Psychological Medicine

cambridge.org/psm

Dissecting the impact of depression on decision-making

Victoria M. Lawlor[1,2], Christian A. Webb[1], Thomas V. Wiecki[3], Michael J. Frank[4], Madhukar Trivedi[5], Diego A. Pizzagalli[1] and Daniel G. Dillon[1]

[1]Center for Depression, Anxiety and Stress Research, McLean Hospital/Harvard Medical School, Belmont, Massachusetts, USA; [2]Emory University, Atlanta, Georgia, USA; [3]Quantopian, Inc, Boston, Massachusetts, USA; [4]Brown University, Providence, Rhode Island, USA and [5]UT Southwestern Medical Center, Dallas, Texas, USA

Original Article

Cite this article: Lawlor VM, Webb CA, Wiecki TV, Frank MJ, Trivedi M, Pizzagalli DA, Dillon DG (2019). Dissecting the impact of depression on decision-making. *Psychological Medicine* 1–10. https://doi.org/10.1017/S0033291719001570

Received: 13 November 2018
Revised: 3 May 2019
Accepted: 11 June 2019

Key words:
Computational modeling; decision-making; depression; drift diffusion model; reward

Author for correspondence:
Daniel G. Dillon,
E-mail: ddillon@mclean.harvard.edu

Abstract

Background. Cognitive deficits in depressed adults may reflect impaired decision-making. To investigate this possibility, we analyzed data from unmedicated adults with Major Depressive Disorder (MDD) and healthy controls as they performed a probabilistic reward task. The Hierarchical Drift Diffusion Model (HDDM) was used to quantify decision-making mechanisms recruited by the task, to determine if any such mechanism was disrupted by depression.

Methods. Data came from two samples (Study 1: 258 MDD, 36 controls; Study 2: 23 MDD, 25 controls). On each trial, participants indicated which of two similar stimuli was presented; correct identifications were rewarded. Quantile-probability plots and the HDDM quantified the impact of MDD on response times (RT), speed of evidence accumulation (drift rate), and the width of decision thresholds, among other parameters.

Results. RTs were more positively skewed in depressed *v.* healthy adults, and the HDDM revealed that drift rates were reduced—and decision thresholds were wider—in the MDD groups. This pattern suggests that depressed adults accumulated the evidence needed to make decisions more slowly than controls did.

Conclusions. Depressed adults responded slower than controls in both studies, and poorer performance led the MDD group to receive fewer rewards than controls in Study 1. These results did not reflect a sensorimotor deficit but were instead due to sluggish evidence accumulation. Thus, slowed decision-making—not slowed perception or response execution—caused the performance deficit in MDD. If these results generalize to other tasks, they may help explain the broad cognitive deficits seen in depression.

Introduction

Depression is characterized by impaired executive function (Snyder, 2013), difficulty sustaining attention (Biringer *et al.*, 2007), and memory problems—including trouble recalling details from encoding (MacQueen *et al.*, 2003) and loss of the positive memory bias typically seen in healthy adults (Burt *et al.*, 1995; Dillon *et al.*, 2013). Indeed, multiple meta-analyses document broad cognitive deficits in depressed adults (Burt *et al.*, 1995; Zakzanis *et al.*, 1998; Snyder, 2013), with problems related to executive function and attention persisting in remission (Douglas and Porter, 2009; Rock *et al.*, 2014). Each cognitive problem may involve separate pathophysiologies, but some processes may be common to most of the tests on which depressed adults show impairment. If so, then a negative effect of depression on those processes would help explain the broad range of cognitive deficits observed. This work aimed to study the impact of depression on one such process—namely, decision-making.

Decision-making is an appealing candidate because most tests of attention, executive function, and memory involve choosing among alternatives (Shadlen and Kiani, 2013). Thus, a negative effect of depression on decision-making would lead to the broad impairments that have been seen. Another reason to focus on decision-making is that computational models can parse it into component processes, providing an opportunity to pinpoint the specific mechanisms affected by depression. In particular, the drift-diffusion model (DDM) has been used for over 40 years to decompose decision-making during recognition memory tests (Ratcliff, 1978), lexical decision tasks (Ratcliff *et al.*, 2004), purchasing games (Krajbich *et al.*, 2012), and many other paradigms (for review, see Ratcliff and McKoon, 2008). The DDM and similar models have also been used to study brain systems that support decision-making in humans (Frank *et al.*, 2015) and non-human primates (Gold and Shadlen, 2007). Finally, prior work has emphasized that by applying the DDM in clinical contexts, it may be possible to uncover deficits in patients that cannot be detected with traditional analysis of response times (RT) and accuracy (White *et al.*, 2010). For these reasons, we elected to use a Bayesian variant of the DDM called the Hierarchical Drift Diffusion Model (HDDM; Wiecki *et al.*, 2013) to study decision-making in depression.

CAMBRIDGE UNIVERSITY PRESS

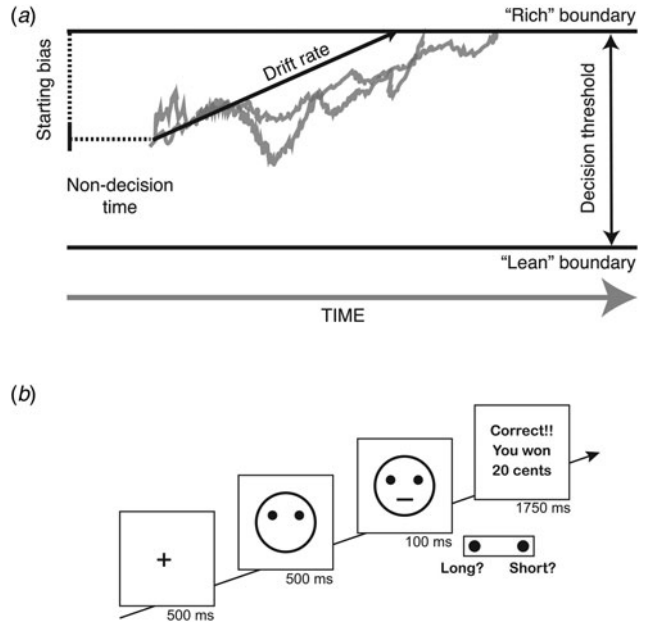

Fig. 1. (*a*) The Hierarchical Drift Diffusion Model (HDDM). The HDDM represents decisions as a process of evidence accumulation towards response boundaries separated by a *decision threshold*. The speed of evidence accumulation is referred to as the *drift rate*. The drift process moves left to right over time, from a starting point that can be midway between the boundaries or shifted towards one to an extent captured by the *starting bias*. The time needed for perception and response execution is captured by the *non-decision time*. In applying the model to PRT data, we mapped the upper and lower boundaries to 'rich' and 'lean', responses, respectively. The HDDM is a Bayesian extension of the original DDM (Ratcliff and McKoon, 2008) that provides enhanced parameter estimation for studies with between-group designs. (*b*) The probabilistic reward task (PRT). On each trial, participants must indicate whether a short (11.5 mm) or long (13.0 mm) mouth was shown. Correct identifications of one length (the 'rich' stimulus) are rewarded three times more frequently than correct identifications of the other length (the 'lean' stimulus).

The HDDM is shown in Fig. 1*a*. Briefly, it conceptualizes decision-making as a process of evidence accumulation. When a participant views two options and must choose between them, the HDDM assumes that the participant sets up boundaries specifying the amount of evidence needed to select one alternative over the other. Next, the participant draws a sample of the evidence for each option, computes the difference between the samples, and then increments a decision variable towards whichever boundary is favored by the difference score. This process is performed repeatedly until the evidence crosses one of the boundaries, at which point the corresponding response is rendered.

For example, several non-human primate studies have used a task in which monkeys must decide if a swirling dot pattern is moving mainly to the left or right (Gold and Shadlen, 2007). Monkeys solve this task by repeatedly drawing samples of evidence from motion-sensitive neurons that encode movement to the left *v.* right, respectively, until the evidence crosses a threshold (boundary) in favor of one direction; at this point, the monkeys respond. The speed of evidence accumulation is called the *drift rate*. Drift rate sets a limit on the speed and accuracy of decision-making, and it can be estimated by applying the HDDM to RT and accuracy data. Importantly, the HDDM also includes a *non-decision time* parameter that captures the time needed to perceive stimuli and execute a response once a decision has been made. Thus, the model can distinguish between factors that affect the speed of evidence accumulation, quantified by drift rate, *v.* those that affect sensorimotor aspects of behavior, quantified by non-decision time.

As a first step towards testing the hypothesis that depression affects decision-making, we applied the HDDM to behavioral data collected from 281 unmedicated adults with Major Depressive Disorder (MDD) and 61 healthy controls as they performed a probabilistic reward task (PRT; Pizzagalli *et al.*, 2005). As described below, on each trial the participants viewed a schematic face onto which a mouth was briefly flashed. The task was to indicate whether the mouth was long or short. Correct responses elicited monetary rewards, and correct identifications of one mouth length (the 'rich' stimulus) were rewarded three times more often than correct identifications of the other length (the 'lean' stimulus). This asymmetric reinforcement rate was used to induce a response bias. Many studies have found that depressed adults develop a weaker response bias than controls (Pizzagalli *et al.*, 2005, 2008; Vrieze *et al.*, 2013; Liu *et al.*, 2016), and we report on this below. However, our main goal was to conduct a fine-grained analysis of the impact of MDD on decision-making.

We expected that MDD would be associated with slow evidence accumulation. To test this account, we first compared RT distributions across the groups. Prior research (Ratcliff and McKoon, 2008) indicates that differences in drift rate have a small effect on fast RTs but a large effect on slow RTs, such that the RT distribution should be more positively skewed in the group with the slower drift rate (because the impact of the drift rate difference is magnified as RT increases). Therefore, we predicted that the RT distribution would be more positively skewed in the MDD group. Next, we fit the HDDM to the data. Our second prediction was that the model would reveal slower drift rates in the MDD group. We also examined the non-decision time parameter to determine whether MDD affected sensorimotor processes.

We also report on *decision threshold* (Fig. 1*a*). Wider thresholds correspond to greater distance between the boundaries, which means more evidence must accumulate for the participant to respond. Manipulations that increase response caution—such as prioritizing accuracy over speed (Ratcliff and Rouder, 1998)—result in slower RTs and wider thresholds. We did not have an *a priori* hypothesis about this parameter, but because depressed adults often respond more slowly than controls in this task (Pizzagalli *et al.*, 2008), wider thresholds in MDD might be expected.

The drift process can also begin from a position shifted towards either boundary to an extent captured by the *starting point bias*. White and Poldrack (2014) found that a response bias in behavior typically maps onto a starting point bias in the DDM: the drift process begins closer to one boundary such that even a little evidence in favor of that option will elicit a response. Given the asymmetric reinforcement rate, we expected the starting point to be biased towards the more frequently rewarded 'rich' boundary. If so, then fast responses should predominantly be rich responses because if the starting point is close to the rich boundary, then the accumulator will need to travel just a short distance to reach it. To our knowledge, no PRT study has examined whether the response bias is stronger for fast *v.* slow RTs, as these considerations predict. Finally, we examined the split-half reliability of our measures to assess internal consistency (Levinson *et al.*, 2017; Luking *et al.*, 2017).

Method

Participants

Study 1

Data were collected from 296 adults with MDD and 40 healthy participants in the multi-site 'Establishing Moderators and

Biosignatures of Antidepressant Response in Clinical Care' (EMBARC) study. EMBARC was a randomized, placebo-controlled trial of sertraline. The goal was to identify predictors of treatment response, thus a variety of measures—including the PRT—was administered before randomization to drug or placebo. Only pre-treatment data were included in this analysis. Depressed participants were outpatients between 18 and 65 who met DSM-IV criteria for MDD, as assessed by the Structured Clinical Interview for DSM-IV Disorders (SCID; First *et al.*, 2002). The SCID was administered by graduate-level clinicians. To limit heterogeneity, depressed participants had to report early onset (before age 30), chronicity (current episode >2 years duration), or recurrence (two or more episodes). Data were collected at Columbia University Medical Center, Massachusetts General Hospital (MGH)/McLean Hospital, University of Texas Southwestern Medical Center, and the University of Michigan. Participants consented to a protocol approved by local Institutional Review Boards. See Trivedi *et al.* (2016) for details.

Study 2

To determine if the results from Study 1 could be replicated, we reanalyzed data from 23 unmedicated adults with MDD and 25 healthy controls, previously published in Pizzagalli *et al.* (2008). The depressed participants were recruited from treatment studies at MGH, whereas controls came from the community. Depressed participants met DSM-IV criteria for current MDD based on the SCID, which was administered by trained psychiatrists. All participants consented to a protocol approved by the Harvard University and Partners Healthcare IRBs. See the original publication for details.

Self-report

Participants provided demographic information and completed the Mood and Anxiety Symptoms Questionnaire (MASQ: Watson *et al.*, 1995); Studies 1 and 2 used 30 and 62 item versions, respectively. Both include scales for anhedonic depression (MASQ-AD) and anxious arousal (MASQ-AA). The shorter version includes a 'general distress' (MASQ-GD) scale, whereas the longer one has scales for general distress due to depression (MASQ-GDD) and anxiety (MASQ-GDA). All participants except the Study 2 controls were administered the 17-item Hamilton Rating Scale for Depression (HRSD; Hamilton, 1960), which was scored by the diagnostic interviewer. Both studies included additional questionnaires not considered here.

Task

E-Prime version 1.1 (Psychology Software Tools, Sharpsburg PA) was used to present the PRT, which is shown in Fig. 1*b*. On each PRT trial, participants view a schematic face for 500 ms. A mouth is then shown for 100 ms, and the task is to indicate by button press whether the mouth was long (13.0 mm) or short (11.5 mm). Correct identifications of one length (the 'rich' stimulus) are rewarded three times more often than correct identifications of the other length (the 'lean' stimulus). In Study 1, participants completed two 100-trial blocks in which they pressed 'c' or 'm' on a keyboard to report seeing the short or long mouth, which served as the rich and lean stimuli, respectively. The PRT was programmed to deliver 40 20-cent rewards per block—30 rich rewards *v.* 10 lean rewards—although there was some variability due to differences in behavior. No feedback was presented on non-reward trials. Study 2 was very similar, except that participants completed three 100-trial blocks, rewards were worth 5 cents, and the rich and lean keys were counterbalanced across subjects. See Pizzagalli *et al.* (2005) for additional task details.

Analyses

Quality Control (QC).

Trials were excluded for extreme RTs (<150 ms, >2500 ms), or if the remaining (log transformed) RT exceeded the participant's mean ± 3S.D. Participants' datasets were excluded if, in any block, there were more than 20 RT outliers, fewer than 24 rich or 7 lean rewards, a rich-to-lean reward ratio lower than 2.5, or lower than 40% correct accuracy. In Study 1, 258 depressed adults and 36 controls passed the QC criteria. Study 2 data are from participants who passed these QC checks.

Quantile-probability plots

We used quantile-probability plots to determine whether the RT distribution was more positively skewed in the MDD group. To generate these plots, we binned responses by RT quantile. The quantiles used, from fastest to slowest, were 0.1, 0.3, 0.5, 0.7, 0.9, and 0.995. Each quantile served as the RT ceiling for its bin, with the previous quantile as the floor. We used the 0.005 quantile as the floor for the 0.1 quantile, so that the sizes of the fastest RT bin (0.100–0.005 = 0.095) and slowest RT bin (0.995–0.900 = 0.095) would be identical. For each bin, we plotted the percent correct and incorrect on the right and left sides of the *x*-axis, respectively, with the mean correct/incorrect RT plotted on the *y*-axis.

HDDM

Computational modeling was performed in Jupyter Notebooks (Kluyver *et al.*, 2016) and fit to trial-level RT and response data following published recommendations (Wiecki *et al.*, 2013). Briefly, the HDDM is initialized with priors that reflect established findings in the literature, and then the Markov Chain Monte Carlo method fits the model to the data by estimating the joint posterior distribution for all parameters. All HDDM parameters were allowed to vary by group. We drew 10 000 samples from the posterior distribution, discarding the first 1000 'burn-in' samples (Kruschke, 2014). Trace and autocorrelation plots were inspected to assess convergence. To evaluate model quality, the estimated parameters were used to generate simulated data (posterior predictive checks). Summary statistics from the actual data fell well within 95% intervals of the simulated data, indicating a good fit. We examined the between-group overlap of the posterior distributions for all parameters, defining significance as less than 5% overlap. Because these are comparisons of Bayesian posterior distributions, we report the HDDM outcomes as q-values rather than p-values.

Signal-detection analyses

We computed the signal detection metrics response bias and discriminability using published formulas (Pizzagalli *et al.*, 2005), analyzing them in *Group x Block* ANOVAs implemented in the R (R Core Team, 2018) package afex (Singmann *et al.*, 2016). The Greenhouse-Geisser correction was applied to all ANOVAs. To determine how signal detection measures related to HDDM parameters, we computed linear mixed models with response bias or discriminability as the dependent variable, HDDM parameters and *Group* as predictors, and *Subject* as a random effect

Table 1. Demographics and self-report data

Variable	Healthy mean (s.d.)	Depressed mean (s.d.)	*P*	Effect size
Study 1				
Gender	21 f, 15 m	172 f, 86 m	0.324	0.06
Age	37.25 (14.71)	36.81 (13.29)	0.853	0.03
Education (years)	15.21 (2.29)	15.07 (2.61)	0.755	0.06
HRSD	0.69 (0.90)	18.60 (4.44)	<0.001	6.71
MASQ-AD	24.89 (6.92)	43.73 (5.45)	<0.001	3.05
MASQ-AA	10.80 (1.08)	17.60 (5.67)	<0.001	2.02
MASQ-GD	12.09 (2.63)	32.36 (8.04)	<0.001	3.80
Study 2				
Gender	11 f, 14 m	10 f, 13 m	0.971	0.01
Age	38.36 (10.76)	43.65 (9.55)	0.079	0.52
% College education	64.00	65.22	0.999	0.01
HRSD	--	19.40 (3.30)	--	--
MASQ-AD	51.52 (12.60)	91.00 (7.60)	<0.001	3.91
MASQ-AA	18.76 (5.19)	25.30 (11.32)	0.016	0.79
MASQ-GDD	15.64 (5.22)	40.70 (10.71)	<0.001	3.15
MASQ-GDA	14.16 (4.34)	23.26 (8.14)	<0.001	1.46

Note. *p*-values reflect between-group *t* tests except for gender and % of participants with college education in Study 2, which were evaluated by chi-square. All tests were two-sided. Effect size: Cramer's V for Gender, otherwise Cohen's *d*. HRSD data were not obtained from Study 2 controls. As detailed in the text, different MASQ versions were used in each study (Study 1: 30 items; Study 2: 62 items)
f, female; m, male; HRSD, Hamilton Rating Scale for Depression (17 items); MASQ, Mood and Anxiety Symptom Questionnaire (AD, anhedonic depression; AA, anxious arousal; GD, general distress; GDD, general distress, depression; GDA, general distress, anxiety).

(Bates *et al*., 2014). *Site* was included as a covariate in all Study 1 analyses.

Individual differences

We computed Pearson correlations between MASQ scores and HDDM parameters in depressed adults, to determine if variability in symptoms was related to decision-making. We also used regressions to determine if there was a group difference in cumulative rewards earned, and to see if such a group difference could be explained by the HDDM parameters. Because the samples were larger in Study 1 and the methods varied somewhat across the studies, we restricted these analyses to Study 1.

Psychometrics

To assess split-half reliability, we computed response bias and discriminability separately for odd and even trials for each participant and ran the HDDM on the odd and even trials to generate two sets of model parameters per person. We then computed Pearson correlations between the odd and even results and applied the Spearman-Brown prophesy formula ($2r/(1+r)$) to quantify internal consistency (Luking *et al*., 2017).

Results

Demographics and self-report

Groups did not differ on age, education or gender (Table 1). As expected, depressed adults reported more anhedonia, higher anxiety, and greater general distress on the MASQ than did controls. Based on HRSD cut-offs (Zimmerman *et al*., 2013), adults with MDD were moderately depressed.

Study 1

Quantile-probability plots

Figure 2 shows the Study 1 quantile-probability plot. This depicts the mean percentage of correct (plotted to the right) and incorrect (plotted to left) responses to the rich (circles) and lean (crosses) stimuli as a function of RT quartile (plotted on the *y*-axis) for the two groups. This figure supports two main conclusions. First, as expected the RT distribution was more skewed in depressed adults. Specifically, although the mean ± s.d. RT in the 0.1 quantile bin, averaged over stimulus type and response accuracy, was 5 ms slower in depressed (242.73 ± 17.22 ms) *v*. healthy (237.37 ± 10.87 ms) participants, $t(239) = 2.23$, $p = 0.031$, Cohen's $d = 0.38$, by the 0.995 quantile bin this group difference had grown to 158 ms (MDD: 841.00 ± 76.33 ms; controls: 683.07 ± 48.54 ms; $t(240) = 15.34$, $p < 0.001$, $d = 2.53$), a 32-fold increase (note that degrees of freedom vary in the quantile-probability analyses as not all participants contributed responses to the more extreme bins). Second, Fig. 2 reveals that response bias was constrained to the fastest 30% of responses. Notice that the circles are farther to the right than the crosses for the 0.100 and 0.300 quantiles, indicating higher accuracy for responses to the rich *v*. lean stimulus. This reflects response bias: when replying quickly, participants pressed 'rich' more than 'lean' and so achieved a higher proportion correct for the rich *v*. lean stimulus. This effect is absent for the remaining quantiles. Thus, the PRT induced a response bias but this was constrained to fast RTs.

To confirm these impressions, *Group x Stimulus* ANOVAs were run on response accuracy for fast (RT ⩽ 0.3 quantile) and slow (RT > 0.3 quantile) responses. As expected from Fig. 2, for fast RTs the *Stimulus* effect was strong, $F(1, 267) = 89.44$, $p <$

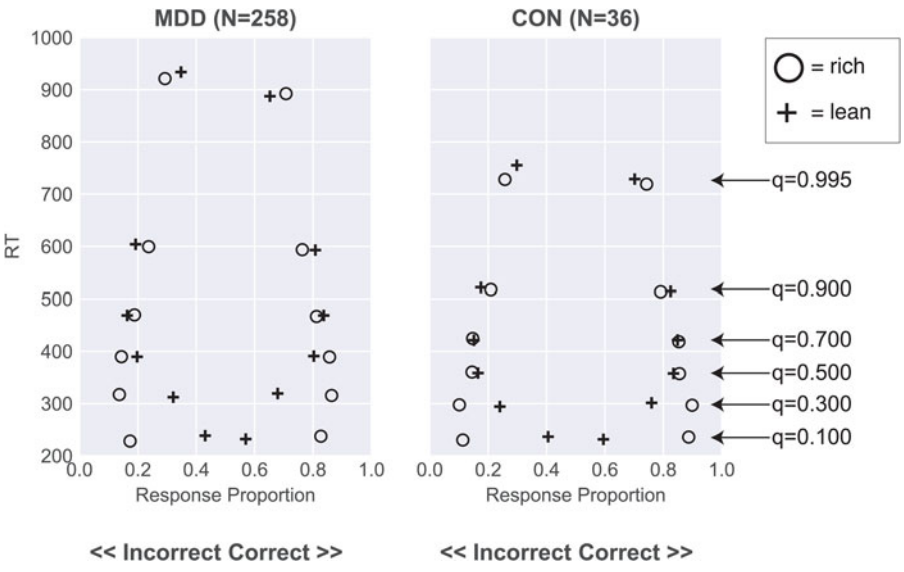

Fig. 2. Quantile-probability plots: Study 1. Percent correct (right) and incorrect (left) for rich (circles) and lean (crosses) stimuli as a function of RT quantiles, for adults with MDD (left column) and healthy controls (right column). The six quantiles are marked on the controls' data; they are shifted upwards on the *y*-axis for the MDD group, with the magnitude of the shift increasing with longer response latency. The effect of stimulus type on accuracy (rich > lean) is restricted to the 0.1 and 0.3 quantiles, indicating that response bias is carried by fast RTs.

0.001, $\eta_p^2 = 0.25$, corresponding to higher mean ± S.E. accuracy for rich (84.15 ± 1.53%) *v.* lean (59.95 ± 1.53%) stimuli. An effect of *Group* was also found, $F(1, 267) = 13.62$, $p < 0.001$, $\eta_p^2 = 0.05$, reflecting lower accuracy in depressed (67.11 ± 1.56%) *v.* healthy (76.98 ± 1.56%) participants. By contrast, for slow RTs no significant results emerged, $Fs < 2.28$, $ps > 0.12$.

HDDM

Figure 3 shows the HDDM results. The posterior distributions revealed slower drift rates ($q < 0.001$) and wider decision thresholds ($q = 0.004$) in depressed adults, but no group differences in starting bias or non-decision time, $qs > 0.17$.

Signal-detection analyses

The response bias and discriminability data are in Figure S1. *Group x Block x Site* ANOVAs yielded no significant results, although there was a trend ($p = 0.06$) for worse discriminability in depressed adults (MDD: 0.59 ± 0.24; controls: 0.67 ± 0.24). The lack of a group difference in response bias was unexpected; exploratory *t* tests against zero revealed reliable biases in both groups (MDD: 0.09 ± 0.16; controls: 0.11 ± 0.13; $ts > 5$, $ps < 0.001$).

Prediction of signal-detection metrics by HDDM parameters

Linear mixed models were used to predict response bias and discriminability with *Group*, *Site*, and HDDM parameters. Starting point bias was the strongest predictor of response bias ($Z = 12.71$, $p < 0.001$), and drift rate was a remarkably strong predictor of discriminability ($Z = 45.56$, $p < 0.001$); these relationships are shown in Figure S2. Response bias was also predicted by drift rate ($Z = -5.17$, $p < 0.001$), and discriminability was also predicted by decision threshold ($Z = 14.81$, $p < 0.001$) and non-decision time ($Z = 2.99$, $p = 0.003$). *Group* did not predict either variable ($ps > 0.23$).

Individual differences

In the MDD group, weak relationships emerged between MASQ-AA scores and drift rate, $r = -0.15$, $p = 0.02$, and between MASQ-GD scores and decision threshold, $r = -0.17$, $p = 0.01$. Neither of these relationships remains significant, however, when a Bonferroni-corrected alpha of 0.004 (0.5/12 comparisons) is applied.

Finally, *Group* predicted cumulative reward, $\beta = -0.13$, $p = 0.029$. Depressed adults received fewer rewards than controls did, although the difference was small (controls: 79.08 ± 1.27; MDD: 78.25 ± 2.22) as the PRT is programmed to equate reward delivery across participants. Adding the HDDM parameters improved the model, $\Delta R^2 = 0.34$, $F(4, 288) = 38.41$, $p < 0.001$, and with these parameters included the effect of *Group* was no longer significant ($\beta = -0.05$, $p = 0.30$). Instead, drift rate ($\beta = 0.39$), non-decision time ($\beta = 0.29$), and starting point bias ($\beta = 0.18$)—but not decision threshold ($\beta = 0.09$)—emerged as predictors of cumulative reward, $ps < 0.001$. The fact that drift rate strongly predicted cumulative reward is sensible because individuals with high drift rates have high discriminability (Figure S2B), which allows them to respond accurately and thus efficiently harvest rewards on rich and lean trials.

Study 2

The next goal was to determine if the quantile-probability plot and HDDM findings from Study 1 would replicate in a dataset characterized by a group difference in response bias. To this end, we reanalyzed data published by Pizzagalli *et al.* (2008), referred to as Study 2.

Quantile-probability plots

Figure 4 shows the quantile-probability plot. As in Study 1, RTs were drastically more skewed in the MDD group. Specifically, while the fastest RTs (0.1 quantile) were 34 ms slower in depressed (298.64 ± 21.09 ms) *v.* healthy (264.31 ± 17.62 ms) adults, $t(34) = 5.22$, $p < 0.001$, $d = 1.77$, by the 0.995 quantile bin this group difference had grown to 339 ms (|MDD: 1365.85 ± 91.57 ms; controls: 1026.81 ± 96.92 ms), $t(36) = 11.08$, $p < 0.001$, $d = 3.60$, a nearly 10-fold increase.

The restriction of response bias to fast RTs was partially replicated. As depicted in Fig. 4, the MDD group showed an accuracy advantage for the rich stimulus in the 0.1 quantile that was reduced in the 0.3 quantile and absent thereafter. This mirrors Study 1. By contrast, in controls the rich > lean accuracy effect was visible at every quantile—notice the consistent horizontal separation between circles and crosses. Accordingly, a *Group x Stimulus* ANOVA on accuracy for fast RTs (⩽0.3 quantile)

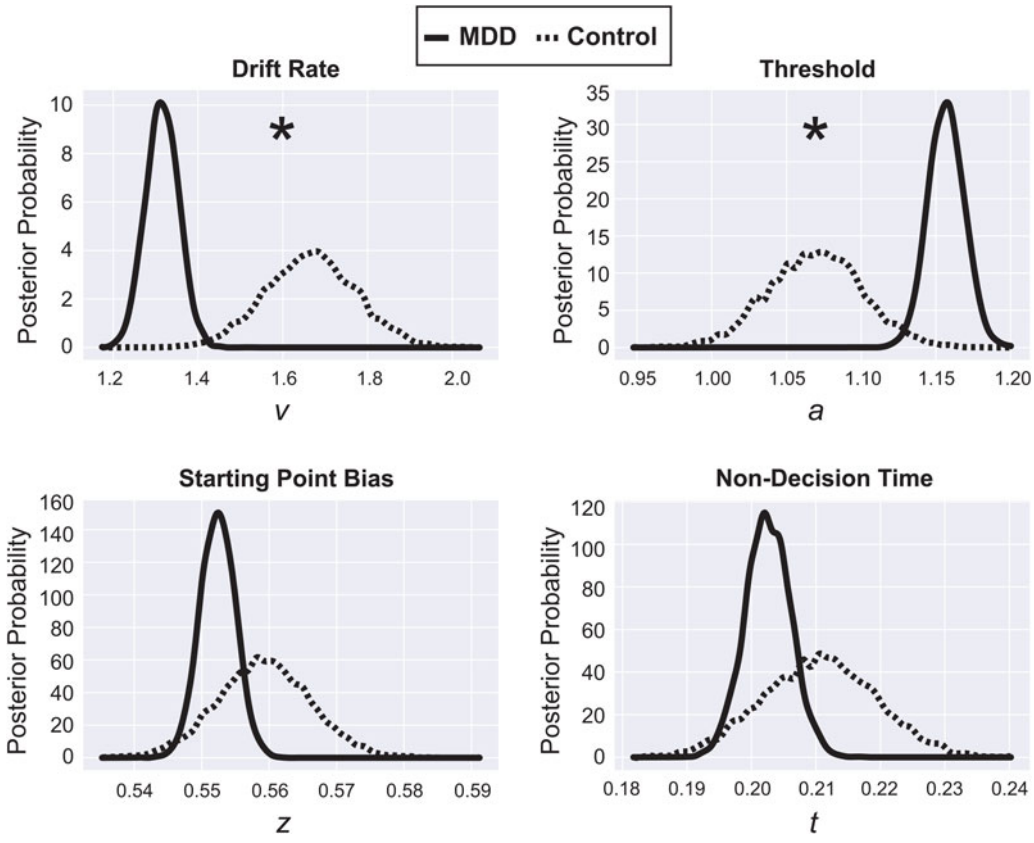

Fig. 3. HDDM results: Study 1. Plots of posterior probabilities for HDDM parameters. Relative to results in the controls (dashed lines), drift rate was reduced and threshold width was increased in the MDD group (solid lines). There were no group differences in starting point bias or non-decision time. *$q < 0.005$.

revealed a main effect of *Stimulus*, $F(1, 41) = 22.40$, $p < 0.001$, $\eta_p^2 = 0.35$, due to higher mean ± S.E. percent correct in response to the rich (89.98 ± 3.47%) *v.* lean (65.99 ± 3.47%) stimulus across both groups. By contrast, an analogous ANOVA on slower RTs (>0.3 quantile) revealed a *Group x Stimulus* interaction, $F(1, 46) = 5.42$, $p = 0.02$, $\eta_p^2 = 0.11$, due to a rich > lean effect that was significant by post-hoc Tukey test only in controls (rich: 86.21 ± 1.83%; lean: 75.05 ± 1.83%; $t = 4.60$, $p < 0.001$).

HDDM

Figure 5 shows that the HDDM again revealed slower drift rates ($q = 0.037$) and higher decision thresholds ($q < 0.001$) in depressed adults. There was no group difference in starting bias or non-decision time, $qs > 0.23$. These results replicate Study 1.

As in Study 1, the strongest predictors of response bias and discriminability were starting point bias ($Z = 3.29$, $p = 0.002$) and drift rate ($Z = 13.26$, $p < 0.001$), respectively (Figure S3). In contrast to Study 1, response bias was also predicted by Group ($Z = -2.12$, $p = 0.040$). As in Study 1, discriminability was also predicted by decision threshold ($Z = 3.58$) and non-decision time ($Z = 3.45$), $ps < 0.001$.

Psychometrics

Split-half reliabilities for Studies 1 and 2 are in Figures S4 and S5. Reliability was good for response bias (SBs > 0.677) and discriminability (SBs > 0.844). For the HDDM, reliability was good for starting bias (SBs > 0.777) and outstanding for all other parameters (SBs > 0.910).

Discussion

This study found support for the hypothesis that decision-making deficits are present in MDD. First, RT distributions were more positively skewed in depressed *v.* healthy adults. Indeed, although there were differences in the fastest RTs, judging by Cohen's *d* the group differences for the slowest RTs were 6.66 (2.53/0.38) and 2.03 (3.60/1.77) times larger in Studies 1 and 2. As detailed in the Introduction, this suggests slow evidence accumulation in MDD. Second, the HDDM returned a result consistent with this interpretation: drift rate was lower in depressed adults. Decision thresholds were also wider in the MDD groups, potentially indicating a cautious response style. However, the effect on drift rate was practically more important as drift rate predicted cumulative reward totals, which were reduced in MDD in Study 1, while decision threshold did not. Neither study returned a group difference in non-decision time, thus the results do not appear to reflect group differences in sensorimotor processes. There were also no group differences in starting point bias, although starting points were consistently shifted away from the midpoint and towards the rich boundary, consistent with the asymmetric reinforcement rate used in the PRT. In summary, this study established a decision-making deficit in MDD across two independent samples, and localized that deficit to the evidence accumulation process, indexed by drift rate.

A goal for future work is to establish the generality of these results. They may reflect a domain-general effect of MDD that would be detectable across various tasks and thus could help explain the broad cognitive deficits seen in MDD. Additional

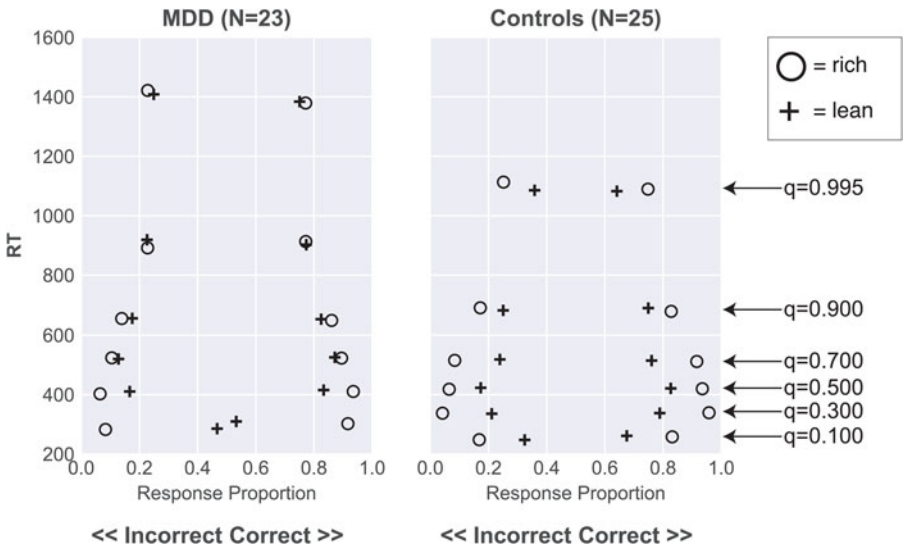

Fig. 4. Quantile-probability plots: Study 2 (Pizzagalli *et al.*, 2008). Percent correct (right) and incorrect (left) for rich (circles) and lean (crosses) stimuli are shown as a function of RT quantiles, for adults with MDD (left column) and healthy controls (right column). The six quantiles are marked on the controls' data; they are shifted upwards on the *y*-axis for the MDD group, with the magnitude of the shift increasing with longer response latency. The effect of stimulus type on accuracy (rich > lean) is apparent in the 0.1 and 0.3 quantiles in the MDD group, but it is evident at every quantile for the controls.

research using different tasks is needed to investigate this possibility, but limited extant work is encouraging. For instance, a recognition memory study identified a drift rate advantage for positive *v.* negative material that was reduced in dysphoric students (White *et al.*, 2009). If similar results can be obtained in MDD, that would increase confidence that depression reliably reduces the speed of evidence accumulation, perhaps especially when stimuli are positively valenced.

However, the results may depend on details of the PRT. In particular, the evidence accumulated during decision-making is influenced by stimulus properties, and drift rates are typically lower for short *v.* long duration presentations (Thapar *et al.*, 2003). It is thus possible that the reduced drift rates in the current MDD groups reflect difficulty extracting high-quality evidence from rapidly presented stimuli. If so, then the results would not generalize to tasks with longer duration stimuli. Additional work is needed to distinguish between these possibilities.

This research also provides new insights into the PRT, which is widely used to study reward processing across different clinical groups (Barch *et al.*, 2017) and non-human animals (Der-Avakian *et al.*, 2013). First, we found that response bias was stronger for fast *v.* slow RTs; in Study 1, the response bias was confined to the 0.1 and 0.3 quantiles, which means that approximately 70% of responses were not biased. Similarly, in both studies starting point bias (in the HDDM) was the strongest predictor of response bias (in the PRT). This helps explain why biased responses are fast: the drift process starts close to the rich boundary such that minimal evidence needs to accumulate before a rich response is made. These findings are consistent with results from White and Poldrack (2014), but they are the first to demonstrate a dependency between response bias and RT in the PRT. Second, the internal consistency of response bias, discriminability, and the HDDM parameters was good to excellent. We conclude that PRT data are reliable, especially if the HDDM is used to extract estimates of underlying processes.

These results also raise several questions. First, given that a reduced response bias is often found in adults who are depressed or at risk for depression, why did a group difference in response bias emerge in Study 2 but not Study 1? The quantile-probability plots provide some insight. White and Poldrack (2014) distinguished between biases confined to response execution (response biases) *v.* stimulus processing (stimulus biases), showing that response biases affect fast but not slow RTs, while stimulus biases affect fast and slow RTs. On this analysis, both Study 1 groups and the Study 2 MDD group showed response biases: they responded 'rich' more than 'lean' only when replying quickly. By contrast, the Study 2 controls developed a stimulus bias, showing a rich > lean accuracy advantage at every quantile. The presence of biased responses for slow RTs in controls but not depressed adults explains why a group difference emerged in Study 2, and the lack of this effect explains why no group difference emerged in Study 1. Of course, this does not explain why only the Study 2 controls developed a stimulus bias, but it suggests that researchers using the PRT should distinguish between response and stimulus bias in order to determine why one *v.* the other is more likely to emerge.

Another question that remains unanswered is, what causes low drift rates in MDD? Two candidate hypotheses include reduced integrity of white matter pathways that enable evidence accumulation from distant brain regions (Madden *et al.*, 2008), or reduction in cortical dopamine levels that can support fast drift rates (Beste *et al.*, 2018). These hypotheses are informed by prior studies linking MDD to white matter abnormalities (Jiang *et al.*, 2017) and dopamine dysfunction (Treadway and Zald, 2011), but they await empirical test. An important future direction is thus to pair DDM-based analysis of behavior with electrophysiological and neuroimaging data collected from patient groups, to study the pathophysiology underlying deficits like the one reported here.

Finally, it is worth underscoring the fact that the HDDM detected effects of MDD on cognition that were not obvious in the traditional PRT analyses. For example, in Study 1 the MDD group showed poorer discriminability than the controls, but only at trend levels. Discriminability depends on a participant's ability to rapidly accumulate evidence in favor of each response option, captured by drift rate, but also on the criterion the participant uses to judge when there is sufficient evidence to make an accurate response, captured by decision threshold. The HDDM revealed that drift rates were markedly lower in depressed *v.* healthy adults, but decision thresholds were also wider. These two results appear to have roughly counterbalanced each other, leading to the weak group difference in discriminability.

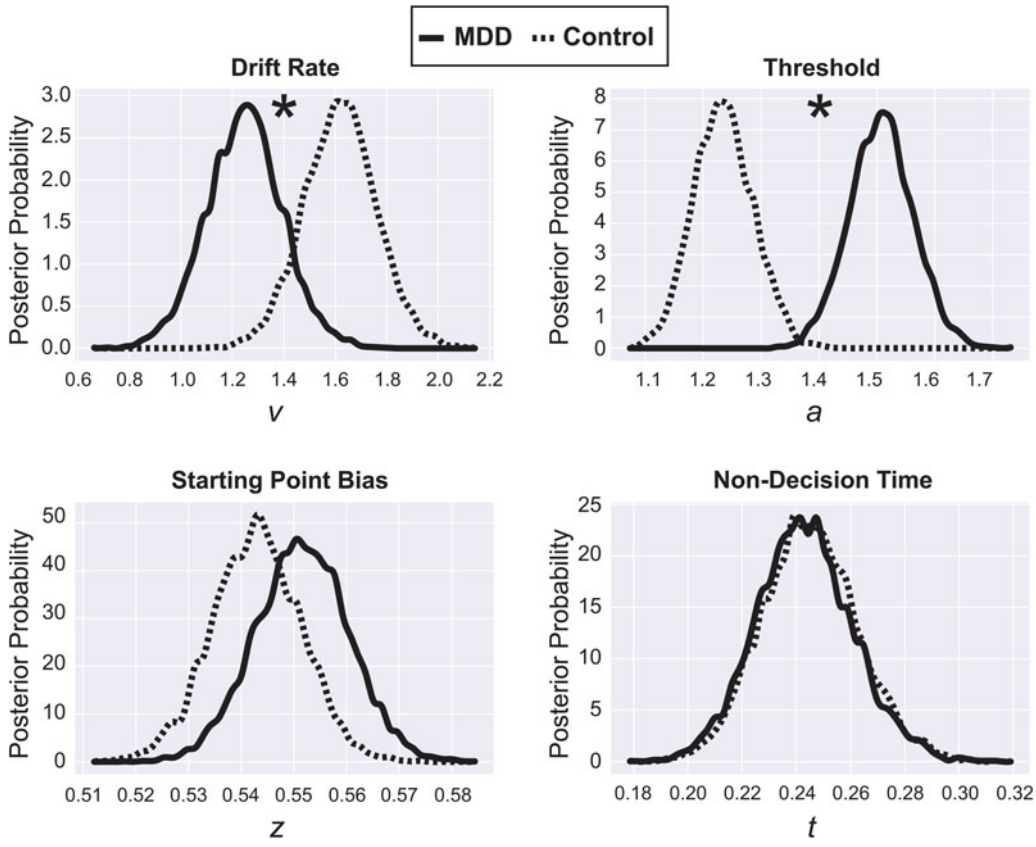

Fig. 5. HDDM results: Study 2 (Pizzagalli *et al.*, 2008). Plots of posterior probabilities for HDDM parameters. Relative to results in the controls (dashed lines), drift rate was reduced and threshold width was increased in the MDD group (solid lines). There were no group differences in starting point bias or non-decision time. *q < 0.038.

Nevertheless, they were clearly dissociable—for example, drift rate predicted cumulative reward totals while decision thresholds did not. This is an example of the explanatory power of computational modeling: it can identify the distinct contributions that different cognitive processes make to behavior.

In conclusion, in two PRT studies, we found evidence of decision-making deficits in unmedicated adults with MDD. These deficits were localized to the evidence accumulation process and were reflected in lower drift rate. It would be valuable to conduct additional studies to determine if these results are specific to tasks that use briefly presented stimuli, such as the PRT, or if they generalize to those that use longer stimulus durations. Moreover, pairing this approach with collection of electrophysiological or neuroimaging data could advance our understanding of pathophysiology. Finally, if we can identify a domain-general decision-making deficit in MDD, then we should gain valuable insight into the broad cognitive deficits that accompany depression.

Declaration of conflicts of interest

In the past 3 years, *Dr Dillon* has provided consulting services to Pfizer, Inc., for projects unrelated to this report. Over the past 3 years, *Dr Pizzagalli* has received consulting fees from Akili Interactive Labs, BlackThorn Therapeutics, Boehringer Ingelheim, Posit Science, and Takeda Pharmaceuticals for activities unrelated to the current research. *Dr Trivedi* reports the following lifetime disclosures: research support from the Agency for Healthcare Research and Quality, Cyberonics Inc., National Alliance for Research in Schizophrenia and Depression, National Institute of Mental Health, National Institute on Drug Abuse, National Institute of Diabetes and Digestive and Kidney Diseases, Johnson & Johnson, and consulting and speaker fees from Abbott Laboratories Inc., Akzo (Organon Pharmaceuticals Inc.), Allergan Sales LLC, Alkermes, AstraZeneca, Axon Advisors, Brintellix, Bristol-Myers Squibb Company, Cephalon Inc., Cerecor, Eli Lilly & Company, Evotec, Fabre Kramer Pharmaceuticals Inc., Forest Pharmaceuticals, GlaxoSmithKline, Health Research Associates, Johnson & Johnson, Lundbeck, MedAvante Medscape, Medtronic, Merck, Mitsubishi Tanabe Pharma Development America Inc., MSI Methylation Sciences Inc., Nestle Health Science-PamLab Inc., Naurex, Neuronetics, One Carbon Therapeutics Ltd., Otsuka Pharmaceuticals, Pamlab, Parke-Davis Pharmaceuticals Inc., Pfizer Inc., PgxHealth, Phoenix Marketing Solutions, Rexahn Pharmaceuticals, Ridge Diagnostics, Roche Products Ltd., Sepracor, SHIRE Development, Sierra, SK Life and Science, Sunovion, Takeda, Tal Medical/Puretech Venture, Targacept, Transcept, VantagePoint, Vivus, and Wyeth-Ayerst Laboratories. All other authors report no biomedical financial interests or potential conflicts of interest.

Supplementary material. The supplementary material for this article can be found at https://doi.org/10.1017/S0033291719001570

Author ORCIDs. Daniel G. Dillon, 0000-0002-1977-700X.

Acknowledgements. The analysis was made possible by funding from McLean Hospital and a grant from NIMH (R00MH094438) awarded to Dr Dillon. Data collection and prior analyses were made possible by additional

funding from NIMH (Pizzagalli: R01MH68376; Trivedi: U01MH092221; McGrath, Parsey, Weissman: U01MH092250). The authors are solely responsible for the content of this manuscript, which does not necessarily represent the official views of the NIH.

Ethical standards. The authors assert that all procedures contributing to this work comply with the ethical standards of the relevant national and institutional committees on human experimentation and with the Helsinki Declaration of 1975, as revised in 2008.

References

Barch DM, Carter CS, Gold JM, Johnson SL, Kring AM, MacDonald III AW, Pizzagalli DA, Ragland JD, Silverstein SM and Strauss ME (2017) Explicit and implicit reinforcement learning across the psychosis spectrum. *Journal of Abnormal Psychology* **126**, 694–711.

Bates D, Mächler M, Bolker BM and Walker SC (2014) Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* **67**, 1–48.

Beste C, Adelhöfer N, Gohil K, Passow S, Roessner V and Li SC (2018) Dopamine modulates the efficiency of sensory evidence accumulation during perceptual decision making. *International Journal of Neuropsychopharmacology* **21**, 649–655.

Biringer E, Mykletun A, Sundet K, Kroken R, Stordal KI and Lund A (2007) A longitudinal analysis of neurocognitive function in unipolar depression. *Journal of Clinical and Experimental Neuropsychology* **29**, 879–891.

Burt DB, Zembar MJ and Niederehe G (1995) Depression and memory impairment: a meta-analysis of the association, its pattern, and specificity. *Psychological Bulletin* **117**, 285–305.

Der-Avakian A, D'souza MS, Pizzagalli DA and Markou A (2013) Assessment of reward responsiveness in the response bias probabilistic reward task in rats: implications for cross-species translational research. *Translational Psychiatry* **3**, e297.

Dillon DG, Dobbins IG and Pizzagalli DA (2013) Weak reward source memory in depression reflects blunted activation of VTA/SN and parahippocampus. *Social Cognitive and Affective Neuroscience* **9**, 1576–1583.

Douglas KM and Porter RJ (2009) Longitudinal assessment of neuropsychological function in major depression. *Australian and New Zealand Journal of Psychiatry* **43**, 1105–1117.

First MB, Spitzer RL, Gibbon M and Williams JBW (2002) *Structured Clinical Interview for DSM-IV-TR Axis I Disorders, Research Version, Patient Edition*. New York: (SCID-I/P) Biometrics Research, New York State Psychiatric Institute.

Frank MJ, Gagne C, Nyhus E, Masters S, Wiecki TV, Cavanagh JF and Badre D (2015) fMRI and EEG predictors of dynamic decision parameters during human reinforcement learning. *Journal of Neuroscience* **35**, 485–494.

Gold JI and Shadlen MN (2007) The neural basis of decision making. *Annual Review of Neuroscience* **30**, 535–574.

Hamilton M (1960) A rating scale for depression. *Journal of Neurology, Neurosurgery, and Psychiatry* **23**, 56–62.

Jiang J, Zhao YJ, Hu XY, Du MY, Chen ZQ, Wu M, Li KM, Zhu HY, Kumar P and Gong QY (2017) Microstructural brain abnormalities in medication-free patients with major depressive disorder: a systematic review and meta-analysis of diffusion tensor imaging. *Journal of Psychiatry & Neuroscience* **42**, 150–163.

Kluyver T, Ragan-Kelley B, Pérez F, Granger BE, Bussonnier M, Frederic J, Kelley K, Hamrick JB, Grout J, Corlay S and Ivanov P (2016) Jupyter notebooks-a publishing format for reproducible computational workflows. In Loizides F and Schmidt B (eds) *Positioning and Power in Academic Publishing: Players, Agents and Agendas*. IOS Press: Amsterdam, pp. 87–90.

Krajbich I, Lu D, Camerer C and Rangel A (2012) The attentional drift-diffusion model extends to simple purchasing decisions. *Frontiers in Psychology* **3**, 193.

Kruschke J (2014) *Doing Bayesian Data Analysis: A Tutorial with R, JAGS, and Stan*. Oxford: Elsevier Science.

Levinson AR, Speed BC, Infantolino ZP and Hajcak G (2017) Reliability of the electrocortical response to gains and losses in the doors task. *Psychophysiology* **54**, 601–607.

Liu WH, Roiser JP, Wang LZ, Zhu YH, Huang J, Neumann DL, Shum DH, Cheung EF and Chan RC (2016) Anhedonia is associated with blunted reward sensitivity in first-degree relatives of patients with major depression. *Journal of Affective Disorders* **190**, 640–648.

Luking KR, Nelson BD, Infantolino ZP, Sauder CL and Hajcak G (2017) Internal consistency of functional magnetic resonance imaging and electroencephalography measures of reward in late childhood and early adolescence. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging* **2**, 289–297.

MacQueen GM, Campbell S, McEwen BS, Macdonald K, Amano S, Joffe RT, Nahmias C, Young LT (2003) Course of illness, hippocampal function, and hippocampal volume in major depression. *Proceedings of the National Academy of Sciences*, USA **100**, 1387–1392.

Madden DJ, Spaniol J, Costello MC, Bucur B, White LE, Cabeza R, Davis SW, Dennis NA, Provenzale JM and Huettel SA (2008) Cerebral white matter integrity mediates adult age differences in cognitive performance. *Journal of Cognitive Neuroscience* **21**, 289–302.

Pizzagalli DA, Jahn AL and O'Shea JP (2005) Toward an objective characterization of an anhedonic phenotype: a signal-detection approach. *Biological Psychiatry* **57**, 319–327.

Pizzagalli DA, Iosifescu D, Hallett LA, Ratner KG and Fava M (2008) Reduced hedonic capacity in major depressive disorder: evidence from a probabilistic reward task. *Journal of Psychiatric Research* **43**, 76–87.

Ratcliff R (1978) A theory of memory retrieval. *Psychological Review* **85**, 59–108.

Ratcliff R and McKoon G (2008) The diffusion decision model: theory and data for two-choice decision tasks. *Neural Computation* **20**, 873–922.

Ratcliff R and Rouder JN (1998) Modeling response times for two-choice decisions. *Psychological Science* **9**, 347–356.

Ratcliff R and Smith PL (2004) A comparison of sequential sampling models for two-choice reaction time. *Psychological Review* **111**, 333–367.

R Core Team (2018) *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. (http://www.R-project.org/).

Rock PL, Roiser JP, Riedel WJ and Blackwell AD (2014) Cognitive impairment in depression: a systematic review and meta-analysis. *Psychological Medicine* **44**, 2029–2040.

Shadlen MN and Kiani R (2013) Decision making as a window on cognition. *Neuron* **80**, 791–806.

Singmann H, Bolker B, Westfall J, Aust F, Højsgaard S, Fox J, Lawrence MA, Mertens U and Love J (2016) afex: analysis of factorial experiments. R package version 0.16-1.

Snyder HR (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.

Thapar A, Ratcliff R and McKoon G (2003) A diffusion model analysis of the effects of aging on letter discrimination. *Psychology and Aging* **18**, 415–429.

Treadway MT and Zald DH (2011) Reconsidering anhedonia in depression: lessons from translational neuroscience. *Neuroscience & Biobehavioral Reviews* **35**, 537–555.

Trivedi MH, McGrath PJ, Fava M, Parsey RV, Kurian BT, Phillips ML, Oquendo MA, Bruder G, Pizzagalli D, Toups M and Cooper C (2016) Establishing moderators and biosignatures of antidepressant response in clinical care (EMBARC): rationale and design. *Journal of Psychiatric Research* **78**, 11–23.

Vrieze E, Pizzagalli DA, Demyttenaere K, Hompes T, Sienaert P, de Boer P, Schmidt M and Claes S (2013) Reduced reward learning predicts outcome in major depressive disorder. *Biological Psychiatry* **73**, 639–645.

Watson D, Weber K, Assenheimer JS, Clark LA, Strauss ME and McCormick RA (1995) Testing a tripartite model: i. Evaluating the convergent and discriminant validity of anxiety and depression symptom scales. *Journal of Abnormal Psychology* **104**, 3–14.

White CN and Poldrack RA (2014) Decomposing bias in different types of simple decisions. *Journal of Experimental Psychology: Learning, Memory, and Cognition* **40**, 385–398.

White C, Ratcliff R, Vasey M and McKoon G (2009) Dysphoria and memory for emotional material: a diffusion-model analysis. *Cognition and Emotion* **23**, 181–205.

White CN, Ratcliff R, Vasey MW and McKoon G (2010) Using diffusion models to understand clinical disorders. *Journal of Mathematical Psychology* **54**, 39–52.

Wiecki TV, Sofer I and Frank MJ (2013) HDDM: hierarchical Bayesian estimation of the drift-diffusion model in python. *Frontiers in Neuroinformatics* **7**, 14.

Zakzanis KK, Leach L and Kaplan E (1998) On the nature and pattern of neurocognitive function in major depressive disorder. *Neuropsychiatry, Neuropsychology, & Behavioral Neurology* **11**, 111–119.

Zimmerman M, Martinez JH, Young D, Chelminski I and Dalrymple K (2013) Severity classification on the Hamilton depression rating scale. *Journal of Affective Disorders* **150**, 384–388.