

Hierarchical Drift Diffusion Modeling of Self-Referential Processing in Social Anxiety and Body
Dysmorphic Disorder

Yuchen Zhao

A thesis
submitted in partial fulfillment of the
requirements for the degree of

Master of Science

University of Washington
2025

Committee:

Angela Fang

Lori Zoellner

Jessica Chen

Program Authorized to Offer Degree:

Department of Psychology

©Copyright 2025

Yuchen Zhao

University of Washington

Abstract

Hierarchical Drift Diffusion Modeling of Self-Referential Processing in Social Anxiety and Body
Dysmorphic Disorder

Yuchen Zhao

Chair of the Supervisory Committee:

Angela Fang

Department of Psychology

Body dysmorphic disorder (BDD) and social anxiety disorder (SAD) share hallmark features of heightened self-consciousness and maladaptive self-referential processing (SRP), yet the cognitive mechanisms underlying SRP in these disorders remain poorly understood. This study employed hierarchical drift diffusion modeling (HDDM) to examine SRP during a self-referential encoding task (SRET) among individuals with BDD, SAD, and healthy controls (HC). Drift rate, a model parameter indexing the efficiency of evidence accumulation, served as a proxy for SRP, with lower drift rates indicating greater decisional inefficiency in processing self-relevant information. A total of 58 participants (15 BDD, 15 SAD, 28 HC) judged positive and negative trait adjectives in self- and other-referential contexts. Both clinical groups endorsed fewer positive and more negative self-descriptors than HC, with no group differences in other-referential trials. Critically, both BDD and SAD showed reduced drift rates relative to HC in both self- and other-referential trials regardless of valence, although deficits were substantially larger for self-referential decisions. Within-group analyses showed marked self–other drift rate reductions in clinical groups. Across patients, disorder-specific symptom severity predicted lower drift rates in self-referential trials even after controlling for depression. No drift rate

differences emerged between BDD and SAD groups, suggesting a shared cognitive process.

Together, these findings integrate computational modeling with experimental task data to reveal that BDD and SAD involve a generalized inefficiency in evaluating social information, most pronounced for self-relevant content, highlighting drift rate as a potential transdiagnostic marker of cognitive impairment.

Introduction

Body dysmorphic disorder (BDD) and social anxiety disorder (SAD) are two closely related mental health conditions that share significant commonalities, particularly in their core features of heightened self-consciousness and fear of negative evaluation (Buhlmann et al., 2015; Dietel et al., 2021; Fang & Hofmann, 2010). BDD is characterized by preoccupation with perceived flaws in physical appearance, often leading to significant emotional distress and functional impairment. SAD is marked by an intense fear of negative evaluation or embarrassment in social situations, leading individuals to avoid these interactions or endure them with substantial distress (Button et al., 2012). Both disorders involve distorted self-perceptions -- BDD focusing on perceived physical flaws, while SAD centers on fears of social incompetence. These concerns often lead to avoidance behaviors such as social withdrawal.

A critical transdiagnostic factor contributing to the persistence of these disorders is self-referential processing (SRP), which involves how individuals interpret and respond to self-related information, including thoughts, emotions, and social feedback (Boehme et al., 2015; Ingram, 1990). SRP can be conceptualized on a continuum, with maladaptive SRP referring to the tendency to interpret and respond to self-related information through a rigid, often negative lens of the self, such as interpreting neutral social feedback as critical, or viewing fleeting thoughts as indicative of core personal flaws. In both BDD and SAD, maladaptive SRP may contribute to negative self-evaluations and an impaired ability to integrate corrective external information, which may reinforce symptoms in BDD and SAD (Iannazzi et al., 2025; Spurr & Stopa, 2002). Evidence further suggests that appearance concerns in BDD extend to global negative self-evaluation and low self-esteem (Buhlmann et al., 2009; Johnson et al., 2018; Kuck et al., 2021), with deficits in emotion recognition being more pronounced in self-referent contexts (Buhlmann et al., 2006). Similarly, Clark and Wells' (1995) model of SAD highlights the role of self-focused attention in maintaining social anxiety symptoms, where individuals

construct a distorted self-perception based on internal cues (e.g., anxiety), assuming these are visible and judged by others, which is a mechanism supported by substantial empirical evidence (Coles et al., 2001; Gaydukevych & Kocovski, 2012; Judah et al., 2016; Norton & Abbott, 2016). Additionally, shared neurobiological mechanisms suggest a role for SRP in maintaining both disorders (Iannazzi et al., 2025), with functional imaging studies indicating dysregulation in brain regions such as the mPFC and insula, which have been associated with SRP (Finlayson-Short et al., 2020; Jamieson et al., 2023; Stein, 2015; Yoon et al., 2019). Moreover, overly rigid SRP may increase the risk of treatment non-response and relapse in depression (LeMoult et al., 2017; Nolen-Hoeksema et al., 2008). Neural evidence suggests that CBT may improve outcomes by altering maladaptive SRP-related activity and connectivity in mPFC–vACC networks (Yoshimura et al., 2014, 2017).

Although theoretical and empirical evidence highlights maladaptive SRP as a core feature of both anxiety and depression, little research has investigated the mechanisms underlying SRP using computational approaches. One commonly used paradigm to study SRP is the Self-Referential Encoding Task (SRET), in which participants judge whether positive (e.g., “honest”) or negative (e.g., “rude”) trait adjectives describe themselves (self-trials) or another person (other-trials; Dainer-Best et al., 2018b; Derry & Kuiper, 1981). People with anxiety and depression are more likely to describe themselves using negative rather than positive words (for a review on depression, see Collins & Winer, 2022; for anxiety, see Dixon et al., 2022; Goldin et al., 2009, Tracy et al., 2021), and they take longer to make negative self-judgments compared to positive ones (Hitchcock et al., 2023). Traditional SRET analyses often focus on reaction times (RTs), typically summarizing performance across conditions by averaging RTs. However, this approach has several limitations. First, behavioral summaries derived from RT averages are frequently inconsistent, which undermines the common assumption that they represent stable, trait-like characteristics (Brown et al., 2020; Hitchcock et al., 2022; McNally, 2019). Second,

averaging RTs obscures individual variability in trial-by-trial responses, ignoring the dynamic nature of decision-making processes (Luu et al., 2000; Noorani, 2014). Third, mean RTs reflect an amalgamation of psychological processes, including both decisional and non-decisional components, such as attentional engagement and motor preparation. As a result, they may lack reliability and fail to isolate specific cognitive mechanisms (Grennan et al., 2023; Haines et al., 2020; Price et al., 2019).

To address these limitations, drift diffusion models (DDMs) offer a powerful alternative by estimating latent parameters that characterize the decision-making process on a trial-by-trial basis (Myers et al., 2022; Ratcliff, 1978). This modeling approach allows for the separation of evidence accumulation from non-decisional components, enabling a more precise and stable analysis of decision-making processes. Importantly, hierarchical drift diffusion models (HDDMs) are particularly well-suited for SRET data, as they can capture individual differences in certain cognitive processes involved in forced choice decisions, which enables inferences about the computational mechanisms associated with SRP (Allison et al., 2021; Beevers et al., 2019; Cataldo et al., 2023; Dainer-Best et al., 2018; Hitchcock et al., 2022; Price et al., 2021). In particular, drift rate is a key parameter in DDMs that reflects the latent rate at which evidence accumulates toward making a decision (Ratcliff & McKoon, 2008; Ratcliff et al., 2016). Because DDM can model trial-by-trial variability, reduce the influence of outlier trials, and disentangle lower-level processes, such as motor and visual components, drift rate may serve as a more rigorous measure of information processing speed than mean reaction time (Voss et al., 2013). Higher drift rates indicate faster and more consistent responses (i.e., evidence accumulates more quickly toward a decision), whereas values closer to zero reflect slower and less consistent responding (Lerche et al., 2017). An increasing number of studies have applied DDM to self-referential tasks, with findings showing that drift rate predicts symptom severity in conditions such as depression and anxiety (e.g., Castagna et al., 2022; Disner et al., 2017;

Price et al., 2021). This computational approach provides a robust framework for disentangling the cognitive and non-cognitive components underlying SRP, paving the way for a more nuanced understanding of maladaptive cognitive processes in these disorders.

In this study, we applied HDDM on SRET to examine SRP as a transdiagnostic mechanism in individuals with BDD and SAD. We chose to examine SRP in SAD and BDD based on theoretical and empirical work highlighting substantial similarities between the two disorders in their underlying psychopathology and neurobiology (Clark & Wells, 1995; Fang & Hofmann, 2010; Hofmann, 2000; Iannazzi et al. 2025; Veale, 2004). HDDM is particularly well-suited to address questions about shared cognitive mechanisms underlying multiple disorders, as it exploits heterogeneity by modeling trial-wise variability. To date, no studies have directly modeled trial-level SRP dynamics in transdiagnostic samples using HDDM, leaving open key questions about whether and how altered evidence accumulation contributes to negative self-bias in these disorders. Moreover, most SRET-based SRP studies compare self-referential trials only with neutral, non-evaluative control trials, without including other-referential conditions. This omission limits the ability to determine whether deficits are specific to SRP or instead reflect a broader inefficiency in evaluation-related information processing about anyone. The present study addresses these gaps by applying HDDM to SRET data incorporating both self- and other-referential trials in individuals with BDD and SAD, with a focus on drift rate as a computational index of aberrant SRP. In the context of the SRET, drift rate reflects the speed and efficiency with which individuals accumulate evidence toward a descriptive judgment, which may be biased or slowed when maladaptive SRP interferes with processing positive or negative self-relevant information. We hypothesized that participants with BDD and SAD would show reduced drift rates in self-referential trials compared to other-referential trials, regardless of word valence. Additionally, these clinical groups were expected to exhibit lower drift rates in self-referential trials for both positive and negative words compared to healthy controls. Finally, we

hypothesized that within the BDD and SAD groups, drift rates would be negatively associated with symptom severity, reflecting a shared transdiagnostic mechanism of altered SRP.

Materials and Methods

Participants

Participants included 30 individuals diagnosed with either primary Social Anxiety Disorder (SAD; $n = 15$) or primary Body Dysmorphic Disorder (BDD; $n = 15$), as well as 28 healthy controls. All participants were recruited through online and print advertisements, clinician referrals, and self-referrals to specialty clinics within a large academic medical center, as part of a larger treatment study examining the brain correlates of SRP (Fang et al., 2022; Fang et al., 2024). Clinical participants met criteria for primary SAD or BDD (LSAS total score ≥ 50 for primary SAD and BDD-YBOCS total score ≥ 20 for primary BDD) and elevated levels of trait self-focused attention based on scoring ≥ 18 on the Public Self-Consciousness Scale Revised (Scheier & Carver, 1985). Healthy controls had no lifetime history of psychiatric disorders and endorsed low levels of trait self-focused attention (≤ 9 on the Public Self-Consciousness Scale Revised). Additional inclusion and exclusion criteria for all groups were associated with the parent study, as follows: ages 18-45, right handed (as assessed by the Edinburgh Handedness Inventory), MRI contraindications, no history of neurological illness, neurosurgery, or significant head trauma, current or past manic or psychotic episodes, no alcohol or substance dependence in the past year, no current cognitive behavioral therapy (CBT) or formal mindfulness/meditation training, no history of ≥ 10 sessions of CBT or formal mindfulness training, and no current use of psychotropic medications (except for clinical participants on a stable regimen for at least two months prior to the first study visit). Patients and controls were matched on demographic variables including age, sex, race, ethnicity, and education. All procedures were approved by the XXXXXXXX (IRB Protocol No: XXXXXXXX), and written informed consent was obtained from all participants prior to starting study procedures.

Measures and Task

Structured Clinical Interview for DSM-5, Research Version (SCID-5-RV). The SCID-5-RV (First et al., 2015) was used during the screening visit to assess primary psychiatric disorders based on DSM-5 criteria. This semi-structured interview was conducted by a licensed psychologist (A.F.).

Liebowitz Social Anxiety Scale (LSAS). The LSAS (Liebowitz, 2013) is a clinician-administered measure designed to assess the severity of social anxiety symptoms over the past week for participants with a primary diagnosis of SAD. It includes two subscales, fear and avoidance, in various social situations, and scores from both subscales are combined to yield an overall severity score. The LSAS showed strong internal consistency in this sample ($\alpha = 0.88$).

Yale-Brown Obsessive-Compulsive Scale Modified for Body Dysmorphic Disorder (BDD-YBOCS). The BDD-YBOCS (Phillips et al., 1997) is a clinician-rated measure that evaluates BDD symptom severity over the past week, focusing on the frequency and functional impact of BDD-related obsessions and compulsions. The scale showed good to excellent internal consistency in this sample ($\alpha = 0.86$).

Beck Depression Inventory- II (BDI-II). The BDI-II (Beck et al., 1996) is a widely used self-report questionnaire that assesses the severity of depressive symptoms experienced over the past two weeks. It includes items measuring cognitive, emotional, and physical aspects of depression, such as sadness, guilt, and fatigue. The scale has demonstrated excellent internal consistency in clinical and non-clinical populations ($\alpha = 0.91$).

Self-Referent Encoding Task (SRET). The SRET (Iannazzi et al., 2025; Kelley et al., 2006) is a two-alternative forced-choice task used to assess SRP (Figure 1). Participants viewed

adjectives on the screen and responded to one of two trial types: *Self* (“Does this adjective describe you?”) or *Other* (“Does this adjective describe Former U.S. President Barack Obama?”), indicating their answers via button press. Obama was selected as the referential “Other” to maintain consistency with prior literature and our own previous work, as presidents and well-known public figures are commonly used in SRET paradigms (Murray et al., 2012). Each trial lasted 2500 ms and displayed a cue word (“SELF” or “OBAMA”) above a central fixation cross as well as a unique adjective (e.g., “POLITE”) below it. The cue and adjective appeared simultaneously for 2000 ms, while the fixation remained onscreen throughout the trial. A total of 270 adjectives were drawn from a pool of normed personality- and appearance-related traits (Anderson, 1968), balanced across valence ($\frac{1}{3}$ positive, $\frac{1}{3}$ negative, $\frac{1}{3}$ neutral), word length, and syllable count. The task included three functional runs, each consisting of 90 trials: 45 *Self*, 45 *Other*. Within each run, 15 *Self* trials, 15 *Other* trials, and 15 fixation-only trials were pseudo-randomly intermixed to ensure each trial type followed every other type equally often. Additional jittered fixation trials (2500–12500 ms) were interspersed to enable estimation of distinct hemodynamic responses per condition. Participants completed the SRET during fMRI. Due to our modest sample size and model complexity, neuroimaging data were not included as a third predictor to avoid overfitting and convergence issues (see also Discussion).

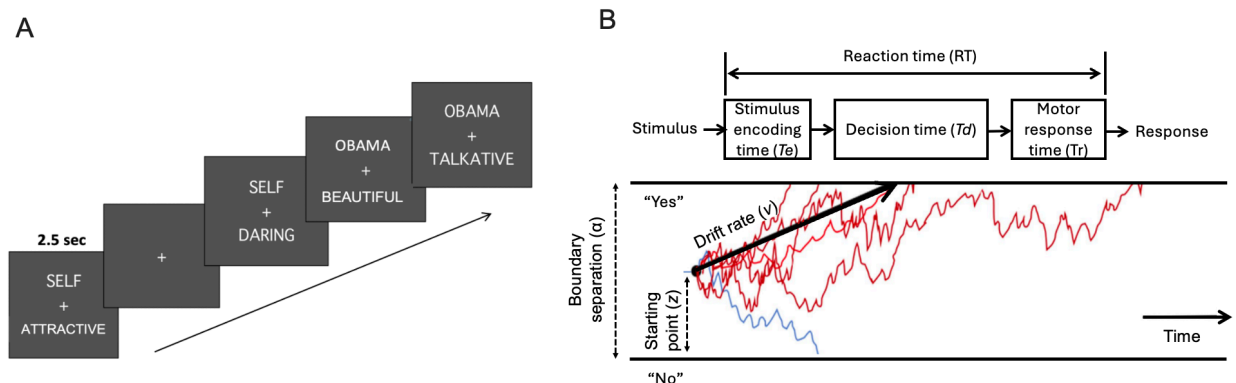


FIGURE 1. (A) The SRET (Self-Referent Encoding Task) design. Participants judge whether the given words are descriptive of themselves or of another person (Former U.S. President Barack

Obama) during the encoding phase. (B) Schematic of the DDM decision-making process for several trials (Myers et al., 2022). Two horizontal lines indicate two responses (“Yes” or “No”). The red and blue lines signify the evidence accumulation process as a noisy process; some trials might reach the same responses at different times and may also reach the opposite boundary, resulting in the opposite answer, depicted here as a blue line. The drift rate (ν) represents the average slope of the evidence accumulation process across numerous trials. A lower drift rate reflects a compromised speed and efficiency with which individuals accumulate evidence toward a judgment, which may indicate that maladaptive SRP interferes with processing positive or negative self- or other-relevant information.

Data Analysis

Models were run with the Hierarchical Drift Diffusion Modeling (HDDM) Python toolbox (Wiecki et al., 2013) with stimulus coding (Ratcliff & McKoon, 2008). All participants ($N = 58$) were analyzed in a single hierarchical within-subject model using effect coding for both valence (negative = -1, positive = 1) and trial type (self-referential = -1; other-referential = 1) to estimate the effects of valence and trial type on drift rate (ν), formalized as: Drift rate $\sim$ valence $\times$ trial type. Outlier trials with reaction times < 200 ms were excluded (0% of trials).

Our approach involved three key steps: (1) Comparing the fit of nested HDDMs of increasing complexity by incorporating two within-subject regressors (i.e., word valence, self- or other-referential trials), their interactions, and additional model parameters (e.g., starting point bias, non-decision time variability); (2) Validating the best-fitting model through model validation, parameter recovery, and quality checks; and (3) Examining the associations between model parameters (such as drift rate), diagnostic groups, and symptom severity, controlling for depressive symptoms given their known influence on SRP.

For each model, 2,000 posterior samples were generated, with the first 1,000 discarded as burn-in. Convergence was assessed via the Gelman-Rubin statistic ($\hat{R} < 1.1$; Gelman & Rubin, 1992) by comparing within-chain and between-chain variance of different runs of the same model, and visual inspection of trace and autocorrelation plots. Model fit was compared using the deviance information criterion (DIC; Spiegelhalter et al., 2002), with lower scores indicating better fit. Predictive checks confirmed parameter recoverability (range: 0.86–0.99, median: 0.92; Table 2).

While group-level models can test between-subject differences, they often lack power to detect within-subject condition effects (e.g., valence/trial type on drift rates) due to overlapping posterior variances across conditions (Wiecki et al., 2013). By pooling all participants into a single hierarchical model, we improved parameter estimation for trial-level effects while retaining the ability to test group and symptom effects post hoc. HDDM does not support simultaneous modeling of within-subject regressors (e.g., trial-level valence/trial type) and between-subject continuous covariates (e.g., symptom severity). To address this, we adopted a two-step approach. First, we estimated condition-specific drift rates (e.g., self-positive, other-negative) for each participant based on the parameter coefficients from the best-fitting model; Second, we extracted maximum a posteriori (MAP) estimates of condition-specific drift rates (e.g., self-negative, other-positive) for each participant from the best-fitting HDDM model and conducted mixed-effects and linear regression analyses to test our hypotheses about group differences and model parameter associations with symptom severity. This approach is widely used in computational psychiatry (e.g., Frank et al., 2015) to disentangle trial-level cognitive processes from individual differences in clinical symptoms. These estimates were analyzed in R (version 4.5.0; R Core Team, 2024) to address two questions: (1) Group differences: a mixed-effect model was conducted to compare drift rates across diagnostic groups (HC, SAD, BDD) and SRET conditions (Self-Positive, Self-Negative, Other-Positive, Other-Negative); (2) Symptom

Relationships: Separate linear regressions examined associations between drift rates and symptom severity (e.g., LSAS & BDD-YBOCS scores) in the clinical group, controlling for depression severity (BDI-II scores). Symptom severity was standardized prior to analysis by dividing each observed score by its maximum possible value to enable cross-group linear regression (i.e., LSAS scores were divided by 144, BDD-YBOCS scores were divided by 48, and BDI-II scores were divided by 63).

Results

Demographic and clinical characteristics

Table 1. Sample Characteristics. BDI-II = Beck Depression Inventory-II. LSAS = Liebowitz Social Anxiety Scale. BDD-YBOCS = Yale-Brown Obsessive-Compulsive Scale Modified for BDD. BABS = Brown Assessment of Beliefs Scale. ^at-statistics reflect differences between social anxiety disorder (SAD) and body dysmorphic disorder (BDD) groups.

Demographic Characteristics	Clinical (n=30) Mean ± SD / n (%)	Control (n=28) Mean ± SD / n (%)	t / χ^2	p
Age	25.4 ± 3.8	25.7 ± 3.9	0.28	0.78
Sex			0.20	0.65
Female	23 (76.7)	20 (71.4)		
Male	7 (23.3)	8 (28.6)		
Race			1.65	0.65
Caucasian	19 (63.3)	18 (64.3)		
Asian	8 (26.7)	5 (17.9)		
Black/African American	2 (6.7)	2 (7.1)		
More than one race	1 (3.3)	3 (10.7)		
Ethnicity			0.09	0.76
Non-Hispanic or Latino	26 (86.7)	25 (89.3)		

Hispanic or Latino	4 (13.3)	3 (10.7)			
Unspecified	0 (0.0)	0 (0.0)			
Highest Level of Education			2.96	0.40	
Some College	3 (10.0)	0 (0.0)			
College Graduate	17 (56.7)	18 (64.3)			
Some Postgraduate	1 (3.3)	1 (3.6)			
Postgraduate/Professional Degree	9 (30.0)	9 (32.1)			
Clinical Characteristics	BDD (n=15)	SAD (n=15)			
	Mean ± SD	Mean ± SD		t^a	p
Age of onset (years)	23.9 ± 2.1	27 ± 4.4	---	2.48	0.02
BDI-II	15.5 ± 12.5	17.0 ± 14.7	0.4 (0.7)	0.29	0.77
LSAS	---	80.3 ± 16.0	---	---	---
BDD-YBOCS	30.2 ± 6.6	---	---	---	---

As shown in Table 1, patients and healthy controls did not differ in age, sex, race, ethnicity, and education. Similarly, patients with SAD and those with BDD were comparable across demographic and clinical characteristics, with the exception of age: individuals with SAD were significantly older than those with BDD (SAD: $M = 27.00$, $SD = 4.41$; BDD: $M = 23.87$, $SD = 2.13$; $t = 2.48$, $p < .05$). Seven patients were taking psychotropic medication at a stable dose for at least two months before study participation, including amphetamine (2 SAD, 1 BDD), escitalopram (1 SAD), and citalopram (3 BDD).

Descriptive statistics

Since SAD and BDD groups did not differ from each other for both endorsement and reaction time in any condition (p 's > 0.05), subsequent group comparisons were conducted between the control and the combined clinical group. For endorsement, participants endorsed positive words much more frequently than negative words across all trial types. A linear mixed-effects model with random intercepts for subjects examined the effects of diagnosis (Clinical vs HC), valence (Pos vs Neg), and trial type (self vs other) on endorsement ratio. There was a significant main

effect of valence, with higher endorsement ratio for positive versus negative trials, $\beta = 0.904$, $SE = 0.036$, $t(168) = 25.29$, $p < .001$. The diagnosis $\times$ trial type interaction was significant, $\beta = 0.231$, $SE = 0.050$, $t(168) = 4.65$, $p < .001$, indicating that the effect of diagnosis differed between Self and Other trials. So was the three-way diagnosis $\times$ valence $\times$ trial type interaction, $\beta = -0.482$, $SE = 0.070$, $t(168) = -6.86$, $p < .001$, indicating that diagnosis effects on Self versus Other trials depended on valence. Other main effects and lower-order interactions were not significant. The largest group differences occurred for negative-self trials.

For reaction time, participants generally responded more slowly to negative words than positive words and faster in other-referential trials than in self-referential trials (mixed-effect model: Valence Beta (SE) = -0.08 (0.01), $p < .001$; Trial Type Beta (SE) = 0.10 (0.01), $p < .001$; Valence $\times$ Trial Type Beta (SE) = -0.05 (0.02), $p < 0.01$). When diagnosis group was included, there was a significant main effect of diagnosis, with patients responding slower than controls, $\beta = 0.102$, $SE = 0.050$, $t(73.55) = 2.05$, $p < .05$. Positive-valence trials were faster than negative-valence trials, $\beta = -0.076$, $SE = 0.021$, $t(168) = -3.59$, $p < .001$, and self-referential trials were slower than other-referential trials, $\beta = 0.059$, $SE = 0.021$, $t(168) = 2.81$, $p < .01$. The diagnosis $\times$ trial type interaction was significant, $\beta = 0.083$, $SE = 0.029$, $t(168) = 2.84$, $p < .01$, indicating that patients were disproportionately slower on self-referential trials. Additionally, the valence $\times$ trial type interaction was significant, $\beta = -0.074$, $SE = 0.030$, $t(168) = -2.50$, $p < .05$, reflecting faster responses on positive-self trials relative to negative-self trials. No other interactions were significant, including the three-way interaction ($p > .05$). Post-hoc pairwise comparisons further clarified that for other-referential trials, patients were slower than controls only in negative trials. (for negative trials: $\beta = -0.102$, $SE = 0.050$, $p < 0.05$; for positive trials: $\beta = -0.096$, $SE = 0.050$, $p > 0.05$). These results indicate that group differences in reaction time were largest on self-referential trials, particularly for positive words.

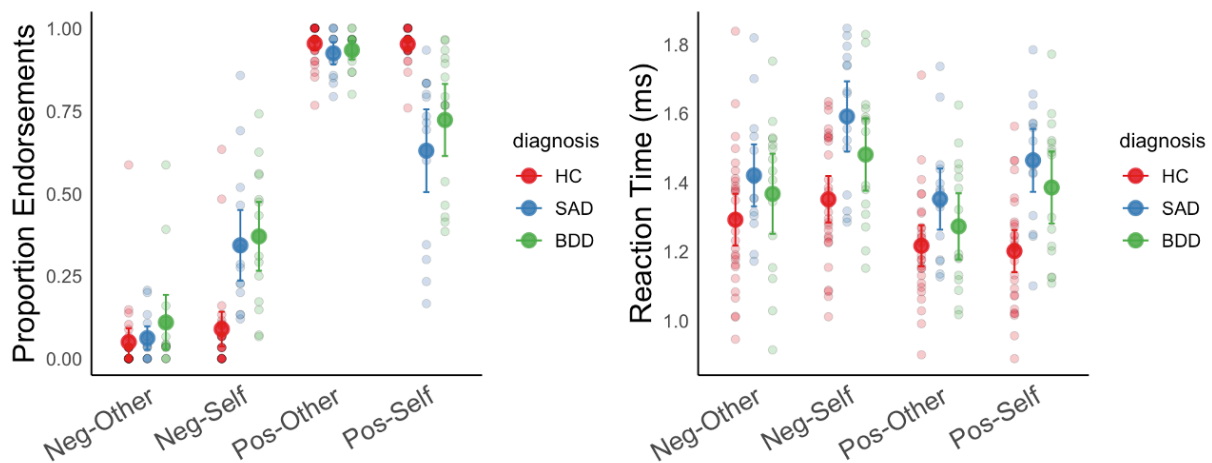


FIGURE 2. Behavioral endorsement (left) and reaction times (right) by trial conditions for positive (blue) and negative (red) words, self- and other- trial conditions (x-axis). Small light points show subject means and darker points show group means; error bars show 95% confidence intervals.

Model fitting and model validation

Five nested models were compared using DIC. As seen in Table 1, HDDM fitting captured these behavioral effects. Compared to a baseline model without regressors, model fit significantly improved after including a regressor for Valence on the drift rate. The fit improved further with the addition of trial type and their interactions with valence. Incorporating a starting-point bias term, which reflects a preference for responding “yes” or “no” regardless of the adjective, also enhanced the model fit. However, adding additional across-trial variability parameters (sv, sz, and st) only slightly increased the model’s DIC. Although the full model had a slightly lower DIC compared to simpler models, we selected Model 3 (adding bias only) as the best-fitting model based on the principle of parsimony.

Table 2 Deviance Information Criterion (DIC) statistics of models examined

Model	DIC	Δ DIC
Null Model	13416	0
+ valence	8131	-5285
+ valence*trial type	7616	-5800
+ z	7538	-5878
+ sv	7544	-5872
+ sz	7537	-5879
+ st	7543	-5873

*Note: DIC values for each model tested, with lower DIC values indicating better model fit. The Δ DIC column represents the difference in DIC values relative to the null model, with negative values representing improvements in fit. Models include the Null Model, models incorporating Valence, the interaction between Valence and Trial Type, and models adding parameters such as bias (z), starting point variability (sv), starting bias variability (sz), and non-decision time variability (st). The best-fitting model, indicated by the lowest DIC, includes both regressors (valence, valence*trial) and starting bias parameters (+z).*

Inferential Statistics

We estimated drift rates for four trial conditions (Positive-Self, Positive-Other, Negative-Self, Negative-Other) based on the MAP estimates of model-based coefficients to assess processing strength independent of response direction (see Figure 4). Negative drift rates indicate higher 'no' response tendencies for negative stimuli. A linear mixed model tested group $\times$ condition

effects on drift rates, with follow-up Bonferroni-corrected comparisons. A Type III mixed ANOVA with Satterthwaite's method was conducted to examine the effects of diagnostic group (HC vs. Clinical) and condition (Neg-Other, Neg-Self, Pos-Other, Pos-Self) on drift rate. There was a significant main effect of diagnostic group, $F(1, 56) = 54.93, p < .001$, a significant main effect of condition, $F(3, 168) = 64.36, p < .001$, and a significant diagnostic group $\times$ condition interaction, $F(3, 168) = 24.25, p < .001$, indicating that both overall drift rates and their modulation by condition differed across groups. Post-hoc group comparisons with Bonferroni correction demonstrated HC had consistently larger drift rates across all trial conditions compared to the clinical group, with the more pronounced HC-Clinical group differences emerging in the self-referential conditions (Neg-Other: difference (SE) = 0.65 (0.15), $t(85.56) = 4.34$; Neg-Self: difference (SE) = 1.31(0.15), $t(85.56) = 8.72$; Pos-Other: difference (SE) = 0.68(0.15), $t(85.56) = 4.53$; Pos-Self: difference (SE) = 1.35(0.15), $t(85.56) = 8.94$; all p 's $< .001$). Drift rate in both SAD and BDD was lower than HC by more than 1.3 units in both positive and negative valence conditions. Although HC also outperformed SAD and BDD in other-referential conditions, the differences were smaller, ranging from 0.58 to 0.73 units (all p 's $< .01$). Again, SAD and BDD groups did not differ from each other in any condition (all p 's $> .05$), suggesting transdiagnostic impairments in self-other referential processing.

Consistent with our hypothesis, within-group comparisons further revealed that the clinical group demonstrated substantially lower drift rates for self-referential compared to other-referential trials, with differences of 0.94 (SE=0.08) units in the negative condition and 0.78 (SE=0.08) units in the positive condition (all p 's $< .0001$). In contrast, HC showed no significant difference in the positive condition and a smaller self-other difference in the negative condition, with a 0.28-unit (SE=0.08) drop for negative-self compared to negative-other trials ($p < .01$). These results highlighted significantly compromised SRP in SAD and BDD compared to HC, and supported our hypothesis regarding group differences in drift rate in self versus other trials.

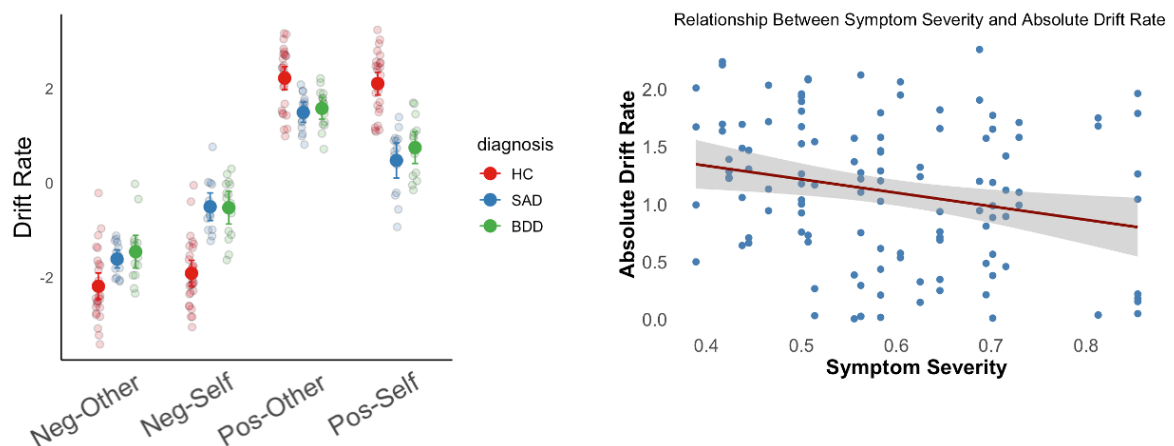


Figure 3. left: MAP estimates of drift rates across diagnosis groups and trial conditions. Drift rates were maximum a posteriori (MAP) estimates of the subject-level traces. Small light points show subject means and darker points show group means; error bars show 95% confidence intervals. Right: Correlation between absolute values of drift rates and symptom severity. Absolute values were used because negative drift rates indicate a tendency to choose “no”, whereas the magnitude (absolute value) reflects the speed of evidence accumulation.

Pearson correlations were performed to examine the associations between absolute drift rates and clinical measures, including depression severity (indexed by mean-centered BDI-II scores) and symptom severity specific to BDD/SAD (indexed by proportional mean-centered LSAS and BDD-YBOCS scores). Significant negative correlations were observed between drift rates and both depression and symptom severity, predominantly in the self-referential conditions.

Specifically, depression severity was only marginally correlated with drift rates in the Positive-Self ($r = -0.34$, $p > .05$) but not in other conditions. Instead, symptom severity was negatively associated with drift rates in the Positive-Self ($r = -.52$, $p < .001$) and Negative-Self ($r = -.48$, $p < .05$) conditions. In contrast, drift rates in other-referential conditions did not show significant correlations with either depression (Positive-Other: $r = -0.30$, $p > .05$; Negative-Other: $r = -0.07$, $p > .05$) or symptom severity (Positive-Other: $r = -0.22$, $p > .05$; Negative-Other: $r = -0.09$,

$p > .05$; see Table 3). Additionally, disorder-specific symptom severity and depression is moderately correlated as expected ($r = 0.54$, $p < .001$). These findings indicate that reduced evidence accumulation speed during self-referential decision-making is closely linked to higher levels of symptom severity in BDD and SAD, highlighting the potential importance of SRP deficits in these disorders.

Table 3. Pearson Correlation Between Drift Rates and Symptom Severity in BDD/SAD groups. Pearson correlation coefficients (r) and p -values (in parentheses) for the relationships between drift rates in four trial conditions, Positive-Self, Positive-Other, Negative-Self, and Negative-Other, and participants' symptom severity (centered proportional LSAS/BDD-YBOCS scores) and depression severity (standardized BDI-II scores).

Drift rate	Positive-Self	Positive-Other	Negative-Self	Negative-Other
symptom severity (p)	-0.52 (0.00)	-0.22 (0.24)	-0.48 (0.01)	-0.09 (0.63)
depression (p)	-0.34 (0.06)	-0.30 (0.11)	-0.27 (0.16)	-0.07 (0.70)

Last, we conducted a series of linear regression analyses to examine associations between symptom severity and drift rate in the clinical group, while controlling for depression severity (Table 4). We found that symptom severity significantly predicted both negative-self drift rates ($\beta = -1.77$, $p < .05$) and positive-self drift rates ($\beta = -1.69$, $p < .05$), with higher severity associated with lower drift rates in both conditions, which supported our hypothesis. Neither negative-other drift rates nor positive-other drift rates were significantly predicted by symptom severity.

Depressive symptoms did not significantly predict drift rates in any condition. The models for self-related conditions explained a moderate proportion of variance (negative self: adjusted $R^2 = .18$, positive self: adjusted $R^2 = .22$), whereas models for other-related conditions showed minimal explanatory power (negative other: $R^2 = -.06$, positive other: adj $R^2 = .03$). These results

suggest that symptom severity in BDD and SAD is more strongly linked to self-relevant perceptual decision-making than to processing of others.

Table 4. Linear regression: symptom severity predicting positive and negative drift rates in self-referential trials during the SRET

Drift rate	Predictor	β	SE	t	p	Lower 95% CI	Upper 95% CI
Negative_Self	depression	-0.02	0.45	-0.04	0.97	-0.93	0.89
	Symptom severity	-1.77	0.74	-2.40	0.02	-3.29	-0.26
Positive_Self	depression	-0.120	0.42	-0.47	0.64	-1.06	0.66
	Symptom severity	-1.69	0.70	-2.42	0.02	-3.12	-0.25
Negative_Other	depression	-0.08	0.54	-0.15	0.88	-1.18	1.02
	Symptom severity	-0.29	0.90	-0.33	0.75	-2.12	1.54
Positive_Other	depression	-0.47	0.40	-1.17	0.25	-1.30	0.36
	Symptom severity	-0.26	0.67	-0.39	0.70	-1.63	1.12
Model	R^2	Adjusted R^2		F		p value	
Negative_Self	0.23	0.18		4.12		0.028	
Positive_Self	0.28	0.22		5.13		0.013	
Negative_Other	0.01	-0.06		0.13		0.88	
Positive_Other	0.10	0.03		1.41		0.26	

Discussion

SRP has increasingly been recognized as a core transdiagnostic mechanism contributing to the onset and maintenance of a wide range of psychiatric disorders. Across conditions such as

SAD, BDD, and depression, disruptions in how individuals evaluate and respond to self-related information appear to underlie key clinical symptoms, including heightened self-consciousness, excessive rumination, and negative self-evaluation. While these behavioral manifestations of SRP dysfunction are well-documented, understanding the cognitive computations driving them has been more elusive. Computational modeling offers a promising method to quantify the latent processes underlying SRP, allowing researchers to move beyond self-report and behavior to identify how individuals weigh self- versus other-related information. This study applied a HDDM to examine SRP in individuals with BDD, SAD, and healthy controls, on a self-referential encoding task, judging whether positive or negative adjectives described themselves or another person (Barack Obama). By modeling trial-level decision data, we aimed to capture the efficiency of evidence accumulation in self- versus other-referential conditions. Including both clinical and healthy groups enabled a transdiagnostic investigation of SRP-related disruptions, moving beyond descriptive behavioral patterns to test for underlying cognitive mechanisms.

Our findings revealed distinct alterations in both self- and other-referential processing among individuals with BDD and SAD, relative to HC. First, consistent with our hypothesis, the clinical group demonstrated significantly lower drift rates than controls. The effect was most pronounced in self-referential trials (positive and negative), but was actually evident across all four conditions, which was unexpected. Within-group comparisons revealed that the difference in drift rate between self- and other-related conditions was significantly larger in the clinical group compared to healthy controls, indicating a more salient disruption in SRP. Importantly, drift rates did not significantly differ between the BDD and SAD groups across any of the four conditions, supporting the idea of a shared transdiagnostic mechanism. Second, lower drift rates were significantly associated with higher symptom severity. Follow-up linear regression analyses indicated that greater symptom severity predicted lower drift rates in self-related trials, but not in other-related trials, within the clinical groups, even after controlling for depression

severity. Together, these findings suggest that individuals with BDD and SAD experience a generalized inefficiency in evidence accumulation, with particularly pronounced disruptions during SRP, likely reflecting a shared transdiagnostic mechanism linked to symptom severity.

Multiple findings from our study suggest differences in evidence accumulation during self-referential decision-making in individuals with BDD/SAD. Specifically, in self-referential trials, we observed higher endorsement ratios for negative words and lower endorsement ratios for positive words in the clinical compared to control group. These results are consistent with a robust literature linking SRET performance with SAD, with previous studies showing reduced endorsement of positive traits and increased endorsement of negative traits compared to healthy controls (Dixon et al., 2022; Goldin et al., 2009; Talmon et al., 2021). Another study found that the typical positivity bias (i.e., greater endorsement of positive words) was diminished or absent among individuals with low and high social anxiety during the recollection of adjectives in self-referential trials (Kalenzaga & Jouhaud, 2018). Moreover, our finding that drift rate was associated with BDD/SAD symptom severity and remained a negative predictor in self-referential trials after controlling for depression severity replicates a similar finding showing inverse relationships between drift rate and social anxiety/depression severity in an adolescent sample (Castagna et al., 2022).

To our knowledge, this is the first study to examine SRP performance in individuals with BDD, yet it is not surprising that the results mirrored those observed in SAD, suggesting a transdiagnostic deficit in SRP across both disorders. While BDD is primarily characterized by excessive preoccupation with appearance, prior research indicates that these concerns often extend to broader negative self-evaluation and low self-esteem (Buhlmann et al., 2009; Johnson et al., 2018; Kuck et al., 2021; Holmes et al., 2025). Notably, deficits in emotion recognition in BDD appear to be particularly pronounced in self-relevant contexts (Buhlmann et al., 2006). Our results are consistent with another recent study employing Bayesian multilevel modeling and

Wiener diffusion models to analyze responses in a sentence word association paradigm (Dietel et al., 2021). That study found that both BDD and SAD participants exhibited shared explicit and implicit interpretation biases toward positive and negative adjectives across both appearance-related and social contexts. Our results lend support to the notion that altered SRP could contribute to interpretation biases observed in both BDD and SAD, particularly through altered endorsement and compromised evidence accumulation in self-relevant contexts. More broadly, the striking similarity in self-referential patterns across these two disorders reinforces the notion of cognitive proximity between BDD and SAD (Fang & Hofmann, 2010; Fang & Wilhelm, 2015). Despite differences in symptom focus, both conditions appear to involve overlapping disruptions in self-concept and information processing. This convergence points to potentially shared underlying mechanisms, such as fear of negative evaluation or core beliefs of inadequacy, which may operate transdiagnostically. Recognizing these commonalities could inform more integrative models of psychopathology and guide the development of transdiagnostic interventions targeting self-referential cognitive distortions, such as mindreading and personalization.

Contrary to our hypothesis, both clinical groups demonstrated lower drift rates in other-referential trials regardless of valence, although the impairment was less pronounced than in self-referential trials. We observed this pattern across BDD and SAD groups, which may indicate inefficient information accumulation not only during SRP, but also when evaluating others, indicating a broader evaluative processing deficit. By far, few existing SRET studies, including those focusing on depression and SAD, have incorporated both self- and other-referential conditions (e.g., Kalenzaga & Jouhaud, 2018), limiting understanding of whether the SRP deficits are really specific to self-contexts or generalize to contexts involving processing information about any person. Research supports self- and other-referential processing as inextricably linked constructs, with one being necessary for the other: our sense of self is

informed by social information from our environment and we evaluate the experience of others using our internal model of ourselves (Forgas & Williams, 2003; Molnar-Szakacs & Uddin, 2013; Mead & Morris, 1934; Uddin et al., 2007). One possibility is that this reflects a shared cognitive vulnerability across SAD and BDD, such as diffuse self-focused attention, cognitive rigidity, or heightened internal noise, that disrupts efficient decision-making even in contexts that are not explicitly self-relevant. For instance, when evaluating others (i.e., Obama), individuals with SAD or BDD may still engage in self-referential thinking, such as “What would I think if someone described me this way?” or “How does this compare to me?” This kind of self-involvement could interfere with task performance regardless of trial type. Another possibility is that the evaluative nature of the task itself triggers heightened fear of social judgment, excessive deliberation, or concern about making mistakes, even when the judgment is directed at someone else. More studies should be conducted to understand the role of self-other processing in these disorders.

This study included depression severity as a covariate, drawing on a substantial body of literature that demonstrates a negative association between drift rate and depression severity in the SRET (Allison et al., 2021; Castagna et al., 2022; Disney et al., 2017; Fritzsche et al., 2013; Goldstein et al., 2015; Hitchcock et al., 2022; Hsu et al., 2020; LeMoult et al., 2017). However, this pattern was not fully replicated in our study: depression severity was marginally associated with drift rate only in the positive-self condition. When both depression and BDD/SAD symptom severity were entered into a regression model, only disorder-specific symptoms significantly predicted drift rates across valence conditions. A key distinction from prior work is that participants in the current study were primarily diagnosed with BDD or SAD, not major depressive disorder; depression, when present, was a comorbid condition. BDD and SAD may be characterized by more narrowly focused self-schemas centered on appearance and social-evaluative concerns, which may exert a stronger and more direct influence on SRET decision-making than the broader self-critical cognitions typical of major depression. In our study, SRP

may correspond to a potential transdiagnostic cognitive mechanism tied specifically to social-evaluative concerns, rather than mood disturbance alone. Another likely contributor to the non-significant depression effect is the substantial shared variance between depression severity and disorder-specific symptoms ($r = 0.54$), with the latter potentially capturing more task-relevant cognitive processes.

One of the main strengths of this study is the use of the HDDM to analyze SRET performance, demonstrating that a standard hierarchical DDM was sufficient to capture group-level differences by extracting latent cognitive parameters from behavioral data that traditional measures may obscure. HDDM has increasingly been applied in SRET research with success (Allison et al., 2021; Beevers et al., 2019; Castagna et al., 2022; Dainer-Best et al., 2018ab; Disner et al., 2017; Hitchcock et al., 2022; Price et al., 2021). Within DDM theory, drift rate is considered a more direct and specific index of SRP than raw reaction time (RT), as it isolates the efficiency of evidence accumulation from non-decisional components, such as perceptual encoding, attention engagement, and motor execution. Previous research has shown that computational model parameters can exhibit substantially higher reliability than behavioral measures (Price et al., 2019). In our study, both RT and drift rate distinguish clinical groups from healthy controls in self-referential trials, indicating convergent sensitivity of these measures when self-processing is disrupted. However, only drift rate captured significant between-group differences in other-referential trials, while RT showed only a marginal group difference. This divergence underscores the advantage of model-based parameters in detecting subtle processing inefficiencies that may be masked in RT due to variability in motor preparation, non-task-specific attentional fluctuations, or measurement noise related to the limited sample size. In other words, while RT can reveal broad performance slowing when disruptions are large (as in self-referential trials), drift rate can detect more fine-grained differences in cognitive efficiency even when overt response speed is unchanged.

Our use of HDDM underscores the distinct links between latent decision-making parameters and SRET performance; however, several study limitations should be acknowledged. First, the Python HDDM package lacks support for simultaneous modeling of within-subject (e.g., valence, self-other) and between-subject (e.g., depression, symptom severity) predictors, which can cause parameter shrinkage and bias group differences (Boehm et al., 2018). We therefore used a two-step approach: fitting HDDM with within-subject predictors to estimate drift rates, then regressing these estimates on between-subject variables. Although this method was used commonly in prior studies (e.g., Allison et al., 2021; Hitchcock et al., 2022), it does not fully account for uncertainty propagation, potentially underestimating errors and inflating Type I errors. Additionally, due to these constraints, we ran a single model across all participants rather than separate group models. While this improves parameter stability and convergence, it may bias estimates and mask group differences. Nonetheless, model validation and parameter recovery indicated good model performance. Future work with more flexible frameworks (e.g., HSSM Python) could address these issues for stronger inferences. Second, only BDD and SAD participants completed disorder-specific symptom scales, while healthy controls were assessed only for depression, limiting symptom severity analyses to clinical groups. Third, to enable cross-group linear regression, we used proportional scores (raw score divided by maximum possible) as proxies for symptom severity. This method risks ceiling effects and the assumption of linearity, treating all symptoms as equally weighted despite likely clinical heterogeneity. Nevertheless, despite the relatively constrained sample size and controlling for depression severity as a covariate, we still observed a predictive relationship between drift rate and BDD/SAD symptom severity, indicating a robust association between SRP inefficiency and disorder-specific symptoms. Fourth, our study included only a single measure of symptom severity. Future research should incorporate more nuanced, multidimensional assessments, such as self-worth, social avoidance, and rumination, and examine their associations with drift

rate. This may reveal additional transdiagnostic symptom dimensions influencing self- and other-referential processing. Lastly, given our modest sample size and model complexity, we did not include neuroimaging data as a third predictor, as this would substantially increase the number of interaction terms and could lead to overfitting or convergence issues. Instead, we focus here on behavioral parameters, and we refer readers to Iannazzi et al. (2025) for neuroimaging results from the same dataset. With a larger sample, future work could clarify the neural mechanisms underlying impaired SRP and link computational parameters to brain activity associated with self-referential cognition. Integrating neuroimaging and computational modeling approaches would be particularly valuable for elucidating the complex neural substrates of SRP and their dynamic relationship with latent cognitive processes.

Reference

- Allison, G. O., Benau, E. M., Asbaghi, S., Pagliacco, 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>
- Beck, A. T., Steer, R. A., Ball, R., & Ranieri, W. F. (1996). Comparison of Beck Depression Inventories-IA and-II in Psychiatric Outpatients. *Journal of Personality Assessment*, 67(3), 588–597. https://doi.org/10.1207/s15327752jpa6703_13
- 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, 212–227. <https://doi.org/10.1037/abn0000405>
- Boehm, U., Annis, J., Frank, M. J., Hawkins, G. E., Heathcote, A., Kellen, D., Kryptos, A.-M., Lerche, V., Logan, G. D., Palmeri, T. J., Van Ravenzwaaij, D., Servant, M., Singmann, H., Starns, J. J., Voss, A., Wiecki, T. V., Matzke, D., & Wagenmakers, E.-J. (2018). Estimating across-trial variability parameters of the Diffusion Decision Model: Expert advice and recommendations. *Journal of Mathematical Psychology*, 87, 46–75. <https://doi.org/10.1016/j.jmp.2018.09.004>
- Boehme, S., Miltner, W. H. R., & Straube, T. (2015). Neural correlates of self-focused attention in social anxiety. *Social Cognitive and Affective Neuroscience*, 10(6), 856–862. <https://doi.org/10.1093/scan/nsu128>
- 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>

Buhlmann, U., Etcoff, N. L., & Wilhelm, S. (2006). Emotion recognition bias for contempt and anger in body dysmorphic disorder. *Journal of Psychiatric Research*, 40(2), 105–111.

<https://doi.org/10.1016/j.jpsychires.2005.03.006>

Buhlmann, U., Teachman, B. A., Naumann, E., Fehlinger, T., & Rief, W. (2009). The meaning of beauty: Implicit and explicit self-esteem and attractiveness beliefs in body dysmorphic disorder.

Journal of Anxiety Disorders, 23(5), 694–702. <https://doi.org/10.1016/j.janxdis.2009.02.008>

Buhlmann, U., Wacker, R., & Dziobek, I. (2015). Inferring other people's states of mind: Comparison across social anxiety, body dysmorphic, and obsessive–compulsive disorders.

Journal of Anxiety Disorders, 34, 107–113. <https://doi.org/10.1016/j.janxdis.2015.06.003>

Button, K. S., Browning, M., Munafò, M. R., & Lewis, G. (2012). Social inference and social anxiety: Evidence of a fear-congruent self-referential learning bias. *Journal of Behavior Therapy and Experimental Psychiatry*, 43(4), 1082–1087.

<https://doi.org/10.1016/j.jbtep.2012.05.004>

Castagna, P. J., Waters, A. C., & Crowley, M. J. (2022). Computational Modeling of Self-Referential Processing Reveals Domain General Associations with Adolescent Anxiety Symptoms. *Research on Child and Adolescent Psychopathology*.

<https://doi.org/10.1007/s10802-022-01012-1>

Cataldo, A. M., Scheuer, L., Maksimovskiy, A. L., Germine, L. T., & Dillon, D. G. (2023).

Abnormal evidence accumulation underlies the positive memory deficit in depression. *Journal of Experimental Psychology: General*, 152(1), 139–156. <https://doi.org/10.1037/xge0001268>

Coles, M. E., Turk, C. L., Heimberg, R. G., & Fresco, D. M. (2001). Effects of varying levels of anxiety within social situations: Relationship to memory perspective and attributions in social

phobia. *Behaviour Research and Therapy*, 39(6), 651–665. [https://doi.org/10.1016/S0005-7967\(00\)00035-8](https://doi.org/10.1016/S0005-7967(00)00035-8)

Clark, D. M., & Wells, A. (1995). A cognitive model of social phobia. In R. G. Heimberg, M. R. Liebowitz, D. A. Hope, & F. R. Schneier (Eds.), *Social phobia: Diagnosis, assessment, and treatment* (pp. 69–93). The Guilford Press.

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.

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.

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>

Dickerson, S. S., & Kemeny, M. E. (2004). Acute stressors and cortisol responses: a theoretical integration and synthesis of laboratory research. *Psychological bulletin*, 130(3), 355.

Disner study is missing here

Fang, A., & Hofmann, S. G. (2010). Relationship between social anxiety disorder and body dysmorphic disorder. *Clinical Psychology Review*, 30(8), 1040–1048. <https://doi.org/10.1016/j.cpr.2010.08.001>

Fang, A., & Wilhelm, S. (2015). Clinical features, cognitive biases, and treatment of body dysmorphic disorder. *Annual Review of Clinical Psychology*, 11, 187-212.

Fernández De La Cruz, L., Enander, J., Rück, C., Wilhelm, S., Phillips, K. A., Steketee, G., Sarvode Mothi, S., Krebs, G., Bowyer, L., Monzani, B., Veale, D., & Mataix-Cols, D. (2021). Empirically defining treatment response and remission in body dysmorphic disorder. *Psychological Medicine*, 51(1), 83–89. <https://doi.org/10.1017/S0033291719003003>

Feusner, J. D., Hembacher, E., Moller, H., & Moody, T. D. (2011). Abnormalities of object visual processing in body dysmorphic disorder. *Psychological Medicine*, 41(11), 2385–2397. <https://doi.org/10.1017/S0033291711000572>

Gaydukevych, D., & Kocovski, N. L. (2012). Effect of self-focused attention on post-event processing in social anxiety. *Behaviour Research and Therapy*, 50(1), 47–55. <https://doi.org/10.1016/j.brat.2011.10.010>

Gelman, A., & Rubin, D. B. (1992). Inference from Iterative Simulation Using Multiple Sequences. *Statistical Science*, 7(4). <https://doi.org/10.1214/ss/1177011136>

Add this reference: Grennan, Zhao, & Fang (2023):

<https://www.sciencedirect.com/science/article/pii/S2211364923000623>

Haines, N., Beauchaine, T. P., Galdo, M., Rogers, A. H., Hahn, H., Pitt, M. A., Myung, J. I., Turner, B. M., & Ahn, W.-Y. (2020). Anxiety Modulates Preference for Immediate Rewards Among Trait-Impulsive Individuals: A Hierarchical Bayesian Analysis. *Clinical Psychological Science*, 8(6), 1017–1036. <https://doi.org/10.1177/2167702620929636>

Hedman, E., Ström, P., Stünkel, A., & Mörtberg, E. (2013). Shame and Guilt in Social Anxiety Disorder: Effects of Cognitive Behavior Therapy and Association with Social Anxiety and Depressive Symptoms. *PLoS ONE*, 8(4), e61713. <https://doi.org/10.1371/journal.pone.0061713>

Hitchcock, P. F., Britton, W. B., Mehta, K. P., & Frank, M. J. (2022). Self-judgment dissected: A computational modeling analysis of self-referential processing and its relationship to trait

mindfulness facets and depression symptoms. *Cognitive, Affective, & Behavioral Neuroscience*.
<https://doi.org/10.3758/s13415-022-01033-9>

Ho, J. T., Preller, K. H., & Lenggenhager, B. (2020). Neuropharmacological modulation of the aberrant bodily self through psychedelics. *Neuroscience & Biobehavioral Reviews*, 108, 526–541. <https://doi.org/10.1016/j.neubiorev.2019.12.006>

Hsu, K. J., McNamara, M. E., Shumake, J., Stewart, R. A., Labrada, J., Alario, A., Gonzalez, G. D. S., Schnyer, D. M., & Beevers, C. G. (2020). Neurocognitive predictors of self-reported reward responsivity and approach motivation in depression: A data-driven approach. *Depression and Anxiety*, 37(7), 682–697. <https://doi.org/10.1002/da.23042>

Hur, J.-W., Shin, H., Jung, D., Lee, H.-J., Lee, S., Kim, G. J., Cho, C.-Y., Choi, S., Lee, S.-M., & Cho, C.-H. (2021). Virtual Reality–Based Psychotherapy in Social Anxiety Disorder: fMRI Study Using a Self-Referential Task. *JMIR Mental Health*, 8(4), e25731. <https://doi.org/10.2196/25731>

Ingram, R. E. (1990). Self-focused attention in clinical disorders: Review and a conceptual model. *Psychological Bulletin*, 107(2), 156–176. <https://doi.org/10.1037/0033-2909.107.2.156>

Jamieson, A. J., Harrison, B. J., Delahoy, R., Schmaal, L., Felmingham, K. L., Phillips, L., & Davey, C. G. (2023). A brain model of altered self-appraisal in social anxiety disorder. *Translational Psychiatry*, 13(1), 344. <https://doi.org/10.1038/s41398-023-02644-7>

Johnson, S., Williamson, P., & Wade, T. D. (2018). A systematic review and meta-analysis of cognitive processing deficits associated with body dysmorphic disorder. *Behaviour Research and Therapy*, 107, 83–94. <https://doi.org/10.1016/j.brat.2018.05.013>

Judah, M. R., Grant, D. M., & Carlisle, N. B. (2016). The effects of self-focus on attentional biases in social anxiety: An ERP study. *Cognitive, Affective & Behavioral Neuroscience*, 16(3), 393–405. <https://doi.org/10.3758/s13415-015-0398-8>

- Kelley, W. M., Macrae, C. N., Wyland, C. L., Caglar, S., Inati, S., & Heatherton, T. F. (2002). Finding the Self? An Event-Related fMRI Study. *Journal of Cognitive Neuroscience*, 14(5), 785–794. <https://doi.org/10.1162/08989290260138672>
- Kuck, N., Cafitz, L., Bürkner, P.-C., Hoppen, L., Wilhelm, S., & Buhlmann, U. (2021). Body dysmorphic disorder and self-esteem: A meta-analysis. *BMC Psychiatry*, 21(1), 310. <https://doi.org/10.1186/s12888-021-03185-3>
- LeMoult, J., Kircanski, K., Prasad, G., & Gotlib, I. H. (2017). Negative Self-Referential Processing Predicts the Recurrence of Major Depressive Episodes. *Clinical Psychological Science*, 5(1), 174–181. <https://doi.org/10.1177/2167702616654898>
- Liebowitz, M. R. (1987). Social Phobia. In D. F. Klein (Ed.), *Modern Trends in Pharmacopsychiatry* (Vol. 22, pp. 141–173). S. Karger AG. <https://doi.org/10.1159/000414022>
- Luu, P., Collins, P., & Tucker, D. M. (2000). Mood, personality, and self-monitoring: Negative affect and emotionality in relation to frontal lobe mechanisms of error monitoring. *Journal of Experimental Psychology: General*, 129, 43–60. <https://doi.org/10.1037/0096-3445.129.1.43>
- 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>
- Meffert, H., Blanken, L., Blair, K. S., White, S. F., & Blair, J. R. (2013). The influence of valence and decision difficulty on self-referential processing. *Frontiers in Human Neuroscience*, 7. <https://doi.org/10.3389/fnhum.2013.00046>
- Moscovitch, D. A. (2009). What Is the Core Fear in Social Phobia? A New Model to Facilitate Individualized Case Conceptualization and Treatment. *Cognitive and Behavioral Practice*, 16(2), 123–134. <https://doi.org/10.1016/j.cbpra.2008.04.002>

Myers, C. E., Interian, A., & Moustafa, A. A. (2022). A practical introduction to using the drift diffusion model of decision-making in cognitive psychology, neuroscience, and health sciences.

Frontiers in Psychology, 13, 1039172. <https://doi.org/10.3389/fpsyg.2022.1039172>

Nolen-Hoeksema, S., Wisco, B. E., & Lyubomirsky, S. (2008). Rethinking Rumination.

Perspectives on Psychological Science, 3(5), 400–424. <https://doi.org/10.1111/j.1745-6924.2008.00088.x>

Noorani, I. (2014). LATER models of neural decision behavior in choice tasks. *Frontiers in*

Integrative Neuroscience, 8. <https://www.frontiersin.org/articles/10.3389/fnint.2014.00067>

Norton, A. R., & Abbott, M. J. (2016). Self-Focused Cognition in Social Anxiety: A Review of the Theoretical and Empirical Literature. *Behaviour Change*, 33(1), 44–64.

<https://doi.org/10.1017/bec.2016.2>

Phillips, K. A., Hollander, E., Rasmussen, S. A., & Aronowitz, B. R. (1997). A severity rating scale for body dysmorphic disorder: development, reliability, and validity of a modified version of the Yale-Brown Obsessive Compulsive Scale. *Psychopharmacology bulletin*, 33(1), 17.

Add this reference: Price, Brown, & Siegle (2019) <https://pubmed.ncbi.nlm.nih.gov/30449531/>

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>

R Core Team (202). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>.

Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review*, 85(2), 59–108.

<https://doi.org/10.1037/0033-295X.85.2.59>

Schreiber, F., Witthöft, M., Neng, J. M. B., & Weck, F. (2016). Changes in negative implicit evaluations in patients of hypochondriasis after treatment with cognitive therapy or exposure therapy. *Journal of Behavior Therapy and Experimental Psychiatry*, 50, 139–146.

<https://doi.org/10.1016/j.jbtep.2015.07.005>

Spurr, J. M., & Stopa, L. (2002). Self-focused attention in social phobia and social anxiety. *Clinical Psychology Review*, 22(7), 947–975. [https://doi.org/10.1016/S0272-7358\(02\)00107-1](https://doi.org/10.1016/S0272-7358(02)00107-1)

Stein, D. J. (2015). Social anxiety disorder and the psychobiology of self-consciousness.

Frontiers in Human Neuroscience, 9. <https://doi.org/10.3389/fnhum.2015.00489>

Timbremont, B., & Braet, C. (2004). Cognitive vulnerability in remitted depressed children and adolescents. *Behaviour Research and Therapy*, 42(4), 423–437. [https://doi.org/10.1016/S0005-7967\(03\)00151-7](https://doi.org/10.1016/S0005-7967(03)00151-7)

Veale, D. (2004). Body dysmorphic disorder. *Postgraduate Medical Journal*, 80(940), 67–71.

<https://doi.org/10.1136/pmj.2003.015289>

Vandekerckhove, J., Tuerlinckx, F., & Lee, M. D. (2011). Hierarchical diffusion models for two-choice response times. *Psychological Methods*, 16(1), 44–62. <https://doi.org/10.1037/a0021765>

Wiecki, T. V., Sofer, I., & Frank, M. J. (2013). HDDM: Hierarchical Bayesian estimation of the Drift-Diffusion Model in Python. *Frontiers in Neuroinformatics*, 7.

<https://doi.org/10.3389/fninf.2013.00014>

Wilhelm, S., Greenberg, J. L., Rosenfield, E., Kasarskis, I., & Blashill, A. J. (2016). The Body Dysmorphic Disorder Symptom Scale: Development and preliminary validation of a self-report scale of symptom specific dysfunction. *Body Image*, 17, 82–87.

<https://doi.org/10.1016/j.bodyim.2016.02.006>

Woody, S. R., Chambless, D. L., & Glass, C. R. (1997). Self-focused attention in the treatment of social phobia. *Behaviour Research and Therapy*, 35(2), 117–129.

[https://doi.org/10.1016/S0005-7967\(96\)00084-8](https://doi.org/10.1016/S0005-7967(96)00084-8)

Yoon, H.-J., Seo, E. H., Kim, J.-J., & Choo, I. H. (2019). Neural Correlates of Self-referential Processing and Their Clinical Implications in Social Anxiety Disorder. *Clinical Psychopharmacology and Neuroscience*, 17(1), 12–24.

<https://doi.org/10.9758/cpn.2019.17.1.12>

Yoshimura, S., Okamoto, Y., Onoda, K., Matsunaga, M., Okada, G., Kunisato, Y., Yoshino, A., Ueda, K., Suzuki, S., & Yamawaki, S. (2014). Cognitive behavioral therapy for depression changes medial prefrontal and ventral anterior cingulate cortex activity associated with self-referential processing. *Social Cognitive and Affective Neuroscience*, 9(4), 487–493.

<https://doi.org/10.1093/scan/nst009>

Supplementary Information

Table 1 Model Parameter Recovery Results. The correlation coefficients (R) between the true parameter values and the recovered parameter estimates for a diffusion model. Parameters include Threshold (a), Reaction Time (rt), Starting Bias (z), Intercept (β), Valence (β), Trial Type (β), and the interaction term Valence $\times$ Trial Type (β).

parameter s	Threshol d (a)	Reaction time (rt)	Starting bias (z)	Intercept (β)	Valence (β)	Trial type (β)	Valence*tria l type (β)
correlation coefficient (R)	0.92	0.92	0.88	0.90	0.99	0.86	0.97

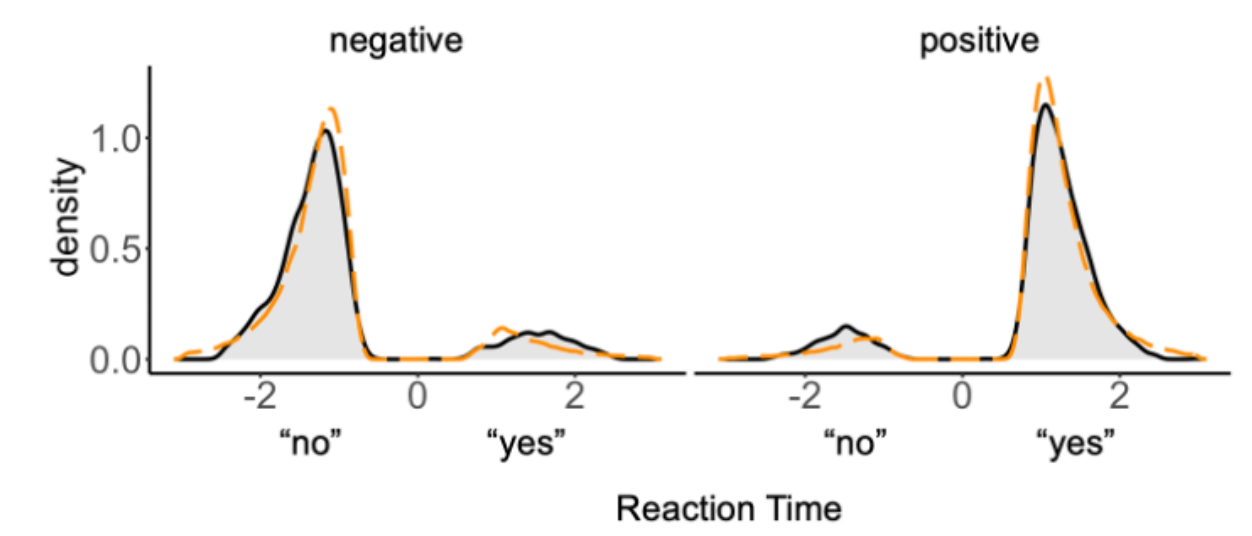


FIGURE 2. Computational model fits the behavioral patterns. Empirical reaction time density plots (black) are displayed by valence, with response type “no” within each valence category mirrored on the left axis and “yes” on the right axis. These plots are overlaid with simulations

from the best-fitting computational model (orange dotted lines). The results indicate that the DDM effectively captures key characteristics of the empirical data, including response frequencies and the overall shape of the reaction time distributions.