the $T_{response}$ component of T_0 invites, again, a tentative compensation account in which the other component of T_0, namely $T_{encoding}$, would be sensitive to Force in the opposite direction. While this hypothesis may seem counter-intuitive, it can be directly tested in our framework by fitting DDM to PMT distributions, as we do in the next section.

Pre-Motor times (PMT)

Linear Mixed Model. The results for PMT were very similar to those reported above for RT with the important exception of the effects of Force. This can be appreciated by comparing the second and fourth columns in Figure 2 and Figure 3. The effect of Force on PMT was opposite to that observed on RT both in accuracy, $\hat{d}_{FC:Acc} = -53$, $CrI = [-79, -27]$ and speed conditions, $\hat{d}_{FC:Speed} = -27$, $CrI = [-46, -8]$. Here the effects of Force were reliably different across SAT conditions, $\hat{d}_{FC:Speed-Acc} = 26$, $CrI = [2, 50]$.

Drift Diffusion Model selection. The model selection procedure applied to PMT selected the same model structure that was selected when the procedure was applied to RT, namely M13 in Table B1.

Effects on T_0 when fitted on PMT. The patterns of effects were similar for both PMT and RT fits, with the following important exceptions (compare the two panels on Table 1). The main difference is that neither Force nor its interaction with SAT appeared to affect the T_0 estimated on PMT (Figure 4 right panel). In addition, we highlight in Figure 5 that the effect of contrast on T_0 interacted with SAT. The effect was in the expected direction in the Speed condition (although the CrI included 0) but centered on 0 in the Accuracy condition.

Effects on decision related parameters when fitted on PMT. Contrary to the results observed with a fit on RT,

DDM Par.	Factor	RT $\hat{d}$	RT 2.5%	RT 97.5%	PMT $\hat{d}$	PMT 2.5%	PMT 97.5%
Drift	(Intercept)	1.79	2.03	1.51	1.91	2.19	1.62
	Contrast	-1.77	-1.46	-2.08	-1.90	-1.55	-2.23
Boundaries	(Intercept)	1.08	1.02	1.15	1.05	0.99	1.10
	SAT	-0.33	-0.40	-0.25	-0.30	-0.38	-0.22
	FC:Acc.	-0.06	-0.14	0.02	-0.17	-0.26	-0.08
	FC:Speed	-0.03	-0.13	0.08	-0.07	-0.19	0.05
	FC:Speed-Acc.	0.03	-0.05	0.11	0.10	0.01	0.18
Bias	(Intercept)	0.50	0.48	0.53	0.49	0.47	0.51
	FC	0.004	-0.001	0.01	0.002	-0.002	0.006
T_0 (ms)	(Intercept)	414.6	391.0	439.1	301.2	280.0	322.8
	SAT	-77.7	-106.1	-49.0	-51.5	-77.7	-22.7
	FC:Acc.	70.6	54.1	87.2	-4.2	-14.6	6.6
	FC:Speed	46.3	26.0	66.2	-5.6	-19.7	9.1
	FC:Speed-Acc.	-24.3	-37.2	-11.3	-1.4	-12.4	9.8
	Contrast:Acc.	1.0	-8.8	11.0	0.0	-9.7	9.4
	Contrast:Speed	-10.2	-21.6	0.7	-8.9	-20.0	2.0
	Contrast:Speed-Acc.	-11.2	-19.3	-2.7	-8.8	-16.6	-1.1
	FC × Contrast:Acc.	-6.0	-16.0	4.1	-2.5	-11.9	6.8
	FC × Contrast:Speed	5.5	-3.6	14.7	6.9	-1.6	15.5
	FC × Contrast:Speed-Acc.	11.5	-1.3	24.7	9.3	-4 .1	21.5

Table 1
Comparison of Estimated differences between conditions levels ($\hat{d}$) across fits on RT and on PMT. Columns labelled $\hat{d}$ refer to the maxima a posteriori from the corresponding marginal posterior distribution. Columns labelled 2.5% and 97.5% refer to the CrI intervals. Colors refer to the sign of the estimate, red for positive values, blue for negative values.

in the fit on *PMT* the boundary parameter was affected by Force and this effect interacted with SAT. Increasing Force resulted in a lower boundary parameter, an effect that was much reduced (if not absent) in the Speed condition (Figure 6 right panel). The other decision parameters, drift and bias, were roughly similar when derived from fits on *PMT* and *RT* (compare the two panels on Table 1).

Summary and discussion of *PMT* observations and their DDM fit. The effects of Contrast were consistent with those observed on *RT*, revealing "opposite" effects on T_0 and drift, including the fact that Contrast interacts with SAT instructions showing that the expected effect is mainly present in the Speed condition only (Figure 5). *PMT* does not include *MT*, which we hypothesized to be strongly linked to $T_{response}$. Therefore, in this analysis, T_0 provides a reasonable estimate of $T_{encoding}$. In sum, encoding processes are orderly affected by Contrast in the Speed condition only.

Conversely, the effect of Force was remarkably different for *RT* and for *PMT*. The linear models revealed Force effects of opposite signs on *RT* and *PMT*, whereby *PMT*s were shorter with stronger force (Figure 2). The DDM fit of *PMT* did not capture this effect on T_0 (Figure 4). This absence of effect is consistent with the assumption that, here, T_0 indexes force-independent encoding processes (i.e. $T_{encoding}$). It undermines the tentative hypothesis (from the section on *MT*) that opposite effects of Force compensate one another on the two components of T_0, $T_{encoding}$ and $T_{response}$. Instead, the DDM fit attributed the effect of Force to the boundary parameter. Boundary decreased with increasing Force requirements, in the Accuracy condition only (Figure 6). This means that the DDM fits on *RT* and on *PMT* lead to different attributions of the Force effect, particularly in the Accuracy condition.

General summary of findings

Altogether when a DDM is fitted on RT we observe the expected results. A manipulation of contrast translates into an effect on the T_0 and the drift rate. A manipulation of force specifically impacts the T_0 while leaving the decision related parameters unchanged. Finally as observed on a regular basis, changing SAT instructions translated into an adjustment of both the decision boundaries and T_0. This last manipulation however also resulted in a change in the expected effect of contrast. When instructions where to favor accuracy over speed the expected decrease in T_0 with the increase in contrast was not found. Fitting the DDM on PMT also proved to be challenging to the interpretation usually made from model parameters as in such case the force manipulation also affects the estimated decision boundaries.

General Discussion

The reported findings help clarifying how faithfully a quantitative model such as DDM can separate decision and non-decision processes. In this General Discussion, we consider the consequences of observing that the response output manipulation changed decision related parameters, in contrast to common expectations. We then re-examine the fact that T_0 did not conform to the predictions regarding visual encoding processes in the accuracy condition. Finally, we discuss more generally why observing different findings across SAT conditions may challenge the assumed generative model, despite the rather canonical setting we used.

T_0 and motor processes

The common assumption that T_0 contains processes related to motor execution or, more specifically, assuming that $T_{response}$ is equal to MT (Luce, 1986; Weindel et al., 2021), entails two predictions. The first one is that T_0 estimated using a DDM fitted on RT should be sensitive to the force manipulation whereas T_0 fitted on PMT (hence corresponding to $T_{encoding}$) should not. As expected MT (measured) and T_0 (estimated) are both influenced by the force required to respond. We found no effect of force on $T_{encoding}$. This pattern of results strengthens the hypothesis that T_0 is the sum of two components, one of which captures motor processes.

The second prediction is that decisional parameters should be the same when estimated on RT *vs.* PMT. This is because motor processes are hypothesized to start after the threshold has been reached. This prediction proved inaccurate, most clearly in the condition combining high force condition and accuracy; there, the threshold parameter was lower in the PMT than in the RT DDM fit (Figure 6). This implies that the EMG onset does not index the end of the accumulation process estimated by DDM.

If the standard fit on RT is taken as a reference to establish decision duration, then the shorter decision durations estimated based on PMT invites the inference that the decision process (*i.e.,* accumulation of evidence) continues beyond EMG onset. In other words, the appropriate interpretation of the response component of T_0 (*i.e.,* $T_{response}$) is it that starts as late as after EMG onset. Evidence that the decision process is pursued during response execution has previously been observed (Buc Calderon, Dewulf, Gevers, & Verguts, 2017; Resulaj, Kiani, Wolpert, & Shadlen, 2009; Selen, Shadlen, & Wolpert, 2012). Crucially, the response settings in those previous studies were very different, *e.g.,* pointing or reaching movements, and response execution took much longer than in (isometric) button presses (for a related discussion see Burle, Roger, Vidal, & Hasbroucq, 2008; Scaltritti, Job, Alario, & Sulpizio, 2020).

In addition to modifying the standard interpretation of $T_{response}$, the previous paragraph begs the question of the interpretability of EMG onsets in the context of DDM. An EMG onset occurring during the evidence accumulation process would challenge our hypothesis that $PMT = T_{encoding} + T_D$. Other aspects of the data help clarifying this issue. Goodness of fit was comparable for PMT and RT fits. PMT could thus be generated by a diffusion process. Moreover, the other decision parameters (drift and bias) were essentially similar across both fits. This would happen if the EMG onset and the end of T_D correspond to the crossing of two different thresholds for the same diffusion process. Such double-threshold hypothesis is congruent with a recent theory proposing that motor execution is determined by an evolving decision variable (Servant et al., 2021; Servant, White, Montagnini, & Burle, 2015).

The unexpected findings occurred in the accuracy SAT condition, only. This could suggest strategical adjustments of the participants across SAT conditions. We come back to this point in the last section.

T_0 and encoding processes

We observed, as expected, a negative relation between contrast and non-decision processes whether estimated on the whole RT (T_0) or the PMT ($T_{encoding}$). Unexpectedly, this was only true when participants had to respond rapidly, but not when accuracy was emphasized. We undertook additional analysis to better understand this discrepancy.

In psychophysical (Harwerth & Levi, 1978) and neurophysiological (Reynaud, Masson, & Chavane, 2012) studies, latencies related to the encoding of visual gratings have been found to be negatively and non-linearly related to stimulus contrast. This well established effect concerns the early ($\sim$ 80 ms) discharge latency of V1 neurons in macaques. Such an early process is not expected to be modulated by response strategy adjustments such as SAT. Therefore, if our estimated $T_{encoding}$ reflects these stages, it should be non-linearly related to the contrast manipulation similarly across SAT levels.

This prediction was tested against external data from Reynaud et al. (2012). These authors measured the latency of the onset of cortical neuron activity in visual area V1 in awake monkeys (described in Appendix C1). Their data provide onset times of V1 neurons, revealing a non-linear increase of onset-latencies across visual gratings of decreasing contrast. Reynaud et al. (2012) then used these latencies to fit a well established neurophysiologically motivated model (Naka & Rushton, 1966) using the inverted equation as in Barthélemy, Fleuriet, and Masson (2010) (summarized in Appendix C). If $T_{encoding}$ really reflects the low level extraction of stimulus features, its latency should follow the same quantitative relationship. We tested whether the modulation of our estimated $T_{encoding}$ matches the modulation observed in V1 (Reynaud et al., 2012) by calculating an R^2 across contrast levels between the centered V1 data and our centered $T_{encoding}$, broken down

by SAT and force conditions (see Figure 5 or Figure C1 in Appendix C for a close-up).

In the condition where speed is emphasized and the force required is weak, there was an almost perfect adjustment ($R^2 = 0.99$) between the model fitted on V1 discharge latencies and our out-of-sample $T_{encoding}$ estimates. When speed is emphasized but the force required is high, we observed a lower but still substantial agreement ($R^2 = 0.63$). Thus, under speed instructions, $T_{encoding}$ likely reflects a meaningful duration pertaining to visual processing.

In contradistinction, the model fitted on V1 discharge latencies did not adjust to $T_{encoding}$ when accuracy was emphasized, neither in the low nor the high force conditions ($R^2 = -0.71$, $R^2 = -1.14$, respectively).[5] Thus, under accuracy instructions, the variations of $T_{encoding}$ are not easily linked to variations of early visual process durations.

Overall, this analysis consolidates the standard interpretation of $T_{encoding}$ durations in speed but undermines it in accuracy.

SAT and non-decision processes

As reported in the two previous sections, the manipulation of SAT instructions played a crucial role for the congruence between experimental and physiological predictions on the estimated non-decision durations. Consistently, these predictions were found to be true mainly when participants where instructed to speed their response. Analysing the effect of SAT on the estimated T_0 is also informative on the nature of this parameter.

Replicating previous studies, T_0 was found to be sensitive to SAT instructions (Palmer et al., 2005; Ratcliff, 2006; Voss et al., 2004; Weindel et al., 2021). This is coherent with the observations made multiple times, including in this manuscript, that motor processes as captured by MT are sensitive to SAT (Spieser et al., 2017; Steinemann et al., 2018; Weindel et al., 2021). It would then be tempting to conclude that motor processes alone drive the SAT effect on the estimated T_0. This is found to be false as the SAT effect on $T_{encoding}$ is of 50 ms, suggesting that most of the SAT effect on T_0 is actually estimated to be in $T_{encoding}$.

Smith and Lilburn (2020) have shown that inferences of SAT effect on T_0 can be influenced by an inappropriate modeling of how the evidence is entering the decision process. Here, we used a canonical stimulus comparison task which we think aligns with the theory behind the DDM. We are thus left with the interpretation that approximately 1/3 of the effect of SAT found on RT (and 2/3 of the effect of SAT on T_O) is of unknown pre-motor origin given the residual SAT effect of 50 ms on $T_{encoding}$. But given the observations that the DDM does not account for all effects of SAT manipulations (Rafiei & Rahnev, 2019, 2021) and that predictions are rejected only in the accuracy condition, we suggest that the DDM is not decomposing decision and non-decision processes as expected by modellers when accuracy is emphasized. This could further fuel the idea that the DDM is the reduced version of an overarching model only reasonable in the condition where RTs are (highly) speeded (Verdonck et al., 2020).

One alternative interpretation can be made given that an increased T_0 is linked to an added non-decision process (Ratcliff & McKoon, 2008). Following this logic and contrary to the common assumption that non-decision processes refer to encoding and motor latencies, an additional pre-motor non-decision stage is present in the accuracy condition. This is actually congruent with a recent finding which shows, using electro-encephalographic data, an additional pre-decision stage specifically in the accuracy condition (van Maanen, Portoles, & Borst, 2021). Together with the observation of decreased boundaries and decreased motor times under speed stress, this would show that SAT is achieved by participants by multiple adjustments therefore definitely breaking the selective influence of SAT hypothesis. If this interpretation turned out to be true, would remain the question of the functional role of the added stage. One could for example suggest that when speed is emphasized, participants are accumulating evidence as soon as evidence enters the system while in accuracy they are using a fixed time to start the accumulation of evidence (respectively the hypothesis of visual short term memory *vs.* a release from inhibition mechanism Smith, Ratcliff, & McKoon, 2014), congruent with the absence of contrast effect on $T_{encoding}$ in accuracy.

Conclusion

The combination of EMG measurement and model fit for a canonical perceptual decision task, questions the usual interpretation of the drift diffusion model in terms of cognitive processes.

The drift diffusion model postulates a partition of the reaction time into decisional and non-decisional times (encoding and response execution durations). However, we show that the EMG onset does not index the end of the accumulation process, contrary to what is commonly assumed. Moreover, when accuracy is emphasized over speed, the model does not allow to recover the encoding time. This may question the validity of the interpretation of DDM parameters in many studies. Therefore, providing decision making models accounting for non-decision times is a major issue for future research.

[5] When computing the R^2 on T_0 estimated from RT rather than from PMT we find the similar results: $R^2 = 0.91$ and $R^2 = 0.88$ in speed, low and high force respectively *vs.* $R^2 = -1.11$ and $R^2 = -0.02$ in accuracy, low and high force. The different levels of adequacy between the V1 data (Reynaud et al., 2012) and DDMs $T_{encoding}$ were further replicated with a different unpublished data set involving a three-level manipulation of contrast.

References

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

Barthélemy, F. V., Fleuriet, J., & Masson, G. S. (2010). Temporal dynamics of 2d motion integration for ocular following in macaque monkeys. *Journal of neurophysiology*, *103*(3), 1275–1282.

Boehm, U., Annis, J., Frank, M. J., Hawkins, G. E., Heathcote, A., Kellen, D., … Wagenmakers, E.-J. (2018). Estimating Between-Trial Variability Parameters of the Diffusion Decision Model: Expert Advice and Recommendations. *PsyArXiv*. doi: 10.17605/OSF.IO/KM28U

Boehm, U., Marsman, M., Matzke, D., & Wagenmakers, E.-J. (2018). On the importance of avoiding shortcuts in applying cognitive models to hierarchical data. *Behavior research methods*, *50*(4), 1614–1631.

Brown, S., & Heathcote, A. (2005). A ballistic model of choice response time. *Psychological review*, *112*(1), 117–128. doi: 10.1037/0033-295X.112.1.117

Buc Calderon, C., Dewulf, M., Gevers, W., & Verguts, T. (2017). Continuous track paths reveal additive evidence integration in multistep decision making. *Proceedings of the National Academy of Sciences*, *114*(40), 10618–10623. doi: 10.1073/pnas.1710913114

Burle, B., Possamaï, C.-A., Vidal, F., Bonnet, M., & Hasbroucq, T. (2002, nov). Executive control in the Simon effect: an electromyographic and distributional analysis. *Psychological research*, *66*(4), 324–36. doi: 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*, 17-21.

Cisek, P., Puskas, G. A., & El-Murr, S. (2009). Decisions in changing conditions: the urgency-gating model. *Journal of Neuroscience*, *29*(37), 11560–11571.

Donders, F. C. (1868). Die schnelligkeit psychischer processe: Erster artikel. *Archiv für Anatomie, Physiologie und wissenschaftliche Medicin*, 657–681.

Dutilh, G., Annis, J., Brown, S. D., Cassey, P., Evans, N. J., Grasman, R. P., … others (2016). The quality of response time data inference: A blinded, collaborative assessment of the validity of cognitive models. *Psychonomic bulletin & review*, 1–19.

Dutilh, G., Wagenmakers, E.-J., Visser, I., & van der Maas, H. L. (2011). A phase transition model for the speed-accuracy trade-off in response time experiments. *Cognitive Science*, *35*(2), 211–250.

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

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. doi: 10.1037/a0039653

Gramfort, A., Luessi, M., Larson, E., Engemann, D. A., Strohmeier, D., Brodbeck, C., … Hämäläinen, M. (2013). MEG and EEG data analysis with MNE-Python. *Frontiers in Neuroscience*(7 DEC). doi: 10.3389/fnins.2013.00267

Harwerth, R. S., & Levi, D. M. (1978). Reaction time as a measure of suprathreshold grating detection. *Vision research*, *18*(11), 1579–1586.

Heathcote, A., & Love, J. (2012). Linear deterministic accumulator models of simple choice. *Frontiers in Psychology*, *3*(AUG), 1–19. doi: 10.3389/fpsyg.2012.00292

Ho, T. C., Brown, S., & Serences, J. T. (2009). Domain general mechanisms of perceptual decision making in human cortex. *Journal of Neuroscience*, *29*(27), 8675–8687.

Kruschke, J. K. (2010). Bayesian data analysis. *Wiley Interdisciplinary Reviews: Cognitive Science*, *1*(5), 658–676.

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*. Retrieved from `http://joss.theoj.org/papers/10.21105/joss.01143` doi: 10.21105/joss.01143

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. doi: 10.3758/s13428-016-0740-2

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. doi: 10.1016/j.bspc.2015.09.004

Luce, R. (1986). Response Times: Their Role in Inferring Elementary Mental Organization. *Oxford University Press New York*(3), 562. doi: 10.1093/acprof:oso/9780195070019.001.0001

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.

Naka, K., & Rushton, W. A. (1966). S-potentials from colour units in the retina of fish (cyprinidae). *The Journal of physiology*, *185*(3), 536–555.

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*, 1075–1100.

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

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

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. doi: 10.1167/5.5.1

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

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. doi: 10.1037/a0036801

Rafiei, F., & Rahnev, D. (2019). Speed-accuracy tradeoff heightens serial dependence. *Journal of Vision*, *19*(10), 289c–289c.

Rafiei, F., & Rahnev, D. (2021). Qualitative speed-accuracy tradeoff effects that cannot be explained by the diffusion model under the selective influence assumption. *Scientific reports*, *11*(1), 1–19.

Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review*, *85*(2), 59–108. doi: 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., & McKoon, G. (2008). The diffusion decision model: theory and data for two-choice decision tasks. *Neural computation*, *20*(4), 873–922. doi: 10.1162/neco.2008.12-06-420

Ratcliff, R., & McKoon, G. (2018). Modeling numerosity representation with an integrated diffusion model. *Psychological review*, *125*(2), 183.

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. doi: 10.1016/j.tics.2016.01.007

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

Ratcliff, R., & Tuerlinckx, F. (2002b). Estimating parameters of the diffusion model: Approaches to dealing with contaminant reaction times and parameter variability. *Psychonomic bulletin & review*, *9*(3), 438–481.

Resulaj, A., Kiani, R., Wolpert, D. M., & Shadlen, M. N. (2009). Changes of mind in decision-making. *Nature*, *461*(7261), 263–266. doi: 10.1038/nature08275

Reynaud, A., Masson, G. S., & Chavane, F. (2012). Dynamics of local input normalization result from balanced short-and long-range intracortical interactions in area v1. *Journal of neuroscience*, *32*(36), 12558–12569.

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. doi: 10.1113/expphysiol.1998.sp004165

Scaltritti, M., Job, R., Alario, F.-X., & Sulpizio, S. (2020). On the boundaries between decision and action: Effector-selective lateralization of beta-frequency power is modulated by the lexical frequency of printed words. *Journal of Cognitive Neuroscience*, *32*(11), 2131–2144.

Schmidgen, H. (2002). Of frogs and men: the origins of psychophysiological time experiments, 1850–1865. *Endeavour*, *26*(4), 142 - 148. doi: https://doi.org/10.1016/S0160-9327(02)01466-7

Selen, L. P. J., Shadlen, M. N., & Wolpert, D. M. (2012). Deliberation in the Motor System: Reflex Gains Track Evolving Evidence Leading to a Decision. *Journal of Neuroscience*, *32*(7), 2276–2286. doi: 10.1523/JNEUROSCI.5273-11.2012

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

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

Servant, M., White, C., Montagnini, A., & Burle, B. (2016, oct). 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. doi: 10.1162/jocna00989

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

Smith, P. L., Ratcliff, R., & McKoon, G. (2014). The diffusion model is not a deterministic growth model: Comment on Jones and Dzhafarov (2014). *Psychological Review*, *121*(4), 679–688. doi: 10.1037/a0037667

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

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

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.

Stevens, S. S. (1961). To honor fechner and repeal his law. *Science*, *133*(3446), 80–86.

Stine, G. M., Zylberberg, A., Ditterich, J., & Shadlen, M. N. (2020). Differentiating between integration and non-integration strategies in perceptual decision making. *Elife*, *9*, e55365.

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

Usher, M., & McClelland, J. L. (2001). The time course of perceptual choice: The leaky, competing accumulator model. *Psychological Review*, *108*(3), 550–592. doi: 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.

van Maanen, L., Portoles, O., & Borst, J. P. (2021). The discovery and interpretation of evidence accumulation stages. *Computational Brain & Behavior*, 1–21.

Verdonck, S., Loossens, T., & Philiastides, M. G. (2020). The leaky integrating threshold and its impact on evidence accumulation models of choice rt.

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

Weindel, G., Anders, R., Alario, F.-X., & Burle, B. (2021). Assess-

ing model-based inferences in decision making with single-trial response time decomposition. *Journal of Experimental Psychology: General*.

Wiecki, T., Sofer, I., & Frank, M. (2016). *Hddm 0.6. 0 documentation.*

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.

Appendix A

From Linear model parameters to estimated effects

All regression models, including the regression on the DDM parameters followed the same factor coding scheme. The SAT factor was coded as a treatment factor (0 for accuracy and 1 for speed). The force condition was coded as a sum contrast (-0.5 for weak and 0.5 for strong force). The factor contrast was centered on the middle value and scaled so that -0.5 represented the lowest contrast and 0.5 the highest. A summary of this factor coding for the regression models is given in the matrices below, the first row represent the original levels, the second row the values on which the regression were estimated :

$$SAT = \begin{pmatrix} Accuracy & Speed \\ 0 & 1 \end{pmatrix}$$

$$Force = \begin{pmatrix} Weak & Strong \\ -0.5 & 0.5 \end{pmatrix}$$

$$Contrast = \begin{pmatrix} 23\% & 37\% & 51\% & 65\% & 79\% & 93\% \\ -.5 & -.3 & -.1 & .1 & .3 & .5 \end{pmatrix}$$

These coding features were chosen to ease the interpretation of the resultant 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 SAT emphasis is on accuracy, and at an intermediate value for the predictors contrast and FC.

Given these coding features and the Bayesian nature of the estimation we can estimate the effect of a factor in a given condition and preserve the uncertainty associated with the effects. E.g. To compute the effect of Force in the speed condition we can add the interaction term $\beta_{SAT \times Force}$ to the estimated β_{Force} in the accuracy condition. As coefficients are estimated using a MCMC procedure this addition is done on each MCMC iteration, allowing to keep the uncertainty around the resulting coefficient.

Units of the g/LMMs parameters. For the LMMs on proportion correct, RT, PMT and MT, the data was transformed prior to the modeling (logit for proportion correct and log for the other variables). Using Monte Carlo Markov Chain (MCMC) processes, we back-transformed the predictions of the linear models for the chosen differences at each iteration, with the exponential for log transformed variables (LMM) or the inverse logit for proportion correct (gLMM). This preserves the uncertainty around the parameter values while reverting them to the natural units of the dependent variables.

Appendix B
Model Selection

As seen in table B1 the DIC criterion almost always favor the complex models over the simpler one. However two patterns are consistent across the models, allowing the boundaries and the bias to vary with force conditions and drift rate to vary with SAT in addition to the contrast always improves the goodness of fit as assessed by the DIC. However when considering the BPIC criterion, initially intended to correct the complexity bias of the DIC, only allowing the variation of force on boundaries and on the bias seems to improve the goodness of fit. Hence based on BPIC we select the model allowing the boundaries and the bias to vary across force condition in addition to the designed base model (respectively M13 and M1 in Table B1). Importantly the results of the model selection is the same for a fit on PMT. The goodness of fit both on RT and PMT as displayed with quantile probability plot (see Figure B2 and B1) is satisfactory in most conditions but the amount of errors is rather misfitted when considering a high force especially in the fit on PMT (that pattern is common across all tested models).

Appendix C
Predictions by V1 neuron activation onset

Reynaud et al. (2012) performed a measurement of the temporal activation of V1 neurons in awake monkey using voltage sensitive dye and the variation of this temporal activation with the contrast of stimuli close to the one used in this study. They then fitted the relationship between onset of V1 neurons activity and contrast with an inverted Naka-rushton equation from Barthélemy et al. (2010) :

$$\tau_c(c) = \tau_{max} + \tau_{shift} \cdot \frac{c_n}{c_n + s_{50}^n}$$

Where c is contrast, τ_{max} and τ_{shift} are respectively the minimum latency observed at highest contrast and the maximum decrease in latency. n is the estimated latency shift exponent, s_{50} the estimated half decay contrast value. For the purpose of our analysis we recovered the values of the parameters estimated by Reynaud et al. (2012) and draw the predictions associated with the mean contrast levels used in our study.

Figure C1 represents the adjustment between point estimate of T_0 and $T_{encoding}$ with the curve predicted by the recovered parameters of Reynaud et al. (2012) for the inverted Naka-Rushton equation. However, since monkey latencies are shorter than those we observed with humans, we centered them by subtracting their mean and adding the mean of the $T_{encoding}$ computed over all contrast levels and participants.

	Bound.	Drift	T_0	Bias	sBias	sDrift	sT_0	$BPIC_{RT}$	DIC_{RT}	$BPIC_{PMT}$	DIC_{PMT}
M1	S	C	C × F × S	1	S	1	1	-12963	13447	-16125	-16595
M2	S	C × S	C × F × S	1	S	1	1	-12931	-13483	-16072	-16618
M3	S	C × F	C × F × S	1	S	1	1	-12866	-13419	-16105	-16645
M4	S	C × S × F	C × F × S	1	S	1	1	-12800	-13467	-16003	-16668
M5	S × F	C	C × F × S	1	S	1	1	-13157	-13661	-16471	-16967
M6	S × F	C × S	C × F × S	1	S	1	1	-13108	-13689	-16423	-16993
M7	S × F	C × F	C × F × S	1	S	1	1	-13042	-13614	-16393	-16959
M8	S × F	C × S × F	C × F × S	1	S	1	1	-12946	-13639	-16272	-16964
M9	S	C	C × F × S	F	S	1	1	-13022	13519	-16132	-16617
M10	S	C × S	C × F × S	F	S	1	1	-12981	-13548	-16075	-16637
M11	S	C × F	C × F × S	F	S	1	1	-12926	-13493	-16115	-16668
M12	S	C × S × F	C × F × S	F	S	1	1	-12860	-13540	-16006	-16687
M13	S × F	C	C × F × S	F	S	1	1	**-13212**	-13729	**-16474**	-16987
M14	S × F	C × S	C × F × S	F	S	1	1	-13150	**-13750**	-16418	**-17007**
M15	S × F	C × F	C × F × S	F	S	1	1	-13095	-13654	-16392	-16976
M16	S × F	C × S × F	C × F × S	F	S	1	1	-13000	-13728	-16284	-16987

Table B1

Summary of the tested models displaying for each model (row) which parameters could vary with experimental conditions (S, F and C respectively for SAT, Force and Contrast, 1 indicates that only 1 estimate was fitted across all conditions). sBias, sDrift and sT_0 refer to the inter-trial variability parameters of the corresponding main parameters. The results in terms of BPIC and DIC are presented in the two last columns.

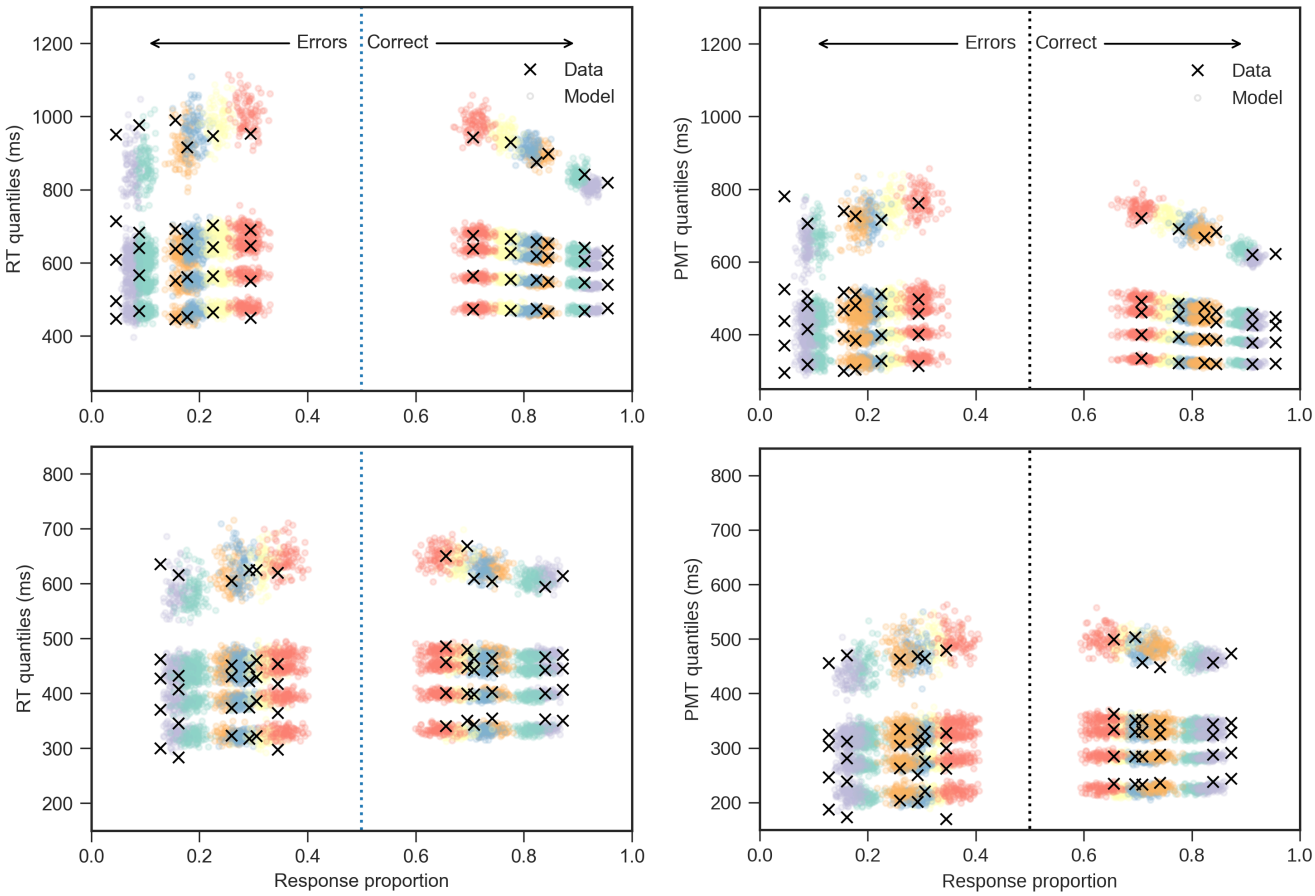

Figure B1. Quantile-probability plots (Ratcliff & McKoon, 2008) for high force based on a fit on RT (left column) and PMT (right column), in the accuracy (upper row) and speed (lower row) conditions, computed from the best fitting model. The X-axis displays obtained response proportion across contrast levels (color coded), 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 250 parameter values from the joint posterior distribution and computing their associated predicted performance. The misfit of the DDM is particularly apparent in the fit on PMT in accuracy (upper right corner), where the DDM clearly predicts a response proportion lower than the one observed on the data.

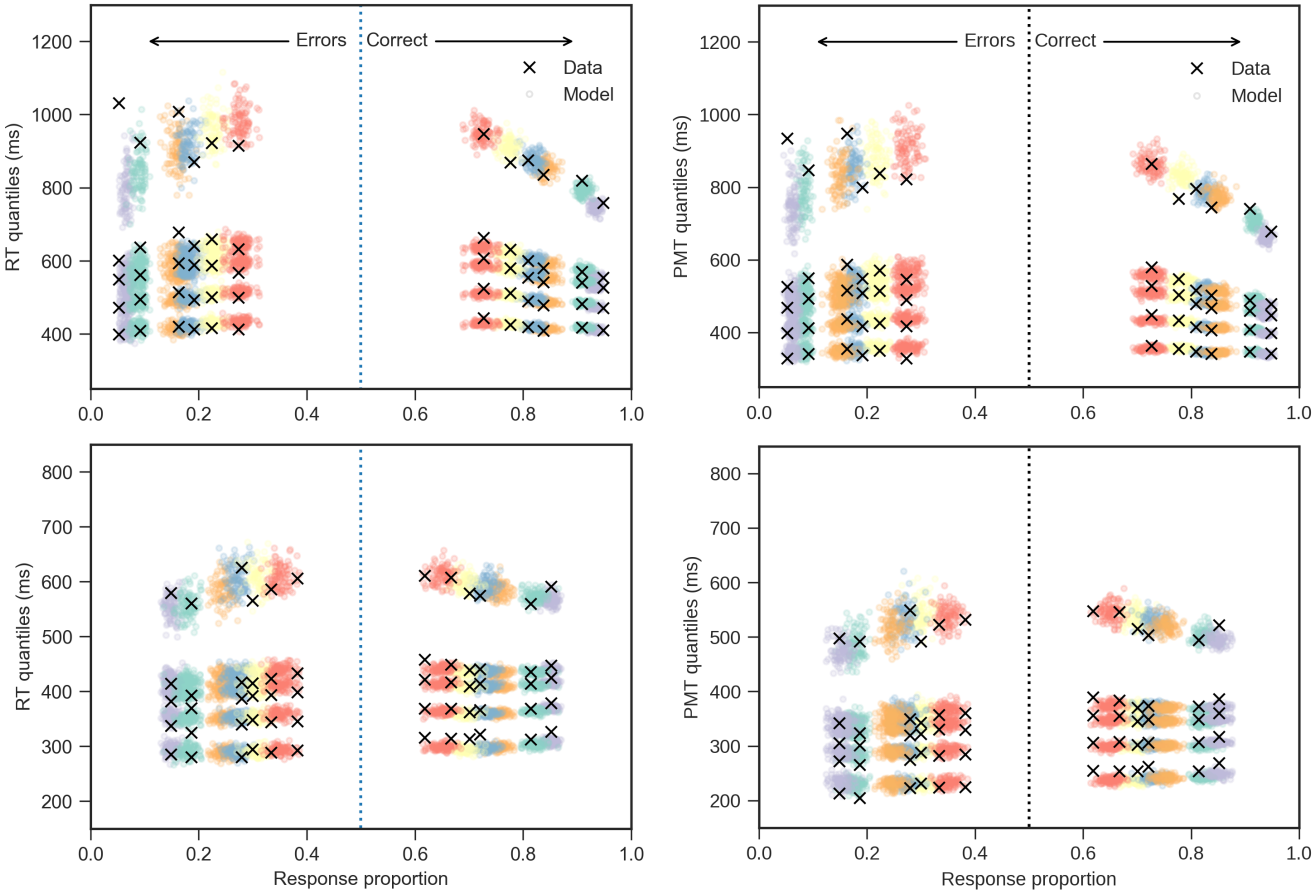

Figure B2. Quantile-probability plots (Ratcliff & McKoon, 2008) for low force based on a fit on RT (left column) and PMT (right column), in the accuracy (upper row) and speed (lower row) conditions, computed from the best fitting model.
The misfit of the DDM is particularly apparent in the fit on PMT in accuracy (upper right corner), where the DDM clearly predicts a response proportion lower than the one observed on the data.

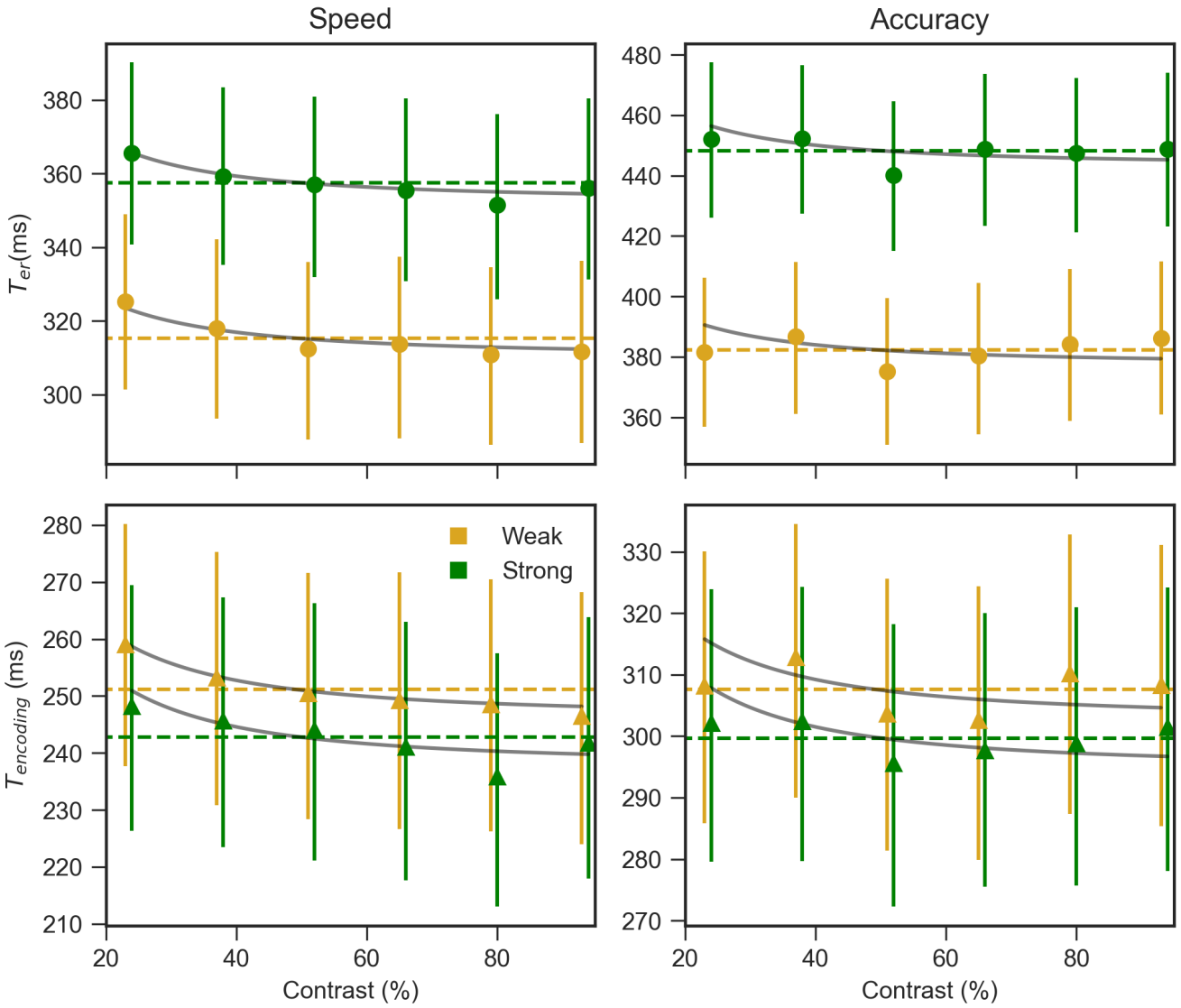

Figure C1. Estimated T_0 (obtained from a DDM fit on *RT*) and $T_{encoding}$ (obtained from a fit on *PMT*) across contrast levels and splitted between SAT and Force conditions. Bars around the point estimate represents 65% CrI of the population mean. The colored lines represent the mean of T_0 or $T_{encoding}$ for each corresponding sub-cell. The grey lines represents the values predicted by the parameters of the inverted Naka-Rushton recovered from Reynaud, Masson, and Chavane 2012. These prediction have been first centered on 0 by subtracting their mean then rescaled by adding the mean of the corresponding sub-cell. Grey and color lines have therefore the same mean in each sub-cell formed by the combination of SAT and Force levels.