Computational modelling of attentional selectivity in depression reveals perceptual deficits

James A. Grange
School of Psychology, Keele University, UK

Michelle Rydon-Grange
Midlands Partnership Foundation NHS Trust, UK

Word Count: 4,497 (Main Body Only)

Abstract

Background. Depression is associated with broad deficits in cognitive control, including in visual selective attention tasks such as the flanker task. Previous computational modelling of depression and flanker task performance showed reduced preprotent response bias and reduced executive control efficiency in depression. In the current study we applied two computational models that account for the full dynamics of attentional selectivity. **Method.** Across 3 large-scale online experiments (one exploratory experiment followed by two confirmatory—and pre-registered—experiments; Total N = 923) we measured attentional selectivity via the flanker task and obtained measures of depression symptomology as well as anhedonia. We then fit two computational models that account for the dynamics of attentional selectivity: The Dual-Stage Two Phase model, and the Shrinking Spotlight model. **Results.** No behavioural measures were related to depression symptomology or anhedonia. However, a parameter of the Shrinking Spotlight model that indexes the strength of perceptual input was consistently negatively associated with the magnitude of depression symtomology. **Conclusions.** The findings provide evidence for deficits in perceptual representations in depression. We discuss the implications of this in relation to the hypothesis that perceptual deficits potentially exacerbate control deficits in depression.

Keywords: Depression; Cognitive Control; Cognitive Modelling

It is an established finding that depression is associated with deficits in cognitive

function (Burt, Zembar, & Niederehe, 1995; McDermott & Ebmeier, 2009; Rock, Roiser, Riedel, & Blackwell, 2014; Snyder, 2013) and executive function (EF) (Snyder, 2013). EFs refer to the set of high-level cognitive processes that control ongoing behaviour so as to allow goal-directed action (Miyake et al., 2000; Norman & Shallice, 1986) by regulating subordinate cognitive processes (see e.g., Logan & Gordon, 2001). The finding that EFs are impaired in depression is perhaps not surprising given that the prefrontal cortex—an essential region for efficient executive functioning (Struss & Knight, 2002)—has been shown to function hypoactively in depressed individuals (Levin, Heller, Mohanty, Herrington, & Miller, 2007).

Assessing the effects of clinical disorders on executive functioning is complicated by the fact that EFs themselves are not directly observable; as they are higher-order processes that act on lower-order processes (such as perceptual processes, for example), one can only infer the influence and efficacy of EFs by observing changes in manifest variables, such as response time (RT; Logan & Gordon, 2001; Miyake et al., 2000). Importantly, though, changes in EF *and* subordinate processes can independently influence manifest variables. This complicates matters because a clinical disorder that negatively affects a subordinate process may lead to prolonged RT in a clinical group in comparison to a control group; the impaired RT could mistakenly be taken as evidence of impaired EFs in the clinical group.

One solution to this *task impurity problem* (Miyake et al., 2000) is to fit computational cognitive models to data from clinical and control groups (Farrell & Lewandowsky, 2010; Lewandowsky & Farrell, 2010). Such models of EF will have parameters that reflect higher-order processes (such as control processes) and lower-order processes (such as the strength of perceptual representations), allowing one to distinguish between effects on EFs and subordinate processes, and how these processes change across conditions and clinical groups.

Modelling the Effects of Depression on Flanker Task Performance

One popular measure of a component of EF—visual selective attention—is the Eriksen flanker task (Eriksen & Eriksen, 1974). In this task, participants are presented with a series of arrows (for example) and must respond to the direction of the central arrow (left vs. right). On some trials, the central arrow is flanked by arrows that point in the opposite direction to that of the target (*incongruent trials*; e.g., >><>>), and on other trials the central arrow is flanked by arrows pointing in the same direction (*congruent trials*; e.g., <<<<<). It is a well replicated finding that incongruent trials are responded to more slowly and with poorer accuracy than congruent trials; this *congruency effect* is thought to reflect the interference caused by the flankers during response selection in the incongruent condition, and the time taken to overcome such interference. The magnitude of the flanker effect can thus be used to assess the efficacy of visual selective attention, with smaller values suggesting better selective attention.

Dillon et al. (2015) recently utilised the flanker task to assess visual selective attention in individuals with major depressive disorder (MDD) compared to a control group. In the behavioural data, Dillon et al. (2015) found those with MDD performed more slowly *but also more accurately* than control participants on incongruent trials. Such speed–accuracy trade-offs are notoriously difficult to interpret (see for example Wagenmakers, van der Maas,

& Grassman, 2007; Wickelgren, 1977): slower RTs suggests *poorer* performance, but better accuracy suggests *better* performance.

Dillon et al. (2015) fitted a computational model of flanker task performance—the Linear Approach to Threshold with Ergodic Rate (LATER) model (Noorani & Carpenter, 2013)—to correct response times for both groups. This model has parameters reflecting three core cognitive processes: prepotent response bias (the degree to which the participant is influenced by distracting flankers); response inhibition (required to resist early responding); and executive control (required to overcome the prepotent response and respond accordingly to the central target). The results of the modelling showed reduced pre-potent response bias and slower executive control in the depressed group. Within the depressed group, there was a significant negative correlation between the speed of executive control and a questionnaire measure of anhedonia.

Modelling Attentional Selectivity

Given the striking speed–accuracy tradeoff reported by (Dillon et al., 2015), it is imperative to take into account both accuracy and response speed when modelling the effects of depression. It has been shown that accuracy data provides essential information—and hence, essential constraints on theoretical modelling—about the dynamics of attentional selectivity in the flanker task. Gratton, Coles, Sirevaag, and Eriksen (1998) used *conditional accuracy functions* (CAFs) to examine the dynamics of attentional selectivity in the flanker task. CAFs are constructed by ordering a participant's RTs for each condition (e.g., congruent trials and incongruent trials separately) from fastest to slowest, and then dividing this data into equally-sized bins. The accuracy for each bin is then plotted against the mean RT for each bin to show how accuracy changes across the RT distribution.

Gratton et al. (1998) used CAFs of flanker task performance to show that attentional selectivity improves with processing time: They reported a large congruency effect in accuracy for the fastest RT bins, but this congruency effect in accuracy reduced as RT increased. This finding is consistent with theoretical accounts of flanker task performance that assume attentional selectivity is relatively poor at early stages of processing, leading to uncertain response selection on incongruent trials as it is heavily influenced by the flanker stimuli; as processing time increases, attentional selectivity improves leading to a reduction of the influence of flanker stimuli on response selection. Note that this theoretical insight could not be established if accuracy data were not accounted for.

The Current Study

The purpose of the present study was to revisit the effect of depression on flanker task performance, but to fit to the data two models—the dual-stage two-phase (DSTP) model of Hübner, Steinhauser, and Lehle (2010) and the shrinking spotlight (SSP) model of White, Ratcliff, and Starns (2011)—that were explicitly designed to jointly account for RT and accuracy performance, and thus could model the full dynamics of attentional selectivity.

Overview of the models. Both the DSTP and the SSP model successfully capture the improvement of attentional selectivity with processing time, but with different theoretical accounts. Figure 1 provides an overview of the main theoretical difference between the two models. A more detailed computational account of the two models is provided in

Table 1

List and brief description of model parameters in the dual stage two phase (DSTP) model and the shrinking spotlight (SSP) model.

Model	Parameter	Description
DSTP	A/B	Height of the boundary for the response selection diffusion process
	C/D	Height of the boundary for the stimulus selection diffusion process
	μ_{TA}	Drift rate for central target during response selection phase 1
	μ_{FL}	Drift rate for flankers during response selection phase 1
	μ_{SS}	Drift rate for stimulus selection
	μ_{RS2}	Drift rate for phase 2 of response selection
	ter	Non-decision time (time taken for perceptual encoding of stimuli & motoric response speed).
SSP	A/B	Height of the boundary for the response selection diffusion process
	ter	Non-decision time (time taken for perceptual encoding of stimuli & motoric response speed)
	p	Perceptual input strength of elements of stimulus display ($+p$ for target and congruent flankers; $-p$ for incongruent flankers).
	rd	Rate of attentional distribution reduction
	sd_a	Initial width of attentional distribution

online supplementary material Appendix A; a brief descriptor of the main parameters in each model is provided in Table 1.

Both models assume that response selection proceeds according to a drift diffusion process: After perceptual encoding of the stimulus, the cognitive system accumulates (noisy) evidence over time toward one of two response boundaries (one representing a correct response, and the other representing an error response). The boundary that is reached by the diffusion process determines the model's response accuracy, and the time taken for the boundary to be reached determines the RT.

Both models assume early stages of response selection are influenced by both the flankers and the central target, but at later stages of processing response selection is primarily driven by attention to the central target. The two models, however, have different assumptions as to how attentional selectivity increases over time, and how this influences response selection. The DSTP model assumes two phases to response selection: In a first phase, attentional selectivity is poor and response selection is influenced by the whole stimulus display; at a discrete point in time, the attentional system selects the central target for more-detailed processing, leading to the second phase of response selection which is influenced solely by the central target. The time taken for the system to select which stimulus to process further is also modelled by a parallel diffusion process (see Figure 1). The SSP model also assumes that early stages of response selection are influenced by the whole stimulus display, but that attentional selectivity improves gradually over time (i.e., the attentional spotlight gradually reduces its diameter) meaning that later stages of response selection are influenced less by the flankers. The contribution of each element in the stimulus display (i.e., the central target and the flankers) is a multiplicative combination of the strength of perceptual input of the stimulus items (represented by model parameter p) and the area of the attentional spotlight currently over the stimulus items.

General Method

We conducted one exploratory experiment, followed by two confirmatory (pre-registered) experiments. In this section, we provide an overview of the general method shared by all three experiments. Where relevant, we also highlight the minor differences between each experiment. All experiments were programmed and delivered us-
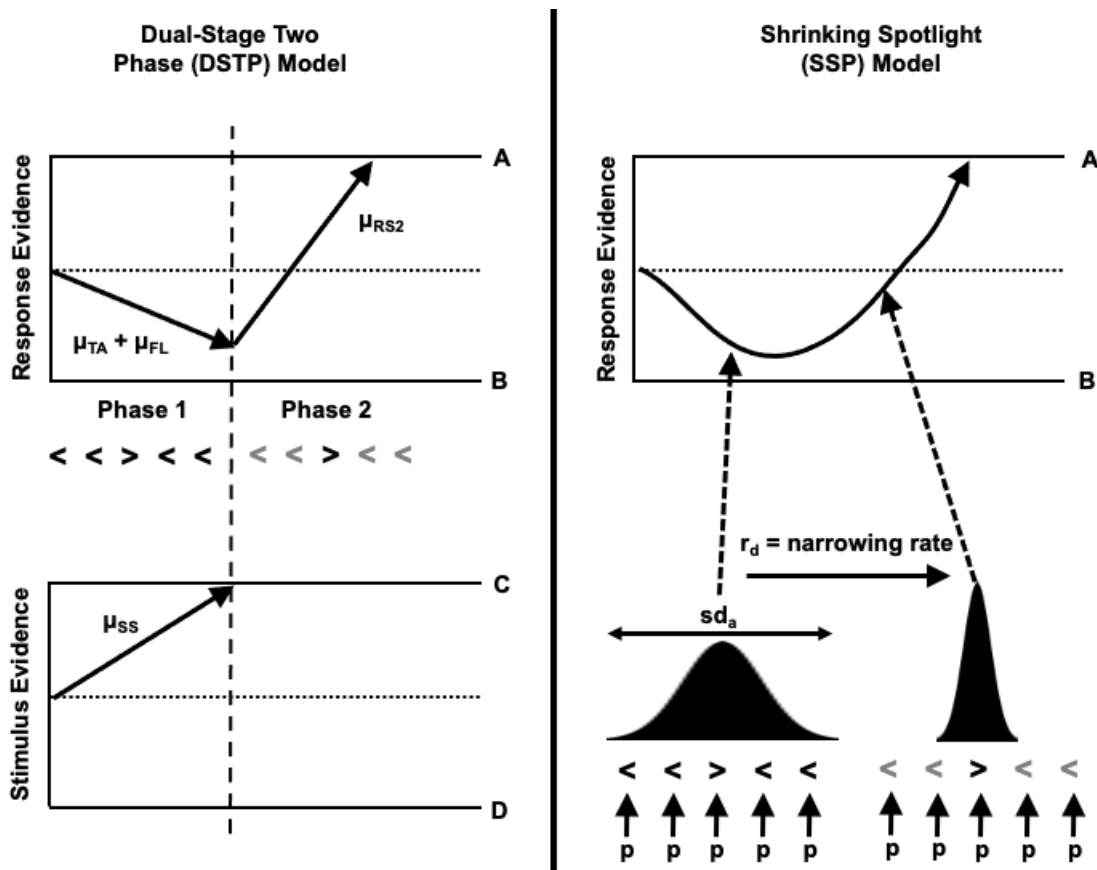
Figure 1. Schematic overview of the response selection processes in both models. **DSTP Model:** The upper panel depicts the diffusion process for response selection, and the lower panel depicts the diffusion process for stimulus selection. The point in time that a stimulus is selected for further processing is depicted by the dashed vertical line. The diffusion process before this point represents the first stage of response selection, and after this point represents the second—more selective—stage of response selection. **SSP Model:** The upper panel depicts the diffusion process for response selection, the drift rate of which is determined by the perceptual input of stimuli that fill within the area of the attentional spotlight, depicted in the lower panel. As time progresses, the attentional spotlight becomes more focussed around the central target, meaning perceptual information from the flankers contribute less to the drift rate of response selection as time increases. Figure available at https://flickr.com/photos/150716232@N04/48957578602 under CC license https://creativecommons.org/licenses/by/2.0/

ing the online behavioural science platform Gorilla^{TM} (https://gorilla.sc; Anwyl-Irvine, Massonnié, Flitton, Kirkham, & Evershed, 2019). Participants were recruited using Prolific Academic (https://prolific.ac). This study received full ethical clearance from the Ethics Panel run by the School of Psychology at Keele University, UK (application number PS–190046). The methods and analytical strategy were pre-registered for Experiments 2 and 3 at https://aspredicted.org (see https://aspredicted.org/3hr8j.pdf and

https://aspredicted.org/8eg68.pdf respectively).

Participants

Using pre-screening tools in Prolific Academic, we limited our sample to participants residing in the UK or USA aged between 18–60. Participants were screened so that they could only use a laptop or desktop machine (i.e., no mobile devices or tablet). After exclusion criteria were applied (see later section in the General Method), the final sample sizes for Experiments 1–3 were 294, 311, and 318 respectively. The demographic information of the final samples across all 3 experiments are shown in Table 2.

Table 2

Demographics data for the sample from all Experiments. Note: SE = standard error of the mean; n = Number.

Item		Experiment 1	Experiment 2	Experiment 3
Final Sample Size		294	311	318
Age: mean Years (SE)		31.30 (0.65)	33.11 (0.64)	35.19 (0.66)
Gender:	n Male (%)	126 (42.86)	106 (34.08)	108 (33.96)
	n Female (%)	159 (54.08)	198 (63.67)	210 (66.04)
	n Other (%)	9 (3.06)	7 (2.25)	0 (0.00)
Education: mean Years (SE)		5.00 (0.15)	5.05 (0.16)	4.78 (0.21)
Depression Diagnosis:	n Yes (%)	91 (30.95)	72 (23.15)	74 (14.78)
	n No (%)	203 (69.05)	239 (76.85)	244 (85.22)
Depression Medication:	n Yes (%)	58 (63.74)	47 (65.28)	46 (62.16)
	n No (%)	33 (36.26)	25 (34.72)	28 (37.84)
Other Medication:	n Yes (%)	29 (9.86)	25 (8.04)	37 (11.64)
	n No (%)	265 (90.14)	265 (91.96)	281 (88.36)
Number of Depressive Episodes:	0 (%)	106 (36.05)	145 (46.62)	165 (51.26)
	1–3 (%)	113 (38.44)	125 (40.19)	113 (35.53)
	4–7 (%)	49 (16.67)	34 (10.93)	30 (9.43)
	8+ (%)	26 (8.84)	7 (2.25)	12 (3.77)

Materials & Flanker Task

Demographics questionnaire. Participants completed a brief questionnaire to measure some demographic information. Participants were asked to: enter their date of birth; select their gender (Male–Female–Other); enter how many years of post-16 education they had; select whether they had a clinical diagnosis of depression (i.e., from a health professional; yes–no); if they did have depression, select whether they were currently taking medication for depression (yes–no); select whether they were taking medication for another mental health problem (yes–no; if yes, participants were asked to indicate what the mental health problem was); enter how many episodes of depression the participant had experienced over the past 2 weeks, regardless of whether they had a formal diagnosis or not[1].

---

[1]The exact wording of this question was *"Regardless of whether you have a diagnosis or not, how many episodes of depression have you had in the past two weeks?"*. We meant this to mean how many discrete times have you felt depressed over the past 2 weeks, rather than the formal diagnostic definition of the word "episode" (i.e., one episode being equivalent to two weeks' duration of depression.)

Quick Inventory of Depressive Symptomatology–Self-Report(QIDS). The QIDS–SR (Rush et al., 2003) is a 16-item self-report questionnaire assessing severity of depression symptoms experienced by respondents over the past seven days. Each item relates to a particular symptom (e.g., feeling sad), and requires the participant to select one response from four options that best describes them (e.g."I almost always feel sad"). The scale shows very good internal consistency and sensitivity to symptom changes (Rush et al., 2003). Scores on the QIDS range from 0 to 27 (low to high depressive symptomatology).

Snaith–Hamilton Pleasure Scale (SHPS). The SHPS (Snaith et al., 1995) is a 14-item self-report questionnaire which probes the participant's ability to experience pleasure; the questionnaire asks participants to consider their response in relation to how they have felt in the last few days. Each item requires participants to read a statement (e.g., " I would find pleasure in my hobbies and pastimes"), and to then select a response to indicate their agreement with the statement (e.g., "Definitely Agree, Agree, Disagree, Definitely Disagree"). We used the scoring method recommended by Franken, Rassin, and Muris (2007) (Definitely Agree = 1, Agree = 2, Disagree = 3, Definitely Disagree = 4). In Experiment 1, the response options were always ordered: Definitely Disagree, Disagree, Agree, Definitely Agree. In Experiments 2 and 3, the response options were changed to match exactly the order and wording presented by Snaith et al. (1995) (e.g., sometimes "Strongly" was used instead of "Definitely"; sometime "Strongly Agree" was the first option etc.).

Attention Checks. In Experiments 2 and 3, we wanted to exclude participants who were demonstrably not paying attention to the questionnaire items. Thus, we used an attention check embedded within the SHPS questionnaire. Specifically, the final item on the SHPS read "It is important that you pay attention to this study. Please select *Disagree*" [2]. Participants who failed to select this response option were excluded from the final sample.

Flanker Tasks. We used the same trial timings for all trials (practice and main blocks) for all experiments. On each trial, a black fixation cross was presented on the centre of the screen for 500 ms; this was immediately followed by presentation of the flanker stimulus, which was displayed until the participant made a response. After a response was registered, the fixation cross for the next trial was presented. The flanker stimuli used in Experiments 1 and 2 were arrows, presented in black font. The central target could face either left or right. This central target was flanked by two arrows on either side; these flankers all faced the same direction as each other. On *congruent* trials, the flankers faced the same direction as the target (e.g., `<<<<<`); on *incongruent* trials, the flankers faced the opposite direction as the target (e.g., `>><>>`). The participants' task was to make a spatially-congruent response as to the direction of the central target, pressing the "Z" key for a left response, and the "M" key for a right response. In Experiment 3, we used the letters "K" and "A" as stimuli (e.g., a congruent trial would be `KKKKK`, and an incongruent trial would be `AAKAA`). The participants' task was to judge the identity of the central letter by pressing the "A" key if the letter was A, and the "K" key if the letter was K. In all experiments, the stimulus from each trial was selected randomly (with replacement) from the total set of four possible stimuli, with the constraint that all stimuli occurred equally often in each block. Participants were asked to respond as quickly and as accurately as possible as soon as the stimulus appeared, using the index finger of each hand for their responses.

---

[2]This wording is recommended by Prolific Academic: http://tiny.cc/AttentionChecks

In all experiments, participants took part in 16 practice trials before commencing the main flanker blocks. The number and duration of the flanker blocks differed across experiments: in Experiment 1, there were 6 blocks of 60 trials; in Experiments 2 and 3 there were 8 blocks of 60 trials. Participants were provided feedback on their accuracy in the practice blocks: on correct trials, a green tick would overlay the central target; on incorrect responses, a red cross would overlay the central target.

Procedure

The same procedure occurred in all 3 experiments. Upon entering the study online, participants were presented with an information screen which provided general information about the study so they could decide whether they wished to take part, at which point participants were presented with a screen on which they provided informed consent.

Participants then completed the brief demographics questionnaire. At this point, the experimental software randomised the participant to one of four presentation orders, which counterbalanced the order of presentation of the questionnaires and the flanker task: (1) flanker task—QIDS—SHPS; (2) flanker task—SHPS—QIDS; (3) QIDS—SHPS—flanker task; (4) SHPS—QIDS—flanker task.

Before completing the flanker task, participants were presented with a full instruction screen that provided complete instructions for how to complete the flanker task. This was followed by a brief practice block before moving on to the main blocks. After each block, participants were invited to take a short (self-paced) break. The QIDS questionnaire was presented on four consecutive screens, with four questions per screen. Participants indicated their response to each item by selecting the relevant response clicking a radio button with their mouse; participants could change their response freely until they proceeded to the next screen. The SHPS questionnaire was presented across 4 screens, with four items on the first three screens, and two items on the last screen. Responses were again indicated via radio buttons. Once participants had completed all elements of the study, they were presented with a debrief screen which provided detailed information about the nature of the experiment.

Quality Checks & Data Exclusion

Before analysing the data, we conducted some quality checks on the data and removed some participants with reference to pre-defined (and pre-registered in Experiments 2 & 3) exclusion criteria. In Experiments 2 and 3, individuals who failed the attention check embedded within the SHPS questionnaire were removed. Responses to the questionnaires were also examined for "straight-lining" (e.g., only selecting the left-most response to all items); participants who showed straight-lining to all questions for both questionnaires were removed.

For the behavioural data, we removed participants who had a mean accuracy lower than 80%. For the RT analysis, error trials were removed; RTs were trimmed by removing RTs shorter than 250ms and longer than 1500ms. In cases where trimming removed more than 25% of an individual's data, that participant was removed.

Results

Analytical Strategy

All inferential analyses utilised a Bayesian regression approach using the `brms` package (Bürkner, 2017) in R (R Core Team, 2017). Default regularising priors from the `brms` package were used throughout. Unless otherwise stated, the response variable in the regressions was modelled as being distributed normally. Predictor variables were considered to contribute meaningfully if their 95% Credible Intervals (CI) did not include zero[3]. Each statistical model was fit using `brms` by running 4 chains of the "no U-turn" sampling (NUTS) of the posterior distribution for each parameter, with 4,000 samples per chain (the first 2,000 samples of which counting as "burn-in"); we inspected the chains to ensure convergence, and all $\hat{R}$ were close to 1.

Questionnaire Scores

There was good variability and ranges of scores across all three experiments for both the QIDS (Experiment 1: range = 0–25, M = 11.27, SD = 6.12; Experiment 2: range = 0–24, mean = 9.27, SD = 5.26; Experiment 3: range = 0–24, mean = 9.15, SD = 5.31) and the SHPS scores (Experiment 1: range = 14–50, mean = 27.91, SD = 7.69; Experiment 2: range = 14–44, mean = 28.02, SD = 6.25; Experiment 3: range = 14–51, mean = 27.08, SD = 6.76). Bayesian regressions showed that QIDS scores were positively predicted by SHPS scores across all three experiments, indicating good convergence of the measures (Experiment 1: $\beta_{Intercept} = 7.63$, 95%CI[6.82, 8.43], $\beta_{SHPS} = 1.11$ [0.95, 1.28]; Experiment 2: $\beta_{Intercept} = -5.29$ [−7.37, −3.16], $\beta_{SHPS} = 0.52$ [0.44, 0.59]; Experiment 3: $\beta_{Intercept} = -4.22$ [−6.08, −2.33], $\beta_{SHPS} = 0.49$ [0.43, 0.56])[4].

In the study of Dillon et al. (2015), participants were included in the "depressed" condition if they had—together with meeting other diagnostic criteria—a QIDS score of 14 or more (which reflects moderate depression); inclusion in the control group required a QIDS score lower than 8. In our samples, there were 109, 67, and 70 who scored above 14 in the QIDS in Experiments 1–3 respectively; these numbers were 88, 122, and 137 for those scoring below 8. Also, our sample contained a reasonable proportion of participants who self-declared a clinical diagnosis of depression (see Table 2). Thus our samples contained a good spread of depression symptomology.

Behavioural Results

Before presenting the results of the computational modelling, we provide an overview of the behavioural results to explore the magnitude of the flanker effect in both RT and error, as well as the relationship between these outcomes and questionnaire scores.

---

[3]Note that Bayesian credible intervals are not to be conflated with frequentist confidence intervals. Unlike the confidence interval, we can make probabilistic statements about the true population value for the predictor variable. If a 95% credible interval does not contain zero, we can state there is a 95% probability that the true population predictor value is not zero.

[4]See Online Supplementary Material Appendix B for plots of these regressions.

Flanker effect in RT & accuracy. For the RT analysis, a Bayesian regression was conducted with trial-level RT as the outcome variable and *congruency* as the predictor variable; the outcome variable was modelled as an ex-Gaussian distribution[5]. There was a large flanker effect in RT in all three experiments (Experiment 1: $\beta_{incongruent} = 48$ [37, 59]; Experiment 2: $\beta_{incongruent} = 53$ [44, 63]; Experiment 3: $\beta_{incongruent} = 30$ [21, 39]). The accuracy analysis utilised a similar Bayesian regression, but the outcome variable (proportion correct) was modelled as a skewed normal distribution. Again, there was a large flanker effect in all three experiments (Experiment 1: $\beta_{incongruent} = -0.01$ [-0.014, -0.006]; Experiment 2: $\beta_{incongruent} = -0.01$ [-0.014, -0.007]; Experiment 3: $\beta_{incongruent}$ = -0.01 [-0.015, -0.008]).

Predicting behavioural data from QIDS and SHPS. We performed a series of Bayesian regressions to explore whether the behavioural flanker effect was predicted by scores on the QIDS and SHPS. Results showed that the RT flanker effect was not predicted by QIDS scores (Experiment 1: $\beta_{QIDS} = 0.02$ [-0.53, 0.58]; Experiment 2: $\beta_{QIDS} = 0.23$ [-0.41, 0.89], Experiment 3: $\beta_{QIDS} = 0.11$ [-0.22, 0.44]) or SHPS scores (Experiment 1: $\beta_{SHPS} = -0.65$ [-1.67, 0.40]; Experiment 2: $\beta_{SHPS} = -0.17$ [-0.71, 0.38], Experiment 3: $\beta_{SHPS} = -$-0.04 [-0.29, 0.21]). The accuracy flanker effect was also not predicted by either QIDS (Experiment 1: $\beta_{QIDS} = -0.0003$ [-0.0010, 0.0005]; Experiment 2: β_{QIDS} = -0.0003 [-0.0009, 0.0003]], Experiment 3: $\beta_{QIDS} = 0.0001$ [-0.0004, 0.0007]) or SHPS scores (Experiment 1: $\beta_{QIDS} = -0.0003$ [-0.0015, 0.0009]; Experiment 2: $\beta_{QIDS} = 0.0000$ [-0.0005, 0.0004]], Experiment 3: $\beta_{QIDS} = 0.0002$ [-0.0001, 0.0006]). See online supplementary material Appendix C for more details.

Modelling Results

We fit the DSTP and SSP models to individual participant data using the `flankr` package (Grange, 2016). The models were fit to trial-level RT and accuracy data. Full details of the fit routine and assessment of the goodness of fits (which were good) for both models can be found in online supplementary material Appendix D. We also report in Appendix D formal model comparison tests of whether one model provided superior fits over the other. We found that the SSP model was superior for Experiments 1 and 2, and the DSTP model was superior for Experiment 3.

To assess whether model parameters were predicted by QIDS and SHPS scores, a series of Bayesian regressions were conducted; the outcome variables in all regressions were modelled as a skewed normal distribution. For ease of exposition, we present the regression coefficients for all models and all Experiments in Table 3; plots of all Bayesian regressions can be found in the Online Supplementary Material Appendix E.

DSTP model. No DSTP parameter was consistently predicted by depression scores (QIDS) across the three experiments. Although the parameter μ_{TA} was negatively associated with depression symptoms in Experiment 1, this did not replicate in Experiments 2 or 3. Likewise, although the parameter μ_{SS} was negatively associated with depression in Experiment 3, we did not find this to be the case in the first two Experiments. However,

---

[5]Response times are not normally distributed (i.e., they are non-Gaussian), and instead are positively skewed. Ex-Gaussian distributions are convolutions of a Gaussian distribution and an exponential distribution (which—together—produce positively skewed distributions), and have been shown to capture response time distributions well.
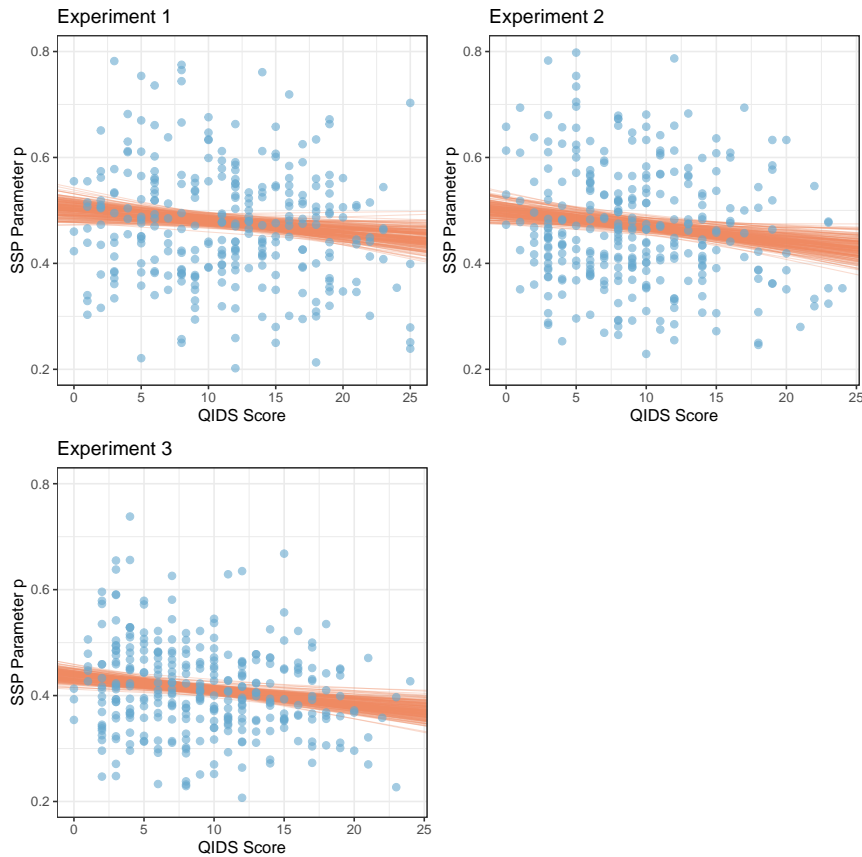
Figure 2. Relationship between the SSP model parameter p and measure of depression from the QIDS questionnaire for Experiments 1–3. Points represent participant data; the lines represent 200 draws from the posterior distribution for the statistical models Experiments 1–3.)

this parameter was negatively associated with anhedonia (SHPS) in Experiments 2 and 3, but this was not evident in Experiment 1. This suggests that the time taken for the cognitive system to select the central target for further processing—in the model, reflected by the drift rate shown in the lower panel of the DSTP schematic shown in Figure 1—is longer for those with higher levels of anhedonia. Parameter μ_{FL} was negatively associated with anhedonia in Experiment 2, but not in Experiments 1 or 3.

SSP model. We found that the SSP model parameter p was consistently (negatively) predicted by depression (QIDS) across all three experiments (the 95%CI for this regression coefficient in Experiment 1 only just included zero, $\beta_{QIDS} = -0.002$ [-0.0043, 0.0001]). This relationship is plotted in Figure 2. This parameter—which reflects the perceptual input strength of each element of the encoded flanker stimulus—was lower in individuals with higher levels of depression symptoms. This parameter was also negatively predicted by measures of anhedonia in Experiments 2 and 3, but not in Experiment 1. No other SSP parameters were predicted by QIDS or SHPS scores.

Table 3

Regression coefficients for the series of Bayesian regressions predicting model parameters (for both the DSTP and the SSP models) from QIDS and SHPS questionnaire scores across all Experiments. The values represent the mean estimate of the posterior distribution for each coefficient, together with their 95% credible interval in square parentheses. Entries in bold face represent regression predictors whose credible interval does not include zero (indicating the presence of an effect).

Experiment	Model	Parameter	QIDS $\beta_{Intercept}$	β_{QIDS}	SHPS $\beta_{Intercept}$	β_{SHPS}
E1	DSTP	A	0.105 [0.097, 0.112]	-0.0001 [-0.0006, 0.0004]	0.106 [0.094, 0.117]	-0.0001 [-0.0005, 0.0003]
		C	0.088 [0.081, 0.096]	0.0004 [-0.0002, 0.0009]	0.087 [0.074, 0.100]	0.0002 [-0.0002, 0.0006]
		μ_{TA}	0.159 [0.144, 0.174]	**-0.0013 [-0.0025, -0.0002]**	0.155 [0.128, 0.181]	-0.0004 [-0.0013, 0.0005]
		μ_{FL}	0.200 [0.179, 0.222]	-0.0007 [-0.0022, 0.0008]	0.212 [0.176, 0.247]	-0.0007 [-0.0019, 0.0005]
		μ_{SS}	0.539 [0.508, 0.570]	-0.0010 [-0.0034, 0.0013]	0.529 [0.472, 0.586]	0.0000 [-0.0020, 0.0019]
		μ_{RS2}	1.240 [1.178, 1.301]	0.0008 [-0.0039, 0.0054]	1.210 [1.102, 1.321]	0.0014 [-0.0025, 0.0051]
		ter	0.288 [0.277, 0.299]	0.0000 [-0.0008, 0.0009]	0.297 [0.277, 0.316]	-0.0003 [-0.0010, 0.0004]
	SSP	A	0.071 [0.067, 0.075]	-0.0001 [-0.0004, 0.0002]	0.072 [0.065, 0.078]	-0.0001 [-0.0003, 0.0002]
		ter	0.345 [0.333, 0.357]	0.0001 [-0.0009, 0.0010]	0.356 [0.334, 0.378]	-0.0004 [-0.0011, 0.0004]
		p	0.502 [0.474, 0.530]	-0.0021 [-0.0043, 0.0001]	0.489 [0.439, 0.539]	-0.0004 [-0.0021, 0.0013]
		rd	0.036 [0.033, 0.040]	-0.0002 [-0.0004, 0.0000]	0.037 [0.031, 0.042]	-0.0001 [-0.0003, 0.0001]
		sda	1.771 [1.692, 1.849]	-0.0023 [-0.0083, 0.0038]	1.807 [1.666, 1.947]	-0.0022 [-0.0069, 0.0026]
E2	DSTP	A	0.104 [0.098, 0.110]	-0.0004 [-0.0009, 0.0000]	0.109 [0.098, 0.120]	-0.0003 [-0.0007, 0.0001]
		C	0.088 [0.081, 0.096]	0.0003 [-0.0002, 0.0008]	0.089 [0.076, 0.101]	0.0001 [-0.0004, 0.0005]
		μ_{TA}	0.148 [0.132, 0.164]	0.0002 [-0.0013, 0.0017]	0.165 [0.131, 0.199]	-0.0005, [-0.0017, 0.0007]
		μ_{FL}	0.206 [0.187, 0.223]	-0.0014 [-0.0031, 0.0003]	0.248 [0.203, 0.291]	**-0.0020 [-0.0035, -0.0004]**
		μ_{SS}	0.539 [0.508, 0.570]	-0.0024 [-0.0052, 0.0004]	0.587 [0.520, 0.654]	**-0.0025 [-0.0049, -0.0002]**
		μ_{RS2}	1.207 [1.151, 1.263]	0.0040 [-0.0013, 0.0093]	1.185 [1.060, 1.313]	0.0021 [-0.0024, 0.0064]
		ter	0.293 [0.285, 0.301]	-0.0003 [-0.0011, 0.0005]	0.290 [0.271, 0.308]	0.0000 [-0.0006, 0.0007]
	SSP	A	0.072 [0.069, 0.076]	-0.0002 [-0.0005, 0.0001]	0.072 [0.065, 0.079]	-0.0001 [-0.0003, 0.0002]
		ter	0.352 [0.344, 0.360]	**-0.0008 [-0.0016, -0.0001]**	0.354 [0.335, 0.373]	-0.0003 [-0.0010, 0.0003]
		p	0.496 [0.473, 0.520]	**-0.0027 [-0.0049, -0.0006]**	0.545 [0.491, 0.601]	**-0.0026 [-0.0045, -0.0008]**
		rd	0.030 [0.027, 0.033]	-0.0001 [-0.0004, 0.0001]	0.030 [0.025, 0.035]	0.0000 [-0.0002, 0.0001]
		sda	1.731 [1.654, 1.806]	0.0012 [-0.0059, 0.0083]	1.736 [1.569, 1.909]	0.0002 [-0.0058, 0.0060]
E3	DSTP	A	0.098 [0.092, 0.104]	-0.0004 [-0.0008, 0.0001]	0.097 [0.084, 0.109]	-0.0001 [-0.0005, 0.0004]
		C	0.082 [0.077, 0.088]	0.0004 [-0.0001, 0.0009]	0.093 [0.082, 0.104]	-0.0003 [-0.0007, 0.0001]
		μ_{TA}	0.120 [0.108, 0.131]	-0.0005 [-0.0015, 0.0004]	0.119 [0.095, 0.140]	-0.0001, [-0.0009, 0.0007]
		μ_{FL}	0.117 [0.104, 0.129]	-0.0006 [-0.0017, 0.0005]	0.133 [0.108, 0.158]	-0.0008 [-0.0017, 0.0001]
		μ_{SS}	0.508 [0.487, 0.529]	**-0.0020 [-0.0040, -0.0001]**	0.546 [0.501, 0.589]	**-0.0021 [-0.0037, -0.0005]**
		μ_{RS2}	1.261 [1.206, 1.313]	-0.0001 [-0.0052, 0.0051]	1.326 [1.214, 1.436]	-0.0025 [-0.0064, 0.0016]
		ter	0.288 [0.279, 0.296]	-0.0003 [-0.0011, 0.0006]	0.281 [0.263, 0.300]	0.0001 [-0.0005, 0.0008]
	SSP	A	0.066 [0.0630, 0.069]	-0.0002 [-0.0005, 0.0001]	0.069 [0.062, 0.075]	-0.0002 [-0.0004, 0.0001]
		ter	0.344 [0.3352, 0.352]	0.0000 [-0.0008, 0.0008]	0.330 [0.312, 0.349]	0.0005 [-0.0002, 0.0012]
		p	0.437 [0.4176, 0.456]	**-0.0026 [-0.0044, -0.0008]**	0.453 [0.415, 0.492]	**-0.0015 [-0.0029, -0.0001]**
		rd	0.052 [0.0473, 0.057]	-0.0003 [-0.0007, 0.0001]	0.052 [0.042, 0.062]	-0.0001 [-0.0005, 0.0003]
		sda	1.802 [1.7212, 1.880]	-0.0030 [-0.0105, 0.0047]	1.794 [1.632, 1.958]	-0.0007 [-0.0067, 0.0052]

General Discussion

The aim of the current study was to extend the findings of Dillon et al. (2015) by fitting two computational models able to account jointly for accuracy and response time performance, as well as the ubiquitous improvement of attentional selectivity with time found in the flanker task. Specifically, we fit the dual-stage two-phase (DSTP) model (Hübner et al., 2010) and the Shrinking Spotlight (SSP) model (White et al., 2011) to data from 3 large-scale online experiments where self-reported measures of depression symptomology (as measured by the QIDS) and anhedonia (as measured by the SHPS) were recorded.

In terms of behavioural data, we did not find any relationship between depression

symptomology or anhedonia on any of the primary dependent variables at the mean level, in contrast to the findings of Dillon et al. (2015). However, in supplementary analysis (see online supplementary material Appendix F) where we grouped participants into high-QIDS (a score of 14 or above) and low-QIDS (a score below 8), group differences emerged in the conditional accuracy functions, with high-QIDS participants showing poorer attentional selectivity at faster responses. As the DSTP and SSP models are fit to conditional accuracy functions (combining RT and accuracy) and cumulative distribution functions (which describe the whole of the correct RT distribution), this analysis suggests that there were effects of depression that the modelling could be useful to describe.

No DSTP model parameter was consistently associated with depression or anhedonia across any of the three experiments. In contrast, however, the SSP model parameter p was consistently negatively associated with depression (and anhedonia in Experiments 2 & 3). This model parameter reflects the strength of the perceptual input of each item in the stimulus display (see Appendix A for more details), suggesting this input strength is weaker in those with higher levels of depression. In experimental validation of the SSP model, the p parameter has been shown to vary systematically with the intensity and contrast of presented visual stimuli (Servant, Montagnini, & Burle, 2014), suggesting it is able to index perceptual index strength accurately.

Do Perceptual Deficits Exacerbate Control Problems in Depression?

Although some evidence exists from psychophysical experiments that depression is associated with deficits in perceptual processing (e.g., Bubl, Kern, Ebert, Bach, & van Elst, 2010; Bubl, van Elst, Gondan, Ebert, & Greenle, 2009; Normann, Schmitz, Furmaier, Do-ing, & Bach, 2007)—and that depression can lead to poorer visual search times (Maekawa, Anderson, de Brecht, & Yamagishi, 2018)—to our knowledge our findings are the first to report perceptual deficits in depression captured by a computational model of higher-order cognition. This finding is of significance because the models capture how perceptual information and higher-order attentional processes interact to produce successful performance. Indeed, this is one of the many advantages of utilising computational models of cognition to probe clinical disorders (see for example White, Ratcliff, Vasey, & McKoon, 2010) because we can tease apart the effects of depression on higher-order processes and lower-order, subordinate, processes. The findings from the current study are striking as they suggest that depression symptomology is not negatively affecting attentional processes per se, but instead suggests depressed participants show deficits in a lower-order, subordinate, process of perceptual representation[6].

This raises the intriguing hypothesis—which should be the subject of future work—that the cognitive control deficits in depression found in recent studies (Burt et al., 1995; McDermott & Ebmeier, 2009; Rock et al., 2014; Snyder, 2013) are problems that are exacerbated by perceptual deficits: If individuals with depression have weaker perceptual representations, then the cognitive system has weaker information with which to work when

---

[6]A reviewer noted—quite rightly—that these results may also be compatible with the view that, in depression, an altered attentional process (e.g., increased attention to depressive ruminations versus external stimuli) may reduce the effectiveness of external perceptual stimuli globally. We are not able to resolve this issue with our data. However, we note that "attentional" parameters in both models (e.g., μ_{SS} in the DSTP model and rd in the SSP model) were not found to be associated with depression severity.

cognitive control is required, leading to less efficient control. Indeed, this hypothesis is not inconsistent with the findings of Dillon et al. (2015), who found significantly reduced prepotent response bias in the depressed group. This parameter reflects the negative influence of flankers on incongruent trials, which push the cognitive system towards an error response. Finding a reduction of this parameter in depressed participants could be interpreted as weaker perceptual representation of the flankers, which would mean they influence behaviour less than in control participants. In the extreme version of this hypothesis, perceptual deficits might exclusively explain control problems in depression.

Future work should combine computational modelling with experimental manipulation of perceptual properties of stimulus displays to test this hypothesis further. In addition, the emotional nature of the stimuli should be considered in future work. We used emotionally neutral stimuli. A substantial body of literature demonstrates that patients with depression show mood-congruent emotion processing bias; specifically, negative stimuli (e.g., negative faces) are processed more rapidly and deeply, whilst processing of positive stimuli seem to be impaired (e.g., Gotlib, Krasnoperova, Yue, & Joormann, 2004; Stuhrmann, Suslow, & Dannlowski, 2011). Furthermore, evidence indicates that these impairments in emotion processing can be reversed with psychotropic treatment (Fu et al., 2007). Future work should thus explore whether our general pattern of results replicate using emotional stimuli.

Limitations

The study has several limitations which should be considered. Firstly, due to the many model parameters, our analyses—and thus our conclusions—rest on the outcome of multiple comparisons, so we should be cautious of our Type-1 error rate. Whilst a valid concern, we believe that our main finding of reduced SSP model parameter p with depression severity is robust: It replicated across (pre-registered) experiments, across depressive symptom measures, and was also evident in the group analysis based on extreme QIDS scores (Appendix F).

Second, we used an online sample rather than using lab-condition testing. However, there is good evidence that online recruitment can provide quality data for cognitive tasks (Crump, McDonnell, & Gureckis, 2013). Our behavioural data quality was also very high (see online supplementary material Appendix C), suggesting that this potential limitation has not impacted on our conclusions.

Finally, we did not recruit a clinical sample; however, we had many participants in all three Experiments who scored above 14 on the QIDS, which Dillon et al. (2015) used as an inclusion criterion in their "depressed" group. Therefore, although not from a clinical sample we had a significant subset in each Experiment who exhibited moderate depression on the QIDS. As we show in online supplementary analysis Appendix F, when we compare this high-QIDS group with a low-QIDS group (QIDS score lower than 8 Dillon et al., 2015), our main conclusions of lower p SSP parameter values in higher depression stands. However, when we grouped participants into those who declared a clinical diagnosis of depression and those who declared no such diagnosis we found no consistent group differences. Although there are many advantages to conducting research on a large non-clinical sample (see for example the Research Domain Criteria set out by the NIMH: https://www.nimh.nih.gov/research/research-funded-by-

nimh/rdoc/index.shtml), further tests of the "perceptual-hypothesis" should be conducted on individuals with diagnosed major depressive disorder. Computational modelling will be essential to address this hypothesis.

References

Anwyl-Irvine, A. L., Massonnié, J., Flitton, A., Kirkham, N., & Evershed, J. K. (2019). Gorilla in our midst: An online behavioral experiment builder. *Behavior Research Methods*. doi: https://doi.org/10.3758/s13428-019-01237-x

Bubl, E., Kern, E., Ebert, D., Bach, M., & van Elst, L. T. (2010). Seeing gray when feeling blue? depression can be measured in the eye of the diseased. *Biological Psychiatry*, *68*, 205–208. doi: https://doi.org/10.1016/j.biopsych.2010.02.009

Bubl, E., van Elst, L. T., Gondan, M., Ebert, D., & Greenle, M. W. (2009). Vision in depressive disorder. *The World Journal of Biological Psychiatry*, *10*, 377–384. doi: https://doi.org/10.3109/15622970701513756

Bürkner, P.-C. (2017). brms: An R package for bayesian multilevel models using Stan. *Journal of Statistical Software*, *80*(1), 1–28. doi: 10.18637/jss.v080.i01

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*, 285–305. doi: https://doi.org/10.1037/0033-2909.117.2.285

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*, e57410. doi: https://doi.org/10.1371/journal.pone.0057410

Dillon, D. G., Wiecki, R., Pechtel, P., Webb, C., Goer, F., Murray, L., ... Pizzagalli, D. (2015). A computational analysis of flanker interference in depression. *Psychological Medicine*, *45*, 2333–2344. doi: http://dx.doi.org/10.1017/S0033291715000276

Eriksen, B. A., & Eriksen, C. W. (1974). Effects of noise leters upon the identification of a target letter in a nonsearch task. *Perception & Psychophysics*, *16*, 143–149. doi: https://doi.org/10.3758/BF03203267

Farrell, S., & Lewandowsky, S. (2010). Computational models as aids to better reasoning in psychology. *Currrent Directions in Psychological Science*, *19*, 329–335. doi: https://doi.org/10.1177/0963721410386677

Franken, I. H. A., Rassin, E., & Muris, P. (2007). The assessment of anhedonia in clinical and non-clinical populations: further validation of the Snaith–Hamilton Pleasure Scale (SHAPS). *Journal of Affective Disorders*, *99*, 83–89. doi: https://doi.org/10.1016/j.jad.2006.08.020

Fu, C. H. Y., Williams, S. C. R., Brammer, M. J., Suckling, J., Kim, J., Cleare, A. J., ... Bullmore, E. T. (2007). Neural responses to happy facial expressions in major depression following antidepressant treatment. *American Journal of Psychiatry*, *164*, 599–607.

Gotlib, I. H., Krasnoperova, E., Yue, D. N., & Joormann, J. (2004). Attentional biases for negative interpersonal stimuli in clinical depression. *Journal of Abnormal Psychology*, *113*, 127–135. doi: https://doi.org/10.1037/0021-843X.113.1.121

Grange, J. A. (2016). flankr: An R package implementing computational models of attentional selectivity. *Behavior Research Methods*, *48*, 528–541. doi: https://doi.org/10.3758/s13428-015-0615-y

Gratton, G., Coles, M. G. H., Sirevaag, E. J., & Eriksen, C. W. (1998). Pre- and poststimulus activation of response channels: A psychophysiological analysis. *Journal of Experimental Psychology: Human Perception & Performance*, *14*, 331–344. doi: https://dx.doi.org/10.1037//0096-1523.14.3.331

Hübner, R., Steinhauser, M., & Lehle, C. (2010). A dual-stage two-phase model of selective attention. *Psychological Review*, *117*, 759–784. doi: https://dx.doi.org/10.1037/a0019471

Levin, R. L., Heller, W., Mohanty, A., Herrington, J. D., & Miller, G. A. (2007). Cognitive deficits in depression and functional specificity of regional brain activity. *Cognitive Therapy & Research*, *31*, 211-233. doi: https://doi.org/10.1007/s10608-007-9128-z

Lewandowsky, S., & Farrell, S. (2010). *Computational modeling in cognition: Principles and practice.* Thousand Oaks, CA: Sage.

Logan, G. D., & Gordon, R. D. (2001). Executive control of visual attention in dual-task situations. *Psychological Review*, *108*, 393–434. doi: https://doi.org/10.1037/0033-295X.108.2.393

Maekawa, T., Anderson, S. J., de Brecht, M., & Yamagishi, N. (2018). The effect of mood state on visual search times for detecting a target in noise: An application of smartphone technology. *PLoS ONE*, *13*, e0195865. doi: https://doi.org/10.1371/journal.pone.0195865

McDermott, L. M., & Ebmeier, K. P. (2009). A meta-analysis of depression severity and cognitive function. *Journal of Affective Disorders*, *119*, 1–8. doi: https://doi.org/10.1016/j.jad.2009.04.022

Miyake, A., Friedman, N. P., Emerson, M. J., Witzkia, A. H., Howertera, A., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex "frontal lobe" tasks: A latent variable analysis. *Cognitive Psychology*, *41*, 49–100. doi: https://doi.org/10.1006/cogp.1999.0734

Noorani, I., & Carpenter, R. H. S. (2013). Antisaccades as decisions: LATER model predicts latency distributions and error responses. *European Journal of Neuroscience*, *37*, 330–338. doi: https://doi.org/10.1111/ejn.12025

Norman, D. A., & Shallice, T. (1986). Attention to action: willed and automatic control of behavior. In R. J. Davidson, G. E. Schwartz, & D. Shapiro (Eds.), *Consciousness and self-regulation: advances in research and theory. (Volume 4.).* Plenum: New York.

Normann, C., Schmitz, D., Furmaier, A., Doing, C., & Bach, M. (2007). Long-term plasticity of visually evoked potentials in humans is altered in major depression. *Biological Psychiatry*, *62*, 373–380. doi: https://doi.org/10.1016/j.biopsych.2006.10.006

R Core Team. (2017). R: A language and environment for statistical computing [Computer software manual]. Vienna, Austria. 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*, 2029–2040. doi: 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*, 573–583.

Servant, M., Montagnini, A., & Burle, B. (2014). Conflict tasks and the diffusion framework: Insight in model constraints based on psychological laws. *Cognitive Psychology*, *72*, 162–195. doi: https://doi.org/10.1016/j.cogpsych.2014.03.002

Snaith, R. P., Hamildon, M, Morley, S., Humayan, A., Hargreaves, D., & Trigwell, P. (1995). A scale for the assessment of hedonic tone: The Snaith–Hamilton Pleasure Scale. *British Journal of Psychiatry*, *167*, 99–103. doi: 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*, 81–132. doi: https://doi.org/10.1037/a0028727

Struss, D. T., & Knight, R. T. (Eds.). (2002). *Principles of frontal lobe function.* Oxford University Press.

Stuhrmann, A., Suslow, T., & Dannlowski, U. (2011). Facial emotion processing in major depression: a systematic review of neuroimaging findings. *Biology of Mood and Anxiety Disorders*, *1*, 10. doi: https://doi.org/10.1186/2045-5380-1-10

Wagenmakers, E.-J., van der Maas, H. L. J., & Grassman, R. P. P. P. (2007). An EZ-diffusion model for response time and accuracy. *Psychonomic Bulletin & Review*, *14*, 3–22. doi: https://doi.org/10.3758/BF03194023

White, C. N., Ratcliff, R., & Starns, J. J. (2011). Diffusion models of the flanker task: Discrete versus gradual attentional selection. *Cognitive Psychology*, *63*, 210–238. doi: https://doi.org/10.1016/j.cogpsych.2011.08.001

White, C. N., Ratcliff, R., Vasey, M. W., & McKoon, G. (2010). Using diffusion models to understand

clinical disorders. *Journal of Mathematical Psychology*, *54*, 39–52. doi: https://doi.org/10.1016/j.jmp.2010.01.004

Wickelgren, W. A. (1977). Speed–accuracy tradeoff and information processing dynamics. *Acta Psychologica*, *41*, 67–85. doi: https://doi.org/10.1016/0001-6918(77)90012-9

Appendix A

Model Details

This section provides a more comprehensive account of the dual-stage two phase (DSTP) model (Hübner et al., 2010) and the shrinking spotlight (SSP) model (White et al., 2011). We describe how each model accounts for the improvement of attentional selectivity with time, as well as a more detailed overview of the latent parameters estimated by the model fitting routine.

DSTP Model

The model assumes that response selection proceeds according to a drift diffusion process, which is shown schematically in Figure A1. After perceptual encoding of the stimulus, noisy evidence is accumulated towards one of two response boundaries, one of which represents the correct response and the other represents the incorrect response (see upper panel of Figure A1). The height of the correct response boundary is determined by the model parameter A, and the height of the incorrect response boundary is set by the model parameter $B = -A$.

At early stages of processing, response selection is poor because it is influenced by both the central target and the flanker stimuli. Specifically, the drift rate—that is, the rate at which the diffusion process rises to one of the two response boundaries—is determined by the additive combination of model parameters μ_{TA} and mu_{FL} representing the contributions of both the target and the flanker stimuli to response selection. On congruent trials, this contribution from the flanker stimuli to the drift rate facilitates response speed and accuracy because the information is congruent with the desired response. However, on incongruent trials, the contribution of mu_{Fl} to the drift rate is detrimental. Specifically, on incongruent trials mu_{Fl} takes on a negative value, and is positive on congruent trials.

In parallel to the response selection process, late attentional processes work to select a single item from the stimulus display for more detailed processing. This stimulus selection phase is also modelled by a diffusion process (see the bottom panel of Figure A1), and thus the time it takes the cognitive system to select a stimulus for further processing is explicitly modelled. Evidence is accumulated in a noisy fashion towards one of two absorbing boundaries, the drift rate of which is determined by the model parameter μ_{SS}. If the upper boundary is reached by the diffusion process, it is assumed that the model has selected the central target for further processing; if the diffusion process reaches the lower panel it is assumed the model has erroneously selected one of the flankers for further processing. The height of the stimulus selection boundary representing a target selection is set to C; the height of the boundary representing selection of a flanker is set to $D = -C$.

If the stimulus selection process finishes before the response selection process, response selection enters its second stage, which is highly selective. Specifically, the drift rate for response selection in Stage 2 is determined solely by which stimulus was selected by the stimulus selection process. If the model selected the central target for processing, the drift rate for Stage 2 of the response selection process increases leading to a sharper rise of the diffusion process towards the correct response boundary (see the example in Stage 2 of the upper panel in Figure A1). If, however, the stimulus selection process erroneously selects one of the flankers for further processing, the drift rate for Stage 2 of the response selection process becomes negative, meaning the diffusion process will march rapidly towards the
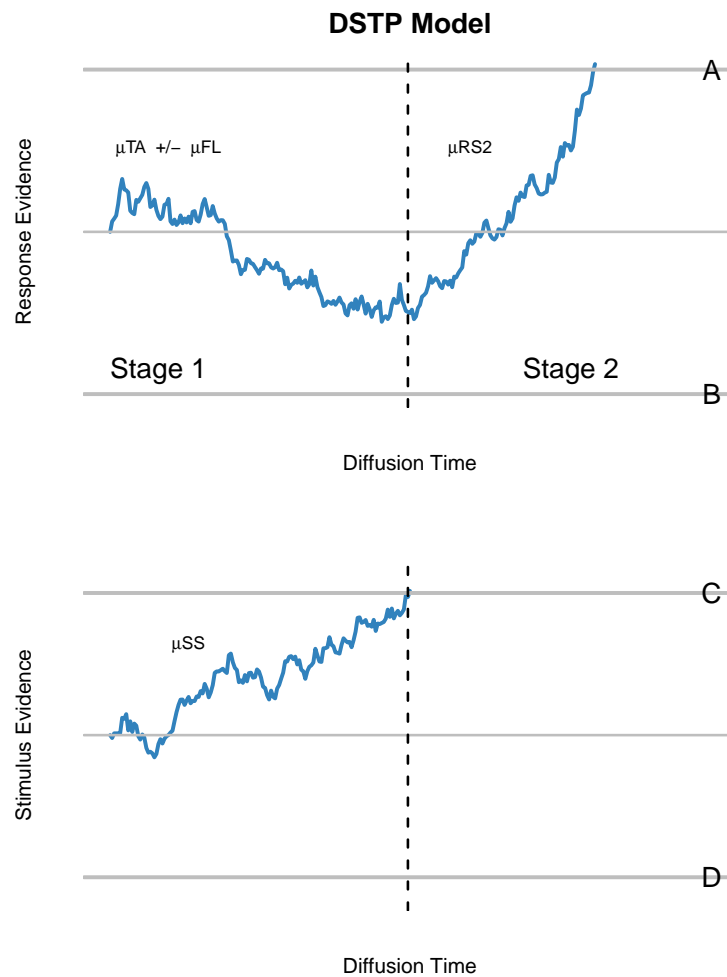
Figure A1. Schematic overview of the parallel drift diffusion processes in the Dual Stage Two Phase (DSTP) model for one trial. The upper panel depicts the diffusion process for response selection, and the lower panel depicts the diffusion process for stimulus selection. The point in time that a stimulus is selected for further processing is depicted by the dashed vertical line. The diffusion process before this point represents the first stage of response selection, and after this point represents the second—more selective—stage of response selection. See text for details.

incorrect response boundary. The drift rate for this second stage of response selection is determined by the model parameter μ_{RS2}, which is set to positive if the target is selected, and negative if a flanker is selected.

The final parameter in the DSTP model is *ter*, which captures all non-decisional components of response time, such as the time taken for motoric responding.

SSP Model

The shrinking spotlight model also assumes that response selection proceeds according to a drift diffusion process (see upper panel of Figure A2), with the same absorbing response boundaries as the DSTP model. The height of the response boundary for the correct response is represented by the model parameter A, and the height of the incorrect response boundary is set to $B = -A$.

Attentional selectivity increases with time, and this increased selectivity leads to an increase in the drift rate of the diffusion process across time—$v(t)$—as seen in Figure A2. The drift rate is determined by the strength of the perceptual input of each stimulus item in the display (i.e., target and flankers) multiplied by the proportion of attention currently being paid to each element in the display. As time progresses, more attention is paid to the central target stimulus, which means that as time progresses the drift rate is determined more by the target and less by the flanker stimuli; the net effect is an increase in attentional selectivity with time.

This change in attentional selectivity is shown schematically in the lower panel of Figure A2. The left plot in the lower panel of Figure A2 shows the distribution of attention across the stimulus display at early stages of processing. The distribution of attentional focus (i.e., the attentional spotlight) is modelled by a normal distribution centered on the target stimulus. The height of the distribution models how much attention is paid to each element in the display. At early stages of processing, a non-trivial amount of attention is paid to the flankers. As time progresses (from left to right in the Figure), the width of this attentional distribution narrows, meaning more attention is paid to the central target.

Formally, this reduction of the width of the attentional distribution is modelled by a reduction of the standard deviation of the normal distribution across time; the standard deviation of the attentional distribution at time t, $sd_a(t)$ is given by

$$sd_a(t) = sd_a - r_d t, \tag{1}$$

where the model parameter r_d captures the rate of the reduction at time t; larger values of r_d thus reflect faster focusing on the central target.

Given the above assumptions, and assuming each stimulus in the display is one unit wide, the total attention being paid to the outer flankers (a_{outer}), the inner flankers (a_{inner}), and the central target (a_{target}) at time t is given by
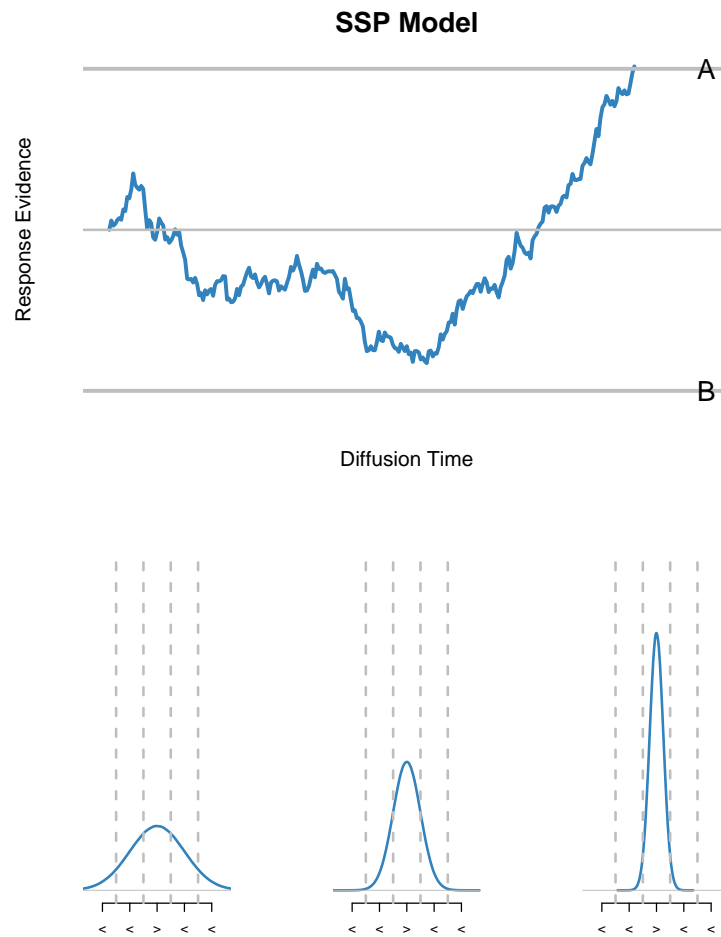
Figure A2. Schematic overview of the diffusion process for response selection in the shrinking spotlight (SSP) model. The upper panel depicts the diffusion process for response selection, the drift rate of which is determined by the perceptual input of stimuli that fill within the area of the attentional spotlight, depicted in the lower panel. As time progresses, the attentional spotlight becomes more focussed around the central target, meaning perceptual information from the flankers contribute less to the drift rate of response selection as time increases.

$$a_{outer}(t) = \int_{1.5}^{\infty} \phi[0, sd_a(t)];$$

$$a_{inner}(t) = \int_{0.5}^{1.5} \phi[0, sd_a(t)]; \tag{2}$$

$$a_{target}(t) = \int_{-0.5}^{0.5} \phi[0, sd_a(t)].$$

Here, ϕ is the density function for the normal distribution with mean 0 and standard deviation of $sd_a(t)$ (see Equation 1). As mentioned earlier, the drift rate for response selection is a multiplicative combination of the perceptual strength, p, of each element in the stimulus display and the amount of attention, a_x, currently being paid to each element. The perceptual strength parameter p for all items in a congruent stimulus is set to positive; for incongruent stimuli, only the central item takes a positive value for p and the flankers take on a value of $-p$.

The drift rate at time t, $v(t)$, is given by

$$v(t) = 2p_{outer}a_{outer}(t) + 2p_{inner}a_{inner}(t) + p_{target}a_{target}(t). \tag{3}$$

Gaussian noise is then added to this drift rate. The SSP model also has a non-decisional parameter ter which is interpreted identically to the DSTP model.

Appendix B

Questionnaire Analysis

This section presents the plots of the Bayesian regressions examining the relationship between QIDS and SHPS scores across all three experiments. As can be seen in Figures B1–B3 there was good variability and ranges of scores across all three experiments for both the QIDS and the SHPS scores. Bayesian regressions showed that QIDS scores were positively predicted by SHPS scores across all three experiments, indicating good convergence (see main text for more details).
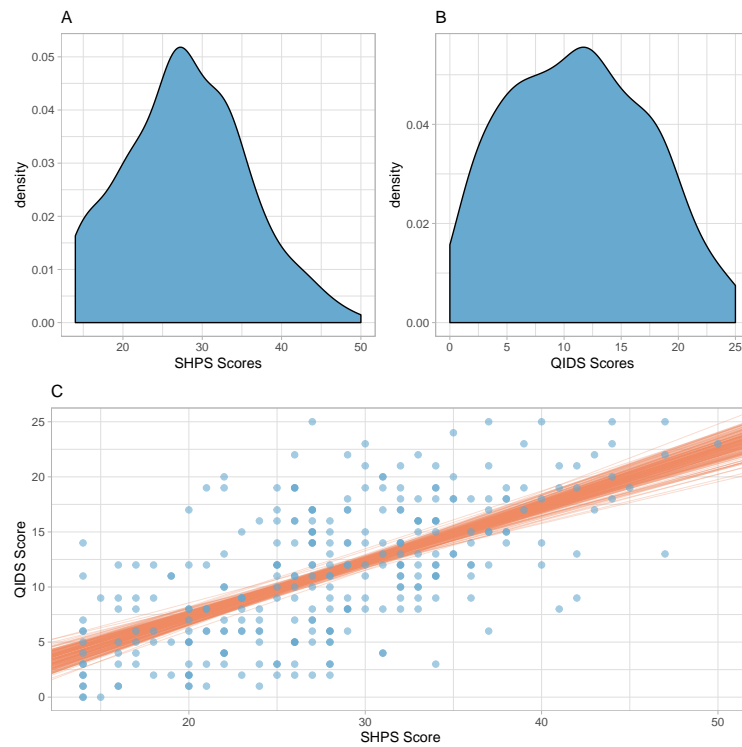
Figure B1. Experiment 1 data. A. Density distribution of scores on the SHPS questionnaire. B. Density distribution of scores on the QIDS questionnaire. C. Points show individual participants' QIDS score plotted against their SHPS score. The lines represent 200 draws from the posterior distribution of a Bayesian linear regression predicting QIDS scores from SHPS scores.
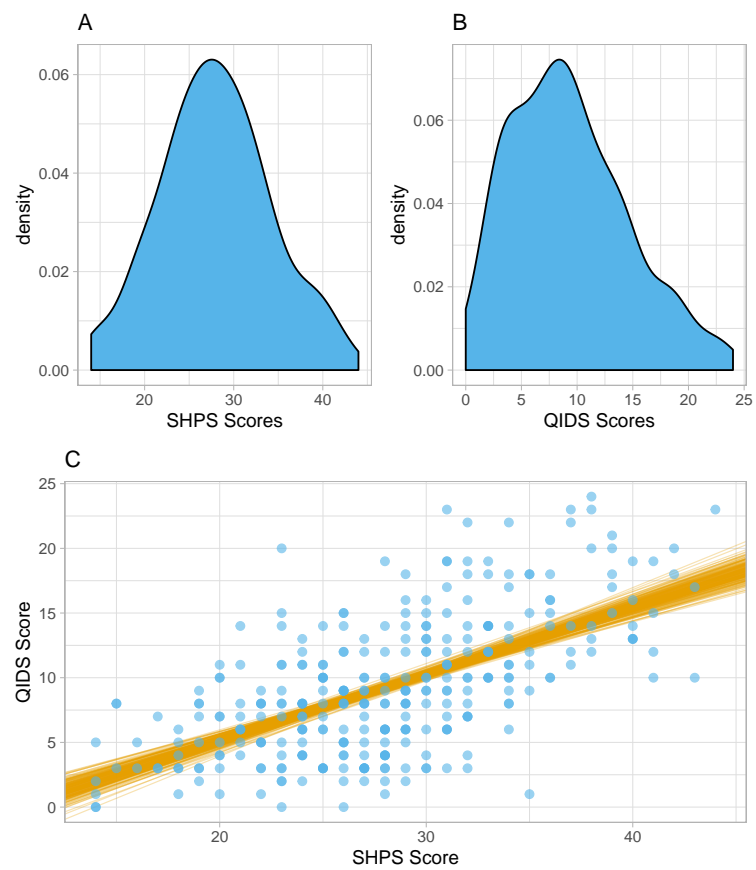
Figure B2. Experiment 2 data. A. Density distribution of scores on the SHPS questionnaire. B. Density distribution of scores on the QIDS questionnaire. C. Points show individual participants' QIDS score plotted against their SHPS score. The lines represent 200 draws from the posterior distribution of a Bayesian linear regression predicting QIDS scores from SHPS scores.
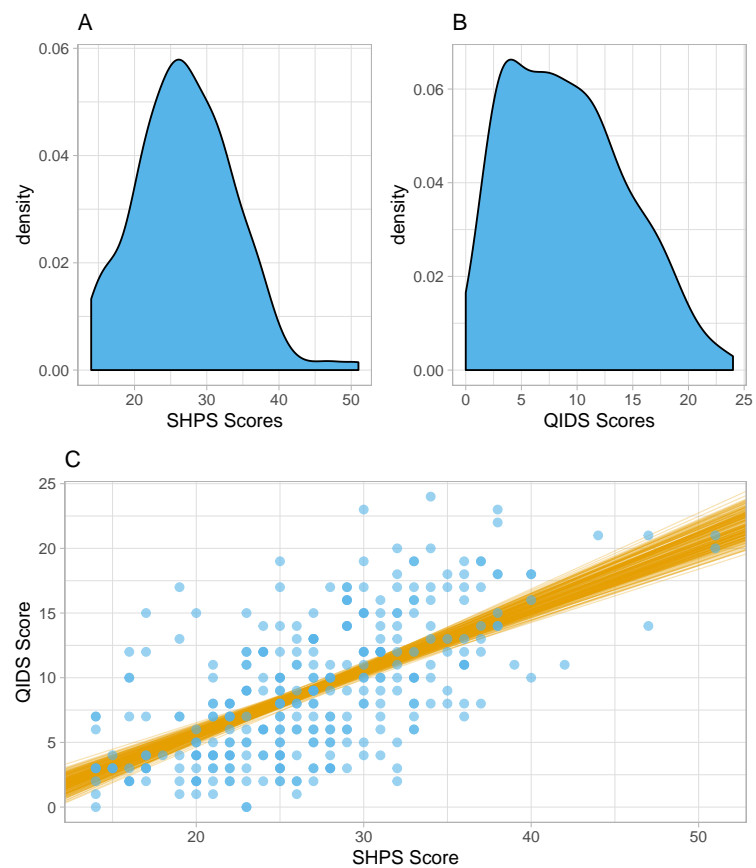
Figure B3. Experiment 3 data. A. Density distribution of scores on the SHPS questionnaire. B. Density distribution of scores on the QIDS questionnaire. C. Points show individual participants' QIDS score plotted against their SHPS score. The lines represent 200 draws from the posterior distribution of a Bayesian linear regression predicting QIDS scores from SHPS scores.

Appendix C

Expanded Detail of Behavioural Data Analysis

In this section we provide additional details of the analyses of the behavioural data, together with the regressions with questionnaire scores. Figures C1–C3 show density distributions of the behavioural data in Experiments 1–3 respectively for both response time and accuracy data. These plots demonstrate good data quality in that there is low variance in the RTs, and excellent accuracy performance.
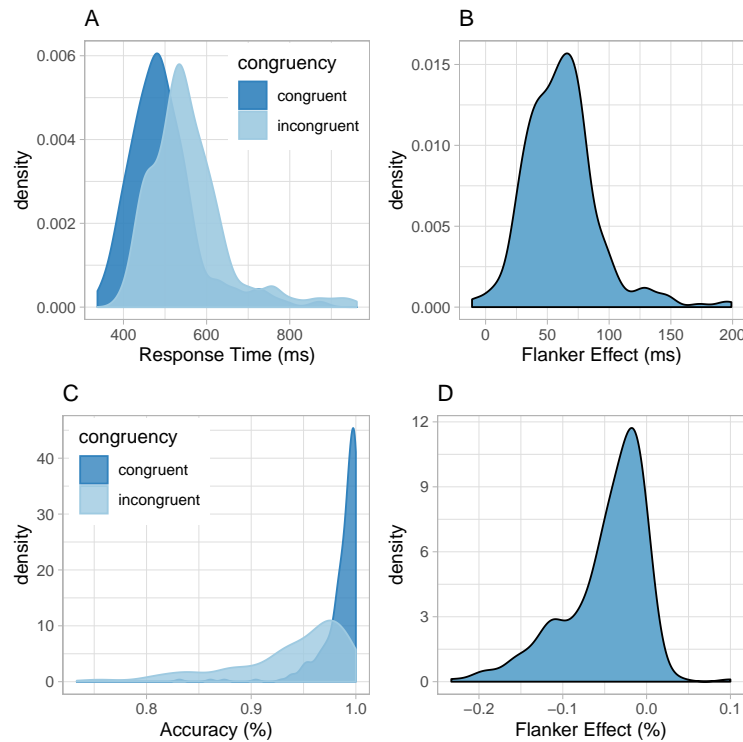
Figure C1. Density distributions of the behavioural data in Experiment 1, collating all trials across all subjects (i.e., density plots of the whole data). Panel A: Density distributions of congruent and incongruent response time (in milliseconds, ms). Panel B: Density distributions of the flanker effect for response time data (RT incongruent – RT congruent). Panel C: Density distribution of the congruent and incongruent accuracy. Panel D: Density distribution of the flanker effect in accuracy.

Table C1 shows all model parameters for the Bayesian regressions conducted predicting behavioural data from questionnaire scores. Note that in this table we also present regressions predicting mean RT and mean accuracy from the questionnaire scores, which is not reported in the main paper. Figures C4–C6 plot the relationship between flanker effects in RT and accuracy and scores on the QIDS and SHPS, together with draws from the posterior distribution of the Bayesian regression model fits.
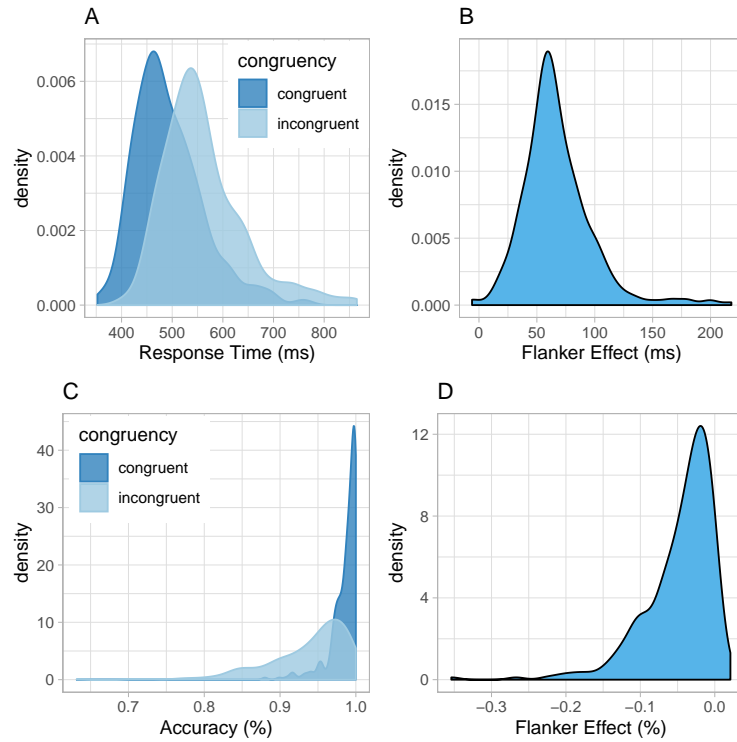
Figure C2. Density distributions of the behavioural data in Experiment 2, collating all trials across all subjects (i.e., density plots of the whole data). Panel A: Density distributions of congruent and incongruent response time (in milliseconds, ms). Panel B: Density distributions of the flanker effect for response time data (RT incongruent – RT congruent). Panel C: Density distribution of the congruent and incongruent accuracy. Panel D: Density distribution of the flanker effect in accuracy.

Table C1
Regression coefficients for the series of Bayesian regressions predicting behavioural dependent variables (DV) from QIDS and SHPS questionnaire scores across all Experiments. The values represent the mean estimate of the posterior distribution for each coefficient, together with their 95% credible interval in square parentheses.

		QIDS		SHPS	
Experiment	DV	$\beta_{Intercept}$	β_{QIDS}	$\beta_{Intercept}$	β_{SHPS}
E1	Mean RT	524.27 [506.95, 541.97]	-0.11 [-1.40, 1.20]	552.71 [523.62, 580.63]	-1.07 [-2.04, -0.09]
	Flanker RT	61.13 [53.77, 68.36]	0.02 [-0.53, 0.58]	63.44 [58.70, 68.17]	-0.65 [-1.67, 0.40]
	Mean Acc.	0.962 [0.957, 0.966]	-0.0002 [-0.0005, 0.0000]	0.959 [0.951, 0.968]	0.0000 [-0.0003, 0.0003]
	Flanker Acc.	-0.047 [-0.057, -0.037]	-0.0003 [-0.0010, 0.0005]	-0.049 [-0.056, -0.043]	-0.0003 [-0.0015, 0.0009]
E2	Mean RT	533.43 [519.16, 548.11]	-0.91 [-2.20, 0.36]	531.29 [501.78, 562.19]	-0.23 [-1.30, 0.80]
	Flanker RT	65.34 [58.31, 72.25]	0.23 [-0.41, 0.89]	72.23 [56.62, 87.76]	-0.17 [-0.71, 0.38]
	Mean Acc.	0.960 [0.956, 0.964]	-0.0001 [-0.0005, 0.0002]	0.960 [0.956, 0.968]	0.0000, [-0.0003, 0.0002]
	Flanker Acc.	-0.048 [-0.055, -0.041]	-0.0003 [-0.0009, 0.0003]	-0.051 [-0.065, -0.037]	0.0000 [-0.0005, 0.0004]
E3	Mean RT	512.69 [499.07, 526.09]	0.20 [-1.05, 1.44]	509.69 [480.57, 538.19]	0.18 [-0.84, 1.22]
	Flanker RT	34.41 [30.79, 37.84]	0.11 [-0.22, 0.44]	36.51 [29.58, 43.52]	-0.04 [-0.29, 0.21]
	Mean Acc.	0.961 [0.957, 0.965]	-0.0002 [-0.0006, 0.0001]	0.962 [0.955, 0.969]	-0.0001, [-0.0003, 0.0001]
	Flanker Acc.	-0.042 [-0.048, -0.037]	0.0001 [-0.0004, 0.0007]	-0.048 [-0.058, -0.037]	0.0002 [-0.0001, 0.0006]
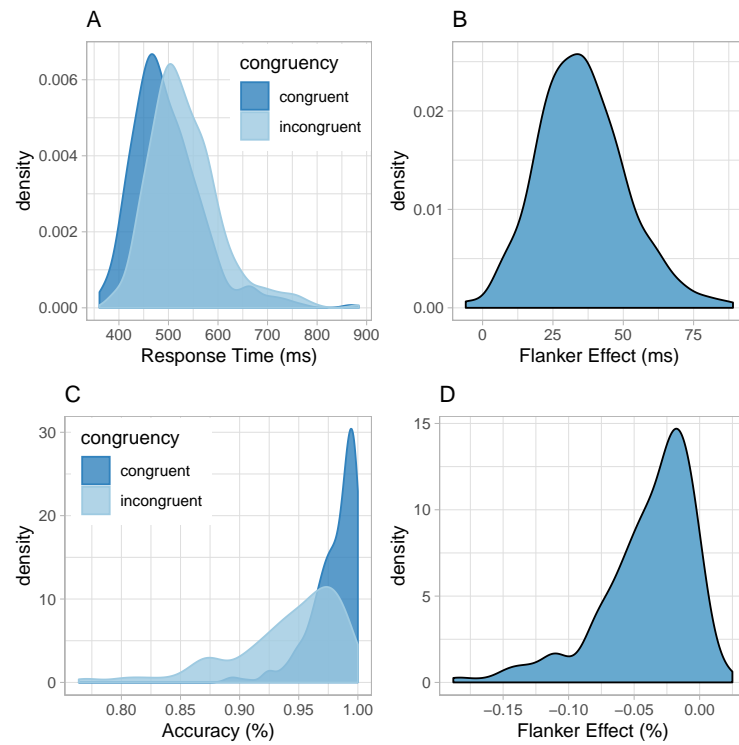
Figure C3. Density distributions of the behavioural data in Experiment 3, collating all trials across all subjects (i.e., density plots of the whole data). Panel A: Density distributions of congruent and incongruent response time (in milliseconds, ms). Panel B: Density distributions of the flanker effect for response time data (RT incongruent – RT congruent). Panel C: Density distribution of the congruent and incongruent accuracy. Panel D: Density distribution of the flanker effect in accuracy.
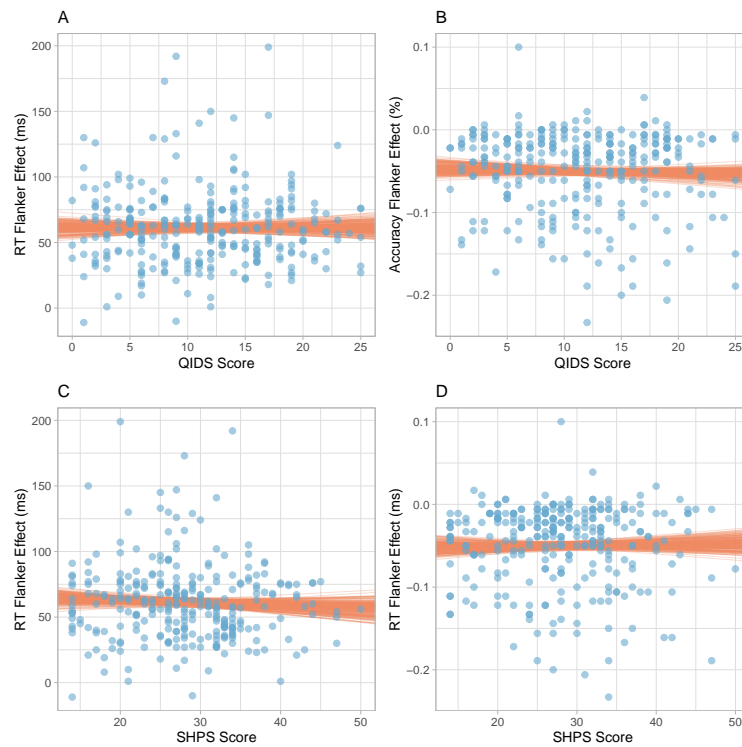
Figure C4. Behavioural flanker effects predicted from questionnaire scores in Experiment 1. Points represent individual participant data; lines represent 200 draws from the posterior distribution of the Bayesian regression model, showing credible estimates of the linear relationship. Panel A: Flanker effect in response time predicted from the QIDS questionnaire. Panel B: Flanker effect in accuracy predicted from the QIDS questionnaire. Panel C: Flanker effect in response time predicted from the SHPS questionnaire. Panel D: Flanker effect in accuracy predicted from the SHPS questionnaire.
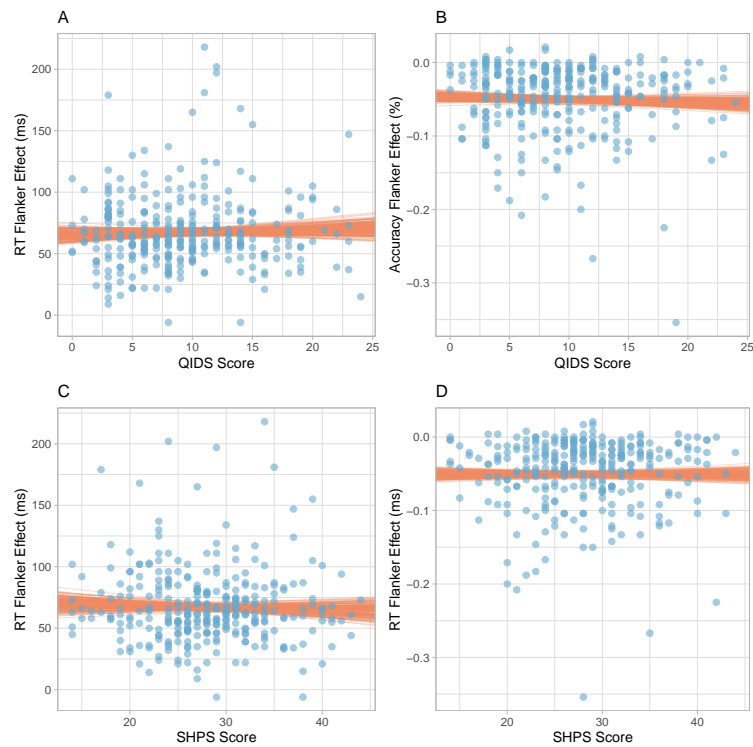
Figure C5. Behavioural flanker effects predicted from questionnaire scores in Experiment 2. Points represent individual participant data; lines represent 200 draws from the posterior distribution of the Bayesian regression model, showing credible estimates of the linear relationship. Panel A: Flanker effect in response time predicted from the QIDS questionnaire. Panel B: Flanker effect in accuracy predicted from the QIDS questionnaire. Panel C: Flanker effect in response time predicted from the SHPS questionnaire. Panel D: Flanker effect in accuracy predicted from the SHPS questionnaire.
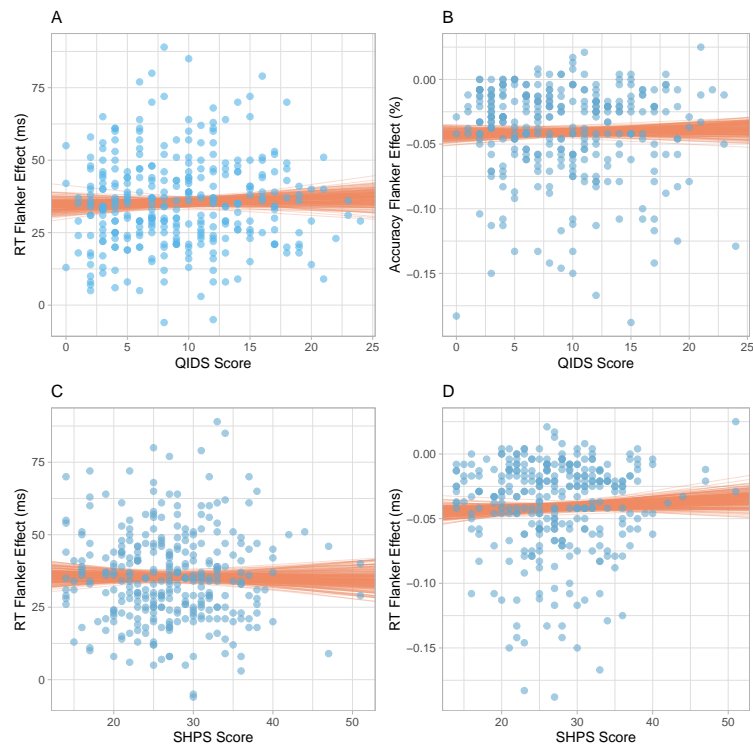
Figure C6. Behavioural flanker effects predicted from questionnaire scores in Experiment 3. Points represent individual participant data; lines represent 200 draws from the posterior distribution of the Bayesian regression model, showing credible estimates of the linear relationship. Panel A: Flanker effect in response time predicted from the QIDS questionnaire. Panel B: Flanker effect in accuracy predicted from the QIDS questionnaire. Panel C: Flanker effect in response time predicted from the SHPS questionnaire. Panel D: Flanker effect in accuracy predicted from the SHPS questionnaire.

Appendix D

Model Fitting and Goodness of Fit Assessment

This section provides more detail on how the computational models were fit to the partici-
pant data together with assessment of the goodness of fit to the behavioural data.

Description of Model Fit Routine

As described in Grange (2016), the models are fitted to response time and accuracy
distributions: specifically, the model is fitted to cumulative distribution functions (CDFs) of
correct response time, and conditional accuracy functions (CAFs), which combine accuracy
and response time information. CDFs and CAFs are created for each condition of congruency
separately, for each participant. CDFs are constructed by finding the response time cut-off
points for the 0.1, 0.3, 0.5, 0.7, and 0.9 quantiles of the correct response time distribution.
Conditional accuracy functions are generated by ordering the complete data (i.e., error and
correct trials) according to the response time (from fastest to slowest). This data is then
partitioned into four bins, each containing 25% of data. For each bin, mean response time
and percent accuracy is calculated. CAFs thus assess how accuracy changes with response
time.

The fit routine in `flankr` finds the set of parameters that generates simulated CDFs
and CAFs that match the participant data. The fit routine aims to minimsie the discrepancy
between the simulated and observed CDFs and CAFs, by aiming to minimise the likelihood
ratio chi-square statistic, G^2:

$$G^2 = 2 \sum_{i}^{J} N p_i ln \left(\frac{p_i}{\pi_i} \right). \tag{4}$$

In Equation 4, p_i is the proportion of observations in the ith bin (i.e., across CDFs
and CAFs) for participants, π_i is the proportion in this bin in the simulated model data,
N is the average number of trials (which is set to 250 in `flankr`, J is the total number of
bins, and ln is the natural logarithm.

The fit routine aims to find the best set of parameters that minimises the G^2 statistic.
Recall that the models were fit to each participant's data individually. In order to avoid
local minima, the fit routine occurred in two stages. In the first stage, we conducted a
broad search of the parameter space by starting the fit routine from 50 random starting
points. During this initial exploration, we simulated 1,000 data points from the model on
each iteration of the fit routine. The best-fitting parameters from this first stage were then
entered as the starting parameters in the final fitting stage, which simulated 50,000 data
points per iteration of the fit routine. The best-fitting parameters from this final stage were
stored as the best parameters for that particular participant. This procedure was used for
each participant individually for both models.

Goodness of Fit Assessment

The goodness of fit of the model to participant data was assessed visually using QQ-
plots of model predictions against participant data. Specifically, for each participant, for
each condition of congruency, and for each model, we simulated data using that participant's
best-fitting model parameters (simulating 50,000 trials); for both the model's simulated data

and the participant's data, we calculated total proportion accuracy, and the 25th, 50th, and 75th quantile of the correct response time distribution. These values were plotted against each other (on four separate plots). This process was then repeated for all participants.

The results of this procedure can be seen in Figures D1–D3 for Experiments 1–3 respectively. In these plots, each point shows the model's prediction plotted against the participant's data; if model predictions are perfect, all data points should lie across the dashed diagonal line. As can be seen, the model predictions (for both the DSTP and SSP) fit the participant data well across all 3 Experiments.

Comparing Model Fit

As both models were fit to individual participant data, we wanted to explore whether one particular model proved superior fits over the other. We therefore wanted to compare the fit statistic G^2 for the DSTP and SSP models across all participants. Recall that lower G^2 values indicate better fit, which can be useful for comparing the fit of two (or more) models; however, the DSTP has more parameters (7) than the SSP model (5), and—all else being equal—models with more parameters fit data better than models with fewer parameters.

We therefore used the Bayesian Information Criterion for binned data—bBIC—which includes a penalty term for the number of parameters in each model; this statistic provides a measure of the goodness of fit whilst controlling for the number of parameters in the model. Lower values of bBIC indicate superior fit; therefore, we wished to compare bBIC scores for the DSTP model and the SSP model across participants.

bBIC is given by

$$\text{bBIC} = -2\left(\sum_i^J Np_i ln(\pi_i)\right) + Mln(N), \tag{5}$$

where M is the number of parameters in the model; all other terms are equivalent to those in Equation 4.

We calculated bBIC for each participant for each model, and plotted them against each other (see the upper panel of Figures D4–D6). Those data points that are plotted above the diagonal represent participants whose data were better explained by the DSTP model, and those below the diagonal represent participants whose data were better explained by the SSP model. The lower panel of Figures D4–D6 is a different representation of the same data; it plots the difference in bBIC scores—delta bBIC—calculated by bBIC(DSTP) - bBIC(SSP). Postive values of delta bBIC represent superior fit of the SSP model, and negative values of bBIC represent superior fit of the DSTP model. These frequency plots show that the SSP was slightly superior across participants for Experiments 1 and 2, but the DSTP was superior across participants in Experiment 3.
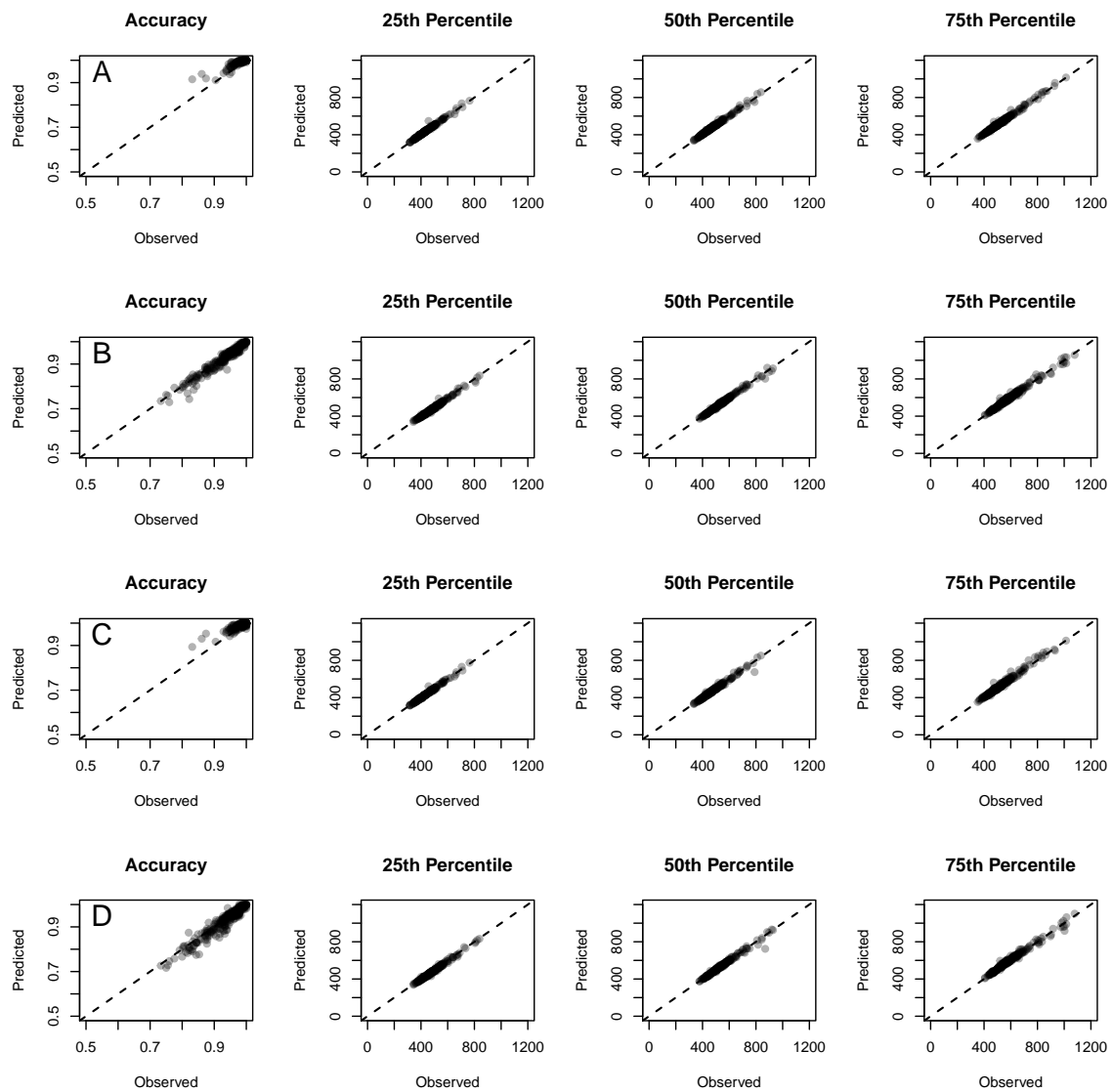
Figure D1. QQ-plots assessing goodness of fit for all models in Experiment 1. Row A: DSTP model, congruent data. Row B: DSTP model, incongruent data. Row C: SSP model, congruent data. Row D: SSP model, incongruent data. The diagonal dashed line represents the theoretical perfect fit.
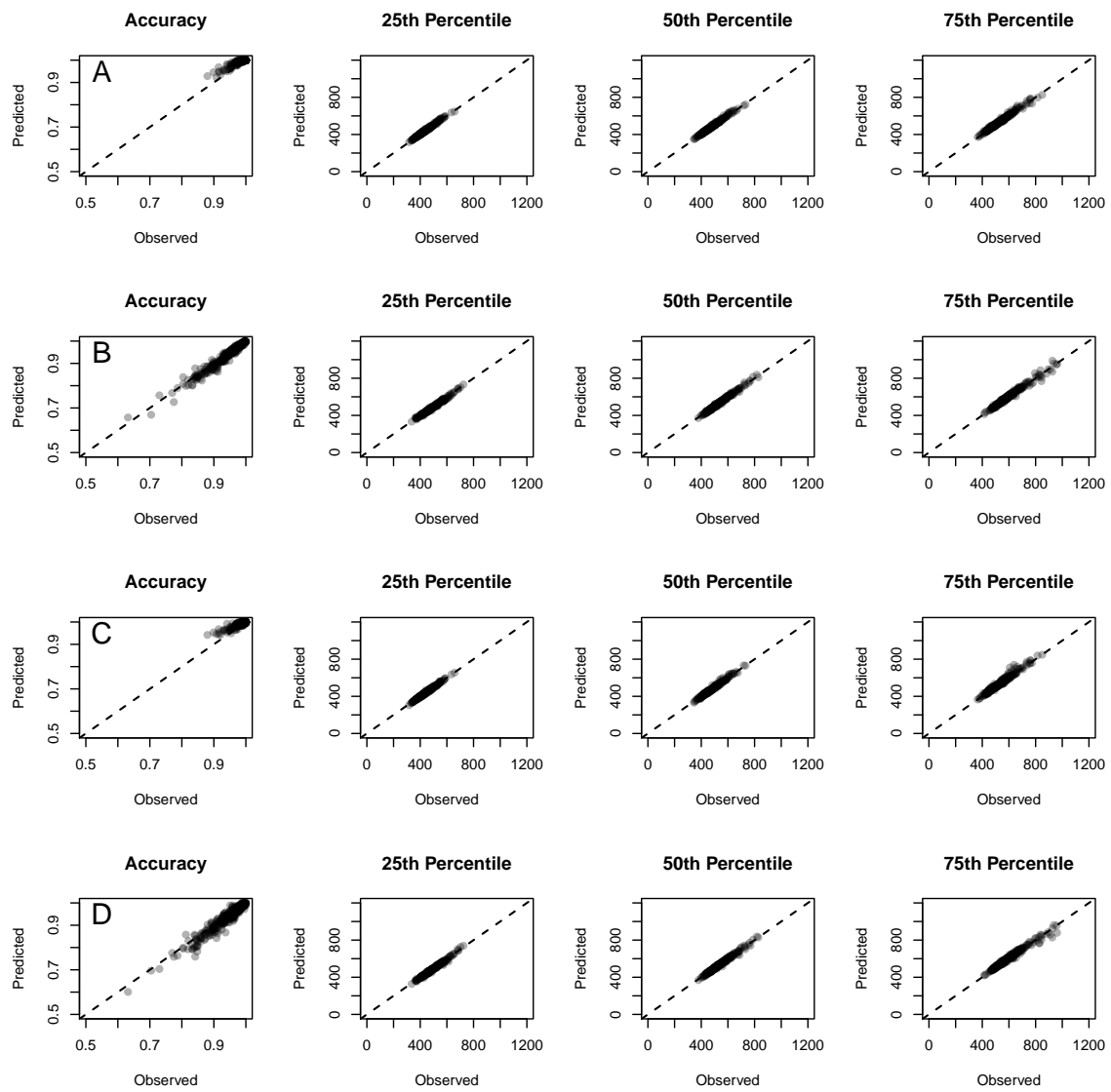
Figure D2. QQ-plots assessing goodness of fit for all models in Experiment 2. Row A: DSTP model, congruent data. Row B: DSTP model, incongruent data. Row C: SSP model, congruent data. Row D: SSP model, incongruent data. The diagonal dashed line represents the theoretical perfect fit.
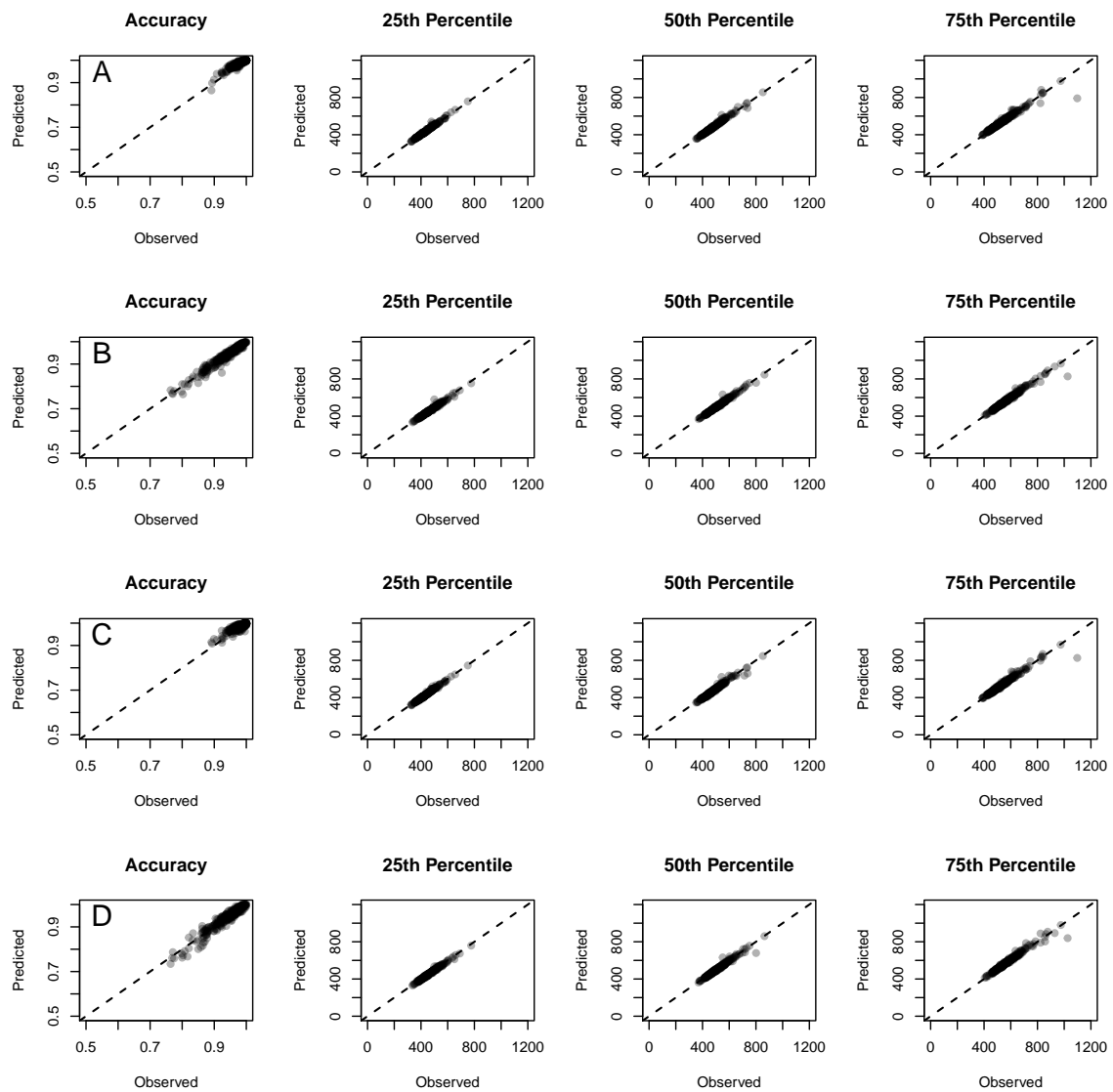
Figure D3. QQ-plots assessing goodness of fit for all models in Experiment 3. Row A: DSTP model, congruent data. Row B: DSTP model, incongruent data. Row C: SSP model, congruent data. Row D: SSP model, incongruent data. The diagonal dashed line represents the theoretical perfect fit.
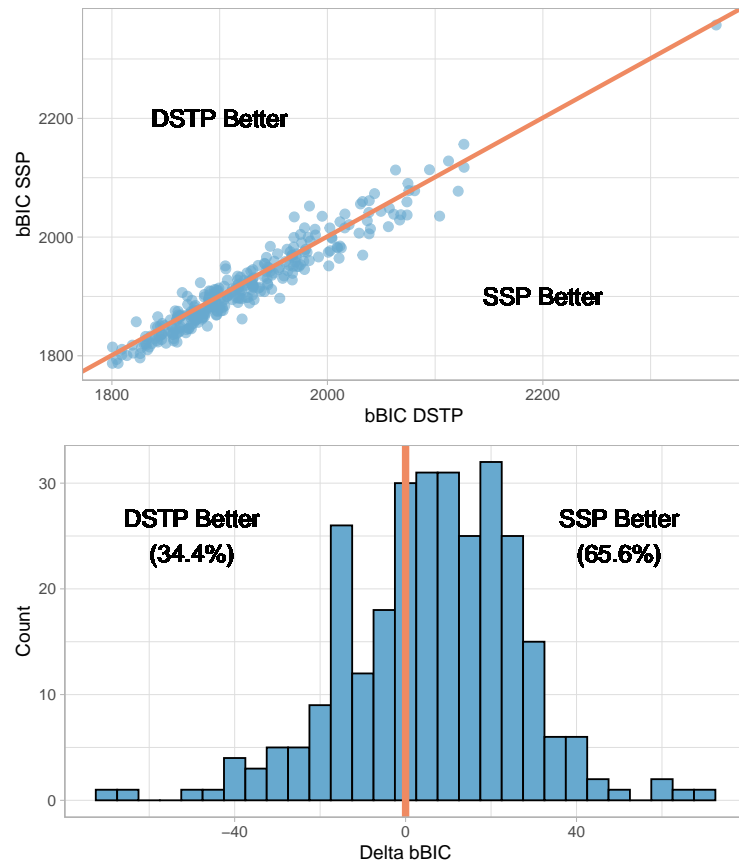
Figure D4. Representation of the assessment of model superiority across all participants in Experiment 1. Upper Panel: bBIC scores for the DSTP fit plotted against the bBIC scores for the SSP fit, across all participants. The diagonal line represents equivalence between both models. Data points below the diagonal represent participants for whom the SSP model fit their data than the DSTP model; the data points above the diagonal represent participants for whom the DSTP provided the superior fit. Lower Panel: The difference between bBIC scores for each model across participants (delta bBIC = bBIC[DSPT] – bBIC[SSP]). Positive values of delta bBIC indicate the SSP model is superior; negative values indicate superiority of the DSTP model. Percentages indicate the proportion of participants whose data were fit better by the respective model.

Figure D5. Representation of the assessment of model superiority across all participants in Experiment 2. Upper Panel: bBIC scores for the DSTP fit plotted against the bBIC scores for the SSP fit, across all participants. The diagonal line represents equivalence between both models. Data points below the diagonal represent participants for whom the SSP model fit their data than the DSTP model; the data points above the diagonal represent participants for whom the DSTP provided the superior fit. Lower Panel: The difference between bBIC scores for each model across participants (delta bBIC = bBIC[DSPT] – bBIC[SSP]). Positive values of delta bBIC indicate the SSP model is superior; negative values indicate superiority of the DSTP model. Percentages indicate the proportion of participants whose data were fit better by the respective model.

Figure D6. Representation of the assessment of model superiority across all participants in Experiment 3. Upper Panel: bBIC scores for the DSTP fit plotted against the bBIC scores for the SSP fit, across all participants. The diagonal line represents equivalence between both models. Data points below the diagonal represent participants for whom the SSP model fit their data than the DSTP model; the data points above the diagonal represent participants for whom the DSTP provided the superior fit. Lower Panel: The difference between bBIC scores for each model across participants (delta bBIC = bBIC[DSPT] – bBIC[SSP]). Positive values of delta bBIC indicate the SSP model is superior; negative values indicate superiority of the DSTP model. Percentages indicate the proportion of participants whose data were fit better by the respective model.

Appendix E

Model Regression Plots

This section presents plots of the Bayesian regressions predicting DSTP and SSP model parameters from QIDS and SHPS scores for Experiment 1–3.
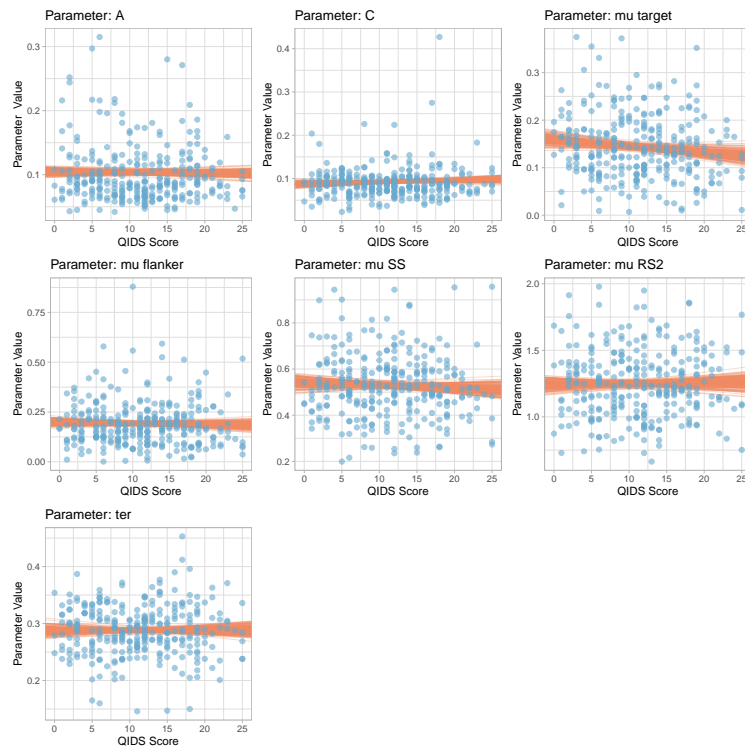
Figure E1. DSTP model parameters predicted from QIDS scores in Experiment 1. Points represent individual participant data; lines represent 200 draws from the posterior distribution of the Bayesian regression model, showing credible estimates of the linear relationship.
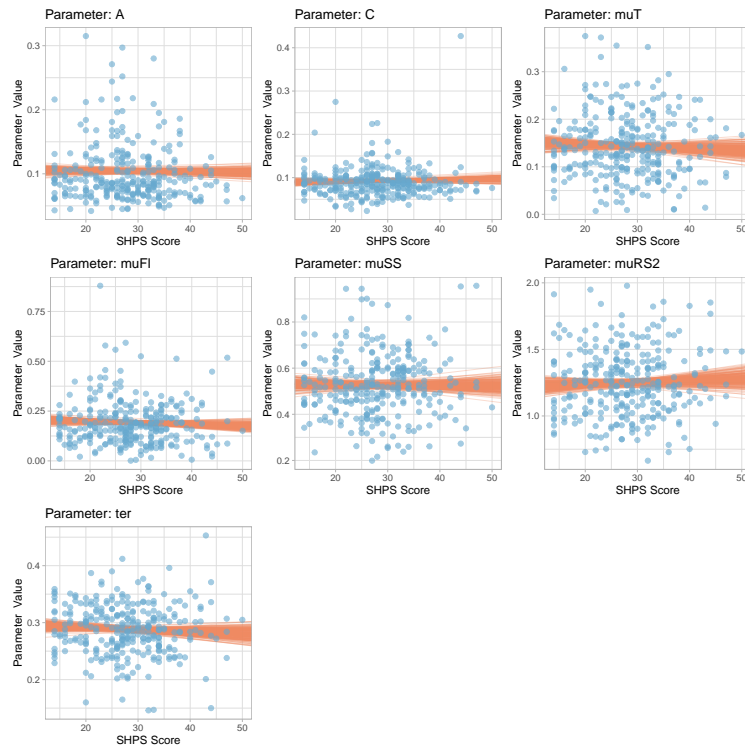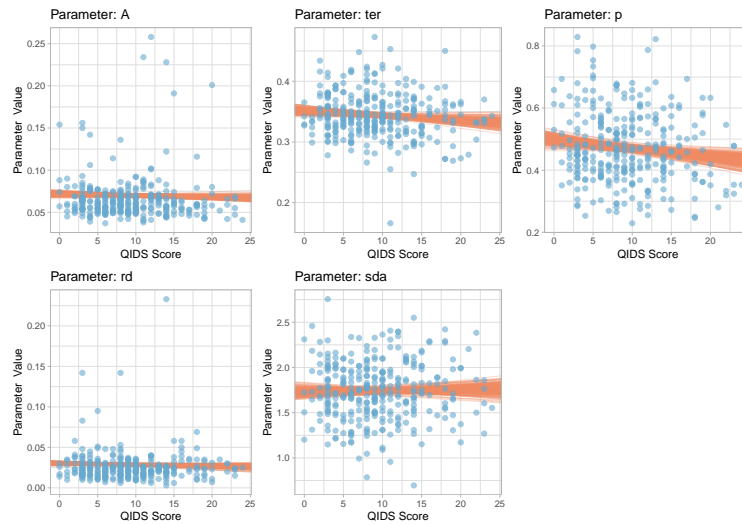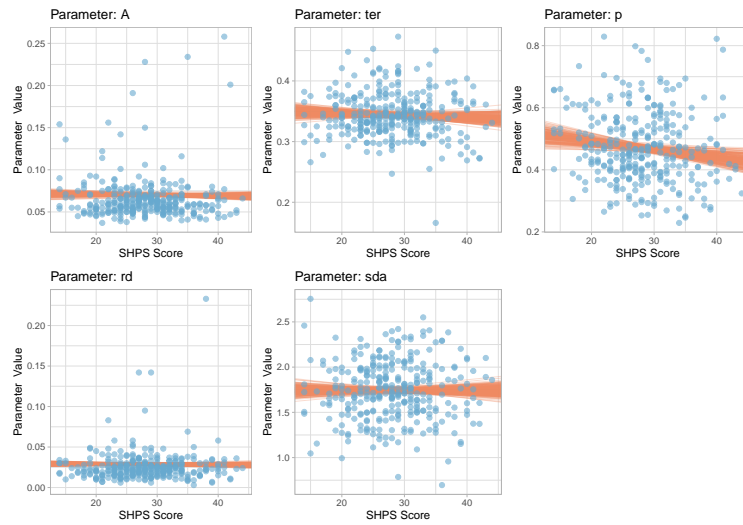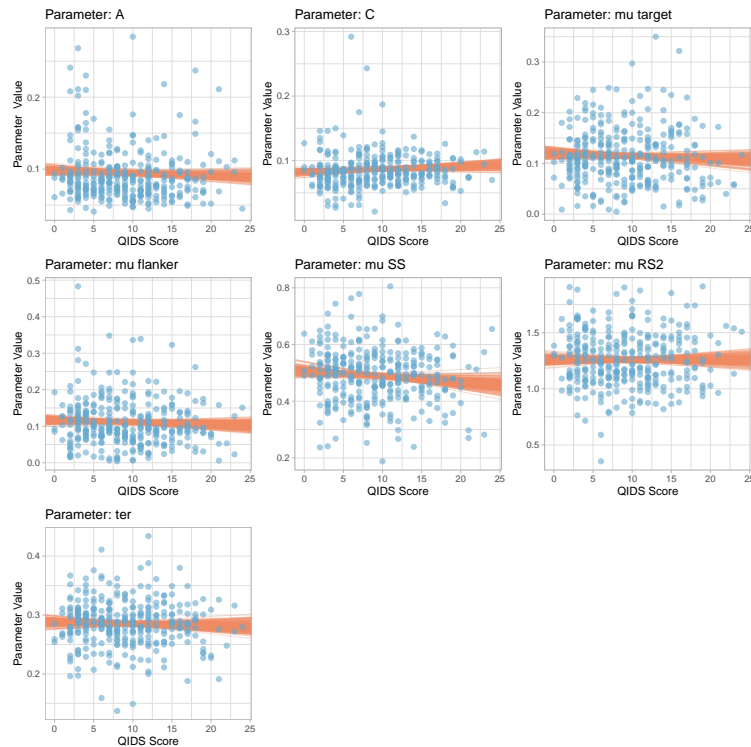
Figure E2. DSTP model parameters predicted from SHPS scores in Experiment 1. Points represent individual participant data; lines represent 200 draws from the posterior distribution of the Bayesian regression model, showing credible estimates of the linear relationship.
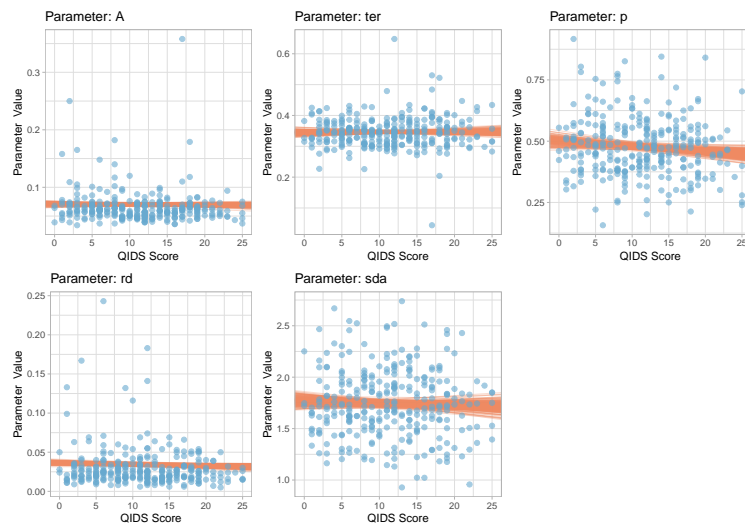
Figure E3. SSP model parameters predicted from QIDS scores in Experiment 1. Points represent individual participant data; lines represent 200 draws from the posterior distribution of the Bayesian regression model, showing credible estimates of the linear relationship.
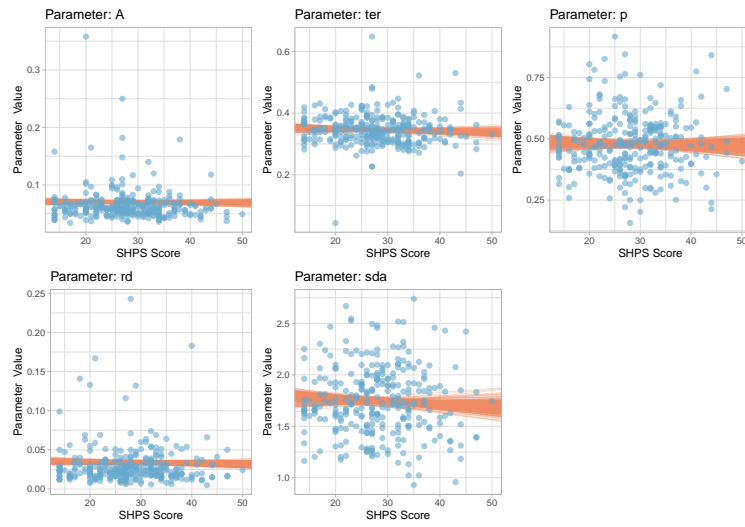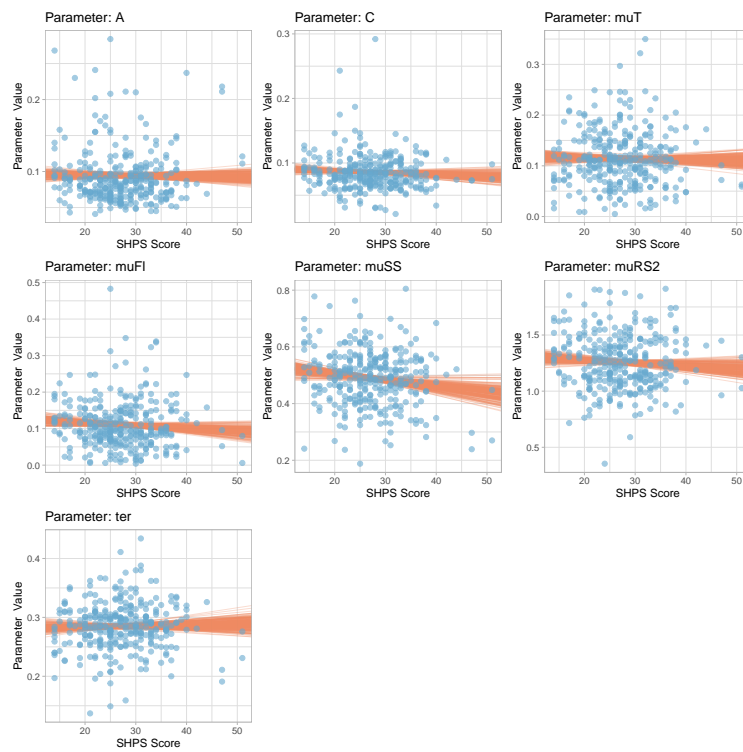
Figure E4. SSP model parameters predicted from SHPS scores in Experiment 1. Points represent individual participant data; lines represent 200 draws from the posterior distribution of the Bayesian regression model, showing credible estimates of the linear relationship.
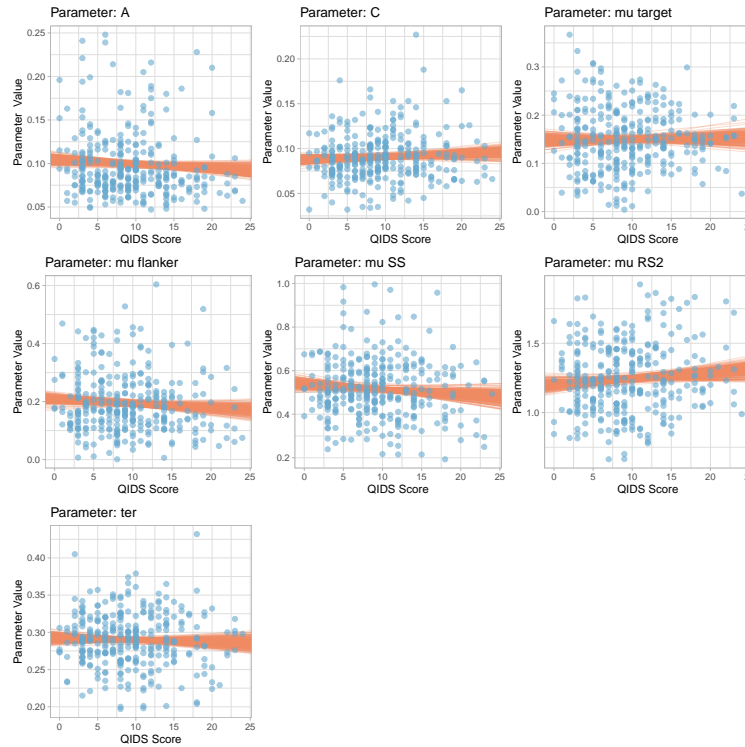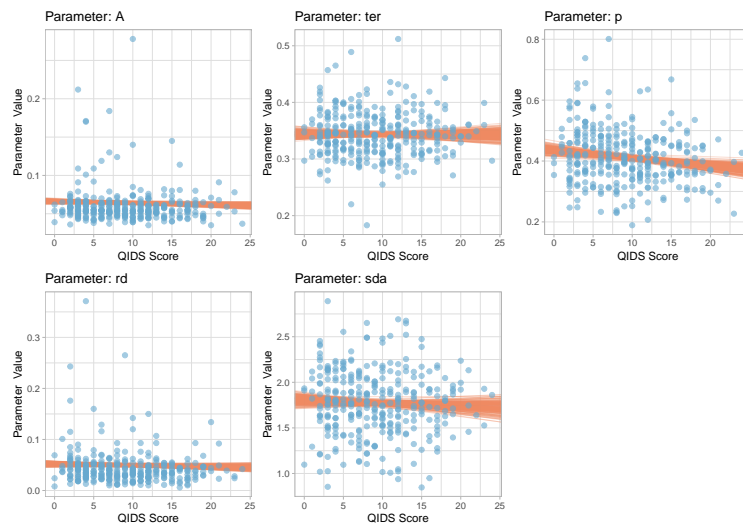

Figure E5. DSTP model parameters predicted from QIDS scores in Experiment 2. Points represent individual participant data; lines represent 200 draws from the posterior distribution of the Bayesian regression model, showing credible estimates of the linear relationship.
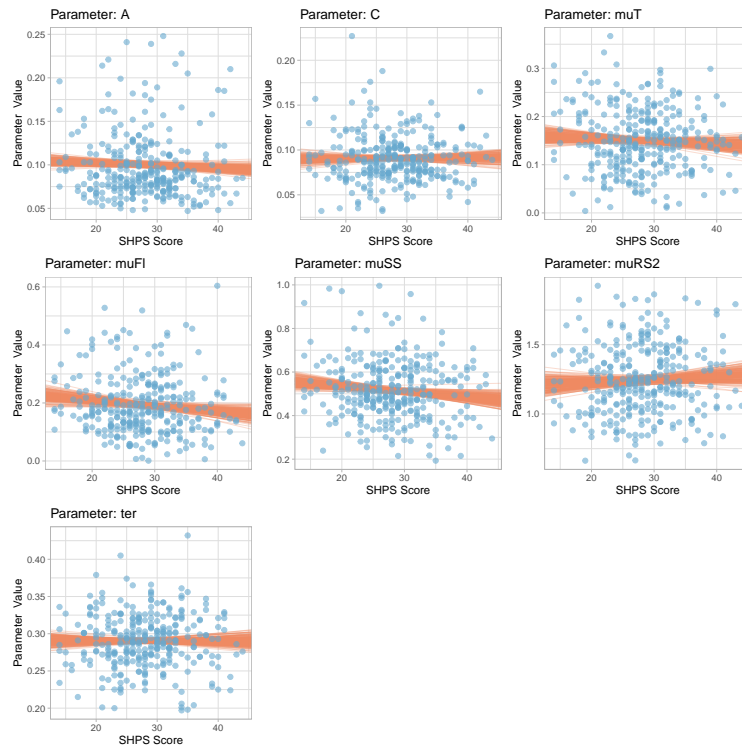
Figure E6. DSTP model parameters predicted from SHPS scores in Experiment 2. Points represent individual participant data; lines represent 200 draws from the posterior distribution of the Bayesian regression model, showing credible estimates of the linear relationship.
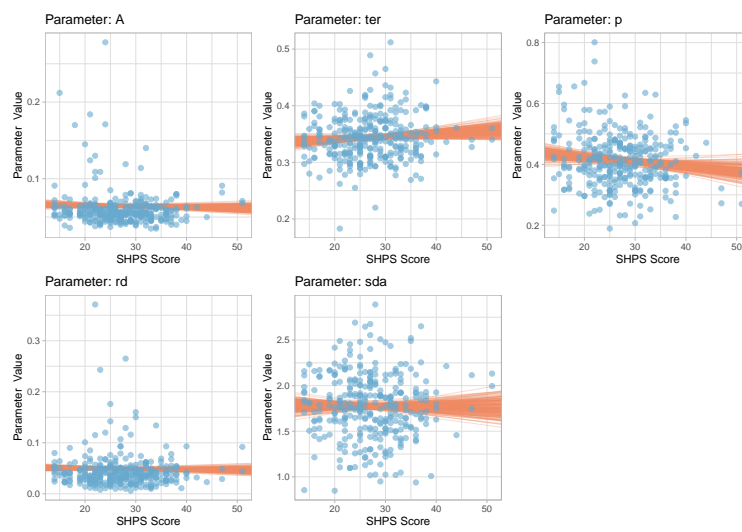

Figure E7. SSP model parameters predicted from QIDS scores in Experiment 2. Points represent individual participant data; lines represent 200 draws from the posterior distribution of the Bayesian regression model, showing credible estimates of the linear relationship.

Figure E8. SSP model parameters predicted from SHPS scores in Experiment 2. Points represent individual participant data; lines represent 200 draws from the posterior distribution of the Bayesian regression model, showing credible estimates of the linear relationship.

Figure E9. DSTP model parameters predicted from QIDS scores in Experiment 3. Points represent individual participant data; lines represent 200 draws from the posterior distribution of the Bayesian regression model, showing credible estimates of the linear relationship.

Figure E10. DSTP model parameters predicted from SHPS scores in Experiment 3. Points represent individual participant data; lines represent 200 draws from the posterior distribution of the Bayesian regression model, showing credible estimates of the linear relationship.

Figure E11. SSP model parameters predicted from QIDS scores in Experiment 3. Points represent individual participant data; lines represent 200 draws from the posterior distribution of the Bayesian regression model, showing credible estimates of the linear relationship.

Figure E12. SSP model parameters predicted from SHPS scores in Experiment 3. Points represent individual participant data; lines represent 200 draws from the posterior distribution of the Bayesian regression model, showing credible estimates of the linear relationship.

Appendix F

Extreme QIDS Sub-Analysis

In this section, we provide a between-groups analysis on our dependent variables by grouping participants who scored 14 or above on the QIDS into a "high-QIDS" group, and those who scored below 8 into a "low-QIDS" group. We chose these QIDS scores as grouping cutoffs as they represent the values used for inclusion in the depressed group and the control group, respectively, in the study of Dillon et al. (2015). In our experiments, there were 109, 67, and 70 who were classified as high-QIDS in Experiments 1–3 respectively; for the low-QIDS groups, we had 88, 122, and 137 participants in Experiments 1–3 respectively.

We first provide an analysis on the behavioural data (i.e., RTs and accuracy) before analysing the model parameters. All analyses utilised Bayesian linear regressions with the respective DV as the outcome variable, and "QIDS group" as a categorical predictor. Regression coefficients for the QIDS-group predictor report the deviation of the low-QIDS group from the high-QIDS group (i.e., positive Beta values represent higher scores for the low-QIDS group in comparison to the high-QIDS group.

Behavioural Data

Density distributions of the outcome variables for the behavioural data for Experiments 1–3 can be seen in Figures F1–F3. The regression parameter values are shown in Table F1.
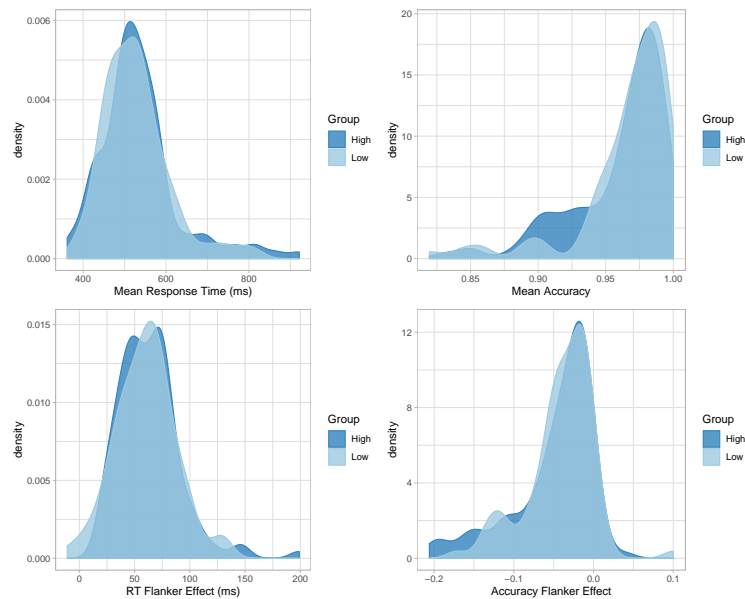
Figure F1. Density distributions of Experiment 1 data split by high QIDS scorers (above 14) and low QIDS scorers (below 8) for the dependent variables mean response time (in milliseconds, ms), mean proportion accuracy, flanker effect (response time, in ms), and flanker effect in accuracy (proportion).

Each plot shows the effect of QIDS grouping on mean response time, mean accuracy, the flanker effect in RT, and the flanker effect in accuracy. No differences were evidence in
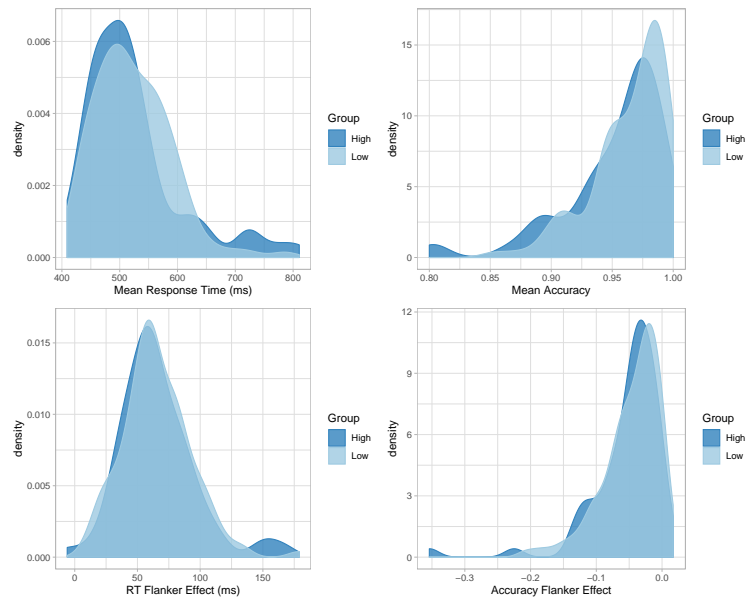
Figure F2. Density distributions of Experiment 2 data split by high QIDS scorers (above 14) and low QIDS scorers (below 8) for the dependent variables mean response time (in milliseconds, ms), mean proportion accuracy, flanker effect (response time, in ms), and flanker effect in accuracy (proportion).

any of the Experiments for the behavioural data.

Although there were no differences at the mean level, it is important to remember that the models are fit to the distributional data (the CDF and the CAF data). Visual inspection of these data suggest that there may be group differences that the models could explain. The cumulative distribution frequency plots of correct RT and the conditional accuracy functions for Experiments 1–3 are shown in Figure F4. There were no consistent differences between groups in the CDF plots; however, for the CAF plots, high-QIDS groups showed larger flanker effects in accuracy for the fastest RT bin, generally caused by poorer accuracy in the incongruent condition for high-QIDS scorers compared to low-QIDS scorers. This was generally true only for the fastest RT bin, with the exception of Experiment 3 where high-QIDS scorers generally had poorer accuracy throughout the whole RT distribution for incongruent trials.

DSTP Model Parameters

Density distributions of group differences on parameter estimates from the DSTP model are shown in Figures F5–F7. The parameter values for the Bayesian regressions are shown in Table F1. There were no consistent effects of QIDS score on any DSTP model parameter, with the exception of the μ_{SS} parameter in Experiment 3, which was lower in the high-QIDS group compared to the low-QIDS group. This reduction suggests that high-QIDS participants were slower to select the central target for further processing, thus delaying the shift from the first phase of response selection to the more-selective second phase of response selection. However, this effect was not evident in any of the other experiments, so
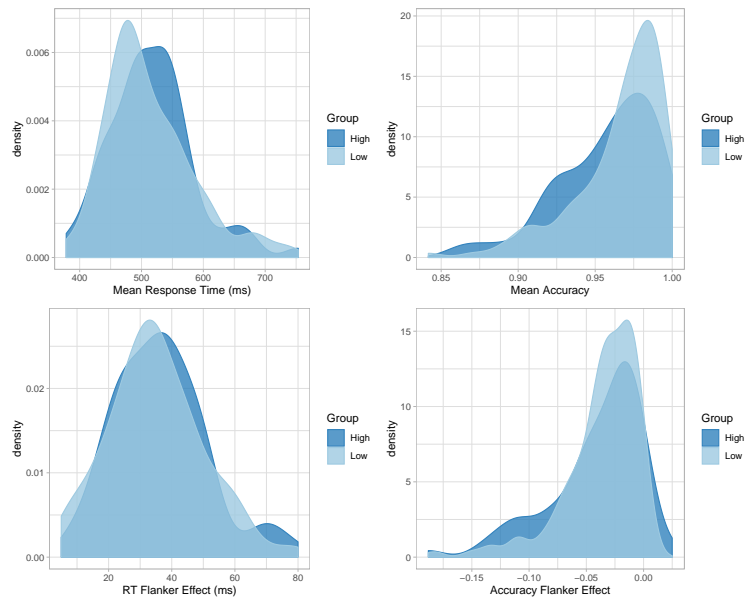
Figure F3. Density distributions of Experiment 3 data split by high QIDS scorers (above 14) and low QIDS scorers (below 8) for the dependent variables mean response time (in milliseconds, ms), mean proportion accuracy, flanker effect (response time, in ms), and flanker effect in accuracy (proportion).

we do not place much weight on this finding.

SSP Model Parameters

Density distributions of group differences on parameter estimates from the SSP model are shown in Figures F8–F10. The parameter values for the Bayesian regressions are shown in Table F1.

Replicating the findings from the regressions reported in the main body of the paper, estimates of model parameter p were lower in the high-QIDS group than the low-QIDS group, although the 95%CI for the regression predictor coefficient in Experiment 1 included zero. This suggests that high-QIDS participants had weaker perceptual representations of the stimulus than those in the low-QIDS group.

The *ter* parameter was also lower in the high-QIDS group than the low-QIDS group for Experiment 2, but this did not replicate in either of the other experiments, so again we do not place much weight on this finding.
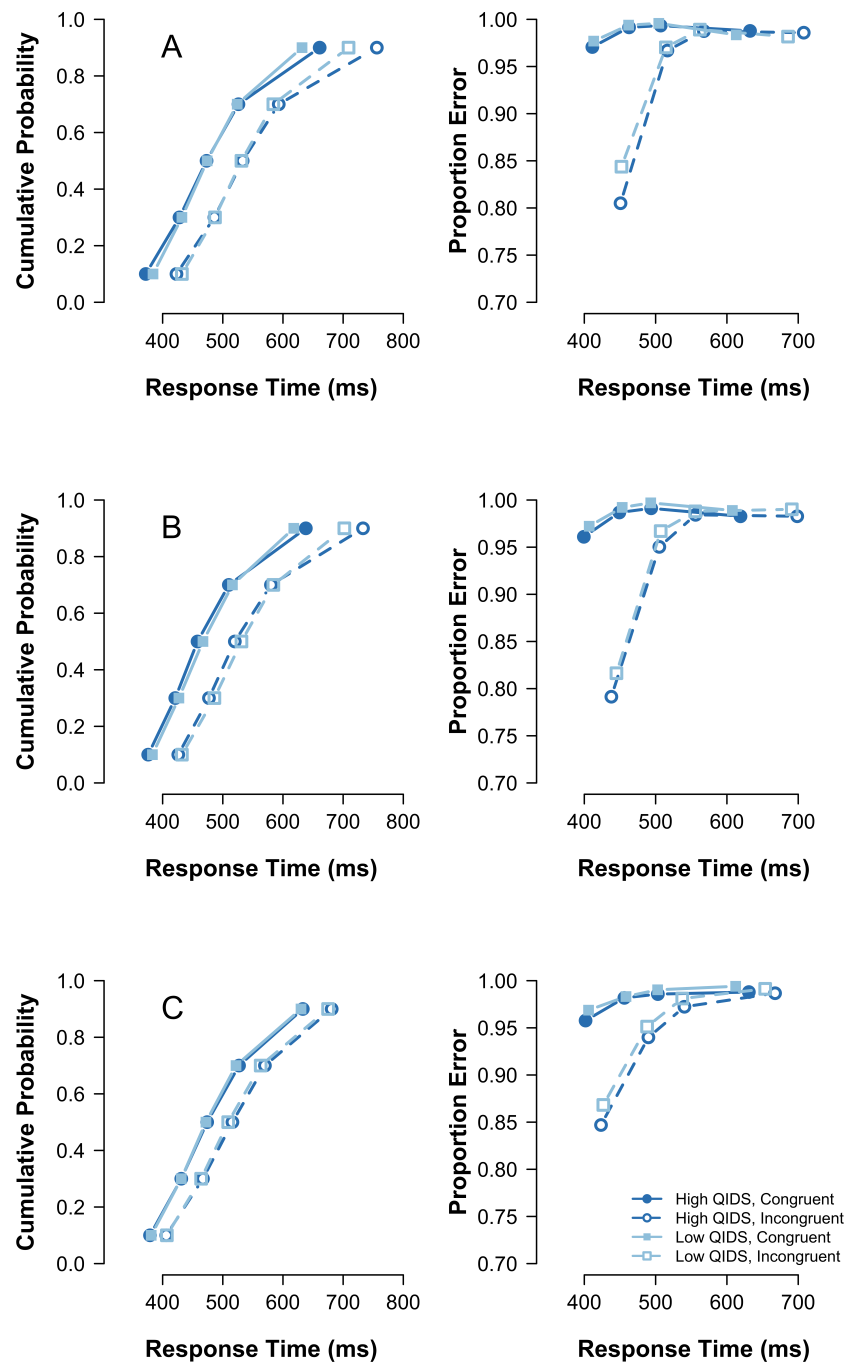
Figure F4. CDFs and CAFs for High and Low QIDS scoring participants. A–C plots Experiments 1–3.)

Table F1

Regression coefficients for the series of Bayesian regressions predicting dependent variable scores (DV) from the categorical predictor QIDS-Group across all Experiments. The values represent the mean estimate of the posterior distribution for each coefficient, together with their 95% credible interval in square parentheses. Entries in bold face represent regression predictors whose credible interval does not include zero (indicating the presence of an effect).

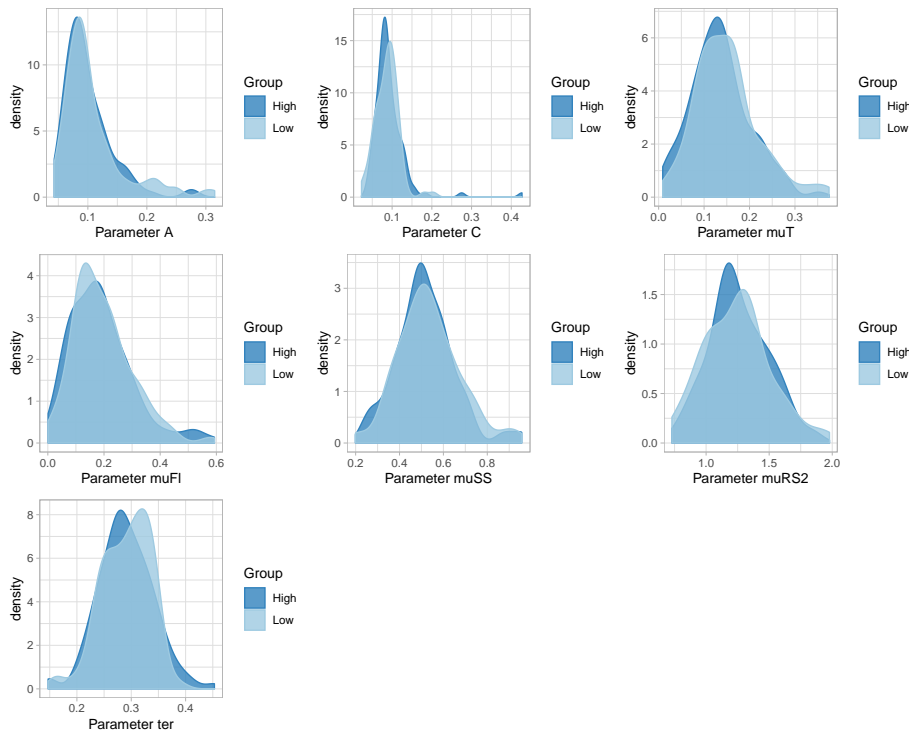|		Experiment 1		Experiment 2		Experiment 3	
Source	DV	$\beta_{Intercept}$	β_{Group}	$\beta_{Intercept}$	β_{Group}	$\beta_{Intercept}$	β_{Group}
Behavioural	Mean RT	529.31 [515.21, 544.17]	0.58 [-19.42, 20.56]	515.77 [501.55, 530.69]	8.01 [-8.24, 24.38]	516.65 [501.83, 531.76]	-5.64 [-22.93, 11.66]
	Mean Acc.	0.958 [0.953, 0.963]	0.003 [-0.001, 0.004]	0.956 [0.950, 0.962]	0.002 [-0.004, 0.008]	0.960 [0.955, 0.965]	0.002 [-0.002, 0.007]
	Fl. RT	63.22 [57.89, 68.38]	-3.11 [-10.67, 4.54]	65.72 [58.77, 72.60]	-0.36 [-8.56, 8.10]	36.46 [32.93, 40.07]	-1.52 [-5.90, 2.85]
	Fl. Acc	-0.05 [-0.06, -0.04]	0.01 [-0.01, 0.02]	-0.06 [-0.06, -0.05]	0.004 [-0.004, 0.012]	-0.04 [-0.05, -0.03]	-0.003 [-0.010. 0.004]
DSTP	A	0.1067 [0.0999. 0.1137]	0.0019 [-0.0063, 0.0102]	0.0953 [0.0888, 0.1017]	0.0064 [-0.0003, 0.0135]	0.0957 [0.0891, 0.1022]	0.0036 [-0.0030, 0.0108]
	C	0.0968 [0.0908, 0.1031]	-0.0070 [-0.0156, 0.0017]	0.0928 [0.0872, 0.0985]	-0.0054 [-0.0122, 0.0015]	0.0886 [0.0822, 0.0945]	-0.0065 [-0.0136, 0.0007]
	mu_T	0.1365 [0.1243, 0.1485]	0.0127 [-0.0051, 0.0305]	0.1588 [0.1420, 0.1747]	-0.0087 [-0.0281, 0.0122]	0.1115 [0.0991, 0.1239]	0.0058 [-0.0086, 0.0208]
	mu_Fl	0.1900 [0.1719, 0.2087]	0.0074 [-0.0173, 0.0312]	0.1746 [0.1538, 0.1960]	0.00246 [0.0003, 0.0502]	0.1101 [0.0964, 0.1236]	0.0077 [-0.0078, 0.0240]
	mu_SS	0.5190 [0.4940, 0.5446]	0.0146 [-0.0229, 0.0526]	0.4941 [0.4633, 0.5248]	0.0262 [-0.0123, 0.0656]	0.4749 [0.4525, 0.4968]	**0.0285 [0.0016, 0.0561]**
	mu_RS2	1.2638 [1.2166, 1.3134]	-0.0227 [-0.0933, 0.0495]	1.2779 [1.2225, 1.3326]	-0.0462 [-0.1142, 0.0220]	1.2869 [1.2257, 1.3480]	-0.0273 [-0.1039, 0.0462]
	ter	0.2902 [0.2812, 0.2986]	0.0003 [-0.0131, 0.0138]	0.2893 [0.2809, 0.2977]	0.0035 [-0.0074, 0.0143]	0.2822 [0.2732, 0.2914]	0.0056 [-0.0054, 0.0163]
SSP	a	0.0723 [0.0682, 0.0767]	0.0012 [-0.0038, 0.0062]	0.0683 [0.0645, 0.0722]	0.0019 [-0.0021, 0.0061]	0.0628 [0.0591, 0.0665]	0.0025 [-0.0011, 0.0064]
	ter	0.3484 [0.3391, 0.3578]	-0.0020 [-0.0169, 0.0123]	0.3352 [0.3271, 0.3437]	**0.0012 [0.0020, 0.0223]**	0.3460 [0.3370, 0.3553]	-0.0011 [-0.0119, 0.0101]
	p	0.4676 [0.446, 0.4915]	0.0277 [-0.0065, 0.0604]	0.4485 [0.4227, 0.4743]	**0.0346 [0.0028, 0.0655]**	0.4022 [0.3806, 0.4229]	**0.0311 [0.0055, 0.0568]**
	rd	0.0321 [0.0289, 0.0353]	0.0026 [-0.0008, 0.0059]	0.0295 [0.0262, 0.0327]	0.0018 [-0.0017, 0.0052]	0.0475 [0.0420, 0.0531]	0.0049 [-0.0006, 0.0108]
	sda	1.7232 [1.6599, 1.7836]	0.0251 [-0.0663, 0.1152]	1.7633 [1.6847, 1.8435]	1.7633 [1.6847, 1.8435]	1.7380 [1.6576, 1.8178]	0.0559 [-0.0401, 0.1513]

Figure F5. Density distributions of Experiment 1 data split by high QIDS scorers (above 14) and low QIDS scorers (below 8) for the parameters of the DSTP model.
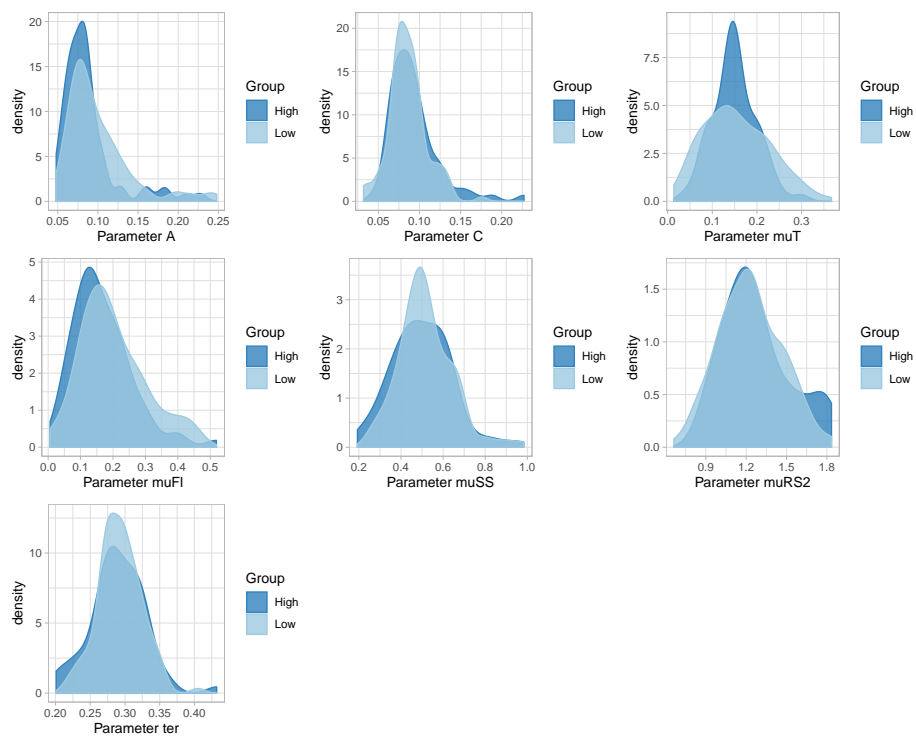
Figure F6. Density distributions of Experiment 2 data split by high QIDS scorers (above 14) and low QIDS scorers (below 8) for the parameters of the DSTP model.
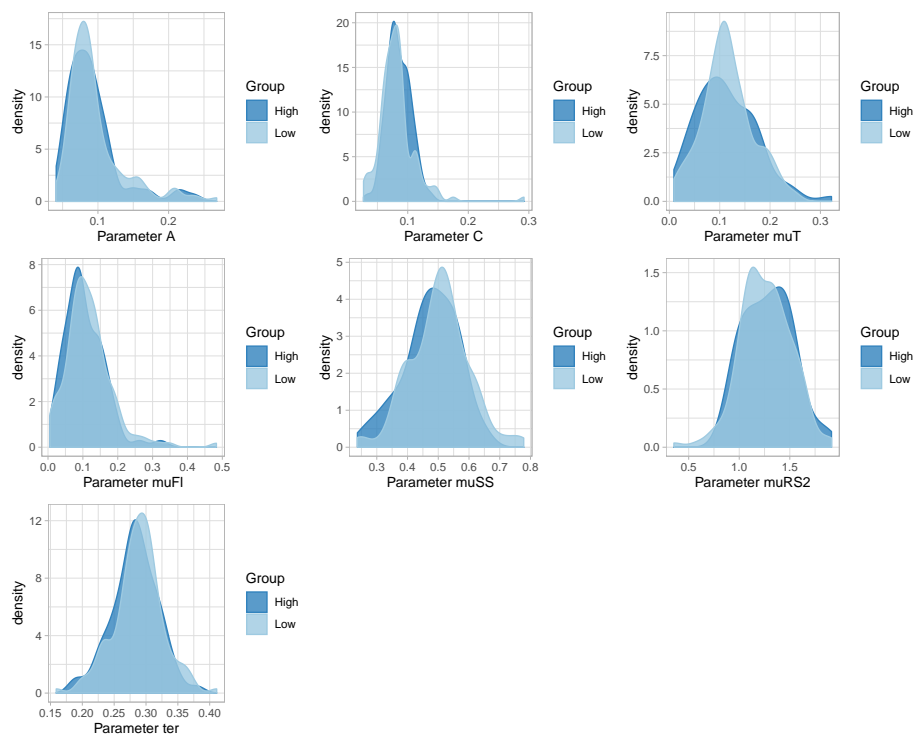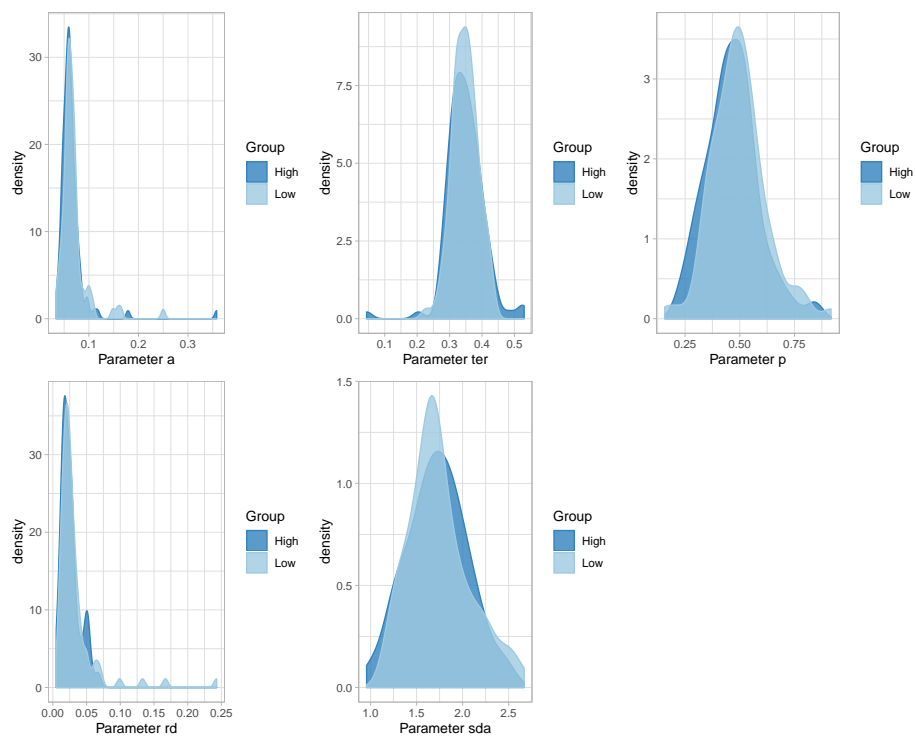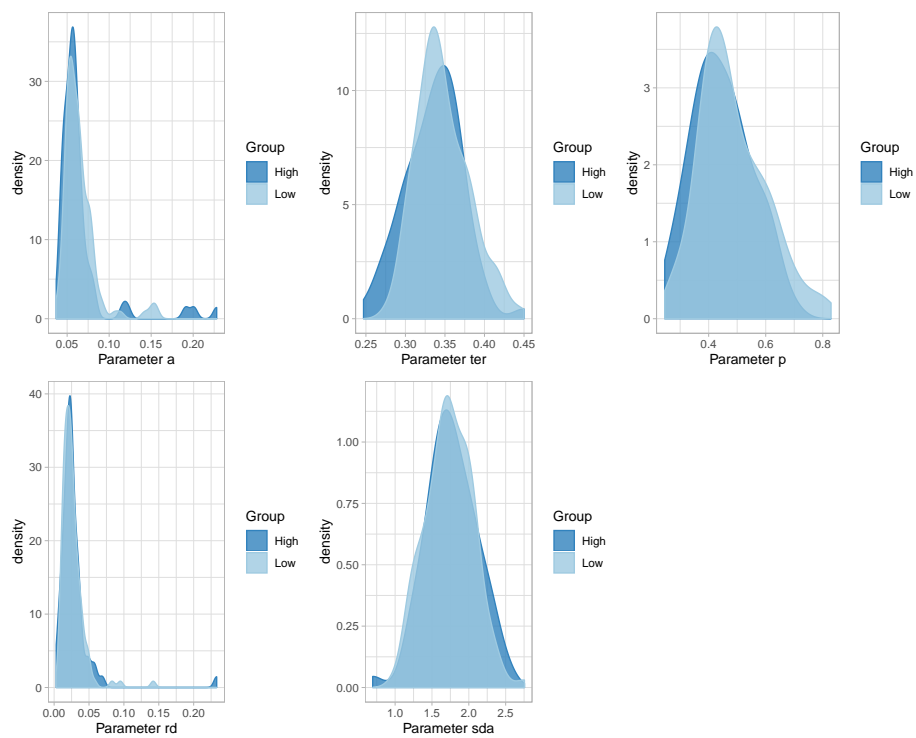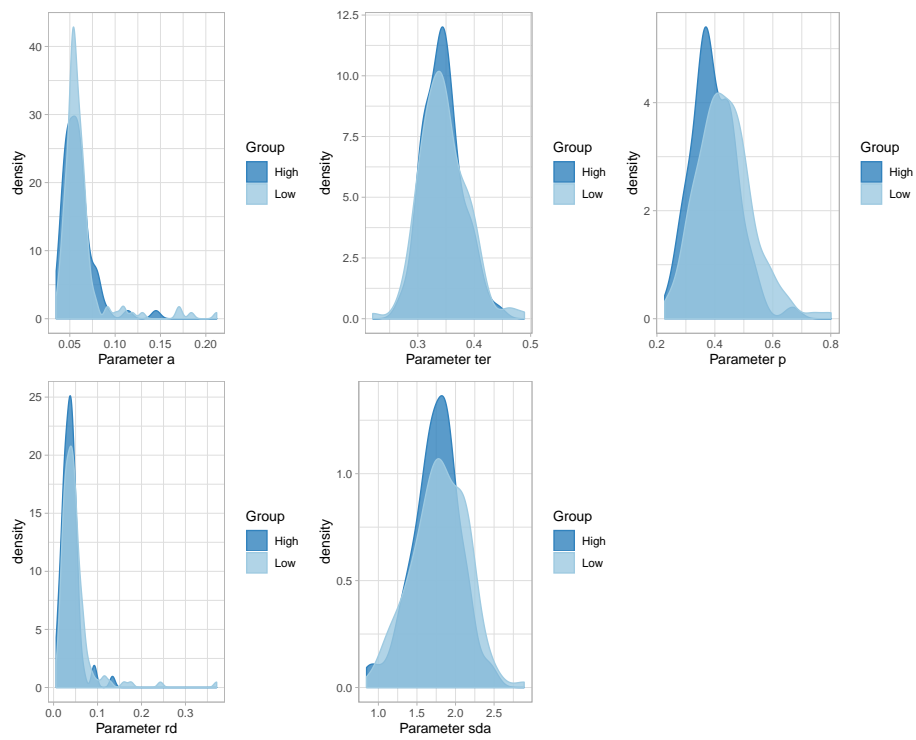
Figure F7. Density distributions of Experiment 3 data split by high QIDS scorers (above 14) and low QIDS scorers (below 8) for the parameters of the DSTP model.

Figure F8. Density distributions of Experiment 1 data split by high QIDS scorers (above 14) and low QIDS scorers (below 8) for the parameters of the SSP model.

Figure F9. Density distributions of Experiment 2 data split by high QIDS scorers (above 14) and low QIDS scorers (below 8) for the parameters of the SSP model.

Figure F10. Density distributions of Experiment 3 data split by high QIDS scorers (above 14) and low QIDS scorers (below 8) for the parameters of the SSP model.

Appendix G

Depression Diagnosis Sub-Analysis

In this section we wished to identify those who declared a clinical diagnosis ("with depression") to those who declared they did not have a clinical diagnosis ("without depression") to observe whether there are group differences on the dependent variables QIDS scores, SHPS scores, response time flanker effect, and accuracy flanker effect, as well as of course the DSTP and SSP model parameters. In our Experiments, we had 30.95%, 23.15%, and 14.78% of participants who declared a depression diagnosis (Experiments 1–3 respectively).

Group differences on QIDS & SHPS

Before looking at group differences on the main dependent variabeles, we were interested in exploring whether those who declared a depression diagnosis scored higher on the QIDS and SHPS. Such a finding would provide reassurance that the self-declarations were truthful. The density distributions for these variables for both groups for all experiments are shown in Figures G1–G3. We modelled the data using a Bayesian regression, with the dependent variable predicted by Group (i.e., "with" vs. "without" depression). Regression coefficients for the Depression-Group report the deviation of the depressed group from the non-depressed group (i.e., positive Beta values represent higher scores for the non-depressed group).
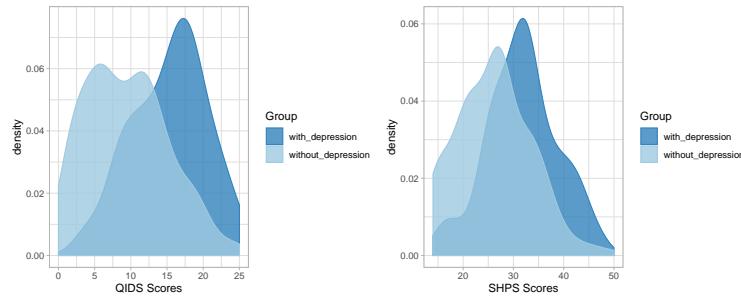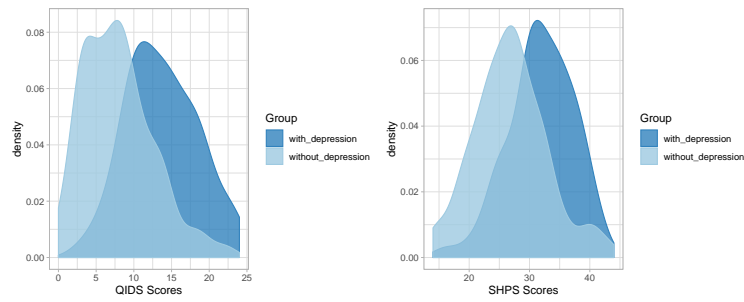

Figure G1. Density distributions of Experiment 1 data split by those declaring a depression diagnosis ("with depression") to those who declare no diagnosis of depression ("without depression") on the dependent variables QIDS scores and SHPS scores.

Experiment 1. For the QIDS questionnaire, those without depression demonstrated *lower* scores ($\beta_{Intercept}$ = 15.31 [14.16, 16.46], β_{Group} = -5.87 [-7.25, -4.46]). For the SHPS questionnaire, those without depression demonstrated lower scores ($\beta_{Intercept}$ = 31.76 [30.30, 33.23], β_{Group} = -5.57 [-7.36, -3.79]). These results suggest that the responses to both questionnaires were sensitive to depression diagnosis.

Experiment 2. For the QIDS questionnaire, those without depression demonstrated *lower* scores ($\beta_{Intercept}$ = 13.49 [12.49, 14.49], β_{Group} = -5.52 [-6.61, -4.46]). For the SHPS questionnaire, those without depression demonstrated lower scores ($\beta_{Intercept}$ = 31.95 [30.59, 33.31], β_{Group} = -5.11 [-6.66, -3.58]). These results suggest that the responses to both questionnaires were sensitive to depression diagnosis.

Experiment 3. For the QIDS questionnaire, those without depression demonstrated *lower* scores ($\beta_{Intercept}$ = 14.02 [13.06, 15.01], β_{Group} = -6.41 [-7.46, -5.39]).

Figure G2. Density distributions of Experiment 2 data split by those declaring a depression diagnosis ("with depression") to those who declare no diagnosis of depression ("without depression") on the dependent variables QIDS scores and SHPS scores.
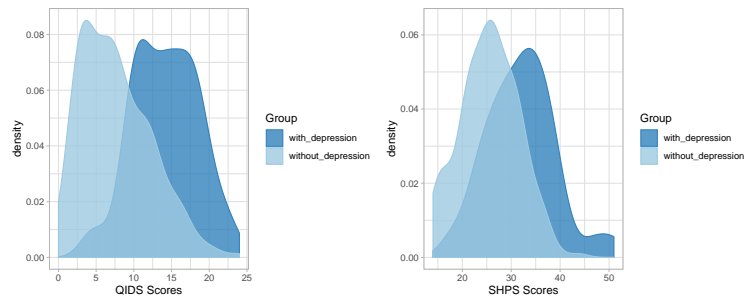

Figure G3. Density distributions of Experiment 3 data split by those declaring a depression diagnosis ("with depression") to those who declare no diagnosis of depression ("without depression") on the dependent variables QIDS scores and SHPS scores.

For the SHPS questionnaire, those without depression demonstrated lower scores ($\beta_{Intercept}$ = 32.06 [30.64, 33.46], β_{Group} = −6.49 [−8.11, −4.88]). These results suggest that the responses to both questionnaires were sensitive to depression diagnosis.

Behavioural Data

Density distributions of the outcome variables for the behavioural data for Experiments 1–3 can be seen in Figures G4–G6. The regression parameter values are shown in Table G1. Each plot shows the effect of Depression grouping on mean response time, mean accuracy, the flanker effect in RT, and the flanker effect in accuracy. No differences were evident.

Distributional data in the form of cumulative distribution functions and conditional accuracy functions for Experiments 1–3 can be found in Figure G7. There were no consistent differences between groups in the CDF plots; however, for the CAF plots, high-QIDS groups showed larger flanker effects in accuracy for the fastest RT bin in Experiment 1, generally caused by poorer accuracy in the incongruent condition for high-QIDS scorers compared to low-QIDS scorers. This was generally true only for the first experiment.
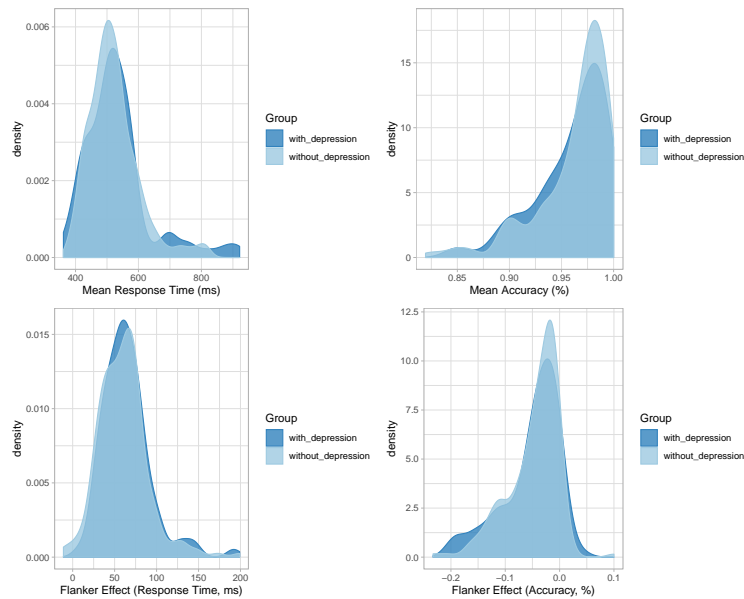
Figure G4. Density distributions of Experiment 1 data split by those declaring a depression diagnosis ("with depression") to those who declare no diagnosis of depression ("without depression") on the dependent variables Mean RT, flanker RT, mean accuracy, flanker accuracy.

DSTP Model Parameters

Density distributions of group differences on parameter estimates from the DSTP model are shown in Figures G8–G10. The parameter values for the Bayesian regressions are shown in Table G1. There were no consistent effects of depression on any DSTP model parameter, with the exception of the μ_{SS} parameter in Experiment 3, which was lower in the depressed group compared to the non-depressed group. This reduction suggests that depressed individuals were slower to select the central target for further processing, thus delaying the shift from the first phase of response selection to the more-selective second phase of response selection. However, this effect was not evident in any of the other experiments, so we do not place much weight on this finding.

SSP Model Parameters. Density distributions of group differences on parameter estimates from the SSP model are shown in Figures G11–G13. The parameter values for the Bayesian regressions are shown in Table C1. No SSP parameter differed according to depression grouping.
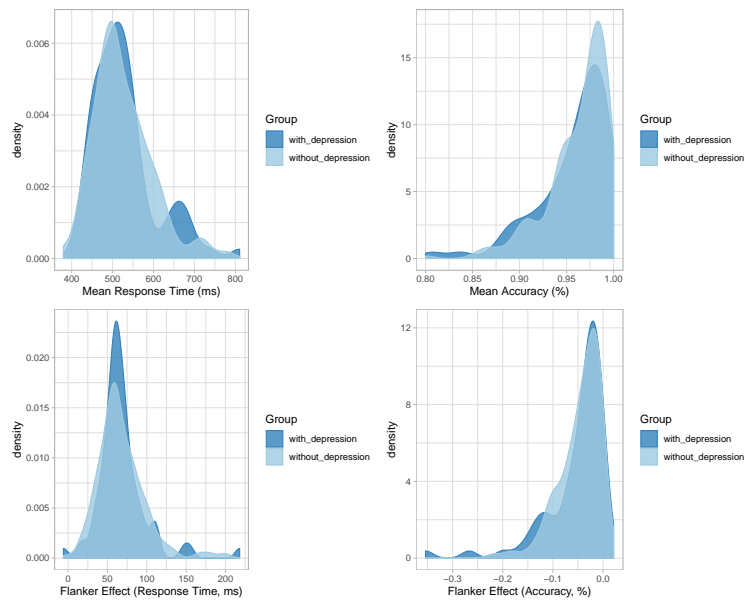
Figure G5. Density distributions of Experiment 2 data split by those declaring a depression diagnosis ("with depression") to those who declare no diagnosis of depression ("without depression") on the dependent variables Mean RT, flanker RT, mean accuracy, flanker accuracy.
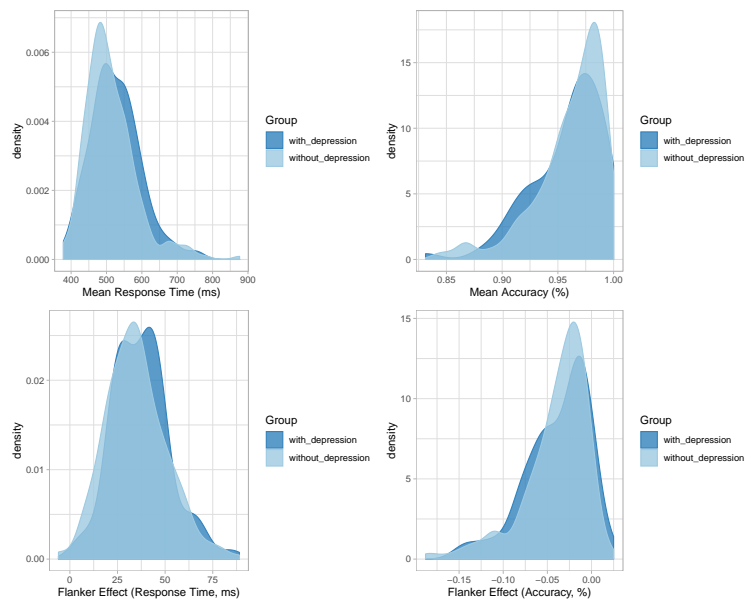

Figure G6. Density distributions of Experiment 3 data split by those declaring a depression diagnosis ("with depression") to those who declare no diagnosis of depression ("without depression") on the dependent variables Mean RT, flanker RT, mean accuracy, flanker accuracy.

Table G1

Regression coefficients for the series of Bayesian regressions predicting dependent variable scores (DV) from the categorical predictor depression diagnosis (with vs. without) across all Experiments. The values represent the mean estimate of the posterior distribution for each coefficient, together with their 95% credible interval in square parentheses. Entries in bold face represent regression predictors whose credible interval does not include zero (indicating the presence of an effect).

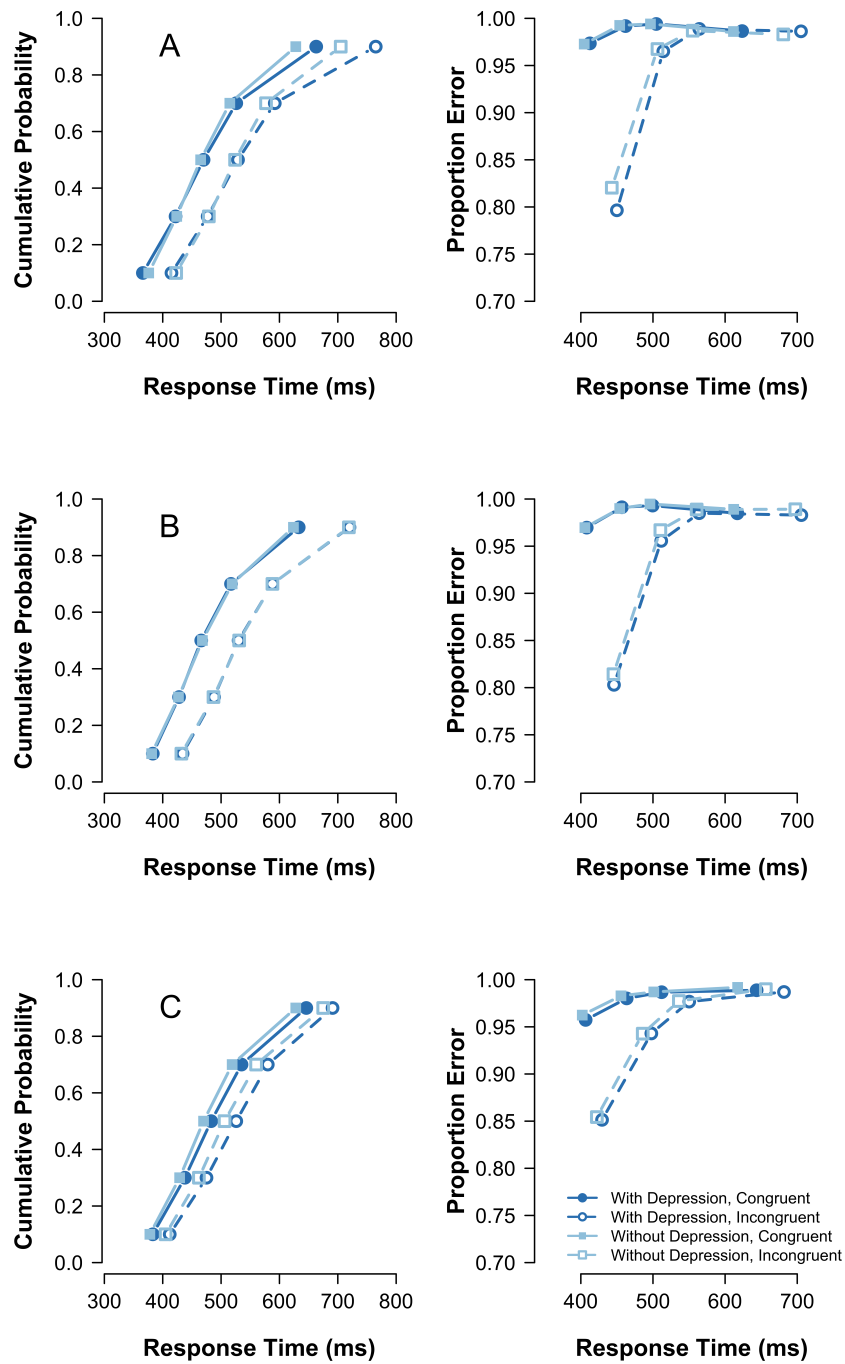|		Experiment 1		Experiment 2		Experiment 3	
Source	DV	$\beta_{Intercept}$	β_{Group}	$\beta_{Intercept}$	β_{Group}	$\beta_{Intercept}$	β_{Group}
Behavioural	Mean RT	521.30 [505.76, 536.68]	2.56 [-14.54, 20.00]	525.62 [510.52, 540.06]	-1.02 [-16.23, 15.21]	524.20 [509.53, 538.58]	-12.82 [28.19, 3.04]
	Mean Acc.	0.958 [0.954, 0.963]	0.001 [-0.003, 0.005]	0.959 [0.954, 0.964]	0.01 [-0.004, 0.004]	0.960 [0.956, 0.964]	-0.002 [-0.006, 0.003]
	Fl. RT	64.21 [58.14, 70.50]	-4.18 [-11.62, 3.26]	67.84 [60.75, 75.04]	-0.41 [-8.43, 7.73]	37.36 [33.75, 40.92]	-2.54 [-6.58, 1.50]
	Fl. Acc	-0.05 [-0.06, -0.04]	0.01 [-0.01, 0.02]	-0.05 [-0.06, -0.05]	0.003 [-0.005, 0.009]	-0.04 [-0.045, 0.033]	-0.003 [-0.001, 0.003]
DSTP	A	0.1026 [0.0962, 0.1090]	0.0014 [-0.0050, 0.0080]	0.0964 [0.0904, 0.1024]	0.0044 [-0.0014, 0.0106]	0.0955 [0.0901, 0.1009]	-0.0012 [-0.0065, 0.0045]
	C	0.0952 [0.0888, 0.1015]	-0.0037 [-0.0107, 0.0035]	0.0907 [0.0849, 0.0966]	0.0002 [-0.0064, 0.0068]	0.0892 [0.0835, 0.0948]	-0.0042 [-0.0105, 0.0021]
	mu_T	0.1403 [0.1270, 0.1536]	-0.0049 [-0.0101, 0.0205]	0.1516 [0.1359, 0.1662]	-0.0017 [-0.0186, 0.0161]	0.1098 [0.0989, 0.1211]	0.0062 [-0.0062, 0.0192]
	mu_Fl	0.1906 [0.1714, 0.2097]	0.0021 [-0.0188, 0.0230]	0.1833 [0.1631, 0.2039]	0.0123 [-0.0095, 0.0346]	0.1082 [0.0956, 0.1210]	0.0036 [-0.0099, 0.0178]
	mu_SS	0.5159 [0.4884, 0.5426]	0.0170 [-0.0168, 0.0512]	0.5078 [0.4766, 0.5389]	0.0111 [-0.0257, 0.0472]	0.4685 [0.4462, 0.4904]	**0.0271 [0.0022, 0.0529]**
	mu_RS2	1.2504 [1.1979, 1.3023]	-0.0224 [-0.0637, 0.0607]	1.2372 [1.1807, 1.2948]	0.0084 [-0.0585, 0.0741]	1.2573 [1.2021, 1.3125]	0.0027 [-0.0632, 0.0660]
	ter	0.2927 [0.2834, 0.3020]	-0.0224 [-0.0637, 0.0607]	0.2976 [0.2889, 0.3061]	-0.0096 [-0.0193, 0.0004]	0.2802 [0.2709, 0.2892]	0.0064 [-0.0040, 0.0171]
SSP	a	0.0693 [0.0655, 0.0731]	0.0012 [-0.0023, 0.0051]	0.0691 [0.0656, 0.0726]	0.0015 [-0.0018, 0.0051]	0.0661 [0.0630, 0.0692]	-0.0022 [-0.0052, 0.0009]
	ter	0.3529 [0.3428, 0.3628]	-0.0105 [-0.0227, 0.0016]	0.3466 [0.3375, 0.3554]	-0.0031 [-0.0131, 0.0069]	0.3460 [0.3370, 0.3555]	-0.0028 [-0.0134, 0.0075]
	p	0.4688 [0.4449, 0.4922]	0.0136 [-0.0155, 0.0429]	0.4533 [0.4289, 0.4818]	0.0206 [-0.0083, 0.0495]	0.3970 [0.3771, 0.4171]	0.0206 [-0.0018, 0.0436]
	rd	0.0330 [0.0299, 0.0360]	0.0020 [-0.0009, 0.0051]	0.0283 [0.0256, 0.0310]	0.0006 [-0.0021, 0.0035]	0.0482 [0.0433, 0.0531]	0.0018 [-0.0028, 0.0070]
	sda	1.7372 [1.6690, 1.8066]	0.0121 [-0.0673, 0.0946]	1.7213 [1.6434, 1.7976]	0.0282 [-0.0573, 0.1142]	1.7621 [1.6833, 1.8408]	0.0165 [-0.0757, 0.1062]

Figure G7. CDFs and CAFs for those participants self-declaring a diagnosis of depression, and those declaring no such diagnosis. A–C plots Experiments 1–3.)
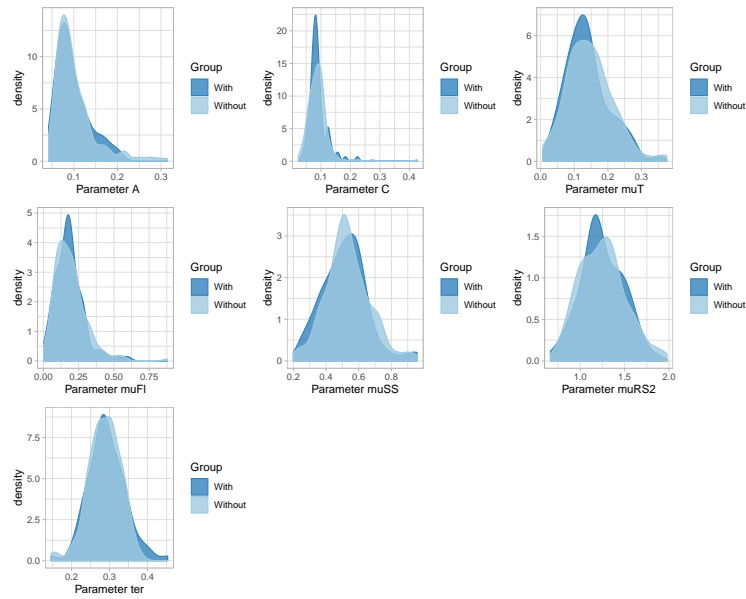
Figure G8. Density distributions of the DSTP model parameters in Experiment 1 split by those declaring a depression diagnosis ("with depression") to those who declare no diagnosis of depression ("without depression").
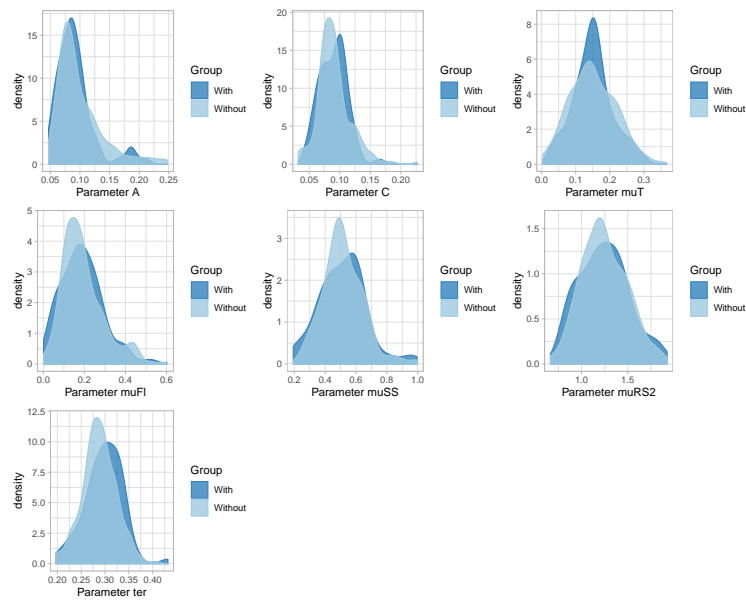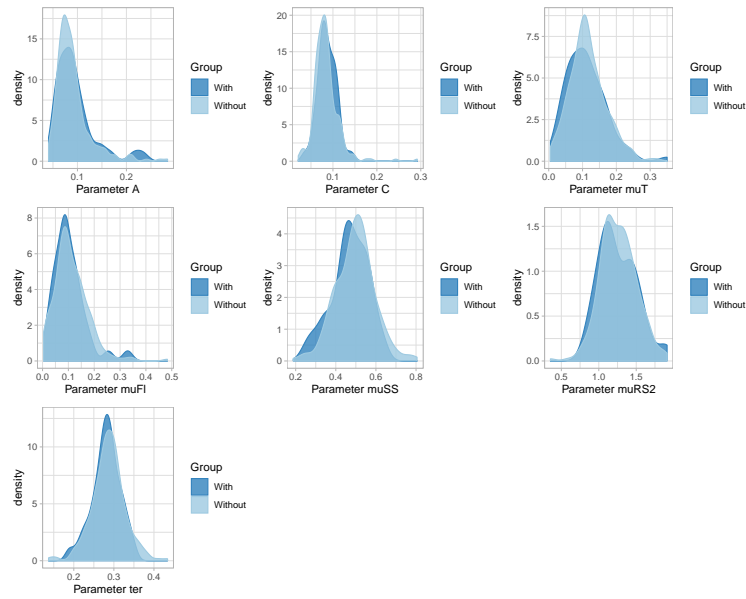

Figure G9. Density distributions of the DSTP model parameters in Experiment 2 split by those declaring a depression diagnosis ("with depression") to those who declare no diagnosis of depression ("without depression").
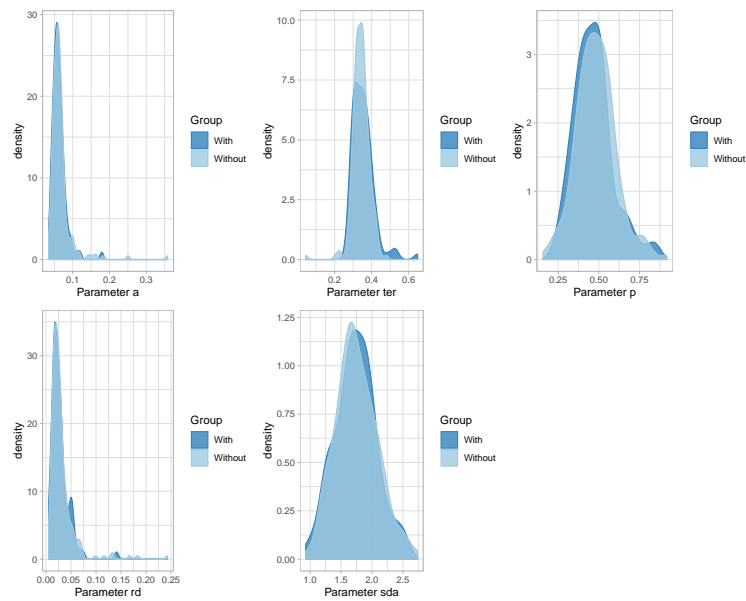
Figure G10. Density distributions of the DSTP model parameters in Experiment 3 split by those declaring a depression diagnosis ("with depression") to those who declare no diagnosis of depression ("without depression").

Figure G11. Density distributions of the SSP model parameters in Experiment 1 split by those declaring a depression diagnosis ("with depression") to those who declare no diagnosis of depression ("without depression").
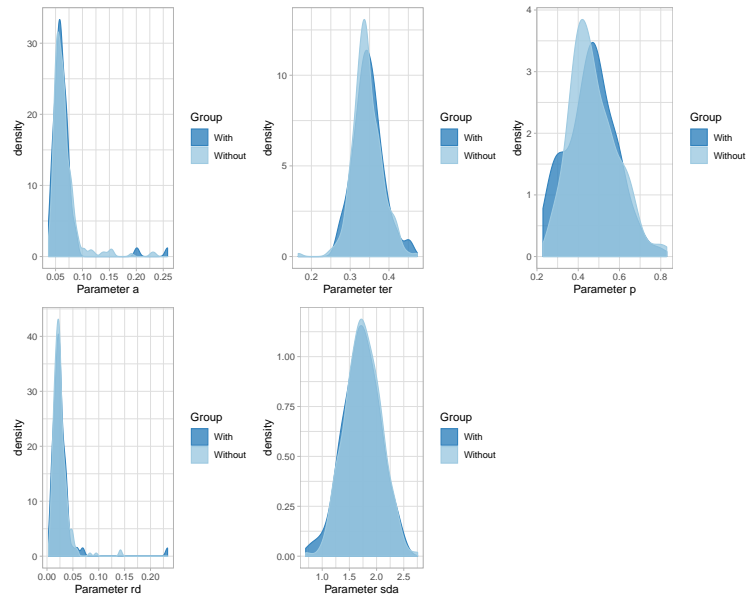
Figure G12. Density distributions of the SSP model parameters in Experiment 2 split by those declaring a depression diagnosis ("with depression") to those who declare no diagnosis of depression ("without depression").
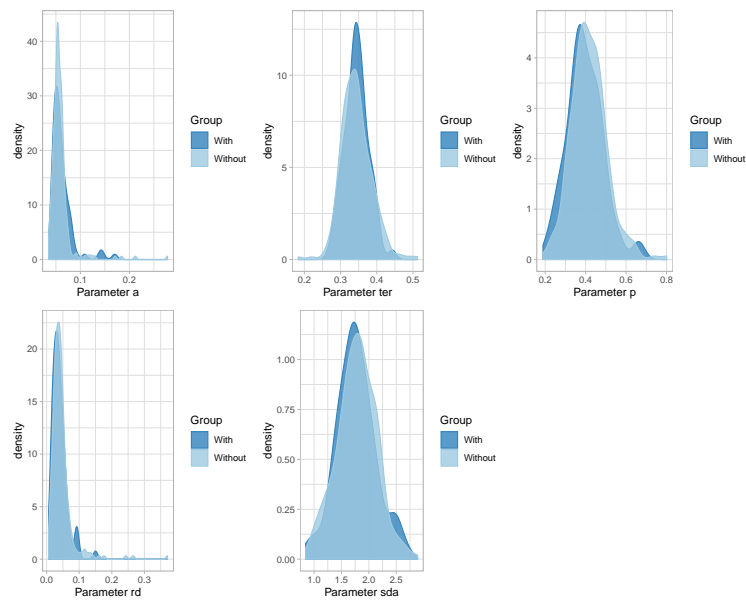

Figure G13. Density distributions of the SSP model parameters in Experiment 3 split by those declaring a depression diagnosis ("with depression") to those who declare no diagnosis of depression ("without depression").