

# Computational modelling of the speed–accuracy tradeoff: No evidence for an association with depression symptomatology

James A. Grange<sup>1</sup>

<sup>1</sup> School of Psychology, Keele University

## Abstract

Successful decision making often requires finding the right balance between the speed and accuracy of responding: Emphasising speed can lead to error-prone performance, yet emphasising accuracy leads to a slowing of performance. Such speed–accuracy tradeoffs (SATs) therefore require establishing appropriate response settings to optimise performance in response to changing environmental demands. Such strategic adaptation of response settings relies on the striatum component of the basal ganglia, an area implicated in depression. The current study explored the association between depression symptomatology and SAT performance. Two experiments presented participants with an SAT paradigm embedded within a simple decision-making task, together with measures of depression symptomatology. Experiment 1 (N = 349) was correlational, whereas Experiment 2 was a two-phase experiment where participants (N = 501) were first pre-screened on depression symptomatology and extreme-low and extreme-high responders (total N = 91) were invited to Phase 2. Behavioural data were modelled with a drift diffusion model. Behavioural data and associated diffusion modelling showed large and robust SAT effects. Emphasising accuracy led to an increase in boundary separation, an increase in drift rate, and an increase in non-decision time. However, the magnitude of the changes of these parameters with SAT instructions was not associated with measures of depression symptomatology. The results suggest that the strategic adaptation of response settings in response to environmental changes in speed–accuracy instructions do not appear to be associated with depression symptomatology.

*Keywords:* Depression; Speed–accuracy tradeoff; Diffusion modeling; Decision making

Depression is associated with a wide range of cognitive deficits (Burt, Zembler, & Niederehe, 1995; McDermott & Ebmeier, 2009; Rock, Roiser, Riedel, & Blackwell, 2014) including higher-level executive functioning (Snyder, 2013) which is essential for goal-directed behaviour. One crucial aspect of goal-directed behaviour is the ability to accurately make

rapid decisions. Many tests of cognition and executive functioning require rapid decisions from participants in the form of choice response time (RT) tasks. For example, in the task switching paradigms (a test of cognitive flexibility; Grange & Houghton, 2014) participants are presented with simple stimuli (e.g., digits) and asked to rapidly switch between two or more simple tasks (e.g., is the stimulus odd or even / is the stimulus lower or higher than five?). Although the switching element is of primary interest to cognitive (Kiesel et al., 2010; Vandierendonck, Liefoghe, & Verbruggen, 2010) and clinical (e.g., Ravizza & Salo, 2014) researchers, rapid decision-making is embedded within the primary task. If depression negatively impacts cognitive processes underlying successful decision making, this could manifest itself in tests of executive functioning (Grange & Rydon-Grange, 2021; Lawlor et al., 2020).

Rapid decision-making is an attractive area to study in depression and other clinical disorders as rich theoretical developments have been made in the understanding of the underlying cognitive processes (Ratcliff & McKoon, 2008). These theories have been developed into computational models allowing the researcher to quantify key latent parameters reflecting the underlying cognitive processes (e.g., Voss, Nagler, & Lerche, 2013; Voss, Voss, & Lerche, 2015). Of interest to the current study, researchers can explore the association between these latent parameters and severity of clinical presentation (White, Ratcliff, Vasey, & McKoon, 2010).

The current study explores the association between depression symptomatology and one essential aspect of decision-making: The ability to trade speed for accuracy. Such *speed-accuracy tradeoffs* (SATs)—the finding that emphasising speed produces poorer accuracy and emphasising accuracy slows responding—are ubiquitous in the laboratory and everyday life (Heitz, 2014) and suggest participants can adopt different response strategies depending on current environmental demands. The question addressed here is to establish the association between depression symptomatology and the ability to shift response strategies.

The paper is structured as follows. First, an overview of a popular computational model of rapid decision-making (the drift diffusion model) is provided. I then discuss how strategic adoption of different response strategies is accounted for by the model. I then present the aims of the current study, and present two experiments probing the association between depression symptomatology and shifts in response strategies.

## The drift diffusion model (DDM)

The DDM (see Figure 1) assumes that a decision in an RT task is the result of noisy accumulation of evidence—modelled as a diffusion process—towards one of two response

---

All raw data, analysis scripts, and code for the modelling can be downloaded from <https://osf.io/48xkh/>  
Correspondence concerning this article should be addressed to James A. Grange, School of Psychology, Dorothy Hodgkin Building, Keele University, Keele, ST5 5BG, United Kingdom. E-mail: [grange.jim@gmail.com](mailto:grange.jim@gmail.com)

boundaries (one reflecting the correct response and the other reflecting an error response<sup>1</sup>). Evidence accumulation continues until one of the response boundaries is met. RT of the model is determined by how long the diffusion process took to reach one of the response boundaries, and accuracy of the model is determined by whether the diffusion process met the correct response boundary.

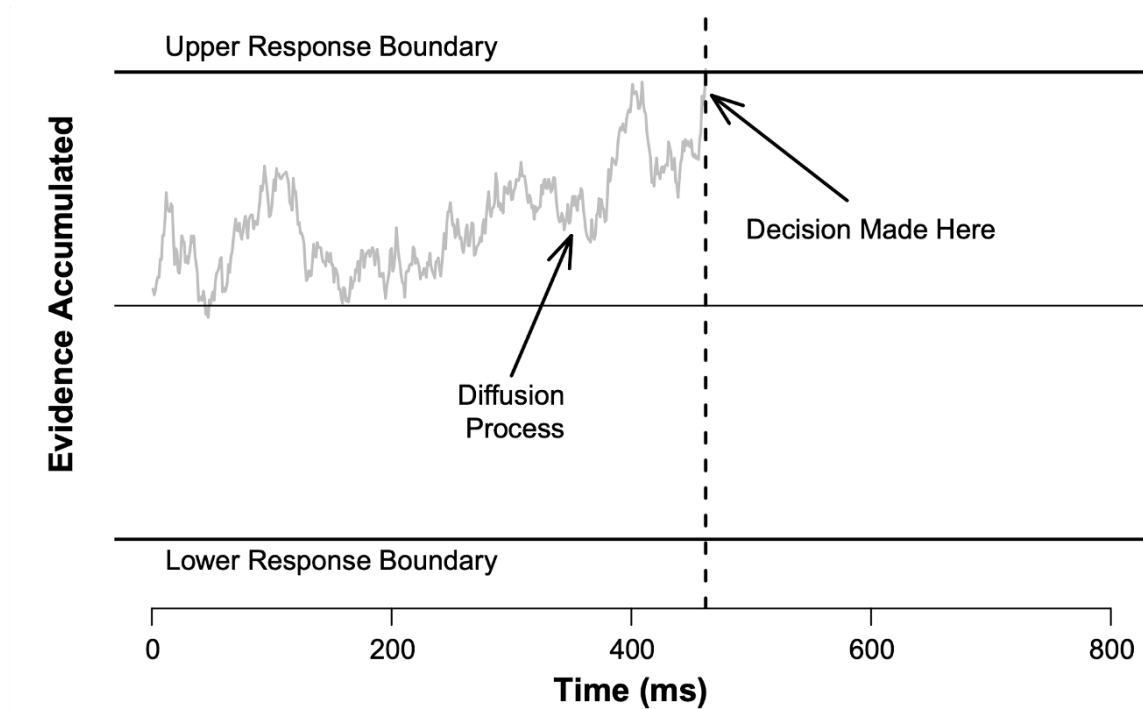


Figure 1. Schematic overview of evidence accumulation in the drift diffusion model. See text for details. Figure available at <https://tiny.cc/diffusionmodel> under CC license <https://creativecommons.org/licenses/by/2.0/>. Note: ms = milliseconds.

The DDM has three main parameters. The *drift rate* represents the average rate of evidence accumulation across trials, with higher values reflecting more rapid accumulation leading to faster responding. The *boundary separation* reflects the height of the response boundaries and determines how much evidence is required before a decision is made. Higher boundary separation requires more evidence to be accumulated, prolonging RT relative to lower boundary separation. However, lowering boundary separation increases the probability of an error due to noise in the diffusion process. The *non-decision* parameter reflects elements of the response unrelated to the decision-making process itself, such as perceptual encoding of the stimulus and motoric responding.<sup>2</sup>

<sup>1</sup>Note that the response boundaries can be treated less generally by representing the actual responses required in the behavioural task regardless of accuracy (e.g., one boundary representing “old” and the other representing “new” in a recognition memory task). In all modelling reported in the current paper, response boundaries are generalised to represent the correct and error response as depicted in Figure 1.

<sup>2</sup>Note that there is an additional parameter capturing the *starting point* of the evidence accumulation, which allows modelling of response bias when one response type is more probable. In the current study, both response options were equiprobable, so the starting point of the accumulation was fixed to start equidistant

The DDM has been successfully applied to a range of clinical questions (see White et al., 2010 for brief review), including depression. Lawlor et al. (2020) applied the DDM to assess the impact of Major Depressive Disorder (MDD) on decision making in two studies. In both, the researchers found that MDD was associated with reduced drift rates and increased boundary separation compared to controls. This extended previous work which found impaired evidence accumulation in MDD using a model of executive control (Dillon et al., 2015; but see Grange & Rydon-Grange, 2021). These results suggest that slowed responding in MDD can be explained by poorer evidence accumulation in the decision-making process together with a more cautious mode of responding.

### **Speed-accuracy tradeoffs (SATs)**

The higher boundary separation found in MDD could reflect the strategic adoption of a more cautious mode of responding, but it could also reflect an inability of MDD participants to engage in a faster (but more error-prone) mode of responding (for a similar argument applied to healthy ageing, see Forstmann et al., 2011). This question cannot conclusively be addressed without inducing SAT behaviour in participants. SATs can be induced in a variety of ways (Heitz, 2014), including via explicit instructions to participants (e.g., in one block of trials emphasise speed, and in another emphasise accuracy). The DDM accounts for SAT by assuming an increase of the boundary separation occurs when accuracy is emphasised (cf., speed) thus demanding more evidence to be accumulated before committing to a response. Studies have shown that emphasising speed over accuracy indeed leads to lower estimates of the boundary separation parameter (see e.g., Forstmann et al., 2011; Ratcliff, Thapar, & McKoon, 2007), but some studies have found that drift rates are also reduced (e.g., Rae, Heathcote, Donkin, Averell, & Brown, 2014).

### **SATs and depression**

Strategic adaptation of response boundary settings could be disrupted because of the involvement of the striatum component of the basal ganglia in SAT tasks, a sub-cortical area also implicated in depression. Neuroimaging has shown that speed-emphasis leads to activation of the striatum as well as the pre-supplementary motor area, and that individual variation in activation of these structures is associated with individual variation in estimates of boundary separation parameters of a version of the DDM (Forstmann et al., 2008, 2011). Research has shown reduced activity in striatal regions in MDD (Hamilton et al., 2018) which can impact goal-directed behaviour due to the interface between the striatum and pre-frontal regions of cortex (Marquand, Haak, & Beckmann, 2017). Indeed, depression has been shown to be associated with a deficit in functional connectivity between striatal and pre-frontal regions (Furman, Hamilton, & Gotlib, 2011; Hamilton et al., 2018; Pan et al., 2017).

Given the link between the involvement of the striatum in SAT behaviour and its disruption due to depression, examination of SAT performance—and formal DDM modelling—could provide important insights into the cognitive profile of depression. Only one study has examined the impact of depression on SAT performance together with DDM

---

to each response boundary.

application. Vallesi, Canalaz, Balestrieri, and Brambilla (2015) compared SAT performance of 20 participants with a current or previous MDD diagnosis with 28 controls. Participants performed a perceptual discrimination task and were cued on a trial-by-trial basis whether to emphasise speed or accuracy in their response. Behavioural results showed the typical SAT pattern of enhanced accuracy but longer RT on accuracy-emphasis trials, but it was not reported whether this behavioural pattern interacted with depression grouping. The DDM results showed an overall reduction of drift rate in the depression group (see also Lawlor et al., 2020). The modelling also showed that accuracy-emphasis increased estimates of boundary separation, although again it was not reported whether this was moderated by depression grouping. However, the results did show that those with MDD maintained a lower boundary separation after speed-emphasis trials compared to when the previous trial emphasised accuracy, suggesting a deficit in strategically updating response settings after emphasising accuracy.

### **The current study**

The current study was designed to further examine the association between depression and the ability to adapt response strategies to SAT instructions. Vallesi et al. (2015) cued participants on a trial-by-trial basis whether to emphasise speed or accuracy; thus, their study required frequent switching between response strategies, which could be disrupted by the well-known deficit in task switching found in depression (Ravizza & Salo, 2014). Thus it is not clear whether the finding reported by Vallesi et al. (2015) of a deficit in MDD participants in updating boundary separation after accuracy emphasis trials was caused by a deficit in the ability to establish appropriate response strategies during SAT (as suggested by Vallesi et al., 2015), or whether it was caused by general deficits in cognitive flexibility in MDD. In addition, due to the lack of a reported interaction in between behavioural SAT performance and depression status—and the lack of a reported interaction of boundary separation parameters overall with depression status—in Vallesi et al. (2015), the evidence for an association between depression and SAT is rather limited. The current study will thus provide more data on this question.

Two experiments were conducted in the current study utilising an SAT paradigm where speed- and accuracy-emphasis occurred in separate blocks. DDM was applied to the data to explore the magnitude of changes in model parameters across speed- and accuracy-emphasis blocks and whether these changes were associated with the magnitude of depression symptomatology. This design allowed me to disentangle the contribution of task switching and the overall ability to set appropriate response strategies (cf., Vallesi et al., 2015) because the relevant speed-accuracy emphasis was constant within blocks.

Experiment 1 was correlational to assess whether changes in outcome variables across speed-accuracy emphasis blocks was associated with the magnitude of depression symptomatology. Experiment 2 was a between-groups design where a large number of participants were first pre-screened on a measure of depression symptomatology, and individuals with very-high and very-low scores were invited back to complete the SAT task.

## Experiment 1

### Method

The experiment was programmed and delivered online using Gorilla (Anwyl-Irvine, Massonnié, Flitton, Kirkham, & Evershed, 2020); participants were recruited via Prolific academic <https://www.prolific.co/>.

**Participants.** The study was approved by the School of Psychology’s ethical review panel at Keele University, and the study was conducted in accordance with the Declaration of Helsinki as revised in 1989. Informed consent was gathered from participants residing in the UK or USA were recruited; the age-range of the sample was restricted to 18–35 because RTs are known to slow due to healthy ageing (Salthouse, 2000). If a wider age range was included it would be hard to find a single RT threshold for the behavioural task that would be suitably challenging for all ages.

An a priori sensitivity analysis (Appendix A) suggested 300 participants would provide over 86% power to detect the effect size of interest ( $r = -0.18$ ); Grange & Rydon-Grange, 2021). 375 participants completed the study, but some were removed due to pre-specified exclusion criteria. 2 were removed due to failing the attention check embedded within the SHPS questionnaire. 17 were removed for failure to retain a session-wise accuracy rate above 75% on the behavioural task. 6 were removed due to having more than 25% of their responses slower than the RT threshold of 500ms. A further 1 participant was removed due to an error with the experimental programme. Behavioural analysis was conducted on 349 participants. 5 participants were removed from the DDM analysis as the model did not fit their data well.

### Materials.

**Quick inventory of depressive symptomatology-self-report (QIDS-SR).** The 16-item self-report version of the Quick Inventory of Depressive Symptomatology, QIDS-SR-16 (Rush et al., 2003), was used. The QIDS has been shown to have excellent internal consistency and validity (Rush et al., 2003). It consists of 16 items assessing the severity of depressive symptoms experienced by participants in the preceding seven days. Items relate to particular aspects of depression symptoms (e.g., “views of self”) and requires participants to select one response from four options that best represents their experience (e.g., “I am more self-blaming than usual”).

**Snaith-Hamilton pleasure scale (SHPS).** The SHPS is a 14-item self-report questionnaire measuring the participants’ ability to experience pleasure “. . . over the last few days” (Snaith et al., 1995). Participants were required to read a statement (e.g., “I would enjoy seeing other people’s smiling faces”) and indicate their agreement with this statement on a four-point scale (Definitely Agree; Agree; Disagree; Strongly Disagree). The scoring method recommended by Franken, Rassin, and Muris (2007) was used.

The items for both questionnaires were presented to participants on the computer screen, and participants selected their response to each item by clicking a radio button with their mouse corresponding to their choice.

**Attention check.** An attention check was embedded within the SHPS to identify participants who were not paying attention to the items. It read “It is important that you pay attention to this study. Please select *Disagree*”. Participants who did not select this

response were deemed to have failed the attention check and were removed before analysis.

**Behavioural Task.** The behavioural task utilised similar stimuli to previous research (e.g., Lawlor et al., 2020; Pizzagalli, Jahn, & O’Shea, 2005; Tripp & Alsop, 1999), but a unique task. The task presents participants with a schematic face within the centre of a square frame in the centre of the screen. The participants’ task was to judge whether the mouth was “long” or “short”. During the instruction phase, participants were informed that the “short” mouth lies entirely within the boundaries of the eyes, whereas the “long” mouth extends to the edge of the eyes. Each trial began with a black fixation cross presented centrally for 500ms. After this, a schematic face minus the mouth appeared for 500 ms, at which point the mouth was presented for 100 ms. After this, the mouth disappeared, and the schematic face remained on screen until a participant made a response. Participants were required to press the “Z” key if the mouth was short, and the “M” key if the mouth was long.

The behavioural task was presented in stages to help the participants learn the task, and then to acclimatise to the speed—accuracy instructions. Participants were first given instructions on how to perform the mouth-length task which asked participants to respond “. . . as quickly and as accurately as possible”. After these instructions, participants completed a 16-trial practice block. In this practice block, participants were given feedback on the accuracy of their responses to help them learn the task. If the participant made a correct response, a green tick was presented centrally for 200ms; if the participant made an error, a red cross was presented centrally for 200ms.

After this initial practice, participants were given instructions about the speed-accuracy emphasis instructions. They were told that they would experience “accuracy-emphasis” and “speed-emphasis” blocks, and that in accuracy-emphasis blocks, they should respond “as accurately as possible without taking too long to respond”. They were told that in speed-emphasis blocks they should respond “as quickly as possible without making too many mistakes. DO NOT GUESS”. After these instructions, participants were presented with a practice session consisting of 12 trials under accuracy-emphasis instructions followed by 12 trials under speed-emphasis instructions.

Response feedback on these trials changed depending on which instructions they were performing under. In accuracy-emphasis trials, if participants made an error they saw the message “INCORRECT!” presented centrally in red font for 1500ms. On speed trials, if participants responded slower than 500ms they saw the message “TOO SLOW!” for 1500ms. 500ms was selected as an appropriate threshold to initiate speed feedback based on performance in a pilot study (see Pilot Study 1 in Appendix B). No feedback was provided for correct responses in accuracy-emphasis blocks, and no feedback was provided for sub-threshold responses in speed-emphasis blocks. (See Pilot Study 2 in Appendix B for evidence these instructions produce the desired behavioural effects.)

After this second practice, participants were then presented with 8 blocks of 48 trials with self-paced rest screens between each block. The relevant emphasis (speed vs. accuracy) alternated every block, beginning with accuracy emphasis.

**Procedure.** Upon entering the study online via Prolific, participants provided informed consent, and then completed a short demographics questionnaire. After this, the

experimental software randomly allocated the participant to one of four presentation orders of study elements: (1) behavioural task-QIDS-SHPS; (2) behavioural task-SHPS-QIDS; (3) QIDS-SHPS-behavioural task; or (4) SHPS-QIDS-behavioural task. Upon completion of all stages, participants were presented with a debrief screen providing detailed information of the study.

**Quality checks and data exclusion.** Participants who failed the attention checks were removed from the analysis. Participants were also removed whose mean accuracy performance across the experiment was below 75%. For response time analysis, error trials were removed, and correct response times were trimmed to only include RTs longer than 150ms and shorter than 5,000ms. For the diffusion modelling, participants whose data were not fit well by the model were removed.

## Results

All of the analysis utilised R (Version 4.0.2; R Core Team, 2020)<sup>3</sup>,

**Questionnaire scores.** There was good variability of scores in the QIDS questionnaire (range=0–24, M=9.08, SD=5.29) and the SHPS questionnaire (range=14–48, M=27.71, SD=6.21). In the QIDS, 77 participants had scores greater than or equal to 14, and 154 had scores lower than 8, which were the thresholds used by Dillon et al. (2015) for inclusion in their major depression and control groups, respectively. Bayesian linear regression confirmed that QIDS scores were predicted by SHPS scores,  $b = 0.47$ , 95% credible interval (CI; 0.39, 0.54).

**Behavioural data.** For the response time analysis, error trials were removed. Correct RTs were trimmed to only include RTs longer than 150ms and shorter than 5,000ms. The analysis of the behavioural data showed the typical effects for speed-accuracy tradeoff instructions: Proportion accuracy was significantly higher in accuracy-emphasis blocks (M=0.944, SE=0.002) than in the speed-emphasis blocks (M=0.860, SE=0.005),  $t(348)=20.26$ ,  $p<.001$ , Cohen’s  $d=1.22$ ; mean RT was shorter in speed-emphasis blocks (M=379ms, SE=2ms) than in accuracy-emphasis blocks (M=421ms, SE=3ms),  $t(348)=18.46$ ,  $p<.001$ ,  $d=0.92$ .

A series of Bayesian linear regressions were performed predicting behavioural measures from QIDS and SHPS scores. Posterior estimates from the statistical models are shown in Table 1. These analyses showed that none of the outcome measures were predicted from questionnaire scores.

In the next stage of analysis difference scores were calculated for both RT and accuracy to quantify shifts in performance between accuracy-emphasis and speed-emphasis blocks. Specifically,  $\Delta_{accuracy}$  was calculated as the mean accuracy in accuracy-emphasis blocks minus the mean accuracy in speed-emphasis blocks. Correspondingly,  $\Delta_{RT}$  was calculated

<sup>3</sup>We, furthermore, used the R-packages *brms* (Version 2.16.1; Bürkner, 2017, 2018), *dplyr* (Version 1.0.7; Wickham et al., 2021), *forcats* (Version 0.5.0; Wickham, 2020), *ggplot2* (Version 3.3.5; Wickham, 2016), *here* (Version 1.0.1; Müller, 2020), *MASS* (Version 7.3.51.6; Venables & Ripley, 2002), *papaja* (Version 0.1.0.9997; Aust & Barth, 2020), *purrr* (Version 0.3.4; Henry & Wickham, 2020), *Rcpp* (Version 1.0.7; Eddelbuettel & François, 2011; Eddelbuettel & Balamuta, 2018), *readr* (Version 1.3.1; Wickham, Hester, & François, 2018), *rtdist* (Version 0.11.2; Singmann, Brown, Gretton, & Heathcote, 2020), *stringr* (Version 1.4.0; Wickham, 2019), *tibble* (Version 3.1.5; Müller & Wickham, 2021), *tidyr* (Version 1.1.4; Wickham, 2021), *tidyverse* (Version 1.3.0; Wickham, Averick, et al., 2019), and *tinylab* (Version 0.1.0; Barth, 2020).



as the mean RT in accuracy-emphasis blocks minus the mean response time in speed-emphasis blocks. A series of Bayesian linear regressions were performed to assess whether the magnitude of shifts in performance with speed-accuracy emphasis instructions was related to depression symptomatology (see Table 1). The results showed no evidence for any associations.

Table 1  
*Posterior estimates from the Bayesian linear regressions predicting various outcome measures from from QIDS and SHPS scores. Note that for the beta estimates, 95% Credible Intervals are provided in square brackets.*

Source	Outcome	QIDS		SHPS	
		Intercept	beta	Intercept	beta
Behavioural	Accuracy	0.904	0.000 [-0.001, 0.001]	0.899	0.000 [-0.001, 0.001]
	RT	398	0.413 [-0.454, 1.265]	397	0.206 [-0.529, 0.950]
	delta-Accuracy delta-RT	0.081 36	0.000 [-0.001, 0.002] 0.720 [-0.131, 1.549]	0.092 34	0.000 [-0.002, 0.001] 0.288 [-0.443, 1.016]
DDM	Boundary Separation	0.774	0.001 [-0.003, 0.005]	0.793	0.000 [-0.004, 0.003]
	Drift Rate	4.273	-0.016 [-0.039, 0.008]	4.542	-0.015 [-0.035, 0.005]
	Non-Decision Time	0.302	0.000 [0.000, 0.001]	0.298	0.000 [0.000, 0.001]
	delta-Boundary Separation delta-Drift Rate delta-Non-Decision Time	0.144 0.901 0.039	0.005 [-0.002, 0.012] 0.011 [-0.018, 0.039] 0.000 [-0.001, 0.001]	0.208 0.889 0.037	-0.001 [-0.006, 0.005] 0.004 [-0.020, 0.028] 0.000 [-0.001, 0.001]

*Note.* DDM = Drift diffusion model. RT = Response time.

**Diffusion modelling.** The diffusion model was fit to individual participant data using the `fast-dm-30` using Kolmogorov-Smirnov parameter optimisation. Boundary separation, drift rate, and non-decision time were allowed to freely vary between speed-accuracy emphasis conditions. Starting point was fixed at 0.5 (equidistant between both response options); all other parameters were either fixed to zero or were not free to vary across conditions (see Voss et al., 2015). 5 participants who showed a significant deviation between observed and model-predicted distributions were removed from inferential analysis. For the remaining participants, the diffusion model provided a good fit to the observed data (see Appendix C).

Results showed that estimates of boundary separation were higher in accuracy-emphasis blocks ( $M = 0.876$ ,  $SE = 0.017$ ) than in the speed-emphasis blocks ( $M=0.689$ ,  $SE=0.009$ ),  $t(343)=10.54$ ,  $p<.001$ ,  $d=0.75$ . Estimates of drift rate were also higher in accuracy-emphasis blocks ( $M=4.63$ ,  $SE=0.07$ ) than in the speed-emphasis blocks ( $M=3.63$ ,  $SE=0.08$ ),  $t(343)=13.17$ ,  $p<.001$ ,  $d=0.74$ . Estimates of non-decision time were higher in accuracy-emphasis blocks ( $M=0.322$ ,  $SE=0.003$ ) than in the speed-emphasis blocks ( $M=0.287$ ,  $SE=0.002$ ),  $t(343)=12.05$ ,  $p<.001$ ,  $d=0.82$ .

Posterior parameter estimates from Bayesian linear regressions are enumerated in Table 1 and visualised in (Figure 2). Again, there was no evidence for an association with questionnaire scores. Similarly to the behavioural data, the next stage of the analysis calculated difference scores for each model parameter between accuracy-emphasis and speed-emphasis blocks. A series of Bayesian linear regressions were performed to assess whether the magnitude of shifts in performance with speed-accuracy emphasis instructions was related to depression symptomatology. The results showed again no evidence for an association.

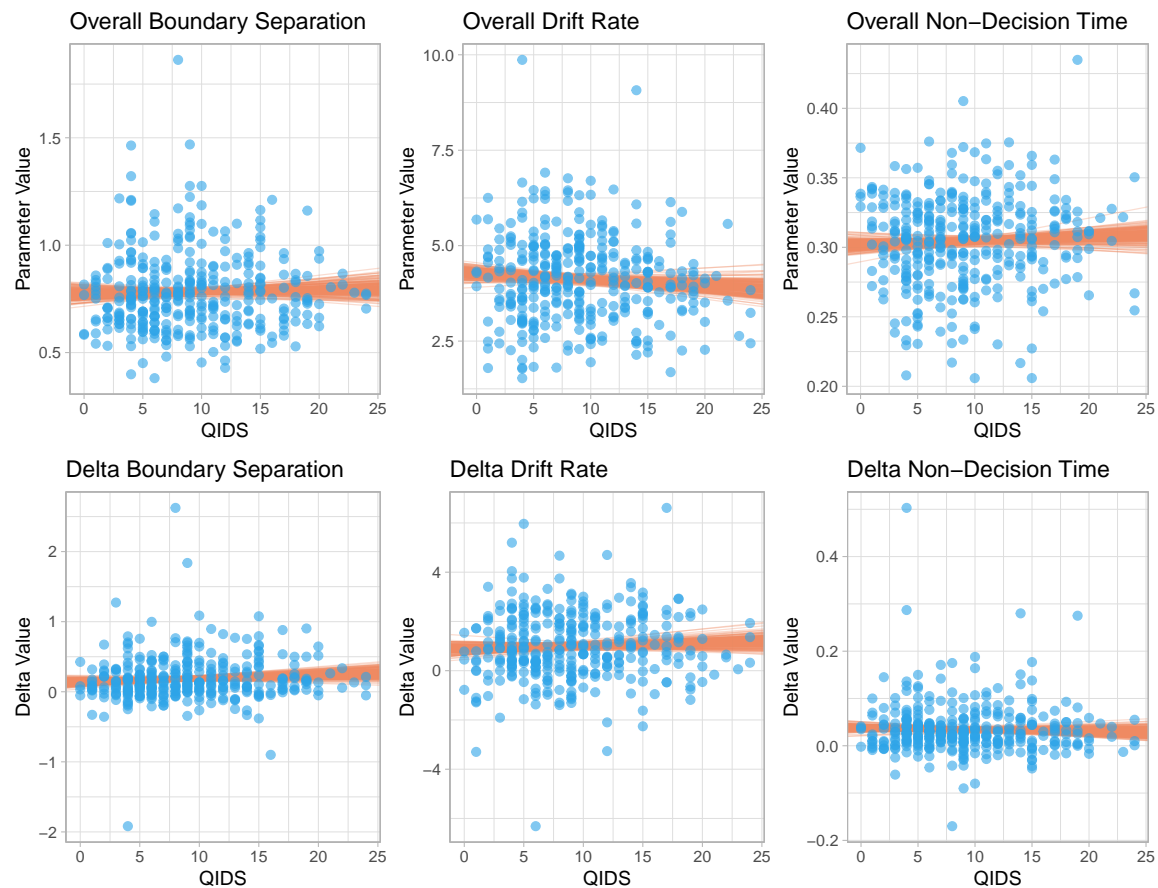
## Experiment 2

Experiment 2 utilised two phases. In Phase 1 a large sample of participants completed the QIDS, In Phase 2, participants with extreme scores were contacted to take part in the same behavioural task as Experiment 1. This approach was taken for Experiment 2 to maximise the opportunity to include participants with extreme QIDS-scores—both at the upper- and lower-end of the scale—within the final sample.

## Method

**Participants.** 501 participants aged between 18–35 residing in the UK or USA completed the QIDS. Of these, 12 failed an attention check. Participants were rank-ordered according to their QIDS score, and the 80 lowest and the 80 highest participants were invited to Phase 2. The invitations were left open for 5 days, after which 49 participants from the low-depression group and 42 participants from the high-depression group had completed Phase 2.

**Materials & Behavioural Task.** In Phase 1 all participants completed the QIDS, and in Phase 2 participants completed both the QIDS and the SHPS, as well as the behavioural task which was identical to Experiment 1.

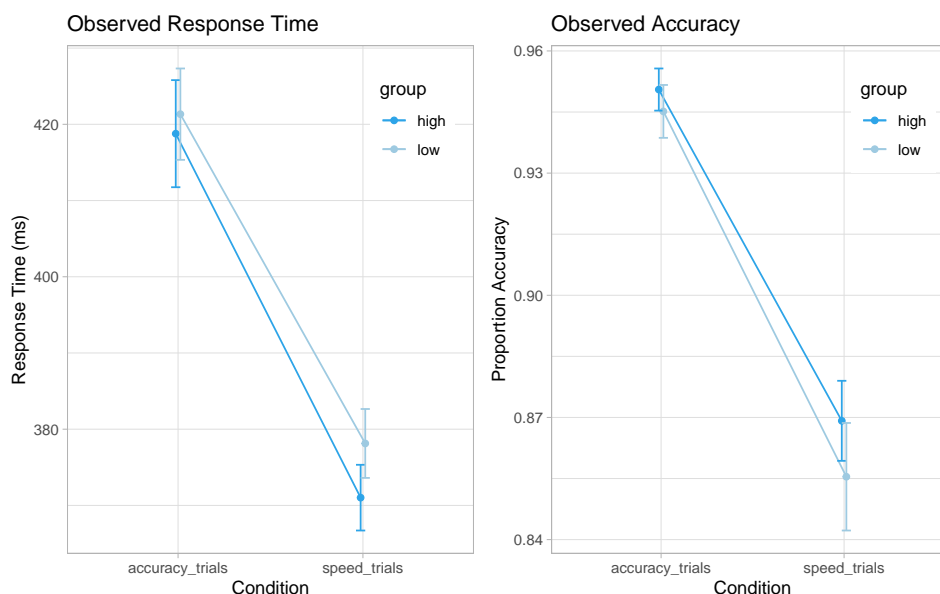


*Figure 2.* Results of the Bayesian linear regression for overall (i.e., averaged across accuracy- and speed-emphasis blocks) diffusion model parameters (Upper Row) and delta parameter values (Lower Row) predicted from QIDS scores. Points represent individual participant data, and the overlaid lines represent draws from the posterior distribution of the statistical model.

## Results

**Questionnaire scores.** The mean QIDS score for the low-depression group was 2.00 (SE=0.19, range=0–5) and for the high-depression group was 17.19 (SE=0.44, range=10–22). The mean SHPS score for the low-depression group was 22.35 (SE=0.78, range=14–34) and for the high-depression group was 33.69 (SE=0.96, range=23–52). Bayesian linear regression confirmed that QIDS scores were predicted by SHPS scores in Phase 2,  $b = 0.70$ , 95%-CI (0.56, 0.84). The product-moment correlation between questionnaire scores in Phase 1 and Phase 2 for the QIDS was  $r=0.985$ .

**Behavioural data.** The behavioural data is summarised in Figure 3. Two separate mixed factorial analyses of variance (ANOVA) were performed on the RT and accuracy data. The ANOVA summary table for all analyses are shown in Table 2. Accuracy-emphasis blocks led to a significant increase in both accuracy and RT. There was no main effect of depression group, and—critical to the current study’s aims—no interaction.



*Figure 3.* Observed mean response time (in milliseconds, ms) and proportion accuracy for both low and high depression groups as a function of speed-accuracy emphasis condition. Error bars denote one standard error around the mean.

**Diffusion modelling.** Model fitting was conducted as in Experiment 1. 3 participants (two from the low-depression group and one from the high-depression group) were removed from analyses as their data were not fit well by the model (see Appendix C). The results of the diffusion model analysis are shown in Figure 4 and the ANOVA summary table for all analyses are shown in Table 2. Accuracy-emphasis blocks led to a significant increase in boundary separation, in drift rate, and in non-decision time. There was no main effect of depression group for any of the DDM parameters, and—critically to the current study’s aims—no interaction.

## Discussion

The current study examined the association between depression symptomatology and the ability to adopt different response strategies in response to speed-accuracy tradeoff instructions. Given the role of the striatum component of the basal ganglia in adapting response strategies (Forstmann et al., 2008, 2011), and given the reported reduced activity of this sub-cortical region in depression (Furman et al., 2011; Hamilton et al., 2018; Pan et al., 2017), the SAT paradigm becomes an important paradigm with which to probe the cognitive profile of depression. In addition, the drift diffusion model (Voss et al., 2015) was used to quantify changes in latent cognitive parameters of decision-making, and to assess whether the magnitude of changes in these parameters with SAT instructions were associated with depression symptomatology.

The results showed no evidence for an association between depression symptomatology and SAT performance. This was true for both the behavioural data (i.e., changes in RT and accuracy) and the DDM parameters: Typical behavioural SAT effects were found across

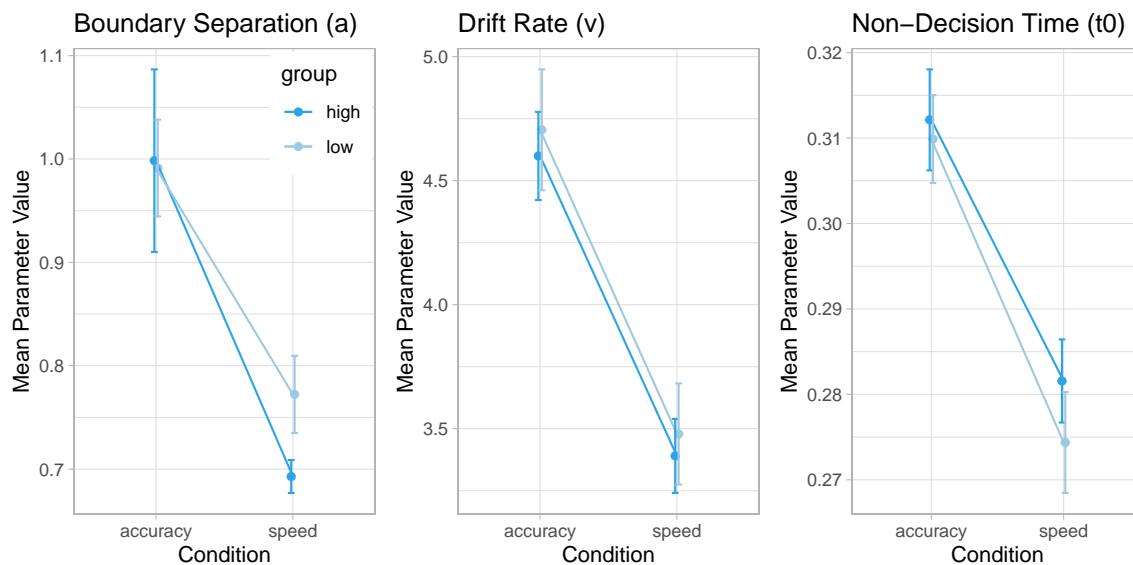


Figure 4. Mean diffusion model parameter estimates for both low and high depression groups as a function of speed-accuracy emphasis condition. Error bars denote one standard error around the mean.

Table 2

ANOVA summary table for Experiment 2 for the behavioural data and parameters of the drift diffusion model (DDM).

Source	Outcome	Effect	F-value	p	Effect Size
Behavioural	Accuracy	Depression Grouping (D)	0.66	.418	.006
		SAT Condition (S)	176.00	<.001	.314
		D * S Interaction	0.42	.518	.001
	RT	D	0.49	.478	.004
		S	147.20	<.001	.273
		D * S	0.37	.545	<.001
DDM	Boundary Separation	D	0.43	.514	.003
		S	27.27	<.001	.126
		D * S	0.74	.392	.004
	Drift Rate	D	0.17	.685	.001
		S	60.00	<.001	.175
		D * S	0.00	.958	<.001
	Non-Decision Time	D	0.58	.448	.004
		S	48.86	<.001	.173
		D * S	0.27	.602	.001

Note. DDM = Drift diffusion model. RT = Response time. Effect size = Generalised eta squared. The degrees of freedom for the Behavioural data are (1, 89), and are (1, 86) for the DDM parameters.

both experiments suggesting the experimental instructions were successful in inducing large effects on response strategies; the DDM fitting (which produced excellent fits to the data) showed that emphasising accuracy (cf., speed) led to an increase in the boundary separation parameter, an increase in drift rate, and an increase in non-decision time (cf., Rae et al., 2014), but the magnitude of changes in these parameters was not associated with depression symptom severity.

These results appear to stand in contrast to previous research. For example, Vallesi et al. (2015)—in a study comparing MDD patients with controls on an SAT paradigm—found some evidence for deficits in the MDD group on updating the boundary separation parameter immediately after speed-emphasis trials. This suggests that this group struggled to adapt their response strategy to a more cautious mode of responding after emphasising speed. However, because this study utilised trial-by-trial cuing of—and hence required rapid switching between—different speed–accuracy strategies, there is the potential that the well-known deficit in task switching found in depression (e.g., Ravizza & Salo, 2014; Snyder, 2013) contributed to the deficit observed by Vallesi et al. (2015). That is, the finding of Vallesi et al. (2015) could be due to problems with rapid shifting of response strategies, a problem with general task switching, or both. The current study removed the requirement for frequent switching between response strategies by separating speed- and accuracy-emphasis instructions into separate blocks. A fruitful avenue for future research would be to compare performance on within-block switching of response strategies (i.e., like Vallesi et al., 2015) and between-block switching of response strategies (like the study) in the same sample to explore the contribution of task switching.

Similar to the current study, Vallesi et al. (2015) reported no interaction between changes in behavioural performance across SAT conditions—nor an interaction between overall changes in the boundary separation parameter across SAT conditions—with depression status. Understanding which aspects of cognition are impacted by depression—and which are *not* impacted—is important to inform the development of interventions aiming to improve cognitive and executive functioning. The current study thus adds to an emerging overall picture of little—if any—impact of depression on SAT performance.

### Limitations

The study utilised online testing rather than a lab-based approach, which may have increased the variability in the behavioural data (due to timing differences in participants' devices, room lighting etc.). However, there is good evidence that online testing provides high-quality data in many cognitive paradigms (see for example Anwyl-Irvine et al., 2020; Crump, McDonnell, & Gureckis, 2013)}. In addition, the behavioural data in the current study showed clear and large effects of SAT manipulations, suggesting the online testing has not impacted the data quality.

The age-range of our samples was restricted to 18–35 so that a single response time deadline could be used in speed-emphasis trials (see Appendix B) without unduly penalising older adults (due to the well-documented general slowing found in older adults, Salthouse, 2000); it remains possible therefore that the lack of an association between SAT performance and depression symptomatology only holds for this age range. Although the use of a single

global response deadline is found in many SAT procedures (see Rae et al., 2014), future research could address the impact of this by utilising individually tailored response deadlines for each participant via a tracking procedure (see Heitz, 2014).

Another limitation is that the current study did not use a clinical sample; however, both experiments showed good variability in depression symptom severity as measured by the QIDS. In Experiment 1, 22.06% of the sample had a QIDS score of 14 or more, which was the cutoff used by Dillon et al. (2015) as an inclusion criterion to their depression condition. In Experiment 2, this concern was directly addressed by pre-screening a large number of participants on depression severity and only inviting back the extreme-low and extreme-high scorers on the QIDS to take part in the SAT paradigm. An advantage of recruiting non-clinical samples is that the sample size of both of the studies reported here were large. Indeed, Experiment 1 had over 86% power to detect even a small association between depression and SAT performance (see Appendix A). Thus although the data provide compelling evidence for a lack of an association, it would be prudent to replicate this study with a clinical sample



**Acknowledgements**

None.

**Declaration of Competing Interest**

None.

**Financial Support**

This research received no specific grant from any funding agency, commercial, or not-for-profit sectors.

### References

- Anwyl-Irvine, A. L., Massonnié, J., Flitton, A., Kirkham, N., & Evershed, J. K. (2020). Gorilla in our midst: An online behavioral experiment builder. *Behavior Research Methods*, *52*(1), 388–407. <https://doi.org/10.3758/s13428-019-01237-x>
- Aust, F., & Barth, M. (2020). *papaja: Prepare reproducible APA journal articles with R Markdown*. Retrieved from <https://github.com/crsh/papaja>
- Barth, M. (2020). *Tinylabels: Lightweight variable labels*. Retrieved from <https://CRAN.R-project.org/package=tinylabels>
- Burt, D. B., Zembar, M. J., & Niederehe, G. (1995). Depression and memory impairment: A meta-analysis of the association, its pattern, and specificity. *Psychological Bulletin*, *117*(2), 285–305. <https://doi.org/10.1037/0033-2909.117.2.285>
- Bürkner, P.-C. (2017). brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software*, *80*(1), 1–28. <https://doi.org/10.18637/jss.v080.i01>
- Bürkner, P.-C. (2018). Advanced Bayesian multilevel modeling with the R package brms. *The R Journal*, *10*(1), 395–411. <https://doi.org/10.32614/RJ-2018-017>
- Crump, M. J. C., McDonnell, J. V., & Gureckis, T. M. (2013). Evaluating Amazon’s Mechanical Turk as a tool for experimental behavioral research. *PLOS ONE*, *8*(3), e57410. <https://doi.org/10.1371/journal.pone.0057410>
- Dillon, D. G., Wiecki, T., Pechtel, P., Webb, C., Goer, F., Murray, L., . . . Pizzagalli, D. A. (2015). A computational analysis of flanker interference in depression. *Psychological Medicine*, *45*(11), 2333–2344. <https://doi.org/10.1017/S0033291715000276>
- Eddelbuettel, D., & Balamuta, J. J. (2018). Extending extitR with extitC++: A Brief Introduction to extitRcpp. *The American Statistician*, *72*(1), 28–36. <https://doi.org/10.1080/00031305.2017.1375990>
- Eddelbuettel, D., & François, R. (2011). Rcpp: Seamless R and C++ integration. *Journal of Statistical Software*, *40*(8), 1–18. <https://doi.org/10.18637/jss.v040.i08>
- Forstmann, B. U., Dutilh, G., Brown, S., Neumann, J., Cramon, D. Y. von, Ridderinkhof, K. R., & Wagenmakers, E.-J. (2008). Striatum and pre-SMA facilitate decision-making under time pressure. *Proceedings of the National Academy of Sciences*, *105*(45), 17538–17542. <https://doi.org/10.1073/pnas.0805903105>
- Forstmann, B. U., Tittgemeyer, M., Wagenmakers, E.-J., Derrfuss, J., Imperati, D., & Brown, S. (2011). The speed-accuracy tradeoff in the elderly brain: A structural model-based approach. *Journal of Neuroscience*, *31*(47), 17242–17249. <https://doi.org/10.1523/JNEUROSCI.0309-11.2011>
- Franken, I. H. A., Rassin, E., & Muris, P. (2007). The assessment of anhedonia in clinical and non-clinical populations: Further validation of the SnaithHamilton Pleasure Scale (SHAPS). *Journal of Affective Disorders*, *99*(1), 83–89. <https://doi.org/10.1016/j.jad.2006.08.020>

- Furman, D. J., Hamilton, J. P., & Gotlib, I. H. (2011). Frontostriatal functional connectivity in major depressive disorder. *Biology of Mood & Anxiety Disorders*, *1*(1), 11. <https://doi.org/10.1186/2045-5380-1-11>
- Grange, J. A., & Houghton, G. (Eds.). (2014). *Task Switching and Cognitive Control*. New York, NY: Oxford University Press.
- Grange, J. A., & Rydon-Grange, M. (2021). Computational modelling of attentional selectivity in depression reveals perceptual deficits. *Psychological Medicine*, 1–10. <https://doi.org/10.1017/S0033291720002652>
- Hamilton, J. P., Sacchet, M. D., Hjørnevik, T., Chin, F. T., Shen, B., Kämpe, R., . . . Gotlib, I. H. (2018). Striatal dopamine deficits predict reductions in striatal functional connectivity in major depression: A concurrent 11C-raclopride positron emission tomography and functional magnetic resonance imaging investigation. *Translational Psychiatry*, *8*(1), 1–10. <https://doi.org/10.1038/s41398-018-0316-2>
- Heitz, R. P. (2014). The speed-accuracy tradeoff: History, physiology, methodology, and behavior. *Frontiers in Neuroscience*, *8*. <https://doi.org/10.3389/fnins.2014.00150>
- Henry, L., & Wickham, H. (2020). *Purrr: Functional programming tools*. Retrieved from <https://CRAN.R-project.org/package=purrr>
- Kiesel, A., Steinhauser, M., Wendt, M., Falkenstein, M., Jost, K., Philipp, A. M., & Koch, I. (2010). Control and interference in task switching: A review. *Psychological Bulletin*, *136*(5), 849–874. <https://doi.org/10.1037/a0019842>
- Kruschke, J. K., & Liddell, T. M. (2018). The bayesian new statistics: Hypothesis testing, estimation, meta-analysis, and power analysis from a bayesian perspective. *Psychonomic Bulletin & Review*, *25*(1), 178–206. <https://doi.org/10.3758/s13423-016-1221-4>
- Lawlor, V. M., Webb, C. A., Wiecki, T. V., Frank, M. J., Trivedi, M., Pizzagalli, D. A., & Dillon, D. G. (2020). Dissecting the impact of depression on decision-making. *Psychological Medicine*, *50*(10), 1613–1622. <https://doi.org/10.1017/S0033291719001570>
- Marquand, A. F., Haak, K. V., & Beckmann, C. F. (2017). Functional corticostriatal connection topographies predict goal-directed behaviour in humans. *Nature Human Behaviour*, *1*(8), 1–9. <https://doi.org/10.1038/s41562-017-0146>
- McDermott, L. M., & Ebmeier, K. P. (2009). A meta-analysis of depression severity and cognitive function. *Journal of Affective Disorders*, *119*(1), 1–8. <https://doi.org/10.1016/j.jad.2009.04.022>
- Müller, K. (2020). *Here: A simpler way to find your files*. Retrieved from <https://CRAN.R-project.org/package=here>
- Müller, K., & Wickham, H. (2021). *Tibble: Simple data frames*. Retrieved from <https://CRAN.R-project.org/package=tibble>
- Pan, P. M., Sato, J. R., Salum, G. A., Rohde, L. A., Gadelha, A., Zugman, A., . . . Stringaris,

- A. (2017). Ventral striatum functional connectivity as a predictor of adolescent depressive disorder in a longitudinal community-based sample. *American Journal of Psychiatry*, *174*(11), 1112–1119. <https://doi.org/10.1176/appi.ajp.2017.17040430>
- Pizzagalli, D. A., Jahn, A. L., & O’Shea, J. P. (2005). Toward an objective characterization of an anhedonic phenotype: A signal-detection approach. *Biological Psychiatry*, *57*(4), 319–327. <https://doi.org/10.1016/j.biopsych.2004.11.026>
- 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., & 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., Thapar, A., & McKoon, G. (2007). Application of the diffusion model to two-choice tasks for adults 75–90 years old. *Psychology and Aging*, *22*(1), 56–66. <https://doi.org/10.1037/0882-7974.22.1.56>
- Ratcliff, R., Thapar, A., & McKoon, G. (2010). Individual differences, aging, and IQ in two-choice tasks. *Cognitive Psychology*, *60*(3), 127–157. <https://doi.org/10.1016/j.cogpsych.2009.09.001>
- Ravizza, S. M., & Salo, R. E. (2014). Task switching in psychiatric disorders. In J. A. Grange & G. Houghton (Eds.), *Task Switching and Cognitive Control* (pp. 300–331). New York, NY: Oxford University Press.
- R Core Team. (2020). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from <https://www.R-project.org/>
- Rock, P. L., Roiser, J. P., Riedel, W. J., & Blackwell, A. D. (2014). Cognitive impairment in depression: A systematic review and meta-analysis. *Psychological Medicine*, *44*(10), 2029–2040. <https://doi.org/10.1017/S0033291713002535>
- Rush, A. J., Trivedi, M. H., Ibrahim, H. M., Carmody, T. J., Arnow, B., Klein, D. N., . . . Keller, M. B. (2003). The 16-Item quick inventory of depressive symptomatology (QIDS), clinician rating (QIDS-C), and self-report (QIDS-SR): A psychometric evaluation in patients with chronic major depression. *Biological Psychiatry*, *54*(5), 573–583. [https://doi.org/10.1016/S0006-3223\(02\)01866-8](https://doi.org/10.1016/S0006-3223(02)01866-8)
- Salthouse, T. A. (2000). Aging and measures of processing speed. *Biological Psychology*, *54*(1), 35–54. [https://doi.org/10.1016/S0301-0511\(00\)00052-1](https://doi.org/10.1016/S0301-0511(00)00052-1)
- Singmann, H., Brown, S., Gretton, M., & Heathcote, A. (2020). *Rtdists: Response time distributions*. Retrieved from <https://CRAN.R-project.org/package=rtdists>
- Snaith, R. P., Hamilton, M., Morley, S., Humayan, A., Hargreaves, D., & Trigwell, P. (1995). A Scale for the Assessment of Hedonic Tone the SnaithHamilton Pleasure Scale. *The*

- British Journal of Psychiatry*, 167(1), 99–103. <https://doi.org/10.1192/bjp.167.1.99>
- Snyder, H. R. (2013). Major depressive disorder is associated with broad impairments on neuropsychological measures of executive function: A meta-analysis and review. *Psychological Bulletin*, 139(1), 81–132. <https://doi.org/10.1037/a0028727>
- Tripp, G., & Alsop, B. (1999). Sensitivity to Reward Frequency in Boys With Attention Deficit Hyperactivity Disorder. *Journal of Clinical Child Psychology*, 28(3), 366–375. <https://doi.org/10.1207/S15374424jccp280309>
- Vallesi, A., Canalaz, F., Balestrieri, M., & Brambilla, P. (2015). Modulating speed-accuracy strategies in major depression. *Journal of Psychiatric Research*, 60, 103–108. <https://doi.org/10.1016/j.jpsychires.2014.09.017>
- Vandierendonck, A., Liefooghe, B., & Verbruggen, F. (2010). Task switching: Interplay of reconfiguration and interference control. *Psychological Bulletin*, 136(4), 601–626. <https://doi.org/10.1037/a0019791>
- Venables, W. N., & Ripley, B. D. (2002). *Modern applied statistics with s* (Fourth). New York: Springer. Retrieved from <http://www.stats.ox.ac.uk/pub/MASS4>
- Voss, A., Nagler, M., & Lerche, V. (2013). Diffusion models in experimental psychology: A practical introduction. *Experimental Psychology*, 60(6), 385–402. <https://doi.org/10.1027/1618-3169/a000218>
- Voss, A., Voss, J., & Lerche, V. (2015). Assessing cognitive processes with diffusion model analyses: A tutorial based on fast-dm-30. *Frontiers in Psychology*, 6, 336. <https://doi.org/10.3389/fpsyg.2015.00336>
- White, C. N., Ratcliff, R., Vasey, M. W., & McKoon, G. (2010). Using diffusion models to understand clinical disorders. *Journal of Mathematical Psychology*, 54(1), 39–52. <https://doi.org/10.1016/j.jmp.2010.01.004>
- Wickham, H. (2016). *Ggplot2: Elegant graphics for data analysis*. Springer-Verlag New York. Retrieved from <https://ggplot2.tidyverse.org>
- Wickham, H. (2019). *Stringr: Simple, consistent wrappers for common string operations*. Retrieved from <https://CRAN.R-project.org/package=stringr>
- Wickham, H. (2020). *Forcats: Tools for working with categorical variables (factors)*. Retrieved from <https://CRAN.R-project.org/package=forcats>
- Wickham, H. (2021). *Tidyr: Tidy messy data*. Retrieved from <https://CRAN.R-project.org/package=tidyr>
- Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., . . . Yutani, H. (2019). Welcome to the tidyverse. *Journal of Open Source Software*, 4(43), 1686. <https://doi.org/10.21105/joss.01686>
- Wickham, H., François, R., Henry, L., & Müller, K. (2021). *Dplyr: A grammar of data manipulation*. Retrieved from <https://CRAN.R-project.org/package=dplyr>

Wickham, H., Hester, J., & Francois, R. (2018). *Readr: Read rectangular text data*. Retrieved from <https://CRAN.R-project.org/package=readr>

## Appendix A - Sample Size Planning

To determine the sample size for Experiment 1, the procedure introduced by Kruschke and Liddell (2018) was employed. The procedure has several steps (visualised schematically in Figure 5) which I outline below.

First, a minimum effect size of interest needs to be established. For the current study, the research question is primarily interested in the relationship between depression severity and various behavioural outcome variables and diffusion model parameters. I used a correlation coefficient of  $r = -0.18$  as the effect size of interest, as this is the correlation found in our previous work (Grange & Rydon-Grange, 2021).

Second, idealised data that conform to this effect size of interest are simulated. For this, I generated data from 250 simulated participants by sampling from a multivariate normal distribution. I sampled data from two measures (each standardized with mean zero and standard deviation of one) assuming the population-level correlation between the measures was  $r = -0.18$ . This simulated data is shown in the left panel of Figure 5.

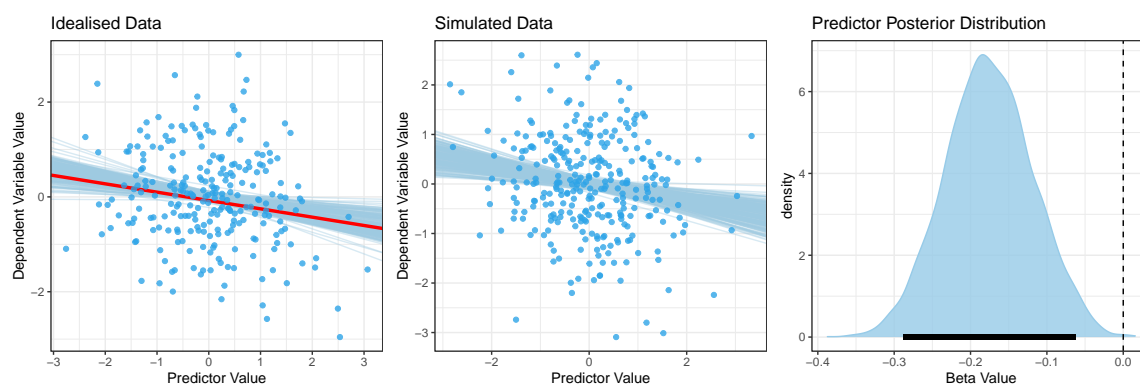


Figure 5. Schematic example of the steps applied to conduct sample size planning for Bayesian linear regression. See text for details.

Then, the statistical model of choice is applied to this simulated data set. The analytical approach in the current study was a multilevel Bayesian regression with one outcome variable and one predictor variable. The posterior distribution of this statistical model provides a set of plausible values for the main statistical parameters of interest (in this case, the main parameter is the slope of the predictor variable). This is shown by the blue lines in the left panel of Figure 5; each line represents a regression slope drawn from the posterior of the Bayesian model. Each regression line is a plausible estimate of the true population-level relationship between the two variables.

Then, new sample data is generated using these plausible parameter values. I generated new samples of participants using 1,000 parameters from the posterior distribution of the model fit to the idealised data. Each sample had  $N = 300$  simulated participants, which was the sample size I wished to assess the adequacy of in this analysis. For example, the red line highlighted in the left panel of Figure 5 shows one draw from the posterior distribution; this draw has a slope of  $-0.176$ . We would therefore simulate new data (again from the

multivariate normal distribution) but use  $r = -0.176$  as the correlation parameter. This new simulated data is shown in the middle panel of Figure 5.

With this new sample data in hand, a new Bayesian regression model is fit to the simulated data. Plausible regression lines from the posterior prediction of this new model fit are shown as blue lines in the middle panel of Figure 5.

We can then assess the posterior distribution of the key parameter of interest; in the current case, the key parameter is the slope of the regression (i.e., the beta value for the predictor). The right panel of Figure 5 shows the posterior distribution of the predictor in the regression fit. We are interested in whether our research goals have been met in this posterior distribution. Specifically, we assess whether 95% of the posterior distribution is different from zero. In the right panel of Figure 5, the 95% Credible Interval (CI) of the posterior is shown as the horizontal black line. As can be seen, the entirety of the CI is below zero; in this case, we would conclude that the true parameter value is non-zero. We would therefore conclude that we have detected an effect.

After this, a new sample from the posterior of the idealised data model fit is selected (i.e., a new red line is chosen) to generate new sample data, and the posterior distribution from the model fit to the new sample data is examined to see whether the 95% CI is non-zero, and so on (for a total of 1,000 iterations). The power analysis then counts the proportion of sample data where the CI was non-zero.

If the intended sample size is sufficient, a true effect should be found in a large proportion of simulated data sets. In our power analysis (assuming  $r = -0.18$  and  $N = 300$ ), 86.2% of simulated samples had a non-zero CI. Put another way, the intended sample size of 300 had a power of 86.2%.



## Appendix B - Pilot Studies

### Pilot 1: Establishing Baseline Performance

In speed-accuracy tradeoff experiments, researchers typically reinforce accuracy and speed instructions with performance feedback on a trial's accuracy or speed (respectively) when a particular criterion is not met. For example, in accuracy-emphasis blocks researchers will typically present feedback (e.g., "ERROR!") when an incorrect response is made. In speed-emphasis blocks, feedback is provided (e.g., "TOO SLOW!") when the participant's response is slower than a pre-determined threshold. The question thus arises of what response time threshold should be used.

The purpose of the current pilot study was to establish baseline task performance in terms of speed and accuracy on the task without speed- or accuracy-emphasis instructions; with this baseline performance in hand, we can then use the pilot data to establish a sensible response time threshold for speed-emphasis blocks in the main speed-accuracy tradeoff Experiments.

#### Method.

**Participants.** 50 participants were recruited via Prolific Academic residing in the UK or USA. I used pre-screening tools within Prolific Academic to limit our sample age range to 18-35. I maintained a tight (and young) age-range because response times are known to slow due to healthy ageing (Salthouse, 2000). If a wider age range was included in the sample it would be hard to find a single response time threshold that would be suitably challenging for older and younger participants alike. Participants could only take part via a laptop or desktop computer.

**Stimuli.** The same stimuli used in the behavioural task reported in the main paper was used in this first pilot study. See Figure 6 for a schematic overview of a single trial in this task.

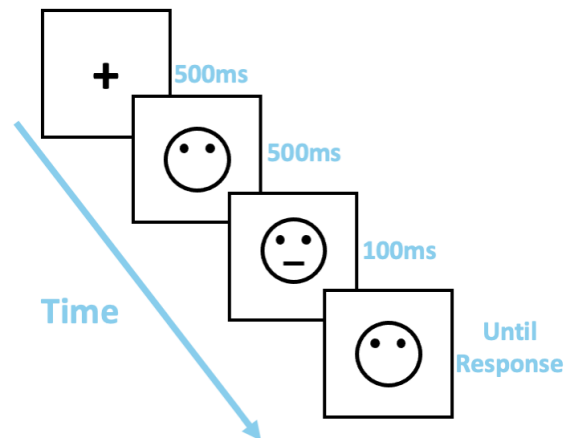


Figure 6. Schematic overview of a single trial in the behavioural task. Note that images are not to scale.

**Procedure.** After providing informed consent, the participants were presented with full written instructions on how to perform the task. Participants then engaged in some practice trials before moving on to the main experimental blocks. On each trial, a black

fixation cross was presented in the centre of the frame for 500ms. After this, the schematic face was presented without the mouth for 500ms, at which point the mouth was presented for 100ms. After this time, the mouth disappeared, and the schematic face remained on screen until the participant made a response. The stimulus was selected randomly on each trial with the constraint that there were equal numbers of long- and short-mouth stimuli within each block.

The participants were required to press the “Z” key if they thought the presented mouth was short, and “M” if they thought the presented mouth was long. Response instructions asked participants to respond as quickly and as accurately as possible as soon as the mouth appeared, and were asked to use the index finger of each hand for the responses. Once a response was registered, the face was removed and the fixation cross for the next trial appeared. During the practice block, participants received feedback on the accuracy of their responses to help them learn the task. If the response was correct, a green tick was presented on top of the schematic face; if the response was incorrect, a red “X” was shown. Feedback was displayed for 200ms.

Participants were presented with 24 practice trials. The main experimental section consisted of 6 blocks of 48 trials. There was the opportunity for a self-paced break between each block. Once all blocks were complete, participants were presented with a debrief screen informing them of the purpose of the study.

**Quality checks and data exclusion.** As the purpose of this pilot study was to establish typical performance in this paradigm, I did not perform any quality checks or exclude any data for the accuracy analysis. Trials in which an error was made were removed from the response time analysis.

**Results.** Density distributions of both the mean proportion accuracy and mean response times are shown in Figure 7. Mean proportion accuracy was 0.935 (SE = 0.007), and mean response time was 453ms (SE = 11ms).

**Establishing a suitable threshold.** Examination of the response time density distribution in Figure 7 suggests that the majority of response times fell below 500ms. Indeed, the mean percentage of trials that fell below 500ms was 76.7% (SE = 2.8%, min = 1.1%, max = 99.6%), suggesting that for the majority of participants 500ms would prove a suitable challenge for a response time threshold.

## **Pilot 2: SAT Manipulation Check**

In this pilot study I introduced speed-accuracy manipulations by emphasising speed and accuracy in separate blocks; this provided an opportunity to test whether the response time threshold established in pilot study 1 was suitable. I also fitted the diffusion model to the data.

Success of this pilot study would be determined by being able to successfully induce speed-accuracy tradeoffs with our method (that is, by observing better accuracy in “accuracy-emphasis” blocks, and shorter RTs in “speed-emphasis” blocks). Establishing which diffusion model parameters are affected by speed-accuracy tradeoffs would also help establish predictions for the main experiments.

### **Method.**

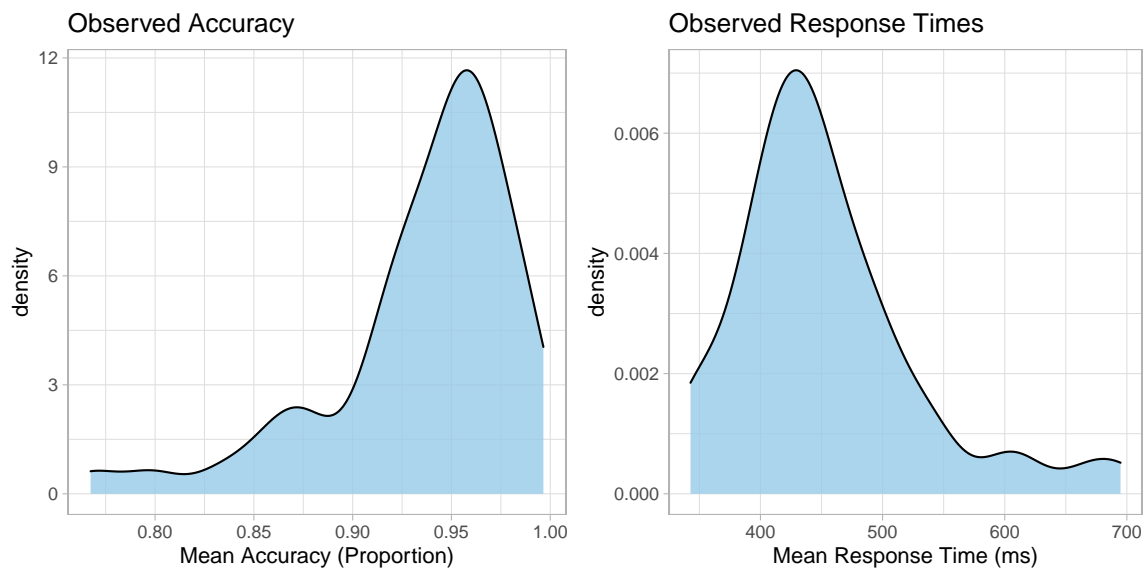


Figure 7. Density distributions of the mean proportion accuracy and mean response times from Pilot Study 1. ms = milliseconds.

**Participants.** 50 participants aged between 18–35 (inclusive) were recruited from Prolific Academic residing in the UK or USA. The same exclusion criteria used in Pilot Study 1 were used. Participants who took part in Pilot Study 1 were not eligible to take part in this study.

**Stimuli.** The same stimuli from Pilot Study 1 were used.

**Procedure.** The task requirements were the same as in Pilot Study 1. However, there were some general changes to the procedure in terms of instructions. I also added an attention check (described below) to make sure participants were reading the instructions; I removed participants who failed this attention check from the analysis.

When participants first entered the study, they were given general instructions that initially did not mention speed-accuracy instructions. Instead, they were given general instructions to learn the response requirements, and were then presented with a practice block to learn the task. During this initial practice block, the instructions asked participants to respond “as quickly and as accurately as possible”. Feedback was presented if participants made an error in the same way as in Pilot Study 1. This initial practice consisted of 16 trials.

After this, participants were given instructions about the speed-accuracy emphasis instructions. They were told that in the main study they would experience “accuracy-emphasis” and “speed-emphasis” blocks. They were told that in accuracy-emphasis blocks, they should respond “... as accurately as possible without taking too long to respond”. They were told that in speed-emphasis blocks they should respond “... as quickly as possible without making too many mistakes. DO NOT GUESS”. After these instructions, participants were presented with a practice session where they were presented with 12 trials under accuracy-emphasis instructions followed by 12 trials under speed-emphasis instructions.

Response feedback on these trials changed depending on which instructions they were performing under. In accuracy-emphasis trials, if participants made an error they saw the message “INCORRECT!” presented centrally in red font for 1500ms. On speed trials, if participants responded slower than 500ms (the threshold established in pilot study 1) they saw the message “TOO SLOW!” for 1500ms. No feedback was provided for correct responses in accuracy-emphasis blocks, and no feedback was provided for sub-threshold responses in speed-emphasis blocks.

After this second practice, participants were then presented with blocks 8 of 48 trials in each with self-paced rest screens between each block. The relevant emphasis (speed vs. accuracy) alternated every block, beginning with accuracy emphasis. Reminders of the relevant speed or accuracy emphasis instructions were presented at the beginning of each block.

**Attention check.** After the fourth block, an attention check was presented. This attention check consisted of a mini-block of 8 trials. However, the instruction screen for this block read: *"In this block, we are applying an attention check to ensure you are attending to these instructions. For all responses please press the B key regardless of the correct response"*. Participants who did not respond to all trials with the B key were deemed to have failed this check.

This attention check was important because the study relies on participants implementing the instructions provided; we wanted to identify participants who were not reading the block instructions carefully so they could be removed from data analysis.

**Quality checks and data exclusion.** Before analysing the data I performed some quality checks to identify participants and trials to remove from the final analysis. 10 participants failed the attention check, and were thus removed from the analysis. I also wished to identify participants who performed the task poorly. For this study, I selected a minimum criterion of 75% accuracy for inclusion. No participant performed below this criterion. For the response time analysis, I removed error trials, and trimmed the correct response times to only include RTs longer than 150ms and shorter than 5,000ms. Although liberal, this response time trimming is important for the diffusion modelling.

**Results.** Density distributions of both the mean proportion accuracy and mean response times are shown in Figure 8. Mean proportion accuracy was 0.935 (SE = 0.007), and mean response time was 453ms (SE = 11ms).

The results showed that proportion accuracy was higher in the accuracy-emphasis blocks ( $M = 0.951$ ,  $SE = 0.005$ ) than in the speed-emphasis blocks ( $M = 0.866$ ,  $SE = 0.011$ ),  $t(39) = 9.94$ ,  $p < .001$ , Cohen's  $d = 1.54$  (95% confidence interval 1.04, 2.05); response time was shorter in the speed-emphasis block ( $M = 382$ ms,  $SE = 5$ ms) than the accuracy-emphasis block ( $M = 425$ ms,  $SE = 7$ ms),  $t(39) = 9.24$ ,  $p < .001$ , Cohen's  $d = 1.12$  (95% confidence interval 0.64, 1.60). Together, these results provide strong evidence that the speed-accuracy emphasis instructions produced the expected speed-accuracy tradeoff in the behavioural data.

**Diffusion modelling.** The diffusion model was fitted to the behavioural data using the fast-dm-30 software (Voss et al., 2015). This software was called via custom in-house

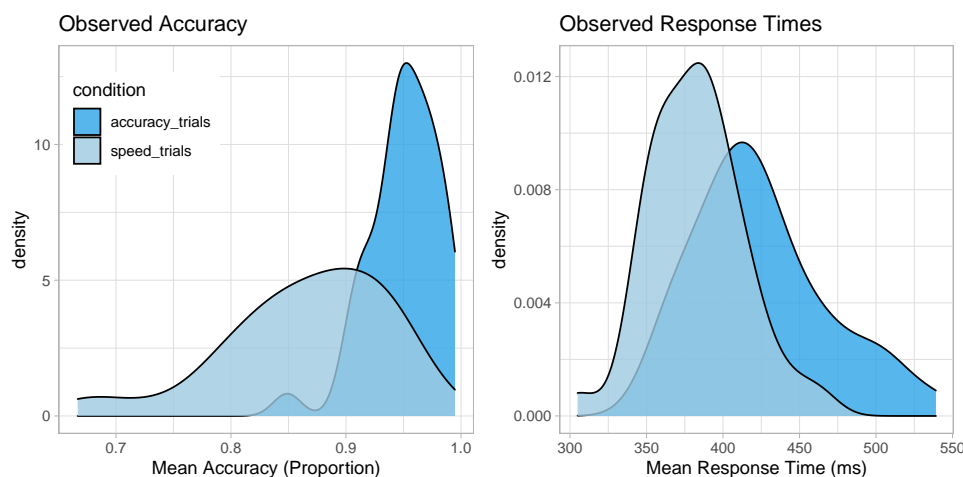


Figure 8. Density distributions of the mean proportion accuracy and mean response times from Pilot Study 2 for the speed- and accuracy-emphasis conditions. ms = milliseconds.

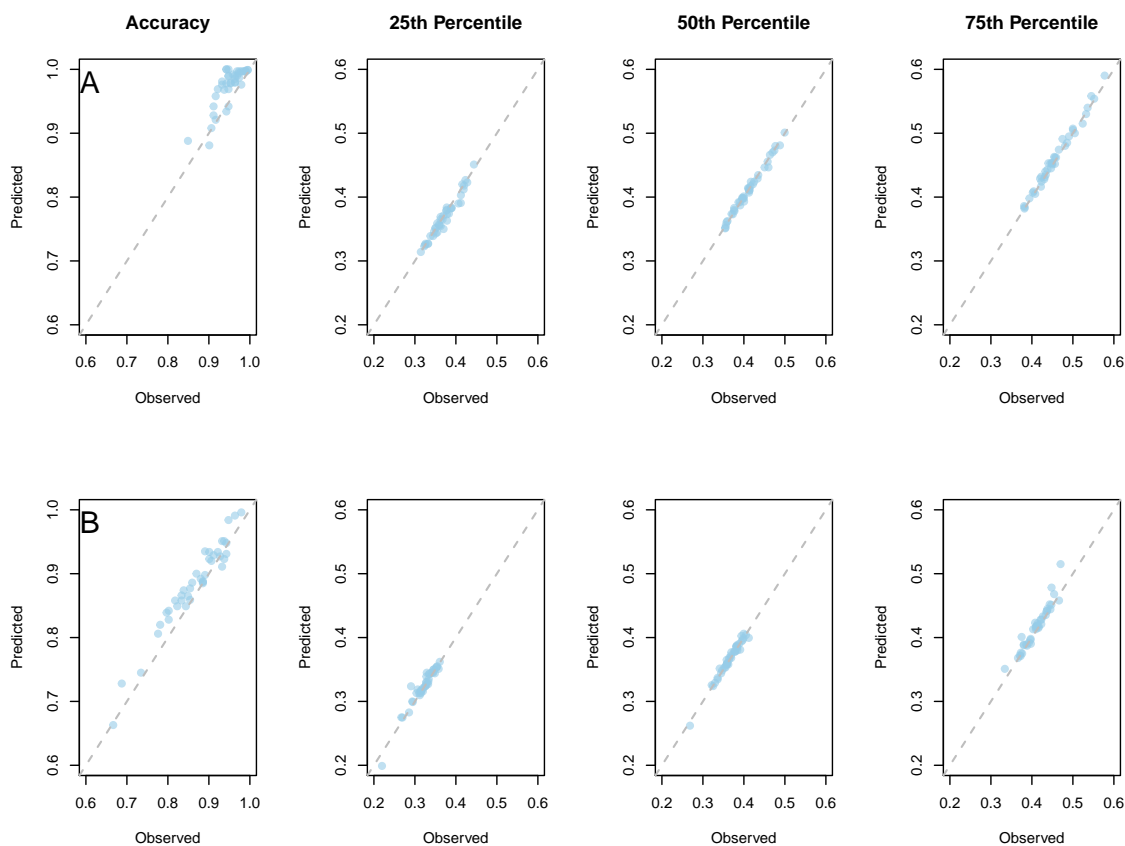
scripts written using R.<sup>4</sup> The full diffusion model has 8 parameters, but I only allowed 3 to vary across speed-accuracy conditions. Specifically, I allowed drift rate, boundary separation, and non-decision time to freely vary between conditions. Non-decision time was not allowed to vary for upper and lower response boundaries (i.e.,  $d$  was fixed at zero). Trial-level variability in drift rate ( $s_v$ ) and starting point of the drift rate ( $s_{zr}$ ) were also fixed to zero (Voss et al., 2015). Trial-level variability in non-decision time was a free parameter, but was not free to vary across speed-accuracy conditions. The model was fit using the Kolmogorov-Smirnov optimisation criterion.

I assessed goodness of fit of the model to individual participant data by plotting each participant’s data against simulated predictions from the diffusion model using the best fitting model parameters for that participant. Proportion accuracy was plotted as well as the 25th, 50th, and 75th percentile of each participant’s response time distribution. If the model fits the data well across all participants, data points will fall along the main diagonal in QQ-plots. The QQ-plots for the current model fit are shown in Figure 9; these plots show a generally good model fit.

Density distributions of the mean parameter estimates across all participants is shown in Figure 10. From this it can be seen that accuracy-emphasis instructions led to an increase in all parameter values in the diffusion model.

For the boundary separation parameter, estimates were higher for the accuracy-emphasis blocks ( $M = 0.933$ ,  $SE = 0.043$ ) than in the speed-emphasis blocks ( $M = 0.692$ ,  $SE = 0.019$ ),  $t(39) = 6.00$ ,  $p < .001$ , Cohen’s  $d = 1.14$  (95% confidence interval 0.66, 1.62). The results also showed that estimates of drift rate were also higher for the accuracy-emphasis blocks ( $M = 4.64$ ,  $SE = 0.172$ ) than in the speed-emphasis blocks ( $M = 3.33$ ,  $SE = 0.172$ ),  $t(39) = 8.63$ ,  $p < .001$ , Cohen’s  $d = 1.21$  (95% confidence interval 0.72, 1.69). Estimates of

<sup>4</sup>Note that I verified the model fit results by using the R package `rtdists` and obtained qualitatively similar results.



*Figure 9.* QQ-plots of participant data against diffusion model predictions in Pilot Study 2. The main diagonal shows where data should fall under a perfect model fit. **A.** Data & model predictions for accuracy-emphasis trials. **B.** Data & model predictions for speed-emphasis trials.

the non-decision parameter were also higher for the accuracy-emphasis blocks ( $M = 0.322$ ,  $SE = 0.005$ ) than in the speed-emphasis blocks ( $M = 0.291$ ,  $SE = 0.006$ ),  $t(39) = 5.21$ ,  $p < .001$ , Cohen's  $d = 0.86$  (95% confidence interval 0.39, 1.33).

**Conclusion.** Together, the results from Pilot Study 2 show that the speed-accuracy instructions and relevant block-level response feedback were successful in inducing results typical of a speed-accuracy tradeoff. The diffusion modelling showed that emphasising accuracy over speed increased estimates of response caution (via the boundary separation parameter), which is a typical finding (Forstmann et al., 2011; Ratcliff et al., 2007; Ratcliff, Thapar, & McKoon, 2010). Emphasising accuracy over speed also increased the rate of evidence accumulation (i.e., the drift rate) and non-decision elements of response time, which is a result reported before (see for example Rae et al., 2014).

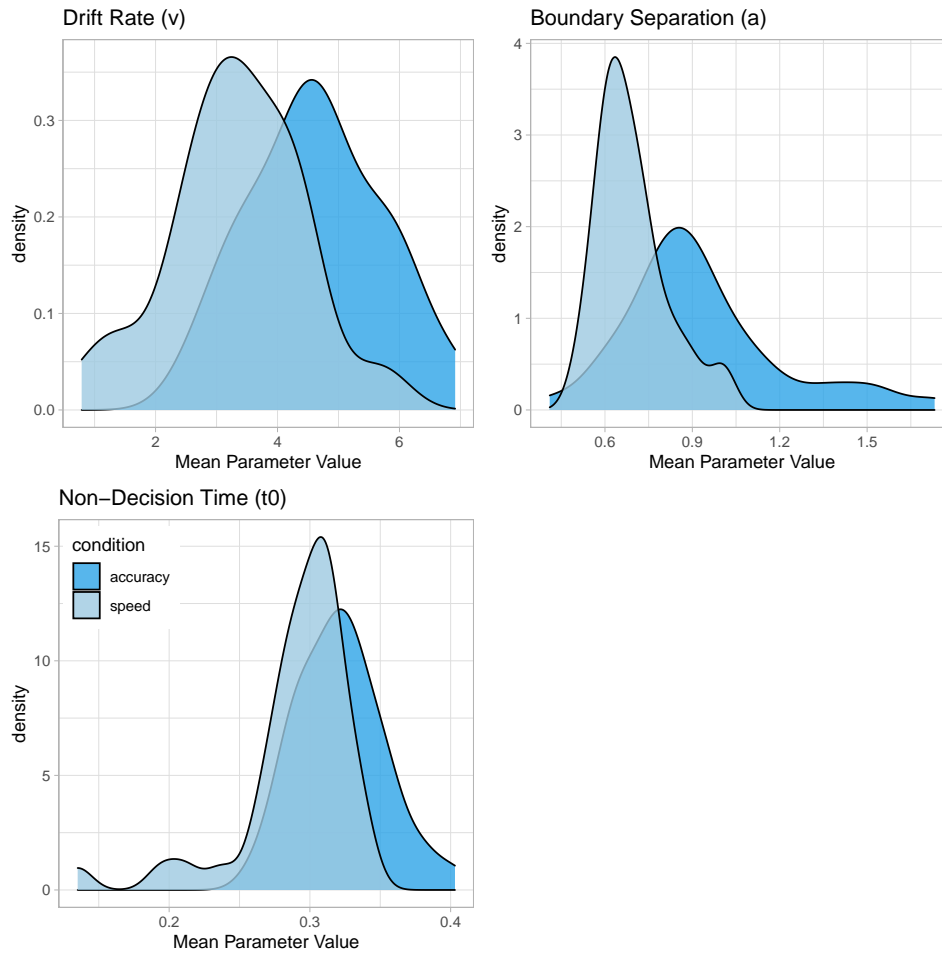
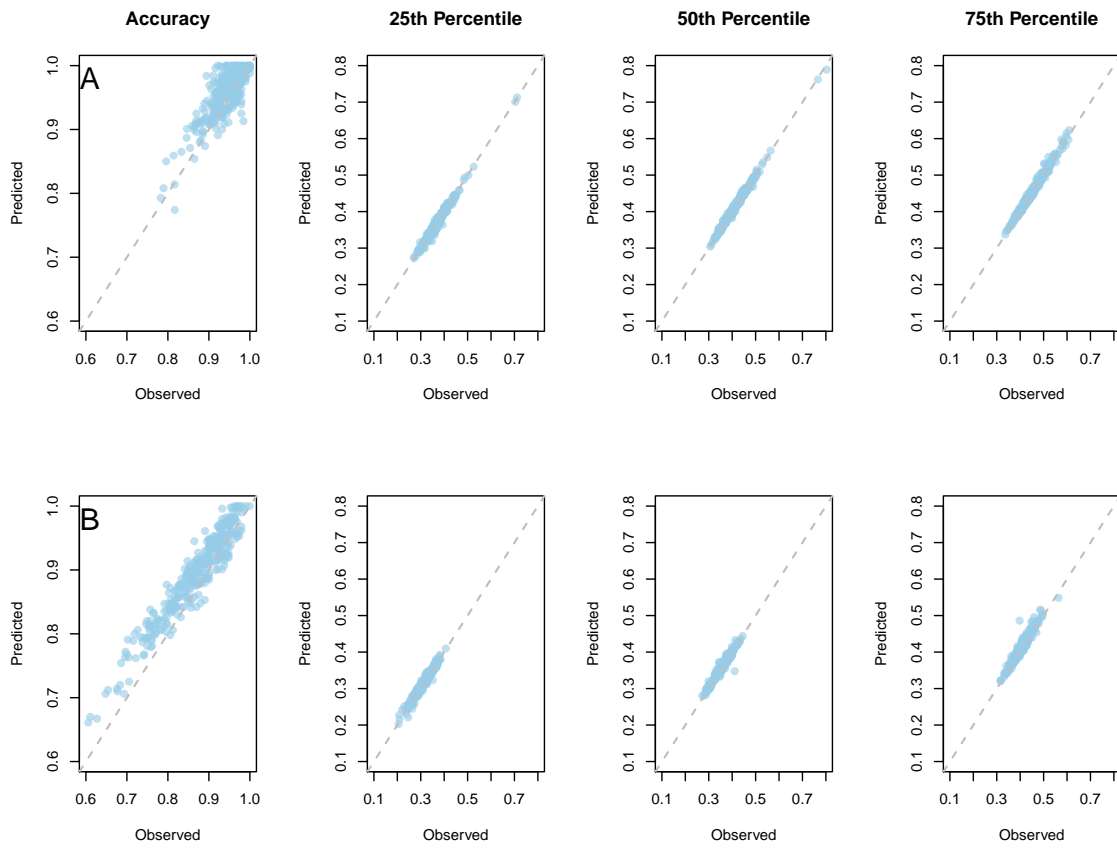


Figure 10. Density distributions of the mean value of each of the diffusion model parameters from Pilot Study 2 for the speed- and accuracy-emphasis conditions.

### Appendix C- Assessment of Model Fit

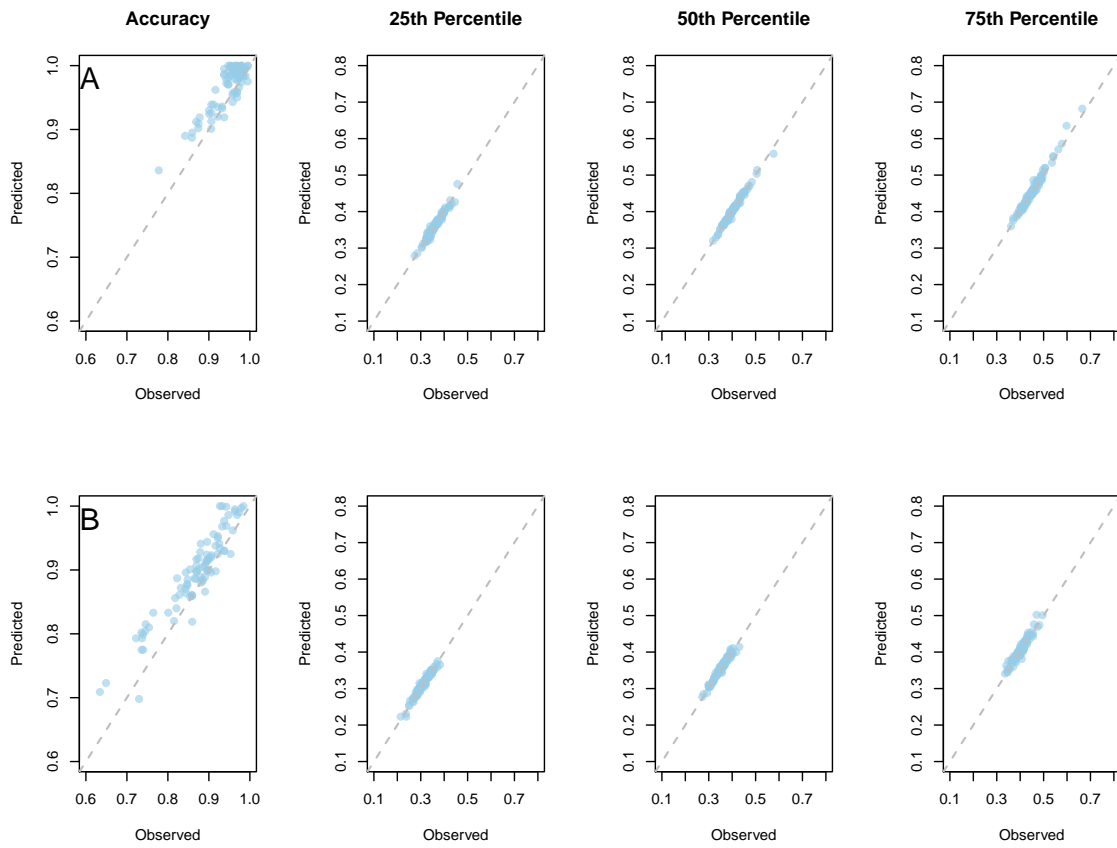
The quality of fit of the diffusion model to participant data was assessed in two stages. In a first stage, we removed participants who showed a significant deviation between observed and predicted distributions via the Kolmogorov-Smirnov test (i.e., if their fit  $p$ -values provided by `fast-dm-30` are below 0.05). 5 participants were removed from Experiment 1, and 3 participants were removed from Experiment 2 due to this criterion.

In a second assessment, each participant's data were plotted against simulated predictions from the diffusion model using the best fitting model parameters for each participant. Overall proportion accuracy was plotted, as well as the 25th, 50th, and 75th percentile of each participant's response time distribution. If the model fits the data well across all participants, data points will fall along the main diagonal in QQ-plots. The resulting QQ-plots—shown in Figure 11 for Experiment 1 and Figure 12 for Experiment 2—show good correspondence between model predictions and participant data.



*Figure 11.* QQ-plots of participant data against diffusion model predictions in Experiment 1. The main diagonal shows where data should fall under a perfect model fit. **A.** Data & model predictions for accuracy-emphasis trials. **B.** Data & model predictions for speed-emphasis trials.





*Figure 12.* QQ-plots of participant data against diffusion model predictions in Experiment 2. The main diagonal shows where data should fall under a perfect model fit. **A.** Data & model predictions for accuracy-emphasis trials. **B.** Data & model predictions for speed-emphasis trials.