



Self-judgment dissected: A computational modeling analysis of self-referential processing and its relationship to trait mindfulness facets and depression symptoms

Peter F. Hitchcock¹ · Willoughby B. Britton² · Kahini P. Mehta^{1,3} · Michael J. Frank^{1,4}

Accepted: 29 August 2022 / Published online: 27 September 2022
© The Psychonomic Society, Inc. 2022

Abstract

Cognitive theories of depression, and mindfulness theories of well-being, converge on the notion that self-judgment plays a critical role in mental health. However, these theories have rarely been tested via tasks and computational modeling analyses that can disentangle the information processes operative in self-judgments. We applied a drift-diffusion computational model to the self-referential encoding task (SRET) collected before and after an 8-week mindfulness intervention ($n = 96$). A drift-rate regression parameter representing positive—relative to negative—self-referential judgment strength positively related to mindful awareness and inversely related to depression, both at baseline and over time; however, this parameter did not significantly relate to the interaction between mindful awareness and nonjudgmentalness. At the level of individual depression symptoms, at baseline, a spectrum of symptoms (inversely) correlated with the drift-rate regression parameter, suggesting that many distinct depression symptoms relate to valenced self-judgment between subjects. By contrast, over the intervention, changes in only a smaller subset of anhedonia-related depression symptoms showed substantial relationships with this parameter. Both behavioral and model-derived measures showed modest split-half and test-retest correlations. Results support cognitive theories that implicate self-judgment in depression and mindfulness theories, which imply that mindful awareness should lead to more positive self-views.

Keywords Computational modeling · Mindfulness · Depression · Self-referent encoding task · Drift diffusion model

Introduction

Cognitive theories of depression posit that people who make frequent and rapid negative—and infrequent and slow positive—self-judgments are vulnerable to depression (Disner et al., 2011; LeMoult & Gotlib, 2019; Wisco, 2009). Similarly, mindfulness-based interventions are thought to increase well-being and mental health by promoting

equanimous, unbiased self-views (Desbordes et al., 2014; Hayes et al., 2011; Vago & Silbersweig, 2012). These converging viewpoints have motivated the development of tasks to disentangle the information processes involved in self-judgment (Britton et al., 2021; Dainer-Best et al., 2018c; Lawrence et al., 2022; Miller et al., 2013; Siegle et al., 2006). The most prominent such task is the self-referent encoding task (SRET), which involves making yes/no decisions about whether negative (e.g., “horrible”) or positive (e.g., “wonderful”) adjectives describe oneself (Dainer-Best et al., 2018b; Derry & Kuiper, 1981).

Recent studies applied computational modeling to the SRET (and similar) tasks, with the goal of pinpointing the information-processing mechanisms (i.e., computations) responsible for individual differences in self-judgments (Allison et al., 2021; Beevers et al., 2019; Cataldo et al., 2022; Dainer-Best et al., 2018a, b; Disner et al., 2017; Price et al., 2021). By forcing specificity about how different psychological mechanisms leave distinguishable signatures in behavioral data, and facilitating convergence between

✉ Peter F. Hitchcock
peter_hitchcock@brown.edu

¹ Department of Cognitive, Linguistic, and Psychological Sciences, Brown University, Providence, RI, USA

² Warren Alpert Medical School, Brown University, Providence, RI, USA

³ Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA

⁴ Carney Institute for Brain Science, Brown University, Providence, RI, USA

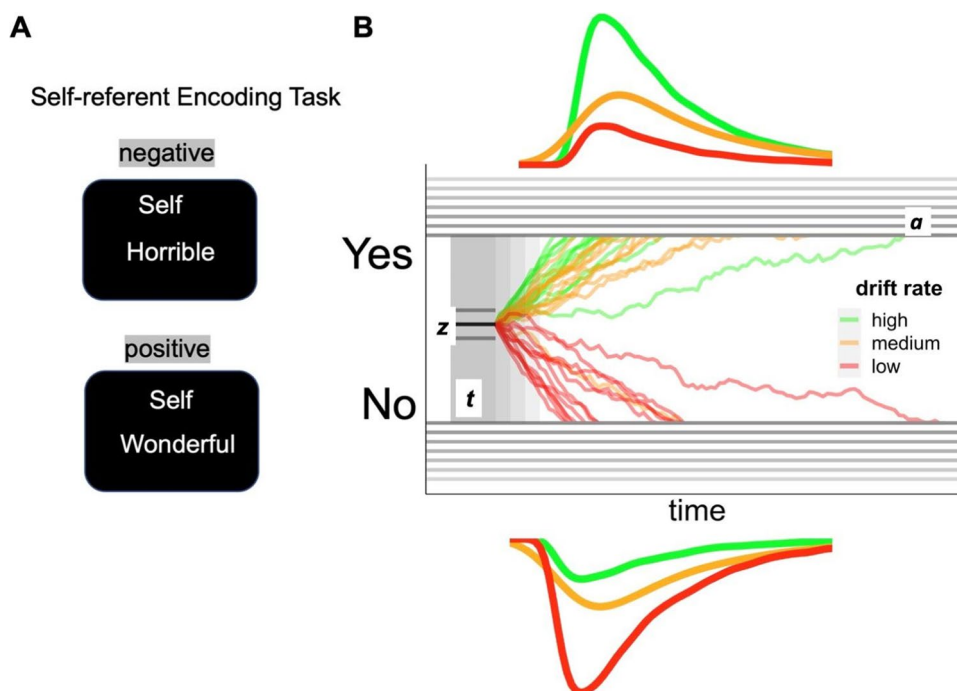


Fig. 1 Self-referent Encoding Task and Drift-Diffusion Model (DDM). **A** In the Self-referent Encoding Task, participants make two-alternative (Yes/No) forced choices about whether negative and positive adjectives describe themselves. **B** Illustration of the DDM, which models choice as a noisy accumulation of evidence over time until a decision (in this case, Yes/No) is reached. The variable noise across trials produces reaction time distributions for both response types (example distributions for Yes and No shown at the top and bottom of the x-axes, respectively). The evidence-accumulation parameter in the DDM is the drift rate; the figure shows some example decision trajectories for high (producing primarily fast Yes responses; note that in the Yes distributions on top, the high distribution is less “fat-tailed” than the others), medium, and low (producing primarily fast No

responses) drift rates. In addition to the drift rate, the DDM has other identifiable decision parameters, labeled a , t , and z (these respectively determine the amount of evidence required to reach a decision, non-decisional time (comprising perceptual processing time and motor execution), and overall bias toward responding Yes/No in absence of any stimulus evidence; shading illustrates different potential values of these parameters). In the current context, we were interested in individual differences in the strength and direction of responding to negative versus positive words in terms of their impact on the drift rate, as assessed via a drift-rate regression parameter. The other parameters served a similar role as nuisance covariates in statistical modeling; they enabled more specific estimates of the parameters of interest. a = threshold; t = nondecision time; z = starting point bias

psychological, neural, and computational constructs (Forstmann et al., 2016; Niv, 2009), computational models enable stronger tests of psychological theories (Guest & Martin, 2021; Hitchcock et al., 2022; Maia et al., 2017; Millner et al., 2020; Robinaugh et al., 2021).

We applied computational modeling to test the predictions of influential theories of mindfulness and depression. Specifically, we fit a drift-diffusion model (DDM) to the SRET (Fig. 1) (Allison et al., 2021; Beevers et al., 2019; Dainer-Best et al., 2018a, b; Disner et al., 2017; Price et al., 2021) and extracted estimates of a drift-rate regression parameter quantifying the relative strength of positive versus negative self-judgment.

We leveraged the mechanistic specificity of our computational modeling analysis to test prominent theories of depression and mindfulness. In particular, cognitive theories of depression propose that a propensity toward negative over positive self-judgments is key to the disorder. This leads to the straightforward prediction that depression severity will

relate inversely to the drift-rate regression parameter. In contrast, prominent mindfulness theories imply diverging predictions about how different facets of mindfulness will influence well-being. Some theories view mindful awareness, which is arguably the foundational skill in mindfulness-based practice, as an unalloyed good. In particular, these accounts argue that mindful awareness leads to natural insight into one’s own behavior, which in turn promotes healthy behavior (Brewer, 2019; Brown & Ryan, 2003; Ludwig et al., 2020). It follows that more mindfully aware people should take more adaptive actions, which should ultimately lead to more positive (and less negative) self-views; thus, awareness should positively relate to the drift-rate regression parameter. By contrast, the influential Monitor and Acceptance Theory argues that mindful awareness only increases well-being when it is accompanied by nonjudgmentalness—and, in fact, that people who are highly aware, yet also highly judgmental, should be the most prone to mental unhealth and ill-being (Lindsay & Creswell, 2017, 2019).

This theory predicts that mindful awareness will show no main effect on self-views; rather, the effect will be qualified by an interaction, with *high-aware, high-judgmental* people having more negative self-views and *high-aware, low-judgmental* people having more positive self-views. We used the drift-rate regression parameter—quantifying the relative strength of positive to negative self-views—to test the diverging predictions of these mindfulness theories.

We conducted analyses on data collected during an 8-week mindfulness intervention study, in which a relatively large sample of individuals ($n = 96$) completed the SRET task at pre- and postintervention. This enabled us to test whether the predictions of depression and mindfulness theories were supported both in terms of baseline individual differences and as these differences changed over time. The investigation of how mindfulness facets and self-referential processes co-vary helps to redress a gap in the literature concerning these relationships, despite the conceptual importance of self-referential processing to many mindfulness theories (Britton et al., 2021).

We also considered how different depression symptoms related to the model-based measures. Much evidence suggests that depression is not a unitary disorder but rather comprises heterogeneous symptoms (Fried, 2015; Fried & Nesse, 2015). This implies that different depression symptoms should show distinct relationships to specific computations (Beevers et al., 2019; Bennett & Niv, 2018; Hitchcock et al., 2022; Maia & Frank, 2011). We built on influential recent work that demonstrated—using a variable-importance metric from a machine-learning model—that the drift rate for negative words in the SRET strongly related to some depression symptoms (e.g., sadness, self-dislike, pessimism), yet barely at all to others (e.g., failure, appetite, crying; Beevers et al., 2019). We extended this past work in three ways. First, given our interest in the conceptual question of which symptoms related to self-referential mechanisms, we used a more conventional measure of the relationship between variables—zero-order correlations—rather than a variable-importance metric. Second, we leveraged our intervention dataset to examine how change in self-referential mechanisms related to changes in different depression symptoms over time. Third, as described, we used model-based regression analyses to examine the relative difference in positive versus negative self-judgment strength—rather than only measuring negative self-judgment strength (as in Beevers et al., 2019)—because depression theories predict not only that negative self-views should be harmful but also that positive self-views should be protective (Disner et al., 2011; LeMoult & Gotlib, 2019; Vanderlind et al., 2020).

A final goal of this study was to estimate the split-half and test-retest reliability of SRET task measures. Computational modeling parameters often show higher reliability than summaries of unmodeled behavioral data (Brown et al., 2020;

Haines et al., 2020; Hitchcock et al., 2017; Lawlor et al., 2020; Mkrtchian et al., 2021; Price et al., 2019; Shahar et al., 2019). This is important, because behavioral summaries often are highly inconsistent—which poses a fundamental challenge to the common assumption that they reflect stable, trait-like characteristics (Brown et al., 2020; McNally, 2019; Parsons et al., 2019; Rodebaugh et al., 2016). One reason that many behavioral measures may be unstable across testing sessions is state variation, such as being more fatigued on one day completing a task than another (Hitchcock et al., 2017; Lilienfeld, 2014). A second reason may be that many behavioral summaries (e.g., mean reaction time) reflect a mix of psychological processes, and such aggregates may be intrinsically unreliable (Haines et al., 2020). Computational modeling can help with both problems: by pinpointing precise processes latent in behavioral data, which may be more stable over time than composites; and allowing variables of no interest, which might be especially susceptible to random state variation, to be factored out of the decision process (Nassar & Frank, 2016). In practice, however, it is always an open empirical question whether applying a given computational model to a specific task will substantially improve its reliability (Pike et al., 2022). A further complication to estimating the reliability of measures is that they can change over time in meaningful ways (Yip et al., 2022; Brown et al., 2021; Konova et al., 2020), such as due to the mindfulness intervention in the current study. Thus, we estimated reliability at a single timepoint (via split-half Intraclass Correlation Coefficients [ICCs] at preintervention) and over time (via test-retest ICCs from pre- to postintervention).

Methods and materials

Procedure

Our analyses sought to: (1) test the predictions of mindfulness and depression theories concerning how these processes would co-vary with positive (vs. negative) self-judgment drift rate, at baseline and as mindfulness and depression changed over time; (2) test which depression symptoms would relate to positive (vs. negative) self-judgment drift rate, at baseline and as these symptoms changed over time; and (3) compare the split-half and test-retest correlations of behavioral and model-derived measures. To fulfill these goals, we conducted a secondary analysis of an intervention dataset from a Mindfulness-Based Cognitive Therapy (MBCT) dismantling study (Britton et al., 2018; Cullen et al., 2021; clinicaltrials.gov identifier: NCT01831362). The primary goals of the clinical trial, which have been evaluated in prior publications (Britton et al., 2018; Brown et al., 2022; Cullen et al., 2021), were to examine whether three variations of MBCT—full MBCT, a variant only employing focused-attention meditation, and a

variant only employing open-monitoring meditation—would have similar or distinct effects on clinical and mindfulness outcome variables as well as behavioral task neural measures (described in Britton et al., 2018; see Lutz et al., 2008, for an overview of these meditation types). All intervention arms were equivalent in dosage and matched on numerous structural features. The results suggested that during the 8-week pre- to postintervention period, the three interventions led to similar improvements in clinical outcome variables (Britton et al., 2018; Cullen et al., 2021) and comparable effects on self-reports in an emotional reactivity task, although with distinct effects on neural measures during the task (Brown et al., 2022).

Our analyses leveraged the full intervention dataset ($n = 96$ treatment completers)—which offered a rare opportunity to assess the research questions concerning change over time in a relatively large longitudinal dataset, wherein many individuals showed substantial change in mindfulness and depression over the measurement period. We also provide decompositions of the analyses conducted over the intervention course in terms of treatment condition in the [Supplemental Material](#), although these analyses should be interpreted with caution given the small sample size in each treatment condition.

Our analyses involved computational modeling of choices and reaction times from the SRET task. A previous study separately reported on recall data, which is not analyzed here, from the SRET in this sample (Alejandre-Lara et al., 2022); electroencephalography also was recorded while participants completed the SRET and will be reported separately. All participants gave informed consent before the clinical-trial study, which was approved by The Brown University Institutional Review Board (#1105000399).

Participants

The clinical trial from which data were drawn had sought to recruit participants representative of people who seek out meditation training through mindfulness-based treatment programs. These individuals are typically experiencing mild-to-severe anxiety and/or depression, which is their impetus for beginning these programs. Hence, participants reported at least mild depression and negative affect at baseline—specifically, scores of ≥ 10 on the IDS-C and ≥ 18 on the negative-affect scale of the PANAS-X (IDS-C = Inventory of Depression Symptomatology; Rush et al., 1996; PANAS-X = Positive and Negative Affect Schedule—last month; Watson & Clark, 1994). Participants were excluded if they were: severely depressed (IDS-C > 48) and/or reported active suicidal ideation; had a history of psychotic, bipolar, borderline, or antisocial disorders, or of organic brain damage; currently met criteria for obsessive-compulsive, panic, posttraumatic stress, eating, or substance abuse disorders;

Table 1 Demographics and Sample Characteristics

Gender, n (%)	
Female	69 (71.88%)
Male	27 (28.13%)
Age, M (SD)	40.16 (12.88)
Race, n (%)	
White	94 (97.92%)
Ethnicity, n (%)	
Hispanic/Latino	6 (6.25%)
Not Hispanic/Latino or Not Reported	90 (93.75%)
Education in years, M (SD)	17.10 (2.69)
Sample characteristics at baseline—diagnosis, n (%); self-report scores, M (SD)	
MDD diagnosis	37 (38.54%)
GAD diagnosis	48 (50.00%)
DASS-Depression	9.58 (7.69)
FFMQ-Acting with Awareness	24.05 (5.98)
FFMQ-Nonjudgment	26.30 (7.66)
FFMQ-Nonreactivity	18.84 (3.84)

Abbreviations. MDD major depressive, GAD Generalized Anxiety Disorder

were unable to read and write in English (because interventions were conducted in English); had a consistent meditation practice; were currently in psychotherapy; or had changed antidepressant medication within the last 2 months. Participants were a fairly homogenous group comprising primarily white (98%) women (73.1%) (Britton et al., 2018; see limitations in “Discussion”). Pre- and postintervention assessments were conducted within 3 weeks of the first and last sessions, respectively. See Britton et al. (2018) for further details on assessment protocols regarding inclusion and exclusion criteria. Table 1 presents demographics and clinical characteristics. The clinical trial is described in detail in (Britton et al., 2018; Cullen et al., 2021).

Materials

Self-report measures

Depression Anxiety Stress Scales (DASS)—Depression subscale Depression symptoms were assessed via the DASS-Depression subscale of the DASS-42. The DASS, which has a scale ranging from 0 (“Did not apply to me at all”) to 3 (“Applied to me very much or most of the time”), has demonstrated convergent and discriminant validity with other depression, anxiety, and negative-affect measures (Brown et al., 1997). We focused on the 14-item subscale due to our goal of testing clinical accounts of depression that implicate valenced self-referential processing in the disorder (example

items: “I felt I wasn’t worth much as a person”; “I felt that life was meaningless”). Internal consistency as measured by Cronbach’s α at the pre- and post-timepoints was in an acceptable range (0.93 at both timepoints).

Five-factor mindfulness questionnaire Trait mindfulness was measured via three subscales of the Five-factor Mindfulness Questionnaire (FFMQ) (Baer et al., 2008), which enabled us to test the predictions of prominent theories of mindfulness. Specifically, we used the nonjudgment (the ability to experience thoughts and feelings without evaluating them as good or bad), nonreactivity (the ability to experience thoughts and feelings entering one’s mind without needing to engage or do anything else in response to them), and acting with awareness (paying attention in the present and behaving with intention, rather than on “auto-pilot”) subscales. The scale of the FFMQ ranges from 1 (never or very rarely true) to 5 (very often or always true). The subscales employed here have demonstrated convergent and discriminant validity (Bohlmeijer et al., 2011; Baer et al., 2008). Internal consistency as measured by Cronbach’s α at the pre- and post-timepoints was in an acceptable range (pre/post: awareness: 0.88/0.90; nonjudge: 0.94/0.94; nonreact: 0.77/0.85).

Consistent with recommendations in the literature, we excluded the Observing subscale of the FFMQ from primary analyses (although we included it in robustness checks reported in the [Supplemental Material](#)). Specifically, the Observing subscale has consistently shown poor psychometric properties in nonmeditators, including a different factor structure than among meditators, possibly due to its terminology taking on different meaning for meditators versus nonmeditators, and mentally healthy versus unhealthy individuals (Aguado et al., 2015; Baer et al., 2006, 2008; Baer, 2019; Britton, 2019; Curtiss & Klemanski, 2014; Gu et al., 2016; Lecuona et al., 2020; Thompson et al., 2022; see also Choi et al., 2021; Hitchcock et al., 2016; Lester et al., 2018). As described further in our discussion of the current study’s limitations, the Act with Awareness scale of the FFMQ also has been criticized for potentially having a different meaning for nonmeditators than meditators, because all items on this subscale are negatively worded, and it has been suggested that nonmeditators may be more likely to deny mindlessness even if they would not endorse intentional awareness (Baer et al., 2011; Van Dam et al., 2009, 2012). Nevertheless, considering the literature as a whole, we believe this issue with the Act with Awareness subscale is substantially less severe than the multiple well-documented issues concerning the Observing subscale; hence, we opted to use the Act with Awareness scale for our primary analyses to test the competing predictions of mindfulness theories. Future studies may wish to draw on in-process work striving to develop versions of the FFMQ that are less subject to response bias and other psychometric issues (Padmanabham et al., 2021; Van Dam et al., 2018; Lecuona et al., 2020).

Behavioral task: self-referential encoding task

The Self-referential Encoding Task (SRET) is a two-alternative, forced-choice task meant to examine the frequency, strength, and efficiency with which negative and positive self-schemas become active (Dainer-Best et al., 2018b; Derry & Kuiper, 1981). The premise of the SRET is that people with relatively stronger negative schemas, for example, will activate these schemas more readily and therefore make faster and more frequent negative self-judgments. According to cognitive accounts (Disner et al., 2011), people who have rapidly activated negative schemas and sluggishly activated positive schemas are vulnerable to depression. On each trial of the SRET, participants were required to decide, as quickly as possible, whether a negative (e.g., “horrible”) or positive (e.g., “wonderful”) word describes them (yes/no; Fig. 1A). Participants completed the SRET interleaved with a control task that also required making two-alternative forced choices to positive or negatively valenced words, simply about whether or not the word was capitalized (e.g., yes was the correct answer for “GREAT” and no for “great”; half of the adjectives were capitalized and half uncapitalized). The condition in a given trial was indicated by the word “self” for an SRET trial or “case” for a control trial, displayed above the adjective. Participants pressed the right arrow key to select yes and left arrow key to select no in both tasks. Control data were only analyzed in this study for quality control purposes (see below).

Trials timed out if participants failed to respond within 3 seconds. Between trials, a fixation cross (+) was presented for 500 ms. Reaction time and response (yes or no) was recorded for each trial unless the participant did not respond within the time limit. The order of control and SRET trials was randomized. The adjectives in the task (59 distinct adjectives for SRET at each timepoint: 30 positive and 29 negative) were selected from a set of self-descriptive, positive and negative adjectives with standardized ratings (Anderson, 1968). Lists were counterbalanced for word valence and length, and participants received lists in a randomized fashion; some participants received adjectives in the SRET condition that others received in the control condition. Words were displayed in a pseudo-randomized order so that trials of different type—for example, SELF-(negative, lower case)—followed trials of different types with equal frequency.

Quality control

We excluded trials with extreme reaction times (<200 ms; this was just a single trial). For 1.84% of trials, participants failed to meet the 3-s response deadline, hence no response was recorded. Our analyses include changes in outcome variables over time and thus we analyzed full-intervention completers ($n = 97$); we excluded one participant who performed near

chance on the control task postintervention (50.91%, where 50% is chance performance), leaving 96 participants for these analyses. Performance on the control task was high among these participants (median correct = 97.50%). A few item responses from the questionnaires were missing (0.21% of responses at both pre- and postintervention for the DASS, and 0.26% and 0.16% missing at pre- and postintervention for the FFMQ); we imputed the subscale mean for these few items.

Computational modeling and statistical analysis

We applied hierarchical Bayesian estimation of the DDM to infer latent mechanisms from behavioral choices and reaction times, using the Python-based HDDM software toolbox (Fengler et al., 2021; Wiecki et al., 2013), which yielded group and subject estimates for each parameter. We tested specific hypotheses concerning the SRET by estimating within-subject effects of word valence, and between-subject effects of treatment time (pre- vs. postintervention), on the drift rate via a model-based regression analysis (drift rate \sim valence * timepoint). Regressor variables were centered before model fitting. Convergence of Bayesian model parameters was assessed via the Gelman-Rubin statistic; all parameters had \hat{R} below 1.1, which does not suggest convergence issues. Model comparison used the Deviance Information Criterion (DIC), a metric of how well one model fits over another that is appropriate for hierarchical Bayesian models (Spiegelhalter et al., 2014). Model validation and quality checks are described in “Results.” Parameter recovery was excellent for the winning DDM model used for primary analyses (*range* = 0.81–0.97; *median* = 0.94; Supplemental Fig. 1).

All other data cleaning and statistical analysis were conducted using R (Version 4.1.2; R Core Team, 2021). Analyses involved ordinary least squares regression models and mixed-effects logistic and linear regression models; mixed-effect models were constructed via the lme4 package (Bates et al., 2014), with *p* values calculated using Satterthwaite’s method to approximate degrees of freedom. All predictors had a variable inflation factor lower than 2, reflecting acceptable collinearity.

The regression models that were used to test depression and mindfulness theories (by relating model-derived estimates to questionnaire-based individual differences) used maximum a posteriori (MAP) estimates of the subject-level traces as the outcome variables. Of note, simulation studies have demonstrated that hierarchical Bayesian modeling tends to improve point estimates of model parameters (Katahira, 2016), including in HDDM (Wiecki et al., 2013), and thus should improve estimation in statistical models, such as those used here that involve the relationship between hierarchically estimated parameters and between-subject measures (Katahira, 2016). Note also that we did not apply any within-group hierarchical

estimation, which would then lead to biased estimation if the within-group shrinkage on estimates was not accounted for in subsequent analyses (Boehm et al., 2018).

We also used subject-level MAP estimates for parameter recovery and for calculating split-half and test-retest ICCs. ICCs used a two-way random effects model for absolute agreement (Hedge et al., 2018; Shrout & Fleiss, 1979). Test-retest reliability statistics for model parameters were conducted on models fit separately on the pre- and post-intervention data. For parameter recovery, for all parameters other than the across-trial variability parameters, we performed the following procedure: Sampled once, for each parameter for each subject, from a uniform range [0.1, 0.9] of the empirically estimated trace; generated synthetic data for that subject with the resulting parameter values (using the ‘gen_rand_data’ function in HDDM) of the same size as their empirical dataset, and using the regressor variables from their actual data; then performed optimization on these synthetic datasets. Parameter recovery was then assessed by taking the MAP estimate of the resulting recovered traces and comparing them to the ground-truth generative parameters.

Model validation via posterior predictive checks (reported below) involved the same method of generating synthetic data from the winning DDM model, with the exception that we sought to sample the traces for each subjects’ parameter estimates more extensively; thus, we used 25 estimates randomly drawn from the entire trace (rather than a uniform draw of a single value). Otherwise, the simulation procedure was the same, in that we used each draw to generate an entire synthetic dataset for that subject.

Results

Self-referential encoding task

Behavior and its reflection in computational models

Participants endorsed positive words much more frequently than negative words at both time points. This difference increased over time (from pre- to postintervention), with no significant main effect of time on responses (Fig. 2A, left; logistic mixed-effects regression model: *valence* *Beta* (*SE*) = 2.18 (0.08), $p < 2e-16$; *valence*time point* *Beta* (*SE*) = 0.18 (0.03), $p < 5e-9$; *timepoint* $p > 0.27$; valence was centered such that *negative* ≈ -1 , *positive* ≈ 1 , and timepoint such that *preintervention* ≈ -1 and *postintervention* ≈ 1 [approximation is due to a small asymmetry from trials removed for quality control, as described in “Methods”]). Participants responded more slowly to negative than positive words, and faster over the course of the intervention, but with no significant change in speeding as a function of valence over the intervention (Fig. 2A, right; *valence* *Beta* (*SE*) = -0.037

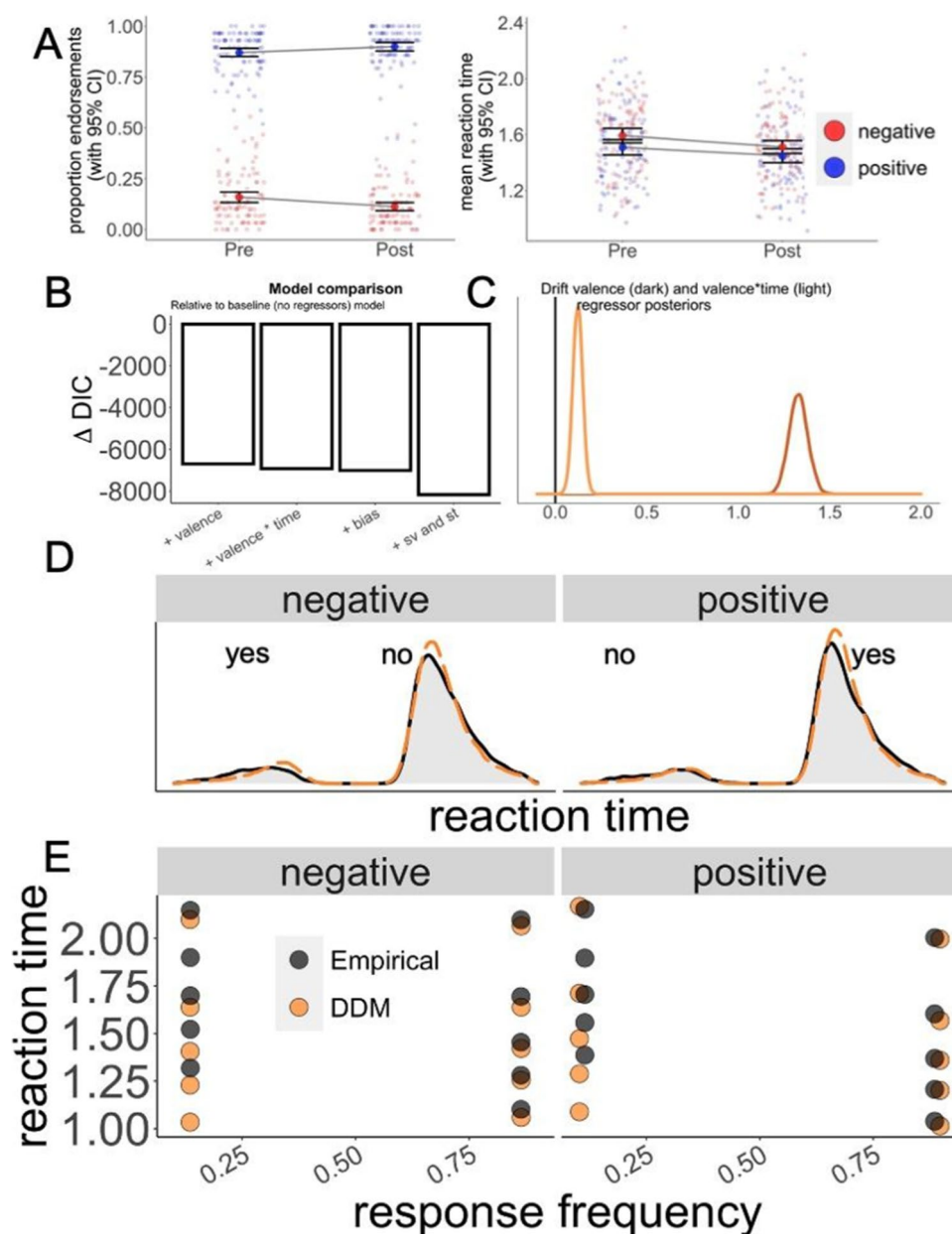


Fig. 2 Behavioral patterns and their capture by computational modeling. **A** Proportion of words endorsed (left) and mean reaction time (right) for positively (blue) and negatively (red) valenced words, pre- and postintervention (x-axis). Small light points show subject means and darker points show group means; error bars show 95% confidence intervals. **B** Model comparison (relative to a baseline model with no regressors) via a metric of model fit appropriate for hierarchical Bayesian models, the DIC; the rightmost model is the “winning” model used for analyses, which includes a valence*timepoint regression parameter, and starting point bias and *sv* and *st* parameters (see *Abbreviations*, this caption). **C** Group posterior distributions for valence (dark orange) and valence*timepoint (light orange) regressors from the hierarchical Bayesian DDM model. The valence variable was coded so that a positive value reflects higher positive, relative to negative, evidence accumulation; the posteriors thus suggest higher strength of positive (over negative) evidence accumulation (as is expected given the endorsement by valence differences in A). **D** Empirical reaction time density plots (black), separated by valence and with the infrequent response types within each valence category (i.e., endorsements of negative words and rejections of positive words) reflected to the left axis. The empirical density plots are shown overlaid with simulations from the winning computational model (orange dotted lines). The plots suggest that the DDM can reasonably capture key features of the empirical data, including response frequencies and the overall shape of the reaction time distributions. **E** A more fine-grained validation of the match between the empirical and model-derived data is given by quantile-probability plots; these show response frequency along the x-axis, with valence-response categories aligned with **D** (e.g., the left-most points correspond to (negative-yes) responses), with the y-axis showing reaction times in the (0.1, 0.3, 0.5, 0.7, 0.9) quantiles within that category. This plot shows that, although the model simulations provide an excellent match to the response frequencies in all categories and the reaction-time quantiles in the common response categories (positive endorsements and negative rejections), there is some mis-specification in the less common response categories, in that the model predicts substantially faster reaction times in the earliest quantiles than occurs in the empirical data. *bias* = starting point bias parameter; *CI* = confidence interval; *DDM* = Drift diffusion model; *DIC* = Deviance Information Criterion; *Pre and Post* = pre- and postintervention; *st* = across-trial nondecision time variability parameter; *sv* = across-trial drift-rate variability parameter

(<0.01), $p < 5e-7$; *timepoint Beta (SE)* = -0.036 (<0.01), $p < 2e-16$; *valence*timepoint* $p > 0.20$; valence and timepoint coded as above).

These behavioral effects were reflected in the DDM model. Relative to a baseline model with no regressor variables, model fit dramatically improved with the inclusion of a regressor of valence on drift rate and improved further when valence was interacted with time point (reflecting the effects of valence and time on responses and reaction times shown in Fig. 2A). The model further improved with the addition of a starting-point bias term (reflecting a preference for responding yes/no irrespective of the adjective) and across-trial variability parameters for the drift rate and nondecision time (Fig. 2B; the winning model [rightmost in the figure] was used for subsequent analyses). Reflecting the main effect of valence on endorsements, the model-derived drift-rate regression parameter, representing the impact of positive over negative words on accumulation toward “yes” responses, was positive and far away from 0; reflecting the comparatively smaller increase in positive (vs. negative) endorsements over time, the *valence * time point* regression posterior also was positive, but of smaller magnitude (*valence posterior*: Fig. 2C, dark orange; maximum a posteriori [MAP] = 1.33, $p < 0 = 0$; *valence * time posterior*: Fig. 2C, light orange; MAP = 0.13, $p < 0 = 0$).

Model validation

The parameters from the winning computational model showed interpretable correlations with various aspects of the behavioral data. At preintervention, subject-level differences between positive and negative endorsements (but not reaction times) positively correlated with subject-level MAP estimates in the valence drift regressor (i.e., those who endorsed more positive words had higher estimates on this parameter; *endorsements*: $r = 0.79$, $p < 2e-16$, *reaction times*: $p > 0.53$). Similarly, the difference from pre- to post-intervention, in this valence-difference in endorsements (but not reaction times) correlated with MAP estimates of the *valence * timepoint* regressor (*endorsements*: $r = 0.64$, $p < 5e-12$; *reaction times*: $p > 0.53$). Subjects’ mean reaction times (marginalizing over valence and time) also strongly correlated with their threshold parameter MAP estimates ($r = 0.69$, $p < 1e-14$). Thus, the drift-valence parameters appeared to primarily reflect individual differences in endorsements as a function of valence, whereas the threshold reflected differences in overall reaction times.

Crucially, the parameters of the winning model not only correlated with different aspects of the behavioral data, but synthetic data using parameter estimates from the winning model (i.e., posterior predictive checks; Gelman et al., 1995)

captured key features of the data, including the shape of the reaction time distributions and the frequency of endorsements in the various valence categories (Frank et al., 2015; Pedersen & Frank, 2020). There was one notable exception: in both valence categories, the model somewhat misspecified infrequent responses in both valence categories (positive and negative; Fig. 2D and E). These results suggest that the model was able to provide an adequate account of the behavioral data, albeit with room for improvement (see [Supplemental Material](#): “Methods: Further Description of Modeling Efforts” for further details on modeling).

Leveraging the model to test the predictions of mindfulness and depression theories

Having established that the winning DDM model provided a reasonable account of the empirical data, we next leveraged it to test influential theories of mindfulness and depression, focusing first on baseline data. Consistent with mindfulness theories that imply mindful awareness should lead to more positive self-views, the drift-rate regression parameter (representing positive over negative evidence accumulation) positively related to FFMQ-ActAware at baseline. However, this parameter did not significantly relate to FFMQ-NonJudge, FFMQ-Nonreact, or (counter to the predictions of Monitor and Acceptance Theory) to the interaction between FFMQ-ActAware and FFMQ-NonJudge at baseline (Fig. 3A, top; *aware Beta (SE)* = 0.10 (0.04), $p = 0.015$; *nonjudge Beta (SE)* = 0.05 (0.05), $p > 0.27$; *nonreact (SE)* = 0.02 (0.04), $p > 0.66$; *aware*nonjudge Beta (SE)* < 0.01 (0.04), $p > 0.91$; all regressors were z-scored). This same pattern was present when regressing within-subject changes in the mindfulness facets on the model-based regressor representing change in positive over negative evidence accumulation over the intervention (Fig. 3B, bottom; *aware Beta (SE)* = 0.06 (0.02), $p = 0.008$; *nonjudge Beta (SE)* = 0.03 (0.02), $p > 0.27$; *nonreact Beta (SE)* = -0.03 (0.02), $p > 0.15$; *aware*nonjudge Beta (SE)* < -0.01 (0.02), $p > 0.63$; all regressors were z-scored). Notably, there was substantial variability in FFMQ-ActAware and FFMQ-NonJudge questionnaire responses, suggesting that the lack of a significant interaction between mindful awareness and nonjudgmentalness was unlikely due to restricted range (Supplemental Fig. 2).

As predicted by cognitive theories of depression, depression was inversely related to the drift-rate regression parameter at baseline (Fig. 3B, left; $r = -0.38$, $p < 5e-4$). The same pattern was present when correlating within-subject changes in depression with the drift-rate-valence-change regressor estimates (Fig. 3B, right; $r = -0.21$, $p = 0.044$). The results were largely consistent across various robustness checks, described in the [Supplemental Material](#).

An advantage of computational-model derived metrics is that they can disentangle latent processes that collectively

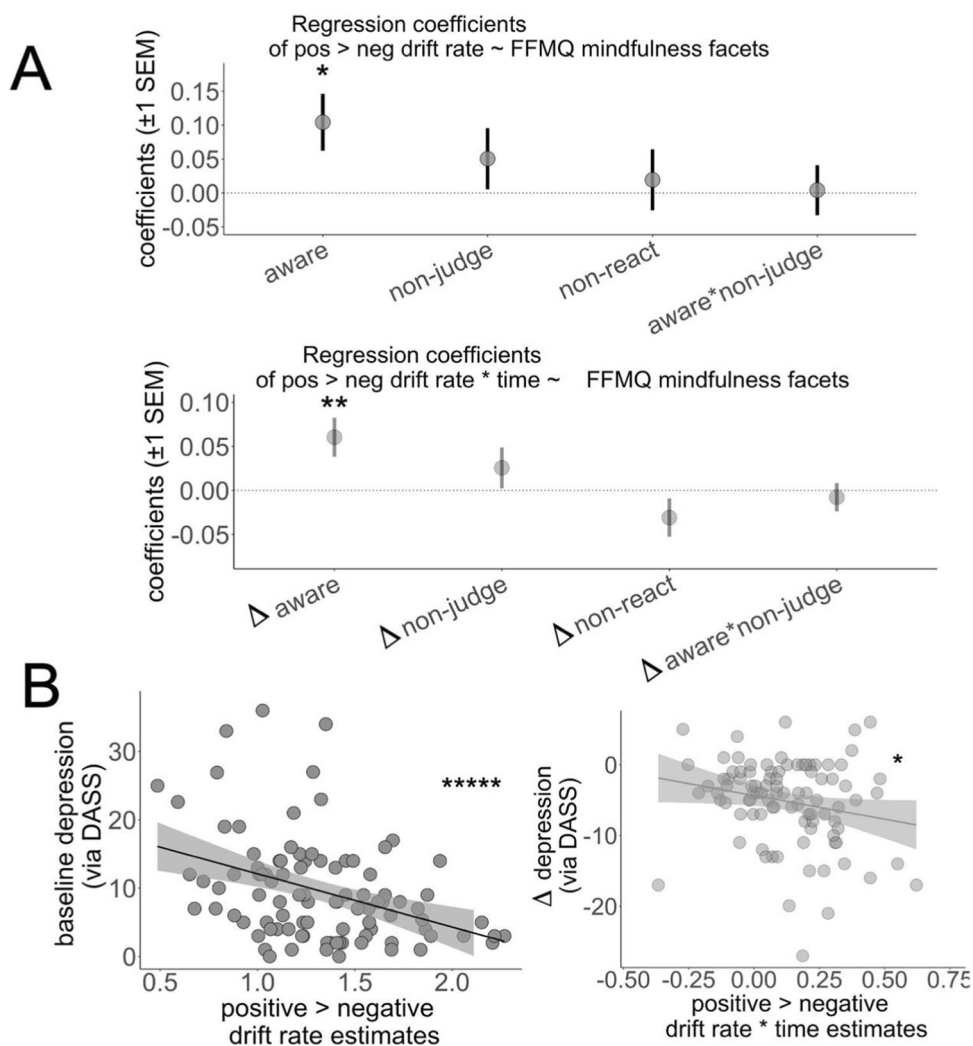


Fig. 3 Statistical results testing influential accounts of depression and mindfulness. **A** Coefficients of (z-scored) mindfulness facets regressed on MAP estimates of the positive-over-negative drift-rate regression parameter (top) and change in mindfulness facets over time (post- minus preintervention) correlated with MAP estimates of the positive-over-negative*timepoint drift-rate regression parameter (bottom). **B** Relationship between subject-level estimates of the positive-over-negative drift-rate parameter and individual differences

in depression (via the DASS-Depression subscale) at baseline (left). Relationship between subject-level estimates of the positive-over-negative*timepoint drift-rate parameter and individual differences in change in depression (via change from post- minus preintervention of the DASS-Depression subscale; right). * $p < 0.05$; ** $p < 0.01$; *** $p < 0.005$; **** $p < 0.001$; ***** $p < 0.0005$. *FFMQ* = Five-Facet Mindfulness Questionnaire; *DASS* = Depression Anxiety Stress Scales

give rise to behavior, which in some cases can lead to stronger correlations with individual-difference measures (Wiecki et al., 2015), paving the way for the use of the model-derived measures in predictive settings, such as treatment selection (Cohen & DeRubeis, 2018). However, in this case, we found that the model-derived, drift-rate regressor parameter did not show substantially stronger relationships with mindfulness and depression scores than comparable summary statistics (namely, analogous behavioral endorsement statistics; see [Supplemental Material](#)).

Depression symptoms and their relationships to model-based measures

We next considered specific depression symptoms, in light of evidence that depression is a heterogeneous disorder (Fried, 2015; Fried & Nesse, 2015). Consistent with heterogeneity, depression symptoms showed variable levels of elevation at baseline, with some symptoms (e.g., “couldn’t get going”; “no initiative”) much more frequently endorsed than others (e.g., “life not worthwhile”; “life meaningless”;

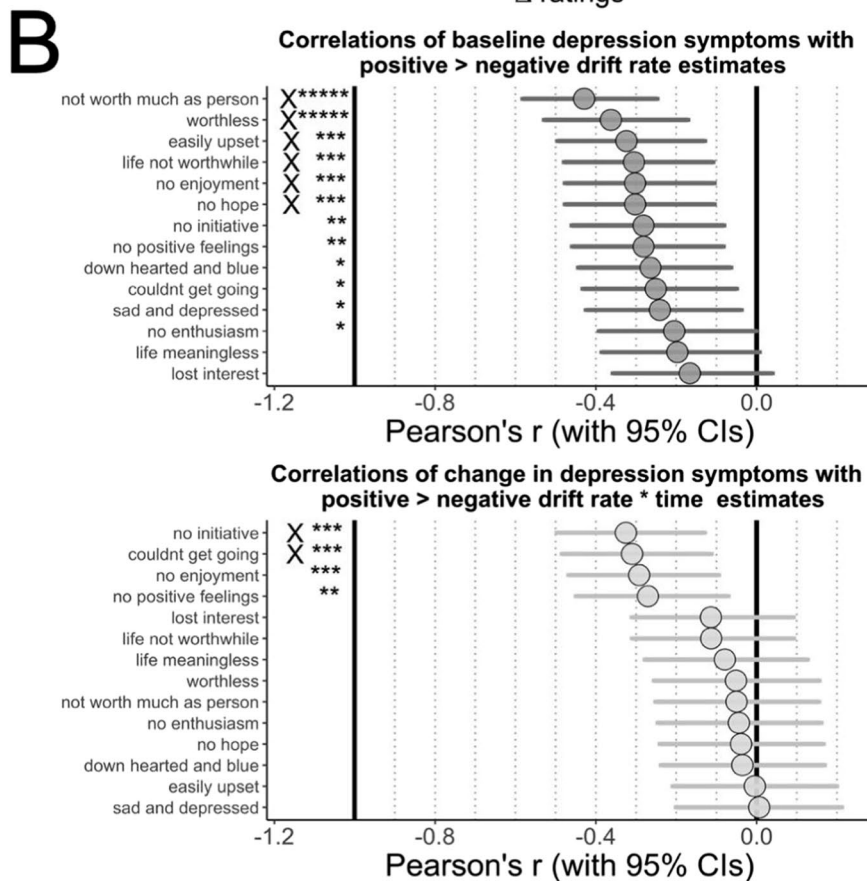
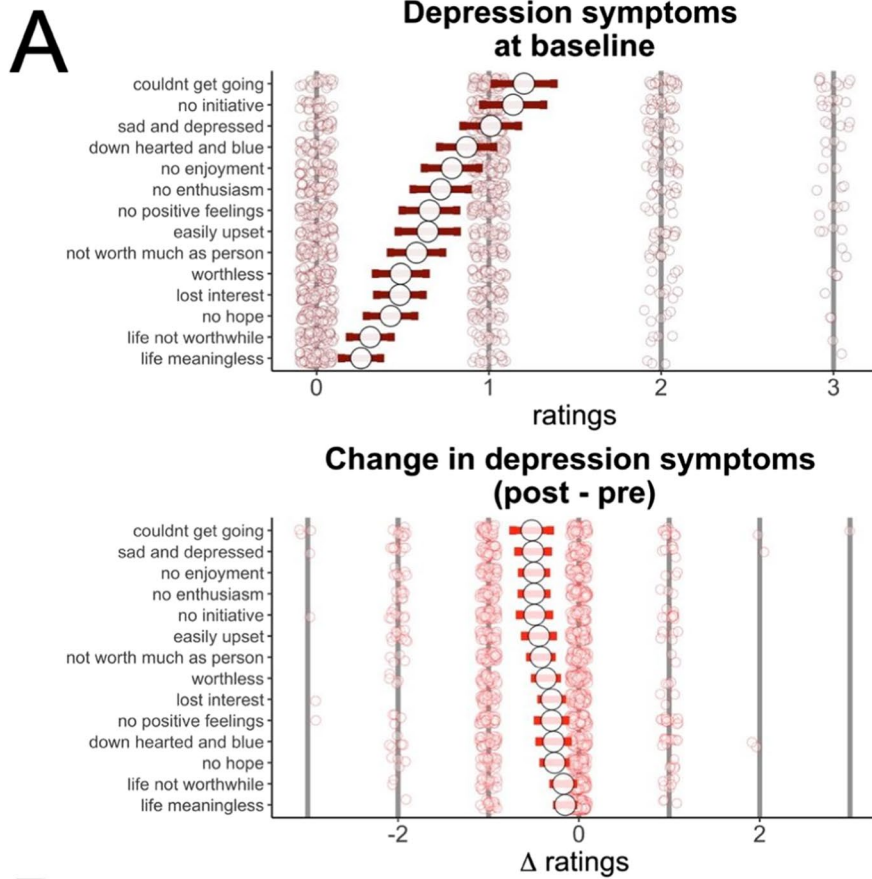


Fig. 4 Depression symptoms (via the DASS-Depression subscale) and their relationships to model-based measures. **A** Depression symptoms at baseline (top) and changes therein from pre- to post-intervention (bottom). Large points and error bars show group means and 95% CIs, and small points show ratings for each individual subject. Vertical gray lines show possible values. **B**. Zero-order correlations between subject-level MAP estimates for the drift-rate valence parameter and baseline symptoms (top) and for this parameter interacted with timepoint and change in symptoms (bottom). Gray points show correlation point estimates and error bars show (uncorrected) 95% CIs. Dotted lines are in intervals of 0.1. “X” denotes significant after Bonferroni correction ($p = 0.05/14$); * $p < 0.05$; ** $p < 0.01$; *** $p < 0.005$; **** $p < 0.001$. *CI* = confidence interval

Fig. 4A, top), although it also is clear that many of the symptoms that were most elevated at the start of the intervention decreased the most over the course of the intervention (Fig. 4A, bottom). At baseline, there was a continuum of relationship strengths between depression symptoms and the drift-rate parameter representing positive (relative to negative) self-judgment strength (Fig. 4B, top). In contrast, a smaller subset of symptom-change scores showed non-trivial correlations with the drift-rate regression parameter representing change in positive (relative to negative) self-judgment strength over time (Fig. 4B, bottom).

Reliability of model parameters

Task summary statistics showed modest estimated reliability (Fig. 5, top), as did several computational model parameters (Fig. 5, bottom). Overall, the test-retest pattern suggests that SRET measures—both behavioral and model-derived—were relatively consistent over the 8-week intervention period, even those that may have been affected by the intervention (although note that the drift-rate positive-valence regression parameter, which may have been especially and variably influenced by the intervention, showed higher consistency in the split-half than test-retest ICC). The results do not suggest a marked difference in the reliability of model parameters compared with behavioral measures.

Conclusions

People who consistently and quickly judge themselves in a negative light, and inconsistently and slowly judge themselves in a positive light, are vulnerable to depression, according to cognitive theories of this disorder (Disner et al., 2011; LeMoult & Gotlib, 2019). Mindfulness theories also propose that self-judgments play a critical role in determining mental health (Vago & Silbersweig, 2012). Mindfulness has been argued to promote adaptive behavior (Ludwig et al., 2020) that should lead to the development of more positive self-views over time. However, there is a paucity of research on how change in mindfulness over time

co-varies with change in self-referential processes (Britton et al., 2021). To examine whether individual differences in mindfulness and depression corresponded to individual differences in self-views, in cross-section and over time, as predicted by these theories, we examined self-referential encoding task (SRET) data collected before and after an 8-week mindfulness intervention ($n = 96$ completers).

We found that people who reported more mindful awareness at baseline, and who showed a greater increase in mindful awareness over time, showed stronger positive (relative to negative) evidence accumulation during self-referential decision-making, as quantified by drift-rate regression parameters. However, these parameters did not significantly relate to the interaction between mindful awareness and nonjudgmentalness. These findings accord with mindfulness theories that suggest that mindful awareness promotes healthy behavior, which should lead people to view themselves more positively (Brewer, 2019; Brown & Ryan, 2003; Ludwig et al., 2020). Yet, they conflict with the predictions that would seem to follow from a prominent theory, Monitor and Acceptance Theory, which posits that mindfulness awareness should only be associated with positive outcomes among low-judgmental people—suggesting that the relationship between mindful awareness and relatively more positive (than negative) self-views should be moderated by judgmentalness (Lindsay & Creswell, 2017, 2019).

At baseline and over time, we also found that depression severity was inversely related to the drift-rate regression parameters representing positive (over negative) evidence accumulation, consistent with cognitive theories that argue negative self-views are key to the disorder (Disner et al., 2011; LeMoult & Gotlib, 2019). Notably, because this result involves a drift-rate regression parameter (where depression was associated with relatively weaker positive relative to negative evidence accumulation), rather than an effect on general drift rate, it cannot be explained by depression being associated with task features or cognitive factors (e.g., task switching, IQ), given that these should have a symmetrical effect on the drift rate (i.e., irrespective of valence).

When considering individual depression symptoms, at baseline, we found a continuum of zero-order correlation values between different symptoms and the drift-rate regression parameter (Fig. 4B, top). This spectrum of relationships appears to conflict with an influential study’s finding of striking sparsity in the variable-importance scores of different depression symptoms in a machine-learning model, which led the authors to conclude that “many symptoms typically measured in depression inventories do not appear to be strongly correlated with important cognitive processes implicated in the maintenance of depression” (Beevers et al., 2019, p. 223). Our results highlight the possibility that sparsity may arise as a result of subtle methodological differences, rather than due to the heterogeneity of depression *per se*. Specifically,

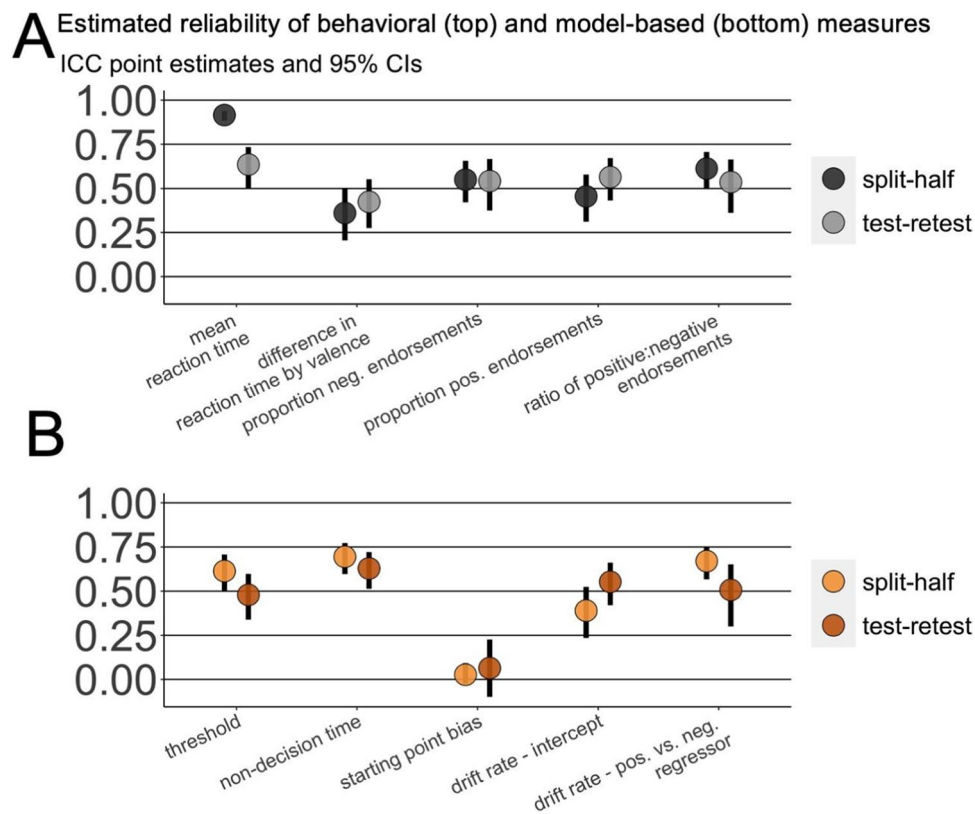


Fig. 5 Split-half and test-retest correlations and 95% CI of behavioral task measures (top) and Drift Diffusion Model (DDM) model-derived measures (bottom). *CI* = confidence interval; *ICC* = intraclass correlation coefficients; *neg.* = negative; *pos.* = positive

two key methodological differences in our study may have contributed to the differences from the results of (Beevers et al., 2019): (1) we correlated depression symptoms with a drift-rate regression parameter that incorporated the *relative* strength of positive to negative self-views—given that the strength of not just negative but also positive self-views are important to depression (Disner et al., 2011; LeMoult & Gotlib, 2019; Vanderlind et al., 2020)—which may have led to stronger relationships with depression symptoms related to anhedonia, and (2) we reported zero-order correlations, given our interest in how different depression symptoms—on their own terms, that is, without controlling for variance shared with other symptoms—related to the SRET measures, rather than via a multivariate machine-learning model. In the [Supplemental Material](#), we demonstrate, using both our empirical data and simulations based on its correlation matrix, that the multivariate machine-learning model used by Beevers et al. (2019) can generate sparse variable-importance scores even given the continuum of zero-order correlation magnitudes shown in Fig. 4B (top).

In terms of computational modeling evaluation, we found that a drift-diffusion model (DDM), which had been used in several prior SRET studies (Allison et al., 2021; Beevers et al., 2019; Dainer-Best et al., 2018a, b; Disner et al.,

2017; Price et al., 2021), was able to capture key features of the data. Specifically, it did an excellent job of capturing endorsement frequencies for positively and negatively valenced words and the shape of the reaction-time distribution for frequently endorsed responses. However, it predicted faster responses in early quantiles than were actually present in the empirical data for infrequent response types (i.e., for endorsements of negatively valenced words, and rejections of positively valenced words).

Finally, although computational model parameters sometimes have markedly higher estimated reliability than behavioral measures (Price et al., 2019), here both behavioral summary and computational modeling measures from the SRET had comparable estimated reliability, at baseline and from pre- to postintervention. Because they can extract mechanisms giving rise to behavior, model-derived metrics also can sometimes be more predictive of individual-difference or clinical-status measures than raw behavioral summary statistics (Geana et al., 2021; Pedersen et al., 2021; White et al., 2009; Wiecki et al., 2015, 2016), supporting the potential clinical utility of computational modeling. The drift rate in particular can give rise to both endorsement and predicted RT distributions and thus in some cases can show enhanced predictive utility (Geana et al., 2021; Pedersen et al., 2021; White et al., 2009; Wiecki et al.,

2015, 2016) while also matching neural dynamics of evidence accumulation leading up to a choice (Smith & Ratcliff, 2004). Yet, we did not find evidence of a substantial distinction in the relationship between mindfulness and depression measures to the drift-rate regression parameters versus comparable summary statistics (differences in endorsement frequencies by valence). This lack of improvement emphasizes the importance of new task designs that can dissect the dynamics of self-referential processes more precisely (as we discuss in more detail below). These may be especially powerful when combined with rigorous computational modeling practices, such as those employed here (e.g., model comparison, model validation).

A key strength of this paper was that we conducted extensive computational model checking, diagnostics, and validation including model comparison, parameter recovery, and posterior predictive checks. These steps, which are essential to computational modeling (Nassar & Frank, 2016; Palminteri et al., 2017; Wilson & Collins, 2019), and the DDM in particular (Fengler et al., 2021; Frank et al., 2015; Stine et al., 2020), enabled us to draw precise conclusions about how well our models could capture the empirical data and how the models fared against each other. A second strength was having longitudinal intervention data of high quantity and quality. Number of participants (completer $n = 96$) and number of SRET trials per participant (trial $n = 118$) was high, and participants appeared to be motivated and paying attention, as suggested by the fact that their performance on an interleaved control task was very high (median correct = 98%).

The rigorous application of sequential-sampling models, such as the DDM, to a valenced self-judgment task paves the way for precise multimodal investigation of this process that may provide detailed insight into its neurocomputational mechanisms (Allison et al., 2021; Lawrence et al., 2022) and thereby facilitate targeted assessment and intervention (Aizik-Reebs et al., 2022; Dainer-Best et al., 2018c). Indeed, algorithmic models, such as the DDM, play a special bridging role in model-based cognitive neuroscience and computational psychiatry, by decomposing behavioral task data into specific mechanisms, which can in turn be linked to neural circuitry and biologically detailed models (Badre et al., 2015; Forstmann et al., 2016; Gläscher & O’Doherty, 2010; Hitchcock et al., 2022; Huys et al., 2016; Turner et al., 2017; Wiecki et al., 2015). This decomposition can therefore accelerate *target identification* in intervention research. Target identification is the first step toward demonstrating *target engagement*: that an intervention changes a neurocomputational target and in turn leads to a predicted change in clinical outcome variables (Insel & Gogtay, 2014; Murphy et al., 2009; Onken et al., 2014). Target identification is especially important in affective (Roberts & Hutcherson, 2019) and self-relevance (Britton et al., 2021) research, because these processes are key targets in many interventions, yet have been difficult to measure precisely.

The idea that sequential-sampling models can provide a useful decomposition of valenced self-judgment can be grounded in the notion that self-judgment decisions involve repeated sampling from memory. In particular, rather than forming a one-off judgment about self-relevance, memory samples are thought to accumulate in the form of a decision variable that evolves until a threshold is reached—at which point an option is selected. In value-based decisions, such as deciding between two snack foods, this process is thought to involve sampling episodic and other memories concerning the two foods, combined with noisy attention to different attributes of those foods, culminating in a choice of one food over another; in the DDM, a higher drift rate is thought to correspond to stronger positive associations for one choice over the other during this integrative sampling process, leading to faster and more consistent decisions (Bakkour et al., 2019; Krajbich, 2019; Shadlen & Shohamy, 2016; but see Frömer et al., 2019). In the SRET, which involves deciding whether an adjective describes oneself (yes or no), the sampling process presumably involves episodic memories and/or cached inferences about oneself; in the DDM, a higher drift rate is assumed to correspond to stronger associations in favor of one option (e.g., yes) over the other.¹

Notably, our finding that the key behavioral differences and self-report correlations in the task were captured by the drift rate is not trivial, because studies often found that individual or group differences are reflected in other DDM components, without necessarily affecting the drift rate. For instance, an intervention study found that active attentional bias training (vs. a sham) influenced nondecision time in a task involving attentional orienting to threat (Price et al., 2019). Healthy aging appears primarily to affect the threshold and non-decision time, rather than the drift rate (Ratcliff et al., 2010; von Krause et al., 2022), whereas IQ appears to influence the drift rate (Ratcliff et al., 2010). Distinct DDM components also covary with changes in different neural circuitry (Forstmann et al., 2016); in perceptual decision-making tasks, the drift rate has tended to co-vary with frontoparietal circuitry (Mulder et al., 2014), but sampling from memory in value-based choices (which may be more similar to self-judgment tasks) are thought to involve hippocampus and a wide swath of association cortex (Biderman et al., 2020; Shadlen & Shohamy, 2016).

The literature on the neurocomputational mechanisms of valenced self-judgment is nascent, but already offers some hints as to how this process could be decomposed, and the potential clinical utility of such a decomposition. In a cross-sectional electroencephalography study with the SRET, individuals with major depressive disorder (vs. healthy controls) showed a relatively

¹ In our implementation, more positive drift rates corresponded to higher evidence favoring yes (over no). Our key results concerned a drift-rate regression parameter (and the change in this variable over time), indexing the relative difference in drift rates between positive and negative adjectives (as well as the change in this parameter over time).

larger late-positive, event-related potential (ERP) for negative than positive words, which correlated with higher endorsement of negative words (Dainer-Best et al., 2017). Interestingly, in a sample of individuals matched on many clinical characteristics, but differing in whether they recently attempted suicide, there were no behavioral or significant drift-rate differences, but heightened activation among the attempter group for negative (versus) positive words in an ERP that peaks ~200 ms, which has been implicated in early monitoring of semantic content (Allison et al., 2021). This may reflect a distinct mechanism than the drift-rate difference found in the current study to relate to depression symptoms, and to inversely relate to mindful awareness and which may have been the same mechanism responsible for the late potential difference among depressed individuals in (Dainer-Best et al., 2017). As noted by Allison et al. (2021), it is plausible that individuals with a heightened early-processing component may require adjunctive treatments to psychotherapy, whereas it is plausible that differences in the later-stage processes could change as a result of psychotherapy. We expect that this developing literature will continue to offer insights as careful computational modeling and experimental design are jointly applied.

Although we believe that our study makes an important contribution to this literature, it also has some key limitations. First, participants were quite homogenous both in terms of demographics (e.g., 98% white), and because the clinical trial from which data were drawn excluded many disorders with which depression is often comorbid (e.g., history of borderline personality disorders, current posttraumatic stress disorder; see full list in “Methods”). It is essential that future studies employ more diverse participants and examine whether the key findings generalize, including not only the overall relationships between mindfulness and depression and the valence drift-rate regressor (and changes in these variables), but also the depression-symptom specific relationships. Second, although all participants began the study with at least mild depression symptoms, even prior to the intervention the large majority of participants endorsed most of the positive adjectives and rejected most of the negative adjectives. This asymmetry in the SRET is common in nonclinically selected populations (Dainer-Best et al., 2018b), whereas samples with more severe psychopathology tend to show a more even number of positive and negative endorsements (Allison et al., 2021). It is plausible that our depression findings will generalize to clinical populations, because dimensional models of depression contend that there is no qualitative distinction among clinical and nonclinical populations (Ruscio, 2019); however, it is important that our findings be replicated in clinical samples to confirm that this is true. Third, the mindful awareness (FFMQ-Aware) and nonjudgmentalness (FFMQ-NonJudge) measures both involve exclusively negative wording, which may lead to “method effects” (Van Dam et al., 2009, 2012; but see also Baer et al., 2011). This potential limitation should be borne in mind when interpreting our results involving these

subscales (see [Supplemental Material](#) for further discussion). Fourth, many of our key analyses extracted *maximum a posteriori* estimates derived from hierarchical Bayesian estimation and then applied these in correlations/regressions. This approach is valid, because hierarchical Bayesian estimation tends to improve such point estimates (Katahira 2016; see more detailed discussion in “Methods”), but estimation using a fully Bayesian approach to the regression/correlations would have allowed for the reporting of posteriors around the estimates. Fifth, because this was a secondary analysis of clinical-trial data, where sample size had been determined at the trial-planning stage, we did not conduct a power analysis before our study to inform sample size for our analyses. Sixth, although our results provided more support for an account that views mindful awareness as leading to positive change, without qualification, rather than one that views mindful awareness as intrinsically qualified by judgmentalness, we believe that more granular analyses of change processes are required to provide convincing support for the former model (Ludwig et al., 2020). For instance, although that model argues for a causal role of awareness in producing healthy behavior (which should lead to more positive self-views), an alternate possibility is that positive changes in self-views lead to less attention-preoccupying negative mental content, which in turn may naturally increase present-moment awareness (Whitmer & Gotlib, 2013). Our findings cannot distinguish between these possibilities. Finally, it is noteworthy that, when depression and mindful awareness (and changes therein) were entered as co-variables in a single model, only one of these variables remained a significant predictor of the drift-rate regression parameter (and change in this parameter), suggesting that shared variance in these variables was partially responsible for the relationships of these variables to this parameter (see “Robustness Checks...” in [Supplemental Material](#)).

In sum, we leveraged a longitudinal dataset and computational modeling to examine how individual differences, and changes, in depression and mindfulness facets related to specific valenced self-judgment computations. We corroborated a key tenet of cognitive theories of depression—that depression inversely relates to the relative strength of positive to negative self-views (at baseline and over time)—and found that mindful awareness was associated with more positive self-judgments (again at baseline and over time). This latter finding makes an especially key contribution to the literature, given the paucity of past research on how self-relevant processes relate to changes in mindfulness over time, despite the centrality of self-relevant processes to theories of the mechanisms of mindfulness interventions (Britton et al., 2021; Vago & Silbersweig, 2012). We also extended influential past research that had suggested that only a small subset of depression symptoms related to valenced self-judgment (Beevers et al., 2019) by applying different methodology and by examining change over time. At baseline, we found evidence of a continuum of relationship strengths between

depression symptoms and valenced self-judgment. We highlighted how methodological differences—rather than the heterogeneity of depression *per se*—may lead to these different findings than those of Beevers et al. (2019). Finally, although we argued that the rigorous application of computational modeling to the SRET is valuable, insofar as it increases mechanistic specificity and thereby paves the way for improved experimental designs and multimodal investigation, computational modeling (relative to behavioral summaries) did not in this case substantially enhance split-half and test-retest reliability, underscoring that it is not a given that computational modeling will do so. This lack of substantial improvement is important to document (Pike et al., 2022) and highlights the utility of developing new experimental designs that—especially in combination with theory-driven computational modeling and multimodal analysis—can advance the assessment and targeted intervention of valenced self-judgment (see Duan et al., 2021; Hadash et al., 2016; Lawrence et al., 2022; Price et al., 2021; Shany et al., 2022 for examples of recent task developments).

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.3758/s13415-022-01033-9>.

Acknowledgments The authors thank The Clinical and Affective Neuroscience Laboratory for help throughout the mindfulness dismantling study.

This work was supported by the National Institutes of Health (NIH) Science of Behavior Change Common Fund Program through an award administered by the National Center for Complementary and Integrative Health UH2AT009145. WBB received funding from NIH grant K23AT006328-01A1. PFH was supported by NIMH grant F32MH123055. MJF was supported by NIMH grants P50MH119467 and R01MH084840-08A. The views presented are solely the responsibility of the authors and do not necessarily represent the official views of the NIH.

Materials Availability Code used to produce the results and figures is available at github.com/peter-hitchcock/britsr.

Declarations

Disclosures Dr. Britton is a MBSR and MBCT teacher and has received financial compensation for this role. Dr. Britton is nominally affiliated with the Mindfulness Center at Brown University, which generates income by offering mindfulness classes to the public. Dr. Britton is the founder of Cheetah House, a RI nonprofit organization that provides information about meditation-related difficulties, individual consultations, and support groups, as well as educational trainings to meditation teachers, clinicians, educators, and mindfulness providers. This interest has been disclosed to and is being managed by Brown University in accordance with its Conflict of Interest and Conflict of Commitment policies. Other authors declare no conflicts of interest.

References

- Aguado, J., Luciano, J. V., Cebolla, A., Serrano-Blanco, A., Soler, J., & García-Campayo, J. (2015). Bifactor analysis and construct validity of the five facet mindfulness questionnaire (FFMQ) in non-clinical Spanish samples. *Frontiers in Psychology*, 6, 404. <https://doi.org/10.3389/fpsyg.2015.00404>
- Aizik-Reebs, A., Amir, I., Yuval, K., Hadash, Y., & Bernstein, A. (2022). Candidate mechanisms of action of mindfulness-based trauma recovery for refugees (MBTR-R): Self-compassion and self-criticism. *Journal of Consulting and Clinical Psychology*, 90(2), 107–122. <https://doi.org/10.1037/ccp0000716>
- Alejandre-Lara, A. L., Canby, N. K., Wesbecher, K. D., Eichel, K., Britton, W. B., & Lindahl, J. R. (2022). How do mindfulness-based programs improve depression symptoms: Selflessness, valence, or valenced self? *Cognitive Therapy and Research*. <https://doi.org/10.1007/s10608-021-10287-5>
- Allison, G. O., Benau, E. M., Asbaghi, S., Pagliaccio, D., Stewart, J. G., & Auerbach, R. P. (2021). Neurophysiological markers related to negative self-referential processing differentiate adolescent suicide ideators and attempters. *Biological Psychiatry Global Open Science*, 1(1), 16–27. <https://doi.org/10.1016/j.bpsgos.2021.04.001>
- Anderson, N. H. (1968). Likableness ratings of 555 personality-trait words. *Journal of Personality and Social Psychology*, 9(3), 272–279. <https://doi.org/10.1037/h0025907>
- Badre, D., Frank, M. J., & Moore, C. I. (2015). Interactionist neuroscience. *Neuron*, 88(5), 855–860. <https://doi.org/10.1016/j.neuron.2015.10.021>
- Baer, R. (2019). Assessment of mindfulness by self-report. *Current Opinion in Psychology*, 28, 42–48. <https://doi.org/10.1016/j.copsyc.2018.10.015>
- Baer, R. A., Smith, G. T., Hopkins, J., Krietemeyer, J., & Toney, L. (2006). Using self-report assessment methods to explore facets of mindfulness. *Assessment*, 13(1), 27–45. <https://doi.org/10.1177/1073191105283504>
- Baer, R. A., Smith, G. T., Lykins, E., Button, D., Krietemeyer, J., Sauer, S., Walsh, E., Duggan, D., & Williams, J. M. G. (2008). Construct validity of the five facet mindfulness questionnaire in meditating and nonmeditating samples. *Assessment*, 15(3), 329–342. <https://doi.org/10.1177/1073191107313003>
- Baer, R. A., Samuel, D. B., & Lykins, E. L. B. (2011). Differential item functioning on the five facet mindfulness questionnaire is minimal in demographically matched meditators and nonmeditators. *Assessment*, 18(1), 3–10. <https://doi.org/10.1177/1073191110392498>
- Bakkour, A., Palombo, D. J., Zylberberg, A., Kang, Y. H., Reid, A., Verfaellie, M., Shadlen, M. N., & Shohamy, D. (2019). The hippocampus supports deliberation during value-based decisions. *eLife*, 8. <https://doi.org/10.7554/eLife.46080>
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2014). Fitting linear mixed-effects models using lme4. In arXiv [stat.CO]. arXiv. <http://arxiv.org/abs/1406.5823>
- Beevers, C. G., Mullarkey, M. C., Dainer-Best, J., Stewart, R. A., Labrada, J., Allen, J. J. B., McGeary, J. E., & Shumake, J. (2019). Association between negative cognitive bias and depression: A symptom-level approach. *Journal of Abnormal Psychology*, 128(3), 212–227. <https://doi.org/10.1037/abn0000405>
- Bennett, D., & Niv, Y. (2018). *Opening Burton's Clock: Psychiatric insights from computational cognitive models*. <https://doi.org/10.31234/osf.io/y2vzu>
- Biderman, N., Bakkour, A., & Shohamy, D. (2020). What are memories for? The Hippocampus Bridges past experience with future decisions. *Trends in Cognitive Sciences*, 24(7), 542–556. <https://doi.org/10.1016/j.tics.2020.04.004>
- Boehm, U., Marsman, M., Matzke, D., & Wagenmakers, E.-J. (2018). On the importance of avoiding shortcuts in applying cognitive models to hierarchical data. *Behavior Research Methods*, 50(4), 1614–1631. <https://doi.org/10.3758/s13428-018-1054-3>

- Bohlmeijer, E., ten Klooster, P. M., Fledderus, M., Veehof, M., & Baer, R. (2011). Psychometric properties of the five facet mindfulness questionnaire in depressed adults and development of a short form. *Assessment, 18*(3), 308–320. <https://doi.org/10.1177/1073191111408231>
- Brewer, J. (2019). Mindfulness training for addictions: has neuroscience revealed a brain hack by which awareness subverts the addictive process? *Current Opinion in Psychology, 28*, 198–203. <https://doi.org/10.1016/j.copsyc.2019.01.014>
- Britton, W. B. (2019). Can mindfulness be too much of a good thing? The value of a middle way. *Current Opinion in Psychology, 28*, 159–165. <https://doi.org/10.1016/j.copsyc.2018.12.011>
- Britton, W. B., Davis, J. H., Loucks, E. B., Peterson, B., Cullen, B. H., Reuter, L., Rando, A., Rahrigh, H., Lipsky, J., & Lindahl, J. R. (2018). Dismantling mindfulness-based cognitive therapy: Creation and validation of 8-week focused attention and open monitoring interventions within a 3-armed randomized controlled trial. *Behaviour Research and Therapy, 101*, 92–107. <https://doi.org/10.1016/j.brat.2017.09.010>
- Britton, W. B., Desbordes, G., Acabchuk, R., Peters, S., Lindahl, J. R., Canby, N. K., Vago, D. R., Dumais, T., Lipsky, J., Kimmel, H., Sager, L., Rahrigh, H., Cheaito, A., Acero, P., Scharf, J., Lazar, S. W., Schuman-Olivier, Z., Ferrer, R., & Moitra, E. (2021). From self-esteem to selflessness: An evidence (Gap) map of self-related processes as mechanisms of mindfulness-based interventions. *Frontiers in Psychology, 12*, 730972. <https://doi.org/10.3389/fpsyg.2021.730972>
- Brown, K. W., & Ryan, R. M. (2003). The benefits of being present: Mindfulness and its role in psychological well-being. *Journal of Personality and Social Psychology, 84*(4), 822–848. <https://doi.org/10.1037/0022-3514.84.4.822>
- Brown, T. A., Chorpita, B. F., Korotitsch, W., & Barlow, D. H. (1997). Psychometric properties of the Depression Anxiety Stress Scales (DASS) in clinical samples. *Behaviour Research and Therapy, 35*(1), 79–89. [https://doi.org/10.1016/s0005-7967\(96\)00068-x](https://doi.org/10.1016/s0005-7967(96)00068-x)
- Brown, V. M., Chen, J., Gillan, C. M., & Price, R. B. (2020). Improving the reliability of computational analyses: Model-based planning and its relationship with compulsivity. *Biological Psychiatry. Cognitive Neuroscience and Neuroimaging, 5*(6), 601–609. <https://doi.org/10.1016/j.bpsc.2019.12.019>
- Brown, V. M., Zhu, L., Solway, A., Wang, J. M., McCurry, K. L., King-Casas, B., & Chiu, P. H. (2021). Reinforcement learning disruptions in individuals with depression and sensitivity to symptom change following cognitive behavioral therapy. *JAMA Psychiatry, 78*(10), 1113–1122. <https://doi.org/10.1001/jamapsychiatry.2021.1844>
- Brown, K. W., Berry, D., Eichel, K., Beloborodova, P., Rahrigh, H., & Britton, W. B. (2022). Comparing impacts of meditation training in focused attention, open monitoring, and mindfulness-based cognitive therapy on emotion reactivity and regulation: Neural and subjective evidence from a dismantling study. *Psychophysiology, 59*(7), e14024. <https://doi.org/10.1111/psyp.14024>
- Cataldo, A. M., Scheuer, L., Maksimovskiy, A. L., Germaine, L. T., & Dillon, D. G. (2022). Abnormal evidence accumulation underlies the positive memory deficit in depression. *Journal of Experimental Psychology. General, 151*(1), 1–12. <https://doi.org/10.1037/xge0001268>
- Choi, E., Farb, N., Pogrebtsova, E., Gruman, J., & Grossmann, I. (2021). What do people mean when they talk about mindfulness? *Clinical Psychology Review, 89*, 102085. <https://doi.org/10.1016/j.cpr.2021.102085>
- Cohen, Z. D., & DeRubeis, R. J. (2018). Treatment selection in depression. *Annual Review of Clinical Psychology, 14*, 209–236. <https://doi.org/10.1146/annurev-clinpsy-050817-084746>
- Cullen, B., Eichel, K., Lindahl, J. R., Rahrigh, H., Kini, N., Flahive, J., & Britton, W. B. (2021). The contributions of focused attention and open monitoring in mindfulness-based cognitive therapy for affective disturbances: A 3-armed randomized dismantling trial. *PLoS One, 16*(1), e0244838. <https://doi.org/10.1371/journal.pone.0244838>
- Curtiss, J., & Klemanski, D. H. (2014). Factor analysis of the five facet mindfulness questionnaire in a heterogeneous clinical sample. *Journal of Psychopathology and Behavioral Assessment, 36*(4), 683–694. <https://doi.org/10.1007/s10862-014-9429-y>
- Dainer-Best, J., Trujillo, L. T., Schnyer, D. M., & Beevers, C. G. (2017). Sustained engagement of attention is associated with increased negative self-referent processing in major depressive disorder. *Biological Psychology, 129*, 231–241. <https://doi.org/10.1016/j.biopsycho.2017.09.005>
- Dainer-Best, J., Disner, S. G., McGeary, J. E., Hamilton, B. J., & Beevers, C. G. (2018a). Negative self-referential processing is associated with genetic variation in the serotonin transporter-linked polymorphic region (5-HTTLPR): Evidence from two independent studies. *PLoS One, 13*(6), e0198950. <https://doi.org/10.1371/journal.pone.0198950>
- Dainer-Best, J., Lee, H. Y., Shumake, J. D., Yeager, D. S., & Beevers, C. G. (2018b). Determining optimal parameters of the self-referent encoding task: A large-scale examination of self-referent cognition and depression. *Psychological Assessment, 30*(11), 1527–1540. <https://doi.org/10.1037/pas0000602>
- Dainer-Best, J., Shumake, J. D., & Beevers, C. G. (2018c). Positive imagery training increases positive self-referent cognition in depression. *Behaviour Research and Therapy, 111*, 72–83. <https://doi.org/10.1016/j.brat.2018.09.010>
- Derry, P. A., & Kuiper, N. A. (1981). Schematic processing and self-reference in clinical depression. *Journal of Abnormal Psychology, 90*(4), 286–297. <https://doi.org/10.1037//0021-843x.90.4.286>
- Desbordes, G., Gard, T., Hoge, E. A., Hölzel, B. K., Kerr, C., Lazar, S. W., Olendzki, A., & Vago, D. R. (2014). Moving beyond mindfulness: Defining equanimity as an outcome measure in meditation and contemplative research. *Mindfulness, 2014*(January). <https://doi.org/10.1007/s12671-013-0269-8>
- Disner, S. G., Beevers, C. G., Haigh, E. A. P., & Beck, A. T. (2011). Neural mechanisms of the cognitive model of depression. *Nature Reviews Neuroscience, 12*(8), 467–477. <https://doi.org/10.1038/nrn3027>
- Disner, S. G., Shumake, J. D., & Beevers, C. G. (2017). Self-referential schemas and attentional bias predict severity and naturalistic course of depression symptoms. *Cognition & Emotion, 31*(4), 632–644. <https://doi.org/10.1080/02699931.2016.1146123>
- Duan, S., Lawrence, A., Valmaggia, L., Moll, J., & Zahn, R. (2021). Maladaptive blame-related action tendencies are associated with vulnerability to major depressive disorder. *Journal of Psychiatric Research, 145*, 70–76. <https://doi.org/10.1016/j.jpsychires.2021.11.043>
- Fengler, A., Govindarajan, L. N., Chen, T., & Frank, M. J. (2021). Likelihood approximation networks (LANs) for fast inference of simulation models in cognitive neuroscience. *eLife, 10*. <https://doi.org/10.7554/eLife.65074>
- Forstmann, B. U., Ratcliff, R., & Wagenmakers, E.-J. (2016). Sequential sampling models in cognitive neuroscience: Advantages, applications, and extensions. *Annual Review of Psychology, 67*, 641–666. <https://doi.org/10.1146/annurev-psych-122414-033645>
- Frank, M. J., Gagne, C., Nyhus, E., Masters, S., Wiecki, T. V., Cavanagh, J. F., & Badre, D. (2015). fMRI and EEG predictors of dynamic decision parameters during human reinforcement learning. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience, 35*(2), 485–494. <https://doi.org/10.1523/JNEUROSCI.2036-14.2015>
- Fried, E. I. (2015). Schematic assumptions have slowed down depression research: Why symptoms, not syndromes are the way forward. *Frontiers in Psychology, 6*, 309. <https://doi.org/10.3389/fpsyg.2015.00309>

- Fried, E. I., & Nesse, R. M. (2015). Depression is not a consistent syndrome: an investigation of unique symptom patterns in the STAR* D study. *Journal of Affective Disorders*, *172*, 96–102. https://www.sciencedirect.com/science/article/pii/S0165032714006326?casa_token=s8wMm6bUv90AAAAA:6k1YXnIi1pL_Plywoqa-GYq7jW157GXMdntf96rQ2hEd_jNMPKDo8nDrNggT wzY_5dl8ntqcbQY
- Frömer, R., Dean Wolf, C. K., & Shenhav, A. (2019). Goal congruency dominates reward value in accounting for behavioral and neural correlates of value-based decision-making. *Nature Communications*, *10*(1), 4926. <https://doi.org/10.1038/s41467-019-12931-x>
- Geana, A., Barch, D. M., Gold, J. M., Carter, C. S., MacDonald 3rd, A. W., Ragland, J. D., Silverstein, S. M., & Frank, M. J. (2021). Using computational modeling to capture schizophrenia-specific reinforcement learning differences and their implications on patient classification. *Biological Psychiatry. Cognitive Neuroscience and Neuroimaging*. <https://doi.org/10.1016/j.bpsc.2021.03.017>
- Gelman, A., Carlin, J. B., Stern, H. S., & Rubin, D. B. (1995). *Bayesian data analysis*. Chapman and Hall/CRC <https://www.taylorfrancis.com/books/mono/10.1201/9780429258411/bayesian-data-analysis-andrew-gelman-john-carlin-hal-stern-donald-rubin>
- Gläscher, J. P., & O'Doherty, J. P. (2010). Model-based approaches to neuroimaging: combining reinforcement learning theory with fMRI data. *Wiley Interdisciplinary Reviews. Cognitive Science*, *1*(4), 501–510. <https://doi.org/10.1002/wcs.57>
- Gu, J., Strauss, C., Crane, C., Barnhofer, T., Karl, A., Cavanagh, K., & Kuyken, W. (2016). Examining the factor structure of the 39-item and 15-item versions of the Five Facet Mindfulness Questionnaire before and after mindfulness-based cognitive therapy for people with recurrent depression. *Psychological Assessment*, *28*(7), 791–802. <https://doi.org/10.1037/pas0000263>
- Guest, O., & Martin, A. E. (2021). How computational modeling can force theory building in psychological science. *Perspectives on Psychological Science: A Journal of the Association for Psychological Science*, *17*(4), 569–585. <https://doi.org/10.1177/1745691620970585>
- Hadash, Y., Plonsker, R., Vago, D. R., & Bernstein, A. (2016). Experiential self-referential and selfless processing in mindfulness and mental health: Conceptual model and implicit measurement methodology. *Psychological Assessment*, *28*(7), 856–869. <https://doi.org/10.1037/pas0000300>
- Haines, N., Kvam, P. D., Irving, L. H., Smith, C., Beauchaine, T. P., Pitt, M. A., Ahn, W.-Y., & Turner, B. (2020). *Learning from the reliability paradox: How theoretically informed generative models can advance the social, behavioral, and brain sciences*. <https://psyarxiv.com/xr7y3/download?format=pdf>
- Hayes, S. C., Strosahl, K. D., & Wilson, K. G. (2011). *Acceptance and commitment therapy, second edition: The process and practice of mindful change*. Guilford Press <https://play.google.com/store/books/details?id=og28CwAAQBAJ>
- Hedge, C., Powell, G., & Sumner, P. (2018). The reliability paradox: Why robust cognitive tasks do not produce reliable individual differences. *Behavior Research Methods*, *50*(3), 1166–1186. <https://doi.org/10.3758/s13428-017-0935-1>
- Hitchcock, P. F., Martin, L. M., Fischer, L., Marando-Blanck, S., & Herbert, J. D. (2016). Popular conceptions of mindfulness: Awareness and emotional control. *Mindfulness*, *7*(4), 940–949. <https://doi.org/10.1007/s12671-016-0533-9>
- Hitchcock, P., Niv, Y., Radulescu, A., & Sims, C. R. (2017). Translating a reinforcement learning task into a computational psychiatry assay: Challenges and strategies. *CogSci*. https://www.researchgate.net/profile/Peter_Hitchcock2/publication/316220935_Translating_a_Reinforcement_Learning_Task_into_a_Computational_Psychiatry_Assay_Challenges_and_Strategies/links/59035bb1aca272116d2fa4d9/Translating-a-Reinforcement-Learning-Task-into-a-Computational-Psychiatry-Assay-Challenges-and-Strategies.pdf
- Hitchcock, P. F., Fried, E. I., & Frank, M. J. (2022). Computational psychiatry needs time and context. *Annual Review of Psychology*, *73*, 243–270. <https://doi.org/10.1146/annurev-psych-021621-124910>
- Huys, Q. J. M., Maia, T. V., & Frank, M. J. (2016). Computational psychiatry as a bridge from neuroscience to clinical applications. *Nature Neuroscience*, *19*(3), 404–413. <https://doi.org/10.1038/nn.4238>
- Insel, T. R., & Gogtay, N. (2014). National Institute of Mental Health clinical trials: New opportunities, new expectations. *JAMA Psychiatry*, *71*(7), 745–746. <https://doi.org/10.1001/jamapsychiatry.2014.426>
- Katahira, K. (2016). How hierarchical models improve point estimates of model parameters at the individual level. *Journal of Mathematical Psychology*, *73*, 37–58. <https://doi.org/10.1016/j.jmp.2016.03.007>
- Konova, A. B., Lopez-Guzman, S., Urmanche, A., Ross, S., Louie, K., Rotrosen, J., & Glimcher, P. W. (2020). Computational markers of risky decision-making for identification of temporal windows of vulnerability to opioid use in a real-world clinical setting. *JAMA Psychiatry*, *77*(4), 368–377. <https://doi.org/10.1001/jamapsychiatry.2019.4013>
- Krajbich, I. (2019). Accounting for attention in sequential sampling models of decision making. *Current Opinion in Psychology*, *29*, 6–11. <https://doi.org/10.1016/j.copsyc.2018.10.008>
- Lawlor, V. M., Webb, C. A., Wiecki, T. V., Frank, M. J., Trivedi, M., Pizzagalli, D. A., & Dillon, D. G. (2020). Dissecting the impact of depression on decision-making. *Psychological Medicine*, *50*(10), 1613–1622. <https://doi.org/10.1017/S0033291719001570>
- Lawrence, A. J., Stahl, D., Duan, S., Fennema, D., Jaekle, T., Young, A. H., Dazzan, P., Moll, J., & Zahn, R. (2022). Neurocognitive measures of self-blame and risk prediction models of recurrence in major depressive disorder. *Biological Psychiatry. Cognitive Neuroscience and Neuroimaging*, *7*(3), 256–264. <https://doi.org/10.1016/j.bpsc.2021.06.010>
- Lecuona, O., García-Garzón, E., García-Rubio, C., & Rodríguez-Carvajal, R. (2020). A psychometric review and conceptual replication study of the five facets mindfulness questionnaire latent structure. *Assessment*, *27*(5), 859–872. <https://doi.org/10.1177/1073191119873718>
- LeMoult, J., & Gotlib, I. H. (2019). Depression: A cognitive perspective. *Clinical Psychology Review*, *69*, 51–66. <https://doi.org/10.1016/j.cpr.2018.06.008>
- Lester, E. G., Murrell, A. R., & Dickson, D. E. (2018). A mixed methods approach to understanding conceptions of mindfulness meditation. *OBM Integrative and Complementary Medicine*, *3*(4), 1–1. <https://www.lidsen.com/journals/icm/icm-03-04-026>
- Lilienfeld, S. O. (2014). The Research Domain Criteria (RDoC): an analysis of methodological and conceptual challenges. *Behaviour Research and Therapy*, *62*, 129–139. <https://doi.org/10.1016/j.brat.2014.07.019>
- Lindsay, E. K., & Creswell, J. D. (2017). Mechanisms of mindfulness training: Monitor and Acceptance Theory (MAT). *Clinical Psychology Review*, *51*, 48–59. <https://doi.org/10.1016/j.cpr.2016.10.011>
- Lindsay, E. K., & Creswell, J. D. (2019). Mindfulness, acceptance, and emotion regulation: Perspectives from Monitor and Acceptance Theory (MAT). *Current Opinion in Psychology*, *28*, 120–125. <https://doi.org/10.1016/j.copsyc.2018.12.004>
- Ludwig, V. U., Brown, K. W., & Brewer, J. A. (2020). Self-regulation without force: Can awareness leverage reward to drive behavior change? *Perspectives on Psychological Science: A Journal of the Association for Psychological Science*, *15*(6), 1382–1399. <https://doi.org/10.1177/1745691620931460>

- Lutz, A., Slagter, H. A., Dunne, J. D., & Davidson, R. J. (2008). Attention regulation and monitoring in meditation. *Trends in Cognitive Sciences*, 12(4), 163–169. <https://doi.org/10.1016/j.tics.2008.01.005>
- Maia, T. V., & Frank, M. J. (2011). From reinforcement learning models to psychiatric and neurological disorders. *Nature Neuroscience*, 14(2), 154–162. <https://doi.org/10.1038/nn.2723>
- Maia, T. V., Huys, Q. J. M., & Frank, M. J. (2017). Theory-based computational psychiatry. *Biological Psychiatry*, 82(6), 382–384. <https://doi.org/10.1016/j.biopsych.2017.07.016>
- McNally, R. J. (2019). Attentional bias for threat: Crisis or opportunity? *Clinical Psychology Review*, 69, 4–13. <https://doi.org/10.1016/j.cpr.2018.05.005>
- Miller, J. M., Schneck, N., Siegle, G. J., Chen, Y., Ogden, R. T., Kikuchi, T., Oquendo, M. A., Mann, J. J., & Parsey, R. V. (2013). fMRI response to negative words and SSRI treatment outcome in major depressive disorder: A preliminary study. *Psychiatry Research*, 214(3), 296–305. <https://doi.org/10.1016/j.psychres.2013.08.001>
- Millner, A. J., Robinaugh, D. J., & Nock, M. K. (2020). Advancing the understanding of suicide: The need for formal theory and rigorous descriptive research. *Trends in Cognitive Sciences*, 24(9), 704–716. <https://doi.org/10.1016/j.tics.2020.06.007>
- Mkrtrchian, A., Valtou, V., & Roiser, J. P. (2021). Reliability of decision-making and reinforcement learning computational parameters. In *bioRxiv* (p. 2021.06.30.450026). <https://doi.org/10.1101/2021.06.30.450026>
- Mulder, Van Maanen, & Forstmann. (2014). Perceptual decision neurosciences—a model-based review. *Neuroscience*. <https://www.sciencedirect.com/science/article/pii/S0306452214006046>
- Murphy, R., Cooper, Z., Hollon, S. D., & Fairburn, C. G. (2009). How do psychological treatments work? Investigating mediators of change. *Behaviour Research and Therapy*, 47(1), 1–5. <https://doi.org/10.1016/j.brat.2008.10.001>
- Nassar, M. R., & Frank, M. J. (2016). Taming the beast: extracting generalizable knowledge from computational models of cognition. *Current Opinion in Behavioral Sciences*, 11, 49–54. <https://doi.org/10.1016/j.cobeha.2016.04.003>
- Niv, Y. (2009). Reinforcement learning in the brain. *Journal of Mathematical Psychology*, 53(3), 139–154. <https://doi.org/10.1016/j.jmp.2008.12.005>
- Onken, L. S., Carroll, K. M., Shoham, V., Cuthbert, B. N., & Riddle, M. (2014). Reenvisioning clinical science: Unifying the discipline to improve the public health. *Clinical Psychological Science*, 2(1), 22–34. <https://doi.org/10.1177/2167702613497932>
- Padmanabham, S. S., Bilgrami, Z. R., Eisenlohr-Moul, T., & Van Dam, N. T. (2021). Refinement and validation of the Balanced Inventory of Mindfulness-Related Skills (BIMS). *Mindfulness*, 12(5), 1208–1223. <https://doi.org/10.1007/s12671-020-01590-5>
- Palminteri, S., Wyart, V., & Koehlin, E. (2017). The importance of falsification in computational cognitive modeling. *Trends in Cognitive Sciences*, 21(6), 425–433. <https://doi.org/10.1016/j.tics.2017.03.011>
- Parsons, S., Kruijt, A.-W., & Fox, E. (2019). Psychological science needs a standard practice of reporting the reliability of cognitive-behavioral measurements. *Advances in Methods and Practices in Psychological Science*, 2(4), 378–395. <https://doi.org/10.1177/2515245919879695>
- Pedersen, M. L., & Frank, M. J. (2020). Simultaneous hierarchical bayesian parameter estimation for reinforcement learning and drift diffusion models: A tutorial and links to neural data. *Computational Brain & Behavior*. <https://link.springer.com/content/pdf/10.1007/s42113-020-00084-w.pdf>
- Pedersen, M. L., Ironside, M., Amemori, K.-I., McGrath, C. L., Kang, M. S., Graybiel, A. M., Pizzagalli, D. A., & Frank, M. J. (2021). Computational phenotyping of brain-behavior dynamics underlying approach-avoidance conflict in major depressive disorder. *PLoS Computational Biology*, 17(5), e1008955. <https://doi.org/10.1371/journal.pcbi.1008955>
- Pike, A. C., Tan, K., Ansari, H. J., Wing, M., & Robinson, O. J. (2022). Test-retest reliability of affective bias tasks. <https://doi.org/10.31234/osf.io/n2fkh>
- Price, R. B., Brown, V., & Siegle, G. J. (2019). Computational modeling applied to the dot-probe task yields improved reliability and mechanistic insights. *Biological Psychiatry*, 85(7), 606–612. <https://doi.org/10.1016/j.biopsych.2018.09.022>
- Price, R. B., Panny, B., Degutis, M., & Griffo, A. (2021). Repeated measurement of implicit self-associations in clinical depression: Psychometric, neural, and computational properties. *Journal of Abnormal Psychology*, 130(2), 152–165. <https://doi.org/10.1037/abn0000651>
- Ratcliff, R., Thapar, A., & McKoon, G. (2010). Individual differences, aging, and IQ in two-choice tasks. *Cognitive Psychology*, 60(3), 127–157. <https://doi.org/10.1016/j.cogpsych.2009.09.001>
- R Core Team (2021). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>
- Roberts, I. D., & Hutcherson, C. A. (2019). Affect and decision making: Insights and predictions from computational models. *Trends in Cognitive Sciences*, 23(7), 602–614. <https://doi.org/10.1016/j.tics.2019.04.005>
- Robinaugh, D. J., Haslbeck, J. M. B., Ryan, O., Fried, E. I., & Waldorp, L. J. (2021). Invisible hands and fine calipers: A call to use formal theory as a toolkit for theory construction. *Perspectives on Psychological Science: A Journal of the Association for Psychological Science*, 16(4), 725–743. <https://doi.org/10.1177/1745691620974697>
- Rodebaugh, T. L., Scullin, R. B., Langer, J. K., Dixon, D. J., Huppert, J. D., Bernstein, A., Zvielli, A., & Lenze, E. J. (2016). Unreliability as a threat to understanding psychopathology: The cautionary tale of attentional bias. *Journal of Abnormal Psychology*, 125(6), 840–851. <https://doi.org/10.1037/abn0000184>
- Ruscio, A. M. (2019). Normal versus pathological mood: Implications for diagnosis. *Annual Review of Clinical Psychology*, 15, 179–205. <https://doi.org/10.1146/annurev-clinpsy-050718-095644>
- Rush, A. J., Gullion, C. M., Basco, M. R., Jarrett, R. B., & Trivedi, M. H. (1996). The Inventory of Depressive Symptomatology (IDS): Psychometric properties. *Psychological Medicine*, 26(3), 477–486. <https://doi.org/10.1017/s0033291700035558>
- Shadlen, M. N., & Shohamy, D. (2016). Decision making and sequential sampling from memory. *Neuron*, 90(5), 927–939. <https://doi.org/10.1016/j.neuron.2016.04.036>
- Shahar, N., Hauser, T. U., Moutoussis, M., Moran, R., Keramati, M., NSPN consortium, & Dolan, R. J. (2019). Improving the reliability of model-based decision-making estimates in the two-stage decision task with reaction-times and drift-diffusion modeling. *PLoS Computational Biology*, 15(2), e1006803. <https://doi.org/10.1371/journal.pcbi.1006803>
- Shany, O., Gurevitch, G., Gilam, G., Dunskey, N., Reznik Balter, S., Greental, A., Nutkevitch, N., Eldar, E., & Hendler, T. (2022). A corticostriatal pathway mediating self-efficacy enhancement. *Npj Mental Health Research*, 1(1), 1–15. <https://doi.org/10.1038/s44184-022-00006-7>
- Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: uses in assessing rater reliability. *Psychological Bulletin*, 86(2), 420–428. <https://doi.org/10.1037/0033-2909.86.2.420>
- Siegle, G. J., Carter, C. S., & Thase, M. E. (2006). Use of fMRI to predict recovery from unipolar depression with cognitive

- behavior therapy. *The American Journal of Psychiatry*, 163(4), 735–738. <https://doi.org/10.1176/appi.ajp.163.4.735>
- Smith, P. L., & Ratcliff, R. (2004). Psychology and neurobiology of simple decisions. *Trends in Neurosciences*, 27(3), 161–168. <https://doi.org/10.1016/j.tins.2004.01.006>
- Spiegelhalter, D. J., Best, N. G., Carlin, B. P., & van der Linde, A. (2014). The deviance information criterion: 12 years on. *Journal of the Royal Statistical Society. Series B, Statistical Methodology*, 76(3), 485–493. <https://doi.org/10.1111/rssb.12062>
- Stine, G. M., Zylberberg, A., Ditterich, J., & Shadlen, M. N. (2020). Differentiating between integration and non-integration strategies in perceptual decision making. *eLife*, 9. <https://doi.org/10.7554/eLife.55365>
- Thompson, J. S., Jamal-Orozco, N., & Hallion, L. S. (2022). Dissociable associations of facets of mindfulness with worry, rumination, and transdiagnostic perseverative thought. *Mindfulness*, 13(1), 80–91. <https://doi.org/10.1007/s12671-021-01747-w>
- Turner, B. M., Forstmann, B. U., Love, B. C., Palmeri, T. J., & Van Maanen, L. (2017). Approaches to analysis in model-based cognitive neuroscience. *Journal of Mathematical Psychology*, 76(B), 65–79. <https://doi.org/10.1016/j.jmp.2016.01.001>
- Vago, D. R., & Silbersweig, D. A. (2012). Self-awareness, self-regulation, and self-transcendence (S-ART): a framework for understanding the neurobiological mechanisms of mindfulness. *Frontiers in Human Neuroscience*, 6, 296. <https://doi.org/10.3389/fnhum.2012.00296>
- Van Dam, N. T., Earleywine, M., & Danoff-Burg, S. (2009). Differential item function across meditators and non-meditators on the five facet mindfulness questionnaire. *Personality and Individual Differences*, 47(5), 516–521. <https://doi.org/10.1016/j.paid.2009.05.005>
- Van Dam, N. T., Hobkirk, A. L., Danoff-Burg, S., & Earleywine, M. (2012). Mind your words: Positive and negative items create method effects on the five facet mindfulness questionnaire. *Assessment*, 19(2), 198–204. <https://doi.org/10.1177/1073191112438743>
- Van Dam, N. T., Bilgrami, Z. R., & Eisenlohr-Moul, T. (2018). Development and validation of the Balanced Inventory of Mindfulness-Related Skills (BIMS). <https://mindrxiv.org/8qdzg/download?format=pdf>
- Vanderlind, W. M., Millgram, Y., Baskin-Sommers, A. R., Clark, M. S., & Joermann, J. (2020). Understanding positive emotion deficits in depression: From emotion preferences to emotion regulation. *Clinical Psychology Review*, 76, 101826. <https://doi.org/10.1016/j.cpr.2020.101826>
- von Krause, M., Radev, S. T., & Voss, A. (2022). Mental speed is high until age 60 as revealed by analysis of over a million participants. *Nature Human Behaviour*, 6(5), 700–708. <https://doi.org/10.1038/s41562-021-01282-7>
- Watson, D., & Clark, L. A. (1994). The PANAS-X: Manual for the positive and negative affect schedule - expanded form [Data set]. In *University of Iowa*. The University of Iowa. <https://doi.org/10.17077/48vt-m4t2>
- White, C., Ratcliff, R., Vasey, M., & McKoon, G. (2009). Dysphoria and memory for emotional material: A diffusion-model analysis. *Cognition & Emotion*, 23(1), 181–205. <https://doi.org/10.1080/02699930801976770>
- Whitmer, A. J., & Gotlib, I. H. (2013). An attentional scope model of rumination. *Psychological Bulletin*, 139(5), 1036–1061. <https://doi.org/10.1037/a0030923>
- Wiecki, T. V., Sofer, L., & Frank, M. J. (2013). HDDM: Hierarchical Bayesian estimation of the drift-diffusion model in python. *Frontiers in Neuroinformatics*, 7, 14. <https://doi.org/10.3389/fninf.2013.00014>
- Wiecki, T. V., Poland, J., & Frank, M. J. (2015). Model-based cognitive neuroscience approaches to computational psychiatry: Clustering and classification. *Clinical Psychological Science*, 3(3), 378–399. <https://doi.org/10.1177/2167702614565359>
- Wiecki, T. V., Antoniadis, C. A., Stevenson, A., Kennard, C., Borowsky, B., Owen, G., Leavitt, B., Roos, R., Durr, A., Tabrizi, S. J., & Frank, M. J. (2016). A computational cognitive biomarker for early-stage Huntington's Disease. *PLoS One*, 11(2), e0148409. <https://doi.org/10.1371/journal.pone.0148409>
- Wilson, R. C., & Collins, A. G. (2019). Ten simple rules for the computational modeling of behavioral data. *eLife*, 8. <https://doi.org/10.7554/eLife.49547>
- Wisco, B. E. (2009). Depressive cognition: self-reference and depth of processing. *Clinical Psychology Review*, 29(4), 382–392. <https://doi.org/10.1016/j.cpr.2009.03.003>
- Yip, S. W., Barch, D. M., Chase, H. W., Flagel, S., Huys, Q. J. M., Konova, A. B., Montague, R., & Paulus, M. (2022). From computation to clinic. *Biological Psychiatry Global Open Science*. <https://doi.org/10.1016/j.bpsgos.2022.03.011>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.