Association between negative cognitive bias and depression: A symptom-level approach

Christopher G. Beevers,
University of Texas at Austin

Michael C. Mullarkey,
University of Texas at Austin

Justin Dainer-Best,
University of Texas at Austin

Rochelle A. Stewart,
University of Texas at Austin

Jocelyn Labrada,
University of Texas at Austin

John J.B. Allen,
University of Arizona

John E. McGeary, and
Providence Veterans Affairs and Brown University Medical School

Jason Shumake
University of Texas at Austin

Abstract

Cognitive models of depression posit that negatively biased self-referent processing and attention have important roles in the disorder. However, depression is a heterogeneous collection of symptoms and all symptoms are unlikely to be associated with these negative cognitive biases. The current study involved 218 community adults whose depression ranged from no symptoms to clinical levels of depression. Random forest machine learning was used to identify the most important depression symptom predictors of each negative cognitive bias. Depression symptoms were measured with the Beck Depression Inventory – II. Model performance was evaluated using

predictive R-squared (R^2_{pred}), the expected variance explained in data not used to train the algorithm, estimated by 10 repetitions of 10-fold cross-validation. Using the Self-Referent Encoding Task (SRET), depression symptoms explained 34% to 45% of the variance in negative self-referent processing. The symptoms of sadness, self-dislike, pessimism, feelings of punishment, and indecision were most important. Notably, many depression symptoms made virtually no contribution to this prediction. In contrast, for attention bias for sad stimuli, measured with the dot-probe task using behavioral reaction time and eye gaze metrics, no reliable symptom predictors were identified. Findings indicate that a symptom-level approach may provide new insights into which symptoms, if any, are associated with negative cognitive biases in depression.

General Scientific Summary:

This study finds that many symptoms of depression are not strongly associated with thinking negatively about oneself or attending to negative information. This implies that negative cognitive biases may not be strongly associated with depression per se, but may instead contribute to the maintenance of specific depression symptoms, such as sadness, self-dislike, pessimism, feelings of punishment, and indecision.

Keywords

cognitive model of depression; symptom importance; machine learning

---

The cognitive model of depression posits that depression symptoms are maintained by negatively biased cognition, particularly negative cognition about the self (Beck, 1967). In this model, the concept of the self-schema—an internal representation of the self and the world around oneself—influences what people attend to, how they interpret new information, and what they remember at a later point in time (Disner, Beevers, Haigh, & Beck, 2011). In depression, these self-schemas tend to be negatively biased, thus prioritizing the processing of incoming negative information. Negatively biased information processing, in turn, is thought to maintain symptoms of depression.

Two cognitive mechanisms at the center of the cognitive model involve negative self-referent processing and negatively biased selective attention. The self-referent encoding task (SRET; Derry & Kuiper, 1981) has been used extensively to measure self-referent processing (Alloy, Abramson, Murray, Whitehouse, & Hogan, 1997a; Dozois & Dobson, 2001; Goldstein, Hayden, & Klein, 2015). When completing the SRET, participants are asked to make binary decisions (yes/no) about whether or not positive and negative adjectives are self-descriptive. In addition to measuring number of word endorsements for positive and negative stimuli, decision-making reaction time can also be assessed. In general, as depression symptom severity increases, adults more easily classify negative words as self-descriptive (Dainer-Best, Lee, Shumake, Yeager, & Beevers, 2018a).

Negatively biased attention is a second cognitive mechanism that has been implicated both theoretically and empirically in depression (Gotlib & Joormann, 2010). The emotional variant of the dot-probe task measures the tendency to preferentially attend to negative stimuli (typically sad images in depression) versus neutral stimuli. This bias can be inferred

behaviorally with reaction time or measured directly by monitoring eye movements. A meta-analysis of 29 studies found a medium effect size difference between depressed and non-depressed participants for negative attention bias, particularly when measured with the dot-probe task (Peckham, McHugh, & Otto, 2010). Similarly, a meta-analysis of eye movements indicates that attention bias in depression is often characterized by reduced maintenance of gaze towards positive stimuli and increased maintenance of gaze towards negative stimuli (Armstrong & Olatunji, 2012).

Although these two cognitive biases appear to be associated with depression, very few studies have investigated whether certain symptoms of depression are more closely linked than others to these cognitive biases. This is an important issue, as depression is far from a homogenous syndrome (Fried, 2017). For example, 48.6% of 3,703 adults enrolled in a depression treatment trial had a unique pattern of symptoms that no one else endorsed (Fried & Nesse, 2015a). Depression symptoms are also not interchangeable indicators of an underlying depression construct, as commonly measured depression symptoms do not typically load on to a single latent factor (Fried et al., 2016b; Osman et al., 1997). Further, total sum scores on several depression inventories do not show measurement invariance over time—that is, the relationship between the depression latent variable and its indicators (i.e., symptoms) is not stable over time (Fried et al., 2016b). For these reasons, associations may be obscured when total sum scores are used to test etiological or maintenance models of depression. Indeed, some have persuasively argued that depression researchers should focus on the measurement of individual symptoms as an alternative to sum scores (Fried & Nesse, 2015a).

Consistent with this idea, an emerging literature indicates that depression symptoms may be differentially linked to cognitive risk factors (Marchetti, Loeys, Alloy, & Koster, 2016), cognitive biases (Marchetti et al., 2018), and psychosocial functioning in adults (Fried & Nesse, 2014). Prior work (Fried, Nesse, Zivin, Guille, & Sen, 2014) suggests some risk factors (i.e., depression history, childhood stress, sex, stressful life events) predict the development of only a subset of depression symptoms (and the associations with symptoms differed across risk factors). Another study revealed that hopelessness was most strongly connected with the depression symptom of pessimism, and had weaker but significant associations with sadness, suicidality, self-dislike, worthlessness, loss of interest, lack of energy, and anhedonia (Marchetti et al., 2016). Few comparable symptom-level studies have been completed for the cognitive biases of self-referent negative processing and negative attention bias (Alloy & Clements, 1998; see Alloy, Just, & Panzarella, 1997b; Beard, Rifkin, & Bjorgvinsson, 2017).

Given that cognitive theories of depression (e.g., Beck & Bredemeier, 2016) tend not to be specific about which symptoms of depression negative cognitive biases are expected to maintain (e.g., models typically conclude that negative attention bias is associated with depression in general), we did not have any specific a priori hypotheses. Indeed, the empirical literature on this question is very sparse. However, there is a substantial research literature examining negative attention bias (Armstrong & Olatunji, 2012) and negative self-referent processing (LeMoult & Gotlib, 2018), as these are central constructs within the cognitive model of depression (cf., Disner et al., 2011). These negative cognitive biases are

actively being examined as potential treatment targets (Beevers, Clasen, Enock, & Schnyer, 2015; Dainer-Best, Shumake, & Beevers, 2018b; Dozois et al., 2009) and work has examined the psychometric properties of tasks used to measure these biases (Dainer-Best et al., 2018a; Rodebaugh et al., 2016). Thus, although few studies have examined correlations between specific symptoms and these tasks, these cognitive processes are thought to have an important role in the maintenance of depression. It currently remains unclear exactly which symptoms (if any) are most closely associated with these cognitive biases.

Therefore, the current study proposes to identify the most important depression symptoms for self-referent negative processing and negative attention bias. Although a number of approaches are available to identify important predictors (Wei, Lu, & Song, 2015), we adopted a machine learning approach, primarily using random forest (Breiman, 2001) and elastic-net regularization (Zou & Hastie, 2005), to identify the most important symptom predictors of each cognitive bias. Machine learning uses statistical learning techniques to identify patterns in observed data to develop a predictive model and then applies this predictive model to new data (Yarkoni & Westfall, 2017). One goal of this approach is to identify reliable predictors of outcomes that are more likely to generalize to new data than typical statistical approaches. This approach is also good for exploratory, data-driven analyses, particularly when there are many predictors for an outcome and no strong hypotheses about which predictors may be particularly important (Hastie, Tibshirani, & Friedman, 2009). Thus, this approach is well suited for the present study, as we did not have firm hypotheses about which symptoms may be the most important predictors of negative self-referent processing and negative attention bias.

Methods

Participants

Participants included 218 adults of European ancestry recruited from the Austin, Texas community. The larger aim of this project is to examine polygenetic associations with neurocognitive phenotypes measured in the current study; thus, recruitment was limited to people with European ancestry for the genetic aims. Consistent with dimensional approaches to the study of psychopathology (Gibb, Alloy, Abramson, Beevers, & Miller, 2004), we recruited individuals who varied in depression symptoms. To ensure recruitment of individuals along the depression continuum, participant screening was monitored on an ongoing basis and recruitment was altered as necessary to obtain an approximately normal distribution of depression. To achieve this goal, individuals on the upper end of the depression continuum were oversampled. We also oversampled men, to help ensure that men are represented at all levels of depression symptom severity.

Use of an "enriched" community sample is a common strategy in psychopathology (Alloy, Abramson, Raniere, & Dyller, 1999) and genetics research (Rutter, 2005) and it has been suggested that including more cases in the tails of the psychopathology distribution can enhance the likelihood of detecting genetic signal (Cuthbert, 2014). Prior work also suggests that negative cognitive bias is dimensional rather than taxonic (Gibb et al., 2004)—that is, negative cognitive biases are present to a greater or lesser extent in all individuals. The degree of cognitive vulnerability present also varies with depression severity—as depression

severity increases, so too does negative cognitive bias. Thus, sampling individuals across the depression continuum should provide adequate coverage of negative cognitive bias.

Inclusion criteria included: 1) 18 to 35 years of age; and 2) ability to speak, read, and understand English sufficiently well to complete informed consent; 3) normal or corrected to normal vision; and 4) European ancestry. Exclusion criteria included: 1) current use of steroidal (e.g., prednisone, dexamethasone) or psychotropic medications; 2) serious medical complications (e.g., cancer, diabetes, epilepsy or head trauma); 3) heavy tobacco use defined as smoking 20 cigarettes per day or > 20 pack years (Kamholz, 2004; World Health Organization, 2011); 4) a score of 2 or greater on the drug symptoms subscale of the Psychiatric Diagnostic Screening Questionnaire (PDSQ, Zimmerman & Mattia, 2001); 5) a score of 2 or greater on the PDSQ alcohol symptoms subscale; 6) a score of 1 or greater on the PDSQ psychotic symptoms subscale; and 7) being at imminent risk of self-harm or harm to others or having a recent history of suicidal behavior (a Columbia-Suicide Severity Rating Scale (Posner et al., 2011) suicidal ideation score of Type 4 or 5 in the past 2 months or suicidal behavior in the past 2 months). Participants received $50 for participation in this assessment.

We selected these exclusion criteria for a number of reasons. Heavy tobacco and alcohol use can impact users' cognitive function (Hall et al., 2015), so we excluded these individuals from participation. Consistent with prior work (Geschwind & Flint, 2015; Mccarthy et al., 2008), people taking psychotropic medications were also excluded even though up to 20% of the US population may be taking a psychiatric medication (Olfson & Marcus, 2009). Finally, to minimize effects of cognitive aging on our outcomes and further reduce sample heterogeneity, we recruited adults from a restricted age range (18 – 35 years old). These decisions limit generalizability of our findings to a degree, so future research will need to determine whether observed associations are found in more heterogeneous samples.

Given these considerations, our sample was on average in the young adult age range, mostly female, non-Hispanic, and Caucasian (see Table 1). The vast majority have never been married and averaged two years of post-secondary education. Only slightly more than 50% had private health insurance and the majority had a family income below $50,000. Approximately one quarter of the sample met criteria for current MDD and approximately half of the sample met criteria for MDD in their lifetime. There was psychiatric comorbidity, as one quarter of the sample had a current anxiety disorder and approximately half of the sample had a current psychiatric disorder of any kind. The correlation between the presence of current MDD and any current anxiety disorder was modest, $r = .36$, 95% CI [0.23, 0.47], indicating that a minority of individuals in the full sample (n = 33, 15.1%) had both MDD and an anxiety disorder. Of those with current MDD, 55% ($n = 33/60$) had a concurrent anxiety disorder. The Internal Review Board at the University of Texas at Austin approved all study procedures (Title: Contribution of Genome-Wide Variation to Cognitive Vulnerability to Negative Valence System, #2016–08-0015).

Assessments

Beck Depression Inventory-II (BDI-II): The BDI-II (Beck, Steer, & Brown, 1996) is a widely used self-report questionnaire that assesses depression severity. The BDI-II consists

of 21 items and measures the presence and severity of cognitive, motivational, affective, and somatic symptoms of depression. Past reports have indicated test-retest reliability and validity is adequate among psychiatric outpatient samples (Beck, Steer, & Carbin, 1988). The BDI-II was used in the depression symptom importance analyses, as it measures a wide variety of depression symptoms and has been used extensively in research that examines cognitive models of depression.

Mini International Neuropsychiatric Interview (MINI): Research assistants trained on diagnostic interviewing completed in-person interviews for eligible participants, using version 7.2 of the Mini International Neuropsychiatric Interview for the DSM-5 (Sheehan et al., 1998). The MINI is a standardized instrument used for brief screenings to diagnose a variety of psychiatric disorders. Research assistants took part in a training during which they learned interview skills, role-played interviews, and reviewed diagnostic criteria. After the workshop, they listened to interviews conducted by experienced researchers and had their initial screening interviews monitored for fidelity. Interviews were audio-recorded with consent from participants throughout the study for fidelity analyses.

Approximately 10% of the interviews were randomly selected and then rated by both study assessors and a licensed clinical psychologist with extensive experience with diagnostic assessment, who was not a rater for any of the study interviews. Diagnostic agreement among the ratings was computed using Fleiss' Kappa for multiple raters. Agreement for current MDD, MDD lifetime, and MDD recurrent was excellent (ks = 1.00, 0.82, 0.93, ps < .0001, respectively). Agreement was similarly high for Panic Disorder Current, Panic Disorder Lifetime, Generalized Anxiety Disorder, and Alcohol Use Disorder Current (ks = 1.00, 1.00, 1.00, 0.86, ps < .0001, respectively). Of all the disorders assessed, Obsessive Compulsive Disorder had the lowest agreement (k = 0.78, p < .0001), although agreement was still quite strong. Agreement for several diagnoses (e.g., Bipolar Disorder, Anorexia, Agoraphobia, Post-Traumatic Stress Disorder, Mood Disorders with Psychotic Features) was not computed because participants with these disorders were not retained in the reliability subset, which prevented examination of agreement. These disorders were not common in the full sample, likely due in part to low prevalence rates in general and study exclusion criteria.

Self-Referent Encoding Task (SRET): The SRET (Derry & Kuiper, 1981) is a computer-based task designed to assess schema-related processing. Participants make decisions about whether positive and negative adjectives are self-descriptive. Participants view the words one at a time and make rapid judgments about whether or not each word presented described themselves following word offset. Participants viewed 26 negative and 26 positive words.[1] Words were selected from a well-validated list of positive and negative self-descriptive adjectives (Doost, Moradi, Taghavi, Yule, & Dalgleish, 1999).

Following five trials to introduce the mechanics of the task to participants, the SRET consisted of three blocks. In each block, all 52 words were displayed once in random order.

---

[1]Positive words were: Awesome, Best, Brilliant, Confident, Content, Cool, Excellent, Excited, Fantastic, Free, Friendly, Fun, Funny, Glad, Good, Great, Happy, Helpful, Joyful, Kind, Loved, Nice, Playful, Pleased, Proud, Wonderful; Negative words were: Alone, Angry, Annoyed, Ashamed, Bad, Depressed, Guilty, Hateful, Horrible, Lonely, Lost, Mad, Nasty, Naughty, Sad, Scared, Dumb, Sorry, Stupid, Terrible, Unhappy, Unloved, Unwanted, Upset, Wicked, Worried.

Words were displayed in white text on a black screen and remained on-screen until participants responded. Participants were told to use the Q or P keys on the keyboard to answer whether the word described them or not. Each trial was followed by a 1,500 ms inter-trial interval. After completing the task, participants were given a number-based task (i.e., short digit span task) to complete for eight minutes. Then, participants were asked to recall as many adjectives as possible from the SRET within a 5-minute time limit.

SRET metrics.—Although the SRET is a simple, straightforward task, it is possible to generate a number of metrics from this task. Our prior work suggests that number of negative words endorsed as self-descriptive, as well as "drift rate" (i.e., ease with which a person reaches a decision about whether or not a word is self-descriptive) for negative words, have very good 1-week test re-test stability, are internally reliable, and are correlated with self-reported depression severity across multiple samples using different depression inventories (Dainer-Best et al., 2018a). Thus, we examined these two metrics in the current study. Notably, word recall metrics were not shown to reliably be associated with depression severity; thus, they are not used here.

Number of negative words endorsed is simply a count of the number of negative adjectives that a person indicated as self-descriptive. Because words were repeated three times, words that were endorsed two or three times were considered endorsed; those that were endorsed fewer times were considered not endorsed. To quantify drift rate, responses on the SRET are examined via a computational model known as the diffusion model. Both reaction time (RTs) and responses were used as input for the drift diffusion model (Ratcliff, 1978; Ratcliff & Rouder, 1998; White et al., 2010), a sequential sampling technique that decomposes responses, RTs, and their distributions into distinct components of decision-making and processing. The diffusion model has been used with the SRET in previous studies (Dainer-Best et al., 2018a; Disner, Shumake, & Beevers, 2017); it assumes that on each trial, evidence is accumulated until one of two response criteria have been met (i.e., whether a given word is self-descriptive).

The relative ease of evidence accumulation is measured by a component referred to as the drift rate. A very positive drift rate indicates that it is easy to categorize such words as self-referential; a drift rate close to zero indicates that it is difficult to categorize such words; and a strongly negative drift rate reflects that evidence accumulation often leads to rejecting a stimulus as self-referential. Consistent with prior work (e.g., Dainer-Best et al., 2018a), drift rate were calculated separately for positive and negative words using the Kolmogorov-Smirnov estimation method. The diffusion model's components were computed with the program fast-dm (A. Voss & Voss, 2007). For more information about how drift rate and other diffusion model components were computed, please see (Ratcliff & McKoon, 2008; Ratcliff, Van Zandt, & McKoon, 1999). For an example of how the diffusion model can be applied to psychopathology research, please see (White, Ratcliff, Vasey, & McKoon, 2010).

The distribution of the two SRET metrics are presented in supplemental materials, section 1.0. The count data for number of negative words endorsed as self-descriptive is highly non-normally distributed, as would be expected for a distribution of counts. The distribution for drift rate for negative words does not strongly deviate from a normal distribution. Given the

distribution of the number of negative words endorsed, we chose statistical learning approaches that either make no assumptions about response distributions (random forest) or appropriately assume the response follows a count distribution (best subsets regression with a negative binomial distribution). Split-half reliability (odd/even trials) using 10,000 bootstrapped samples was excellent for number of negative words endorsed as self-descriptive and very good for drift rate, rho = .91, 95% CI [.89, .94] and rho = .76, 95% CI [.69, .84], respectively.

Emotional dot-probe task with eye tracking: Gaze location and pupil size were measured by a video-based eye-tracker (EyeLink 1000 Plus Desktop Mount; SR Research, Osgoode, ON, Canada) at a rate of 500 Hz. Pupil area was assessed in a centroid pupil-tracking mode with a monocular setup (25-mm lens, 500 Hz sampling), using participants' dominant eye. Stimulus presentation and data acquisition were controlled by E-Prime (Psychology Software Tools, Pittsburgh, PA) and Eyelink software respectively. Stimuli were presented on a 23.6-inch CRT monitor (ViewPixx; VPixx Technologies, Quebec, Canada), at a screen resolution of 1920×1080 pixels (120 Hz refresh rate). Responses were recorded using a Logitech F310 Gamepad (Logitech, Romanel-sur-Morges, Switzerland).

Stimuli consisted of images of happy (12), sad (12), and neutral (24) facial expressions from the Karolinska Directed Emotional Faces database (KDEF; (Lundqvist, Flykt, & Öhman, 1998). Each emotional face was paired with the neutral expression of the same actor. Stimuli (subtending 2° by 4° visual angle) were presented to the left and right sides of the visual field against a grey background (RGB: 110,110,110), with a center-to-center distance of 480 pixels (5° by 4° visual angle). Stimulus-pairs were randomly presented four times within each block to counterbalance their location. To minimize variation of luminance throughout the task, stimulus backgrounds were removed and replaced to match the display background color. Stimuli did not vary in terms of luminance (12.0 cd/m2).

The experiment consisted of 192 trials (96 trials x 2 blocks) lasting approximately 20 min. Participants were seated in an illuminated room (12.0 cd/m2), 60 cm from the computer screen. Ocular dominance was determined using a modified version of the near-far alignment test (Miles, 1930). Before the task, a thirteen-point calibration routine used to map eye position to screen coordinates. Calibration was accepted once there was an overall difference of less than 0.5° of visual angle between the initial calibration and a validation retest.

After completing calibration, participants were informed that the task would soon begin and all instructions would be presented on the monitor. Participants were instructed to view the images naturally. Further, they were also instructed to look at the fixation cross prior to each trial in order to standardize the starting location of their gaze. At the start of each block, a single-point drift correction was conducted to ensure that calibration was consistent throughout the task.

The task began with a series of 20 practice trials, using KDEF images not included in the test stimuli. Each trial began with the appearance of a central fixation cross (FC). Participants were required to maintain their gaze on the central fixation cross (subtending 2°

by 2° visual angle) for a duration window of 500 msec before continuing to the next trial. If central fixation was not detected within 2000 msec, online single-point drift correction procedures were conducted before continuing to the next trial.

Following offset of the fixation cross, a stimulus-pair appeared for 1000 msec. Following stimulus offset, a probe appeared (the letter "Q" or "O") in place of one of the stimuli. The location was randomized with equal frequency. Participants were asked to indicate the location of the probe by pressing the left trigger when seeing "Q" and the right trigger when seeing "O" (reversed for left-handed participants). Latency to respond was recorded. After their response, the probe disappeared, before beginning the next trial. The inter-trial interval was 1000 msec. The two task blocks were completed sequentially with a self-paced break between them.

Dot-probe metrics.—Attention bias using reaction time data was operationalized according to standard conventions (Mogg, Holmes, Garner, & Bradley, 2008) as well as the more recently developed trial level bias scores (TLBS; Zvielli, Vrijsen, Koster, & Bernstein, 2016) using an R package developed in our lab called *itrak* (https://github.com/jashu/itrak). Prior to creating bias scores, we first excluded incorrect responses, responses that were faster than 200ms, longer than 1500ms, or were 3 median absolute deviations beyond the individual's median. Using these standard cut-offs, 9.8% of the trials were identified as invalid.

There are two types of trials on the dot-probe task that are central to quantifying attention bias: congruent and incongruent trials. For congruent trials, the negative stimulus and probe are presented in the same location. For instance, if the negative stimulus is presented on the left, then the probe (O or Q) is also presented on the left. For incongruent trials, the probe and negative stimulus are in opposite locations. For instance, if the negative stimulus is presented on the left, then the probe (O or Q) is presented on the right. A person with a tendency to attend to negative stimuli should be faster to identify the probe location for congruent trials compared to incongruent trials, as the latter requires shifting of attention to the opposite side of visual field.

The traditional mean bias score is simply obtained by taking the mean reaction time of congruent trials and subtracting it from the mean reaction time of incongruent trials. Positive scores are interpreted as an attention bias directed towards negative stimuli, whereas negative scores indicate a bias away from negative stimuli. However, Zvielli et al. (2016) proposed that attention bias is dynamic and fluctuates over time, so using overall mean reaction time to congruent and incongruent trials may miss important information. Rather than using bias scores that are averaged across all trials, they developed TLBS that create difference scores between congruent and incongruent trials that are within close temporal proximity to each other (e.g., within 5 trials of each other). This method creates a series of bias scores throughout the course of the task and putatively captures dynamic changes in attention bias.

The TLBS approach generates several different bias score metrics (e.g., peak bias, bias away from stimuli, variability in bias), although in this article we focus on the TLBS toward sad

stimuli. A TLBS toward sad stimuli is obtained by calculating the mean of all the bias scores that are greater than 0 for trials involving sad stimuli. That is, bias scores where the reaction time for an incongruent trial was longer than the reaction time for a nearby congruent trial indicate a bias towards a sad stimulus at that point in the task. Bias scores are computed throughout the task (creating a time series of fluctuating bias scores) and the mean of all bias scores greater than 0 is used to represent the TLBS towards sad stimuli.

Our approach to creating the TLBS parameters was highly consistent with Zvelli et al. (2016) with one exception. Rather than computing a bias score for every trial by comparing it to the most temporally proximal trial of opposite type, we used a weighted-trials method that calculates the weighted mean of all trials of opposite type, with closer trials weighted more heavily than more distant trials (as a function of the inverse square of trial distance). In other words, each congruent trial is subtracted from the weighted mean of all incongruent trials, and the weighted mean of all congruent trials is subtracted from each incongruent trial.[2]

The *itrak* package used to calculate the TLBS metrics defaults to the weighted method but can also generate metrics using the nearest neighbor method. Results in the current study are very similar using either method. Split-half reliability (odd/even trials) using 10,000 bootstrapped samples was very low for traditional bias, rho = .01, 95% CI [−.14, .17], but very good for TLBS toward sad stimuli, rho = .80, 95% CI [.74, .87].

For eye movement data, we processed the data in a similar fashion. Missing data were very sparse for the eye tracking data in part because eye tracking data were only considered missing if there were no eye tracking data for a given trial. Thus, it is important to consider the total amount of eye tracking data available. This is plotted in the supplemental materials, section 1.1, which indicates that the vast majority of trials had at least 0.5 sec of total fixation time. Importantly, depression severity was not correlated with amount of missing data (r = −.032; see supplemental materials section 1.2). Rather than making an arbitrary decision about a cut-point for an acceptable amount of missing eye tracking data, we created means that are weighted by total fixation time. This should allow us to analyze all available data and reduce the influence of trials where total fixation time is relatively low.

During the dot-probe task, it was noted by research assistants that some participants did not overtly fixate on the face stimuli. This could be expected, as the dot-probe instructions do not explicitly ask participants to view the face stimuli, but instead ask participants to identify whether the subsequent probe is an O or a Q. Thus, we examined the percentage of trials where the participant did not attend to either face stimulus (see supplemental materials

---

[2]The two methods yield highly similar TLBS numbers, but we used the weighted method for two reasons: 1) in the event that a trial of one type is equidistant from two trials of opposite type, the nearest-trial method arbitrarily chooses one over the other; under this circumstance, the mean of the two trials may be a more valid point of comparison. 2) The nearest-trial method frequently double-counts the same incongruent (IT) - congruent (CT) trial subtraction. For example, consider the sequence IT IT CT CT: under the nearest-trial method, the interior IT-CT pair will result in duplicate TLBS calculations for these two trials. This double counting results in brief but frequent periods where the TLBS time series is completely flat. Under the weighted method, these calculations will be non-identical because a trial is not subtracted directly from another single trial, but rather from uniquely weighted means of all trial pairs.

section 1.21). Using the median as the measure of central tendency, participants did not look at either face stimulus on approximately 5% of the trials.

Given that we are interested in measuring attention for negative stimuli, we restricted analyses to participants who consistently directed fixations toward either face stimuli, operationalized as having at least one fixation on a sad or neutral face for 80% or more of their trials. As a result of this criterion, sample size was reduced from 215 (we were completely unable to obtain eye movement data for 3 participants) to 165. In order for our analyses to be consistent across behavioral and eye gaze measures (that is, using data from the same participants across attention bias analyses), we used this same gaze requirement for the behavioral reaction time data as well. Thus, $n = 165$ for eye gaze *and* reaction time metrics on the dot-probe task.

Participants dropped because of poor quality eye tracking data (n = 53) did not differ from those retained ($n = 165$) in terms of depression symptom severity ($t = -0.42$, $df = 90.67$, $p = 0.67$), age ($t = 1.24$, $df = 76.34$, $p = 0.22$), or gender ($\chi^2 = 1.02$, $df = 1$, $p = 0.31$). Among those retained for dot-probe analyses, there was a small correlation between number of valid trials and depression severity ($r = -0.17$, $p = .03$). People with higher depression severity tended to have fewer valid trials. However, all retained participants had at least 70 trials with good data, except for two participants who had 50 and 53 good trials, respectively. Thus, all participants retained for dot-probe analyses had a sufficient number of trials.

Eye movement data were filtered with a heuristic algorithm (Stampe, 1993) and then classified as saccades or fixations based on whether or not the velocity and acceleration of gaze exceeded the following standard thresholds: saccade velocity threshold of 35 degrees / second and saccade acceleration threshold of 9500 degrees / second^2. Any samples not contained in a saccade were by definition a fixation. For each trial, we computed total gaze time for the emotional stimuli, neutral stimuli, and center of the screen (i.e., location of fixation cross). We computed analogous gaze bias metrics as for the behavioral reaction time data: mean gaze bias (i.e., mean gaze time for sad stimuli – mean gaze time for neutral stimuli on a trial by trial basis) and percentage of trials where total gaze time was greater for sad than neutral stimuli. These indices were approximately normally distributed, as shown in supplemental materials section 1.3. Split-half reliability (odd/even trials) using 10,000 bootstrapped samples was low for gaze bias and percentage of trials where gaze was greater for sad than neutral stimuli, rho = .16, 95% CI [−.01, .32] and rho = .14, 95% CI [−.02, .30], respectively. Nevertheless, we decided to retain these dot-probe metrics for analyses given their ubiquity in the literature; however, we are using them with caution (except for TLBS towards sad stimuli) given their poor psychometric characteristics.[3]

Learning Algorithm and Tuning Parameters

We selected a popular machine learning algorithm—the random forest (Breiman, 2001)—as it performs well with samples of this size (< 300). Random forest regression builds multiple

---

[3]We also completed symptom importance analyses with the other TLBS metrics (bias away from sad stimuli, variability in attention bias). Those analyses are presented in supplemental materials section 1.4. Importantly, no reliable symptom predictors of those TLBS metrics were identified.

complex regression trees, each fit using only a random subset of predictors and a random sample of observations, and then averages across these trees to make a consensus prediction of the outcome (negative cognitive bias, in the current study). The random sampling built into the random forest algorithm forces the consideration of combinations of weaker predictors that might otherwise be overlooked, and averaging models fit to different subsets of data has the effect of extracting those predictive relationships that are most robust (i.e., those that are evident no matter how one partitions the data). Because regression trees are built by a series of recursive binary splits (e.g., forming two different predictions based on whether a variable falls above vs. below a cutoff value), random forests have several desirable features: they automatically capture interaction and nonlinear effects, and their performance is not sensitive to distributional assumptions or the presence of outliers.

Most machine learners (e.g., gradient boosting machines) require "tuning" of several hyperparameters to achieve good predictive performance, which is typically done by trying out hundreds of different combinations of hyperparameter values by fitting hundreds of models and selecting the one that achieves the best cross-validation performance. As noted by other researchers (Cawley & Talbot, 2010; Varma & Simon, 2006), overfitting from model selection will be most severe when the sample of data is small and the number of hyperparameters is large. For small data sets, Cawley and Talbot (2010) recommend avoiding hyperparameter tuning altogether by using an ensemble approach, such as random forests, that performs well without tuning.

While random forests have a few hyperparameters that could be tuned in principle, the default values for these typically work very well in practice (Breiman, 2001; Hastie et al., 2009; James, Witten, Hastie, & Tibshirani, 2013). Thus, we used the following model parameters: 1) The number of trees was set to 500; 2) The depth to which trees are grown was set to terminate if the additional split resulted in fewer than 5 observations in the terminal node; 3) The number of variables that are randomly sampled when determining a split was set to the recommended value for regression problems, which is 1/3 of the candidate predictors.

Alternative Learning Algorithms

To confirm that results were not confined to a particular statistical approach (Munafò & Smith, 2018), we conducted parallel analyses using a statistical learning model with different assumptions and techniques. We chose linear regression models regularized by elastic net (Zou & Hastie, 2005). These are fit in the same way as ordinary linear models except, instead of just finding the beta coefficients that minimize the residual sum of squares, the fitting procedure also shrinks the overall magnitude of the beta coefficients. This coefficient shrinkage makes it possible to fit stable models using all of the candidate predictors, even if some of them are highly correlated with one another, without overfitting the models to the training data. The amount of shrinkage is governed by the elastic net penalty, which is a blend of two penalties: the L1 'lasso' penalty (sum of absolute values of all coefficients) and the L2 'ridge' penalty (sum of squared values of all coefficients). Two parameters, which are tuned using cross-validation, control how this penalty is applied: α,

the proportion of L1 vs. L2 regularization in the penalty, and λ, the factor by which the penalty is multiplied.

For each outcome, we searched over 100 possible values of λ (autogenerated by the model fitting program) and 3 possible values of α: 0.01 (favoring the inclusion of most variables), 0.99 (favoring sparsity), and 0.5 (an equal mix of both L1 and L2). To obtain an unbiased estimate of test error, we used a nested cross-validation procedure (Varma & Simon, 2006). That is, an outer cross-validation was utilized for estimating test error and an inner cross-validation (nested within the training folds of the outer cross-validation) was used for tuning the optimal number of predictors. The final model retained had the smallest number of parameters that is within one standard error of the model with the lowest cross-validation error (the "1-SE rule"). This 1-SE rule approach tends to produce sparse models but results are expected to generalize well to new data compared to more permissive models.

Finally, the number of negative words endorsed as self-descriptive on the SRET is a count variable. Such variables often have many zeros or very low scores. Linear models, which assume that prediction errors are normally distributed, cannot fit these data very well. A potentially better way to fit such data is with a negative binomial regression, a generalized linear model which accounts for overdispersion—that is, when the variance of the distribution is greater than the mean. Thus, when examining count-related outcomes, we will use best subset regression using a negative binomial distribution.

This approach tests all possible combination of predictors and then chooses the model that performs best under cross-validation. Notably, because these models are very computationally expensive, for this alternative model only, we will limit the number of predictors in these models to no more than 10 predictors. For example, as a model goes from 10 to 20 predictors, it goes from fitting 1,024 models to fitting 1,048,576 models. Thus, this algorithm rapidly becomes impractical as one moves beyond 10 predictors and, for most computers, outright impossible beyond 40 predictors. Nevertheless, we believe that capping the search to the 10 best depression symptoms predictors should still identify the most important symptoms.

Prediction Metrics and Cross-Validation

We used 10-fold cross validation to estimate a *predictive* R^2 (R^2_{pred}), which is the fraction of variance in previously unseen data that the model is expected to explain. This allows for examining how well the model performs on cases that it was not trained on. Cross-validation therefore provides an approximation of how well the model would generalize to new data, as a model's cross-validated performance will typically converge with performance in out-of-sample data (Yarkoni & Westfall, 2017). The 10-fold cross validation was repeated 10 times using different randomized partitions of the data. Within each repetition, 10 models were fit, each trained to 90% of the data and used to predict the outcomes for the 10% of cases that were withheld. An R^2_{pred} was then calculated based on the residual errors of the holdout predictions and then averaged across the 10 repetitions.

Calculation of Variable Importance Scores

For the random forest, predictor importance was quantified as the percent mean increase in the residual squared prediction error on the out-of-bag (OOB) samples when a given variable was randomly permuted. In other words, if permuting a variable substantially worsens the quality of the prediction, then it is important; if permutation has no impact on the prediction, then it is not important. For the elastic net, variable importance was quantified as the absolute value of the standardized regression coefficient for each predictor. To facilitate comparison between predictors and the different statistical methods, importance scores were scaled so that the importance scores of all variables sum to 1.

All analyses were implemented in R (version 3.5). Our code made extensive use of the *tidyverse* (Wickham, 2018) packages *dplyr*, *purrr,* and *tidyr*. Figures were generated using the packages *ggplot2* (Wickham, 2009) and *gridExtra* (Auguie, 2017). The *randomForest* (Liaw & Wiener, 2002) and *glmnet* (Friedman, Hastie, & Tibshirani, 2010) packages were used to implement the machine learning ensembles. We also used two R packages developed in-house, *itrak* and *beset* (https://github.com/jashu). All analysis code and results reported in this article are presented in the supplemental materials.

Procedure

Participants first completed screening via online questionnaires. Participants not excluded following online screening were then contacted, inclusion/exclusion criteria were confirmed, study procedures were reiterated, and an in-person appointment was scheduled. Written informed consent was obtained during this in-person session. Participants completed a four-hour, in-lab appointment. During this appointment, participants completed the dot-probe, SRET, and BDI-II. Other self-report, task-based, biological, and neurophysiological data were collected as part of this protocol; however, those data will be examined in other reports. Study data were collected and managed using an electronic data capture tool hosted at The University of Texas (Harris et al., 2009). The Institutional Review Board at the University of Texas at Austin approved all study procedures.

Results

SRET: Drift Rate

Random forest.—Using 10-fold cross validation repeated 10 times with all 21 BDI-II symptoms entered as predictors, out-of-sample R^2_{pred} = 40.2% (SE = 0.054, Min = 0.392, Max = 0.423) for drift rate for negative words. As seen in Figure 1A, the most important symptom was sadness. The relative importance score for sadness was nearly 2 times greater than any other symptom. Other important symptoms included self-dislike, pessimism, worthlessness, punishment, and indecision. The remaining symptoms made relatively small contributions to the prediction of drift rate for negative words.

Figure 1B shows partial dependence plots, which indicate the relationship between each symptom and drift rate for negative words, holding all other predictors constant at their mean. This figure reveals that individuals who report no sadness (i.e., report a 0 on this item) have a low drift rate, indicating it was relatively easy to reject negative words as self-

descriptive. Participants with sadness scores greater than 0 had a relatively higher drift rate. These individuals still tended to endorse negative words as not self-descriptive, but with much less efficiency. Drift rate then remained stable at higher levels of sadness, suggesting that drift rate did not become more pronounced as sadness increased. Self-dislike showed a similar pattern, although it had a more modest impact on drift rate. Pessimism and worthlessness had a weak linear relationship with drift rate. The relationship between drift rate and many of the other symptoms was relatively flat, indicating the absence of a strong relationship.

Based on visual inspection of the variable importance scores in Figure 1A, symptoms after the first seven appeared to make a minimal contribution to the prediction of drift rate for negative words. Thus, our final model compared the variance explained by the full model containing all depression symptoms to a model with only the top seven symptoms. For this reduced model, out-of-sample R^2_{pred} = 39.1% (SE = 0.046, Min = 0.370, Max = 0.406) for drift rate for negative words. A comparison of R^2_{pred} between the full and reduced models indicated that the full model only explained 1.2% (95% CI [−0.075, 0.054]) more variance in drift rate.

Alternative model.—Elastic net regression was also used to identify the most important symptom predictors for drift rate for negative words. Using 10-fold cross validation repeated 10 times and a one standard error criterion to select the best model, out-of-sample R^2_{pred} = 34.1% (SE = 0.027, Min = 0.322, Max = 0.353) for drift rate for negative words. This model retained 13 symptoms (the remaining 8 symptoms had standardized regression coefficients that were not different from 0). As seen in supplemental materials section 2.0, the most important symptom was sadness. Other important symptoms included self-dislike, punishment, and pessimism. Guilt, indecision, and loss of pleasure were also important predictors. The remaining symptoms made relatively small (but non-zero) contributions to the prediction of drift rate for negative words. The partial dependence plots (supplemental materials section 2.0) indicate that the linear association between each symptom and drift rate for negative words generally decreased as symptom importance decreased.

A comparison of R^2_{pred} for the full random forest and elastic net models indicated that the random forest explained 6.1% (95% CI [−0.027, 0.140]) more variance in drift rate; however, the 95% CI for this difference contained 0, suggesting variance explained in the models were not significantly different from one another. Indeed, their predictions were highly consistent with each other, r = .88 (see supplemental materials, section 2.1).

SRET: Endorsements of negative words as self-descriptive.

Random Forest.—Using 10-fold cross-validation repeated 10 times with all 21 BDI-II symptoms entered as predictors, out-of-sample R^2_{pred} = 45.3% (SE = 0.056, Min = 0.433 Max = 0.470) for number of negative words endorsed. As seen in Figure 2A, the most important symptom was sadness. Other important symptoms included punishment,

worthlessness, self-dislike, guilt, and suicide. The remaining symptoms made relatively small contributions to the prediction of endorsements of negative words.

Figure 2B shows partial dependence plots for each of the important symptoms. Similar to drift rate, individuals with no sadness endorsed few negative words as self-descriptive. There was then a strong increase in number of negative words endorsed if any sadness was endorsed (i.e., 1 or greater on that item). Number of negative words endorsed continued to increase as sadness increased. The symptoms of worthlessness, self-dislike, and suicide had a similar relationship, although the effect was less pronounced. Punishment, guilt, and sleep change appeared to have a linear relationship with number of negative words endorsed as self-descriptive, such that as symptom severity increased so too did endorsements of negative words. The relationship between drift rate and many of the other symptoms was relatively flat, indicating the absence of a strong relationship.

Based on visual inspection of the variable importance scores, the symptoms after the first six appeared to make a minimal contribution to the prediction of endorsements of negative words. Thus, our final model compared the variance explained by the full model containing all depression symptoms to a model with only the top six symptoms. For this reduced model, out-of-sample R^2_{pred} = 46.9% (SE = 0.055, Min = 0.447, Max = 0.490) for drift rate for negative words. A comparison of R^2_{pred} between the full and reduced models indicated that full model explained 1.6% (95% CI [−0.032, 0.066]) *less* variance in negative endorsements.

Alternative model.—Best subset regression with a negative binomial distribution was used to identify the most important symptom predictors for number of negative words endorsed. Using 10-fold cross validation repeated 10 times, out-of-sample R^2_{pred} = 35.0% for number of negative words endorsed as self-descriptive, log likelihood = −609.6 on 9 *df*, AIC = 1237.1. This model retained 7 symptoms. The symptoms with the strongest standardized regression coefficients included sadness (b = 0.33), guilt (b = 0.12), punishment (b = 0.16), self-dislike (b = .11), suicidal thoughts (b = 0.21), indecision (b = .07) and change in sleep (b = .16). The cross-validated model that was within one standard error of the best model explained 29.0% of the out-of-sample variance, log likelihood = −624.8 on 4 *df*, AIC = 1257.6, and retained two symptoms, sadness (b = .60) and feelings of punishment (b = .25). See supplemental materials section 2.2 for more detail.

Dot-Probe: Reaction Time Metrics

Random Forest.—Using 10-fold cross-validation repeated 10 times with all 21 BDI-II symptoms entered as predictors, out-of-sample R^2_{pred} was < 0% (Mean = −0.124, SE = 0.064, Min = −0.174 Max = −0.095) for traditional reaction time attention bias for negative stimuli. Thus, no depression symptoms were reliably associated with the traditional reaction time metric for negative attention bias.[4]

---

[4]Note that unlike the train R^2, which is bounded between 0 and 1, it is possible for the cross-validation R^2 to be negative, as is the case when the number of predictors is 0, which corresponds to the null, or intercept-only, model, which is simply the mean of the train sample. Given that the train sample's mean cannot predict the test sample better than the test sample's own mean, the predictive R^2 for the null model can never be positive. More generally, the predictive R^2 will be negative whenever a model is bad (i.e., instead of

The same analysis was performed for TLBS towards sad stimuli. Using 10-fold cross-validation repeated 10 times with all 21 BDI-II symptoms entered as predictors, out-of-sample R^2_{pred} was < 0% (Mean = −0.033, SE = 0.077, Min = −0.091, Max = −0.008). Thus, no depression symptoms were reliably associated with the TLBS metric for bias towards negative stimuli.

Alternative models.—Elastic-net regression was performed for both the traditional dot-probe and TLBS towards sad stimuli metrics and, in both cases, the models explained less than 0% of the out-of-sample variance and no reliable symptom predictors were identified. See supplemental materials section 3.0 for more information.

Dot-Probe: Eye Gaze Metrics

Random Forest.—Two parallel attention bias metrics using eye gaze were examined next. Using 10-fold cross-validation repeated 10 times with all 21 BDI-II symptoms entered as predictors, out-of-sample R^2_{pred} was < 0% (Mean = −0.152, SE = 0.052, Min = −0.183, Max = −0.125). Thus, no depression symptoms were reliably associated with the eye gaze metric for negative attention bias. A similar pattern was observed for percentage of trials where eye gaze was directed towards sad stimuli: using 10-fold cross-validation repeated 10 times, out-of-sample R^2_{pred} was < 0% (Mean = −0.114, SE = 0.058, Min = −0.159, Max = −0.085).

Alternative models.—Elastic-net regression was performed for both the eye gaze bias and percentage of trials where eye gaze was directed towards sad stimuli. In both cases, the models explained less than 0% of the out-of-sample variance and no reliable symptom predictors were identified. See supplemental materials section 3.1 for more information.

Bivariate correlations.—Given the lack of cross-validated associations between reaction time or eye gaze metrics and depression symptoms, our final analysis examined bivariate correlations among attention bias metrics and depression symptom total score. As seen in Table 2, the BDI-II total score was correlated with the TLBS measure of attention bias towards sad stimuli ($r = .22$, $p = .001$). A linear regression with BDI-II total score as a predictor and TLBS bias towards sad stimuli as the outcome indicated that depression total score explained 4.8% of the variance in TLBS towards sad stimuli $F(1, 163) = 8.294$, $p = .004$. Using 10-fold cross-validation repeated 10 times, out-of-sample $R^2_{pred} = 2.8\%$ (SE = .039, Min = 0.009, Max = 0.04). Thus, these data suggest that the bivariate correlation between TLBS bias towards sad stimuli and total score depression severity is relatively modest but this relatively small association may generalize to new data.[5]

---

reducing uncertainty when making predictions, the model adds to it). This can also happen with non-null models if they include too few observations and/or too many useless predictors.

[5]Although analyses focus on negative cognitive bias, results for positive cognitive biases are presented in the supplemental materials, section 4.0. In general, results for positive cognitive biases were consistent with associations observed between depression symptoms and negative cognitive biases (only the associations were in the opposite direction).

Discussion

The current study is among the first to examine associations between different BDI-II symptoms of depression in adults recruited from the community and two common negative cognitive biases: self-referent negative processing and biased attention towards sad stimuli. Several general themes emerged from the findings. Across all importance analyses, at least half of the measured depression symptoms contributed little or not at all to the prediction of negative self-referent processing. Some symptoms, such as sadness and self-dislike, were much more important to the prediction than others (e.g., failure, appetite). Further, there were no reliable, cross-validated, symptom predictors of negative attention bias, measured with reaction time or eye tracking. Importantly, all analyses used more than one machine learning method in combination with 10-fold cross-validation repeated 10 times to examine the generalizability of the findings.

The most important symptoms for the prediction of drift rate (i.e., the ease with which a person could endorse or reject a word as self-descriptive) for negative words using random forest machine learning included: sadness, self-dislike, pessimism, worthlessness, punishment, indecision, and agitation. Notably, the elastic-net model, which tends to produce sparser models, found that sadness, self-dislike, punishment, and pessimism were the most important symptoms, followed by guilt, indecision, and loss of pleasure. Thus, even though these methods use different statistical approaches, there was considerable agreement between the symptoms identified as most important. It appears that sadness, self-dislike, pessimism, feelings of punishment, and indecision may be particularly relevant to the determination of whether a negative word is self-descriptive or not.

Number of negative words endorsed as self-descriptive is another related but not redundant measure of negative self-referent processing (Dainer-Best et al., 2018a). Random forest models indicated once again that sadness, punishment, worthlessness, self-dislike, and guilt were among the most important symptoms for this model. In contrast, many symptoms appeared to have a relatively small (or non-existent) contribution, such as crying, low libido, loss of energy, feeling like a failure, fatigue, loss of interest, being self-critical, and change in appetite. A similar set of findings emerged for the elastic net analyses as well. Thus, once again, these findings highlight the importance of studying individual symptoms of depression.

It is noteworthy that many of the most important symptoms for predicting negative self-referent processing have also been identified as the most central symptoms in the depression symptom network. Specifically, when strength of symptoms is used to identify central symptoms (i.e., individual symptoms that are most strongly associated with other symptoms), network analyses in adult clinical samples have identified sadness (Beard et al., 2016; Fried, Epskamp, Nesse, Tuerlinckx, & Borsboom, 2016a; Santos, Fried, Asafu-Adjei, & Ruiz, 2017), loss of interest (Bringmann, Lemmens, Huibers, Borsboom, & Tuerlinckx, 2014), and fatigue (van Borkulo et al., 2015) as central symptoms. Further, a recent network analysis in a non-selected sample of adolescents found that sadness, self-dislike, pessimism, and loneliness were the most central symptoms of adolescent depression (Mullarkey, Marchetti, & Beevers, 2018). It is notable that some of the most central symptoms of

depression identified in other samples also appear to be important predictors of self-referent negative processing.

Further, in addition to sadness, several of the most important symptom predictors of negative self-referent processing are consistent with Beck's cognitive model of depression (Beck, 1967): self-dislike, pessimism, and worthlessness. Some of these connections are straightforward, as it follows that someone who has high self-dislike would be more likely to rapidly endorse negative adjectives as self-descriptive than a person who is low in self-dislike. Although studies have linked the related construct of a pessimistic attributional style to negative self-referent processing (Alloy et al., 2012; Molz Adams, Shapero, Pendergast, Alloy, & Abramson, 2014), we could find no studies that have specifically examined associations between worthlessness and self-referent negative processing.

Another important finding from the current study was that negative attention bias, measured with reaction time or eye tracking, was not reliably associated with any of the depression symptoms. This finding is consistent with prior work documenting that negative attention bias (in contrast with memory bias) measured with different approaches was not strongly correlated with symptom severity across four studies involving 463 adult participants (Marchetti et al., 2018). Similarly, in a study that used best subsets regression to identify SRET and dot-probe metrics associated with concurrent depression severity, number of positive and negative words endorsed from the SRET were the best predictors of total score depression severity. Negative attention bias was not strongly related to concurrent depression severity ($r = .20$, similar to the effect observed in the current study); however, the single best predictor of depression symptom change, assessed weekly over a six-week period, was negative attention bias measured with eye tracking (Disner et al., 2017).

The current study and past findings documenting weak associations between depression severity and negative attention bias stand in contrast to meta-analyses documenting group differences between those with and without current Major Depressive Disorder (MDD). For instance, a meta-analysis of 29 studies found a medium effect size difference between depressed and non-depressed participants for negative attention bias, particularly when measured with the dot-probe task. Similarly, a second meta-analysis of eye movements indicates that attention bias in depression is characterized by reduced maintenance of gaze towards positive stimuli and increased maintenance of gaze towards negative stimuli (Armstrong & Olatunji, 2012).

How does one reconcile these discrepant findings? One possibility is that negative attention bias does not track closely with concurrent symptom severity. That is, attention bias may differ between groups with and without MDD, but it is not closely linked to more subtle differences in depression symptoms, even when symptoms are considered in isolation, as in the current study. This may be due in part to measurement issues, as attention bias is notoriously difficult to measure with high reliability (Rodebaugh et al., 2016), even when using novel assessments of attention bias that take into account moment-by-moment shifts in attention (Kruijt, Field, & Fox, 2016; Zvielli et al., 2016). Indeed, in the current study, split-half reliability was poor for traditional attention bias metrics measured with reaction time or eye gaze. Further, the lack of a reliable association between attention bias and depression

symptom severity may also partly stem from the possibility that findings from studies of groups (e.g., depressed versus non-depressed) may not generalize well to individual differences (e.g., depression severity) (Fisher, Medaglia, & Jeronimus, 2018).

In an effort to examine generalization of the current study's findings, all results were cross-validated, which provides an estimate of generalization performance (Yarkoni & Westfall, 2017). Thus, we are relatively confident that sadness, self-dislike, pessimism, feelings of punishment, and indecision may be particularly important depression symptoms for negative self-referent processing. It was also notable that although there was a significant bivariate correlation between depression severity and TLBS towards sad stimuli, that effect was reduced almost in half during cross-validation (4.8 to 2.8% variance explained), suggesting that the bivariate correlation may be sample specific but unlikely to robustly generalize to new data. Nevertheless, we believe this is illustrative of how useful this technique may be for clinical science and strongly encourage others to cross-validate their findings in the future, which may help the field move towards a more cumulative and replicable clinical science (Munafò et al., 2017; Tackett et al., 2017).

Finally, results from this study highlight several important measurement issues. First, many symptoms typically measured in depression inventories do not appear to be strongly correlated with important cognitive processes implicated in the maintenance of depression. Use of a sum score across heterogeneous symptoms may be adding unwanted noise to the depression symptom assessment in these circumstances (Fried & Nesse, 2015a). Investigators may consider taking a symptom-level approach or using sum scores for subsets of depression symptoms most relevant to the construct of interest.

Further, it seems highly likely that other etiological or maintaining factors, in addition to negative cognitive biases, may also have differential associations with depression symptoms. For instance, efforts to identify the genetic etiology of depression may be undermined to a degree by searching for genetic variants that are associated with a diagnosis that involves heterogeneous depression symptoms that can potentially have more than a thousand different symptom configurations (Fried & Nesse, 2015a) rather than searching for variants associated with specific depression symptoms (Flint & Kendler, 2014). Indeed, there is preliminary evidence that depression symptoms may differ in heritability (Pearson et al., 2016). Future work on the etiology and maintenance of depression could potentially enhance model specificity by examining whether or not theoretical models are consistent across different symptom dimensions. Such research may also be informative for subsequent prevention or intervention development.

Second, many depression symptoms measured in widely used inventories may overlap with the constructs that are being correlated with these depression inventories. For instance, self-dislike was strongly associated with the tendency to endorse negative words as self-descriptive. This runs the risk of tautology—symptoms are correlated with the cognitive maintaining factor because they are measuring the same construct. The depression symptom of sadness, by far the most important depression symptom, does not suffer from this issue as it is conceptually distinct from the cognitive maintaining factors. Sadness may therefore be a

better symptom outcome for studies investigating maintaining factors that are difficult to distinguish from some of the measured depression symptoms.

Third, in the current study, symptoms were measured with single items. This is not psychometrically ideal, as the depression scale used in the current study (and most other research) was developed and typically validated based on the sum score of all the items (Beck et al., 1988). Future studies should consider using inventories that measure distinct symptoms with multiple items, such as the Inventory of Depression and Anxiety Symptoms (Watson et al., 2008). Using such assessments should help improve reliability of symptom measurement, but it will still be important to avoid generating sum scores across heterogeneous symptoms (Fried & Nesse, 2015b).

The current study has several limitations that should be mentioned. This study was limited to examining associations between depression symptoms and two putative cognitive maintaining factors: self-referent cognition and attention bias. Future studies should examine associations between specific depression symptoms and other important cognitive processes, such as interpretation biases (Everaert, Podina, & Koster, 2017), autobiographical memory (Hallford, Austin, Takano, & Raes, 2018), or rumination (Whitmer & Gotlib, 2013). The current study also relied on one commonly used self-report assessment of depression to estimate associations between cognitive maintaining factors and depression symptoms. Different questionnaires and different modalities (e.g., interview) may have arrived at different conclusions. Finally, the sample was limited to adults who self-identified as European ancestry (for purposes of future genetic analyses), were not taking psychotropic medication, and were not screened out for higher levels of smoking, psychotic symptoms, or alcohol intake. Our findings may not generalize well to people with different demographic features; indeed, it is possible that exclusionary criteria may have limited the number of somatic symptoms observed in the current study and increased associations between negative cognitive bias, affective, and cognitive symptoms of depression. Additional work is needed to address these issues.

Despite these limitations, this study provides a number of interesting findings. First, many depression symptoms that are typically assessed in depression inventories (e.g., sleep, appetite, feeling like a failure) were not important predictors of negative self-referent processing. The most important depression symptom for negative self-referent processing was sadness. Other important symptoms included self-dislike, pessimism, feelings of punishment, and indecision. Further, cross-validated results found no reliable depression symptom predictors of negative attention bias, regardless if attention bias was measured with reaction time or eye tracking. These findings align with recent evidence suggesting that negative attention bias as measured with the dot-probe task may not track with concurrently measured depression symptom severity (Marchetti et al., 2018). Nevertheless, this study highlights the importance of cross-validating results and that negative self-referent processing (as measured with the SRET) may be most relevant for the maintenance of specific symptoms of depression, such as sadness. Future studies that examine self-referent processing in depression may want to focus on these key symptoms, so that research can more precisely identify the mechanisms that maintain specific symptoms of depression.

Doing so may allow for the development of theories and treatments that more effectively understand and treat specific symptoms of depression.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

References

Alloy LB, & Clements CM (1998). Hopelessness Theory of Depression: Tests of the Symptom Component. Cognitive Therapy and Research, 22(4), 303–335. 10.1023/A:1018753028007

Alloy LB, Abramson LY, Murray LA, Whitehouse WG, & Hogan ME (1997a). Self-referent information-processing in individuals at high and low cognitive risk for depression. Cognition & Emotion, 11(5–6), 539–568. 10.1080/026999397379854a

Alloy LB, Abramson LY, Raniere D, & Dyller IM (1999). Research methods in adult psychopathology In Kendall PC, Butcher JN, & Holmbeck GN (Eds.), Handbook of research methods in adult psychopathology (pp. 466–498). Hoboken, NJ: John Wiley & Sons Inc.

Alloy LB, Black SK, Young ME, Goldstein KE, Shapero BG, Stange JP, et al. (2012). Cognitive vulnerabilities and depression versus other psychopathology symptoms and diagnoses in early adolescence. Journal of Clinical Child & Adolescent Psychology, 41(5), 539–560. 10.1080/15374416.2012.703123 [PubMed: 22853629]

Alloy LB, Just N, & Panzarella C (1997b). Attributional Style, Daily Life Events, and Hopelessness Depression: Subtype Validation by Prospective Variability and Specificity of Symptoms. Cognitive Therapy and Research, 21(3), 321–344. 10.1023/A:1021878516875

Armstrong T, & Olatunji BO (2012). Eye tracking of attention in the affective disorders: A meta-analytic review and synthesis. Clinical Psychology Review, 32(8), 704–723. 10.1016/j.cpr.2012.09.004 [PubMed: 23059623]

Auguie B (2017). Miscellaneous Functions for "Grid" Graphics [R package gridExtra version 2.3]. Retrieved from https://cran.r-project.org/web/packages/gridExtra/index.html

Beard C, Millner AJ, Forgeard MJC, Fried EI, Hsu KJ, Treadway MT, et al. (2016). Network analysis of depression and anxiety symptom relationships in a psychiatric sample. Psychological Medicine, 46(16), 3359–3369. 10.1017/S0033291716002300 [PubMed: 27623748]

Beard C, Rifkin LS, & Bjorgvinsson T (2017). Characteristics of interpretation bias and relationship with suicidality in a psychiatric hospital sample. Journal of Affective Disorders, 207, 321–326. 10.1016/j.jad.2016.09.021 [PubMed: 27741470]

Beck AT (1967). Depression: Clinical, Experimental, and Theoretical Aspects. New York: Harper & Row.

Beck AT, & Bredemeier K (2016). A Unified Model of Depression: Integrating Clinical, Cognitive, Biological, and Evolutionary Perspectives. Clinical Psychological Science, 1–24. 10.1177/2167702616628523

Beck AT, Steer RA, & Brown GG (1996). Manual for the Beck Depression Inventory-II. San Antonio, TX: Psychological Corporation.

Beck AT, Steer RA, & Carbin MG (1988). Psychometric properties of the Beck Depression Inventory: Twenty-five years of evaluation. Clinical Psychology Review, 8(1), 77–100.

Beevers CG, Clasen PC, Enock PM, & Schnyer DM (2015). Attention bias modification for major depressive disorder: Effects on attention bias, resting state connectivity, and symptom change. Journal of Abnormal Psychology, 124(3), 463–475. 10.1037/abn0000049 [PubMed: 25894440]

Breiman L (2001). Random Forests. Machine Learning, 45, 5–32. 10.1007/978-94-010-0664-4_8

Bringmann LF, Lemmens LHJM, Huibers MJH, Borsboom D, & Tuerlinckx F (2014). Revealing the dynamic network structure of the Beck Depression Inventory-II. Psychological Medicine, 45(04), 747–757. 10.1017/S0033291714001809 [PubMed: 25191855]

Cawley GC, & Talbot NLC (2010). On over-fitting in model selection and subsequent selection bias in performance evaluation. Journal of Machine Learning Research, 11(Jul), 2079–2107.

Cuthbert BN (2014). Translating intermediate phenotypes to psychopathology: the NIMH Research Domain Criteria. Psychophysiology, 51(12), 1205–1206. 10.1111/psyp.12342 [PubMed: 25387702]

Dainer-Best J, Lee HY, Shumake JD, Yeager DS, & Beevers CG (2018a). Determining optimal parameters of the self-referent encoding task: A large-scale examination of self-referent cognition and depression. Psychological Assessment. 10.1037/pas0000602

Dainer-Best J, Shumake JD, & Beevers CG (2018b). Positive imagery training increases positive self-referent cognition in depression. Behaviour Research and Therapy, 111, 72–83. 10.1016/j.brat. 2018.09.010 [PubMed: 30321746]

Derry PA, & Kuiper NA (1981). Schematic processing and self-reference in clinical depression. Journal of Abnormal Psychology, 90(4), 286–297. 10.1037/0021-843X.90.4.286 [PubMed: 7264058]

Disner SG, Beevers CG, Haigh EAP, & Beck AT (2011). Neural mechanisms of the cognitive model of depression. Nature Reviews Neuroscience, 12(8), 467–477. 10.1038/nrn3027 [PubMed: 21731066]

Disner SG, Shumake JD, & Beevers CG (2017). Self-referential schemas and attentional bias predict severity and naturalistic course of depression symptoms. Cognition & Emotion, 31(4), 632–644. 10.1080/02699931.2016.1146123 [PubMed: 26901406]

Doost HTN, Moradi AR, Taghavi MR, Yule W, & Dalgleish T (1999). The development of a corpus of emotional words produced by children and adolescents. Personality and Individual Differences, 27(3), 433–451. 10.1016/S0191-8869(98)00253-0

Dozois DJA, Bieling PJ, Patelis-Siotis I, Hoar L, Chudzik S, McCabe K, & Westra HA (2009). Changes in self-schema structure in cognitive therapy for major depressive disorder: a randomized clinical trial. Journal of Consulting and Clinical Psychology, 77(6), 1078–1088. 10.1037/a0016886 [PubMed: 19968384]

Dozois DJ, & Dobson KS (2001). Information processing and cognitive organization in unipolar depression: Specificity and comorbidity issues, 110(2), 236–246. 10.1037//0021-843X

Everaert J, Podina IR, & Koster EHW (2017). A comprehensive meta-analysis of interpretation biases in depression. Clinical Psychology Review, 58, 33–48. 10.1016/j.cpr.2017.09.005 [PubMed: 28974339]

Fisher AJ, Medaglia JD, & Jeronimus BF (2018). Lack of group-to-individual generalizability is a threat to human subjects research. Proceedings of the National Academy of Sciences of the United States of America, 127, 201711978–10. 10.1073/pnas.1711978115

Flint J, & Kendler KS (2014). The Genetics of Major Depression. Neuron, 81(3), 484–503. 10.1016/j.neuron.2014.01.027 [PubMed: 24507187]

Fried EI (2017). The 52 symptoms of major depression: Lack of content overlap among seven common depression scales. Journal of Affective Disorders, 208, 191–197. 10.1016/j.jad.2016.10.019 [PubMed: 27792962]

Fried EI, & Nesse RM (2014). The impact of individual depressive symptoms on impairment of psychosocial functioning. PLoS ONE, 9(2), e90311 10.1371/journal.pone.0090311 [PubMed: 24587318]

Fried EI, & Nesse RM (2015a). Depression is not a consistent syndrome: An investigation of unique symptom patterns in the STAR*D study. Journal of Affective Disorders, 172, 96–102. 10.1016/j.jad.2014.10.010 [PubMed: 25451401]

Fried EI, & Nesse RM (2015b). Depression sum-scores don't add up: why analyzing specific depression symptoms is essential. BMC Medicine, 13(1), 72 10.1186/s12916-015-0325-4 [PubMed: 25879936]

Fried EI, Epskamp S, Nesse RM, Tuerlinckx F, & Borsboom D (2016a). What are "good" depression symptoms? Comparing the centrality of DSM and non-DSM symptoms of depression in a network analysis. Journal of Affective Disorders, 189, 314–320. 10.1016/j.jad.2015.09.005 [PubMed: 26458184]

Fried EI, Nesse RM, Zivin K, Guille C, & Sen S (2014). Depression is more than the sum score of its parts: individual DSM symptoms have different risk factors. Psychological Medicine, 44(10), 2067–2076. 10.1017/S0033291713002900 [PubMed: 24289852]

Fried EI, van Borkulo CD, Epskamp S, Schoevers RA, Tuerlinckx F, & Borsboom D (2016b). Measuring depression over time … Or not? Lack of unidimensionality and longitudinal measurement invariance in four common rating scales of depression. Psychological Assessment, 28(11), 1354–1367. 10.1037/pas0000275 [PubMed: 26821198]

Friedman J, Hastie T, & Tibshirani R (2010). Regularization paths for generalized linear models via coordinate descent. Journal of Statistical Software, 33(1), 1–22. [PubMed: 20808728]

Geschwind DH, & Flint J (2015). Genetics and genomics of psychiatric disease. Science, 349(6255), 1489–1494. 10.1126/science.aaa8954 [PubMed: 26404826]

Gibb BE, Alloy LB, Abramson LY, Beevers CG, & Miller IW (2004). Cognitive vulnerability to depression: a taxometric analysis. Journal of Abnormal Psychology, 113(1), 81–89. 10.1037/0021-843X.113.1.81 [PubMed: 14992660]

Goldstein BL, Hayden EP, & Klein DN (2015). Stability of self-referent encoding task performance and associations with change in depressive symptoms from early to middle childhood. Cognition & Emotion, 29(8), 1445–1455. 10.1080/02699931.2014.990358 [PubMed: 25530070]

Gotlib IH, & Joormann J (2010). Cognition and depression: current status and future directions. Annual Review of Clinical Psychology, 6, 285–312. 10.1146/annurev.clinpsy.121208.131305

Hall FS, Der-Avakian A, Gould TJ, Markou A, Shoaib M, & Young JW (2015). Negative affective states and cognitive impairments in nicotine dependence. Neuroscience and Biobehavioral Reviews. 10.1016/j.neubiorev.2015.06.004

Hallford DJ, Austin DW, Takano K, & Raes F (2018). Psychopathology and episodic future thinking: A systematic review and meta-analysis of specificity and episodic detail. Behaviour Research and Therapy, 102, 42–51. 10.1016/j.brat.2018.01.003 [PubMed: 29328948]

Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, & Conde JG (2009). Research electronic data capture (REDCap)—A metadata-driven methodology and workflow process for providing translational research informatics support. Journal of Biomedical Informatics, 42(2), 377–381. 10.1016/j.jbi.2008.08.010 [PubMed: 18929686]

Hastie T, Tibshirani R, & Friedman J (2009). The Elements of Statistical Learning. New York, NY: Springer Science & Business Media 10.1007/978-0-387-84858-7

James G, Witten D, Hastie T, & Tibshirani R (2013). An introduction to statistical learning (Vol. 103). New York, NY: Springer New York 10.1007/978-1-4614-7138-7

Kamholz SL (2004). Pulmonary and cardiovascular consequences of smoking. The Medical Clinics of North America, 88(6), 1415–30– ix–x. 10.1016/j.mcna.2004.07.008 [PubMed: 15464105]

Kruijt A-W, Field AP, & Fox E (2016). Capturing dynamics of biased attention: Are new attention variability measures the way forward? PLoS ONE, 11(11), e0166600 10.1371/journal.pone.0166600 [PubMed: 27875536]

LeMoult J, & Gotlib IH (2018). Depression: A cognitive perspective. Clinical Psychology Review. 10.1016/j.cpr.2018.06.008

Liaw A, & Wiener M (2002). Classification and regression by Random Forest. R News, 2(3), 18–22.

Marchetti I, Everaert J, Dainer-Best J, Loeys T, Beevers CG, & Koster EHW (2018). Specificity and overlap of attention and memory biases in depression. Journal of Affective Disorders, 225, 404–412. 10.1016/j.jad.2017.08.037 [PubMed: 28850855]

Marchetti I, Loeys T, Alloy LB, & Koster EHW (2016). Unveiling the structure of cognitive vulnerability for depression: Specificity and overlap. PLoS ONE, 11(12), e0168612 10.1371/journal.pone.0168612 [PubMed: 27992548]

Mccarthy MI, Abecasis GR, Cardon LR, Goldstein DB, Little J, Ioannidis JPA, & Hirschhorn JN (2008). Genome-wide association studies for complex traits: consensus, uncertainty and challenges. Nature Reviews Genetics, 9(5), 356–369. 10.1038/nrg2344

Mogg K, Holmes A, Garner M, & Bradley B (2008). Effects of threat cues on attentional shifting, disengagement and response slowing in anxious individuals. Behaviour Research and Therapy, 46(5), 656–667. 10.1016/j.brat.2008.02.011 [PubMed: 18395185]

Molz Adams A, Shapero BG, Pendergast LH, Alloy LB, & Abramson LY (2014). Self-referent information processing in individuals with bipolar spectrum disorders. Journal of Affective Disorders, 152–154, 483–490. 10.1016/j.jad.2013.07.039

Mullarkey MC, Marchetti I, & Beevers CG (2018). Using network analysis to identify central symptoms of adolescent depression. Journal of Clinical Child & Adolescent Psychology, 00(00), 1–13. 10.1080/15374416.2018.1437735

Munafò MR, & Smith GD (2018). Robust research needs many lines of evidence. Nature, 553(7689), 399 10.1038/d41586-018-01023-3 [PubMed: 29368721]

Munafò MR, Nosek BA, Bishop DVM, Button KS, Chambers CD, Sert, du NP, et al. (2017). A manifesto for reproducible science. Nature Publishing Group, 1(1), 1–9. 10.1038/s41562-016-0021

Olfson M, & Marcus SC (2009). National Patterns in Antidepressant Medication Treatment. Archives of General Psychiatry, 66(8), 848–856. 10.1001/archgenpsychiatry.2009.81 [PubMed: 19652124]

Osman A, Downs WR, Barrios FX, Kopper BA, Gutierrez PM, & Chiros CE (1997). Factor structure and psychometric characteristics of the beck depression inventory-II. Journal of Psychopathology and Behavioral Assessment, 19(4), 359–376. 10.1007/BF02229026

Pearson R, Palmer RHC, Brick LA, Mcgeary JE, Knopik VS, & Beevers CG (2016). Additive genetic contribution to symptom dimensions in major depressive disorder. Journal of Abnormal Psychology, 125(4), 495–501. 10.1037/abn0000161 [PubMed: 27124715]

Peckham A, McHugh R, & Otto M (2010). A meta-analysis of the magnitude of biased attention in depression. Depression and Anxiety, 27(12), 1135–1142. [PubMed: 21049527]

Posner K, Brown GK, Stanley B, Brent DA, Yershova KV, Oquendo MA, et al. (2011). The Columbia-Suicide Severity Rating Scale: initial validity and internal consistency findings from three multisite studies with adolescents and adults. American Journal of Psychiatry, 168(12), 1266–1277. 10.1176/appi.ajp.2011.10111704 [PubMed: 22193671]

Ratcliff R, & McKoon G (2008). The diffusion decision model: theory and data for two-choice decision tasks. Neural Computation, 20(4), 873–922. 10.1162/neco.2008.12-06-420 [PubMed: 18085991]

Ratcliff R, Van Zandt T, & McKoon G (1999). Connectionist and diffusion models of reaction time. Psychological Review, 106(2), 261–300. [PubMed: 10378014]

Rodebaugh TL, Scullin RB, Langer JK, Dixon DJ, Huppert JD, Bernstein A, et al. (2016). Unreliability as a threat to understanding psychopathology: The cautionary tale of attentional bias. Journal of Abnormal Psychology, 125(6), 840–851. 10.1037/abn0000184 [PubMed: 27322741]

Rutter M (2005). Environmentally mediated risks for psychopathology: Research strategies and findings. Journal of the American Academy of Child and Adolescent Psychiatry, 44(1), 3–18. 10.1097/01.chi.0000145374.45992.c9 [PubMed: 15608539]

Santos H, Jr., Fried EI, Asafu-Adjei J, & Ruiz RJ (2017). Network Structure of Perinatal Depressive Symptoms in Latinas: Relationship to Stress and Reproductive Biomarkers. Research in Nursing & Health, 101(5 Pt 1), 3747–11. 10.1002/nur.21784

Sheehan DV, Lecrubier Y, Sheehan KH, Amorim P, Janavs J, Weiller E, et al. (1998). The Mini-International Neuropsychiatric Interview (M.I.N.I.): the development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10. The Journal of Clinical Psychiatry, 59 Suppl 20, 22–33.

Stampe DM (1993). Heuristic filtering and reliable calibration methods for video-based pupil-tracking systems. Behavior Research Methods, Instruments, & Computers, 25(2), 137–142. 10.3758/BF03204486

Tackett JL, Lilienfeld SO, Patrick CJ, Johnson SL, Krueger RF, Miller JD, et al. (2017). It's time to broaden the replicability conversation: Thoughts for and from clinical psychological science.

Perspectives on Psychological Science, 12(5), 742–756. 10.1177/1745691617690042 [PubMed: 28972844]

van Borkulo C, Boschloo L, Borsboom D, Penninx BWJH, Waldorp LJ, & Schoevers RA (2015). Association of Symptom Network Structure With the Course of Depression. JAMA Psychiatry, 72(12), 1219–1226. 10.1001/jamapsychiatry.2015.2079 [PubMed: 26561400]

Varma S, & Simon R (2006). Bias in error estimation when using cross-validation for model selection. BMC Bioinformatics, 7(1), 91 10.1186/1471-2105-7-91 [PubMed: 16504092]

Voss A, & Voss J (2007). Fast-dm: a free program for efficient diffusion model analysis. Behavior Research Methods, 39(4), 767–775. [PubMed: 18183889]

Watson D, O'Hara MW, Chmielewski M, McDade-Montez EA, Koffel E, Naragon K, & Stuart S (2008). Further validation of the IDAS: Evidence of convergent, discriminant, criterion, and incremental validity. Psychological Assessment, 20(3), 248–259. 10.1037/a0012570 [PubMed: 18778161]

Wei P, Lu Z, & Song J (2015). Variable importance analysis: A comprehensive review. Reliability Engineering & System Safety, 142, 399–432. 10.1016/j.ress.2015.05.018

White CN, Ratcliff R, Vasey MW, & McKoon G (2010). Anxiety enhances threat processing without competition among multiple inputs: A diffusion model analysis. Emotion, 10(5), 662–677. 10.1037/a0019474 [PubMed: 21038949]

Whitmer AJ, & Gotlib IH (2013). An attentional scope model of rumination. Psychological Bulletin, 139(5), 1036–1061. 10.1037/a0030923 [PubMed: 23244316]

Wickham H (2009). ggplot2: Elegant Graphics for Data Analysis. New York: Springer-Verlag.

Wickham H (2018). tidyverse: Easily install and load the "Tidyverse." Retrieved December 19, 2017, from

World Health Organization. (2011). WHO report on the global tobacco epidemic, 2011: Warning about the dangers of tobacco. Geneva : World Health Organization.

Yarkoni T, & Westfall J (2017). Choosing prediction over explanation in psychology: Lessons from machine learning. Perspectives on Psychological Science, 12(6), 1100–1122. 10.1177/1745691617693393 [PubMed: 28841086]

Zimmerman M, & Mattia JI (2001). The Psychiatric Diagnostic Screening Questionnaire: Development, reliability and validity. Comprehensive Psychiatry, 42(3), 175–189. 10.1053/comp. 2001.23126 [PubMed: 11349235]

Zou H, & Hastie T (2005). Regularization and variable selection via the elastic net. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 67(2), 301–320. 10.1111/j. 1467-9868.2005.00503.x

Zvielli A, Vrijsen JN, Koster EHW, & Bernstein A (2016). Attentional bias temporal dynamics in remitted depression. Journal of Abnormal Psychology, 125(6), 768–776. 10.1037/abn0000190 [PubMed: 27505407]
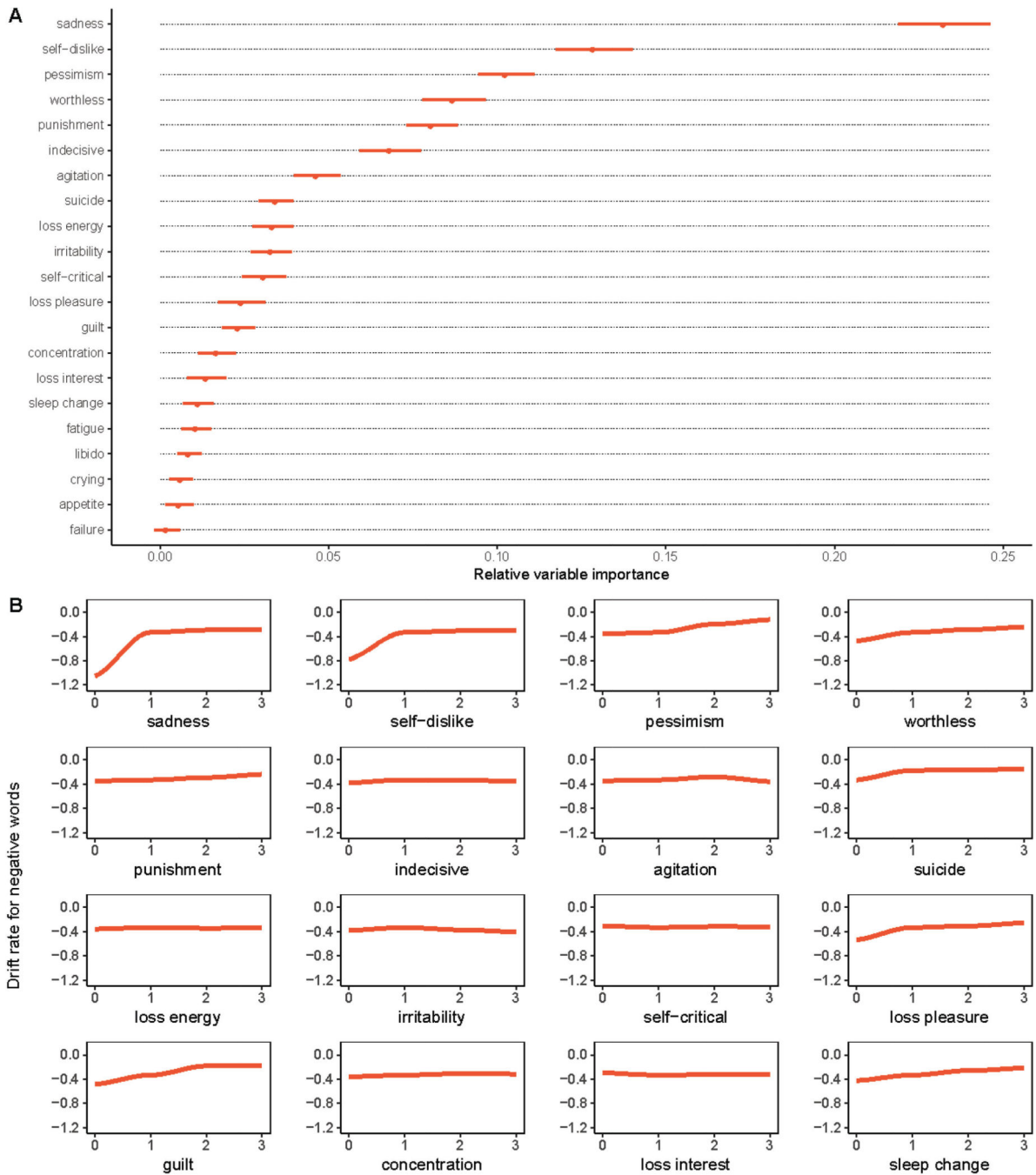
Figure 1.
Symptom importance (A) and partial dependence (B) plots for the random forest model for drift rate for negative words on the SRET. For plot A, dots indicate mean relative importance and lines indicate range across the 10 repetitions of 10-fold cross-validation. Similarly, for plot B, the red line indicates the mean for the 10 repetitions and the grey lines reflect the partial dependence associations observed for each repetition.
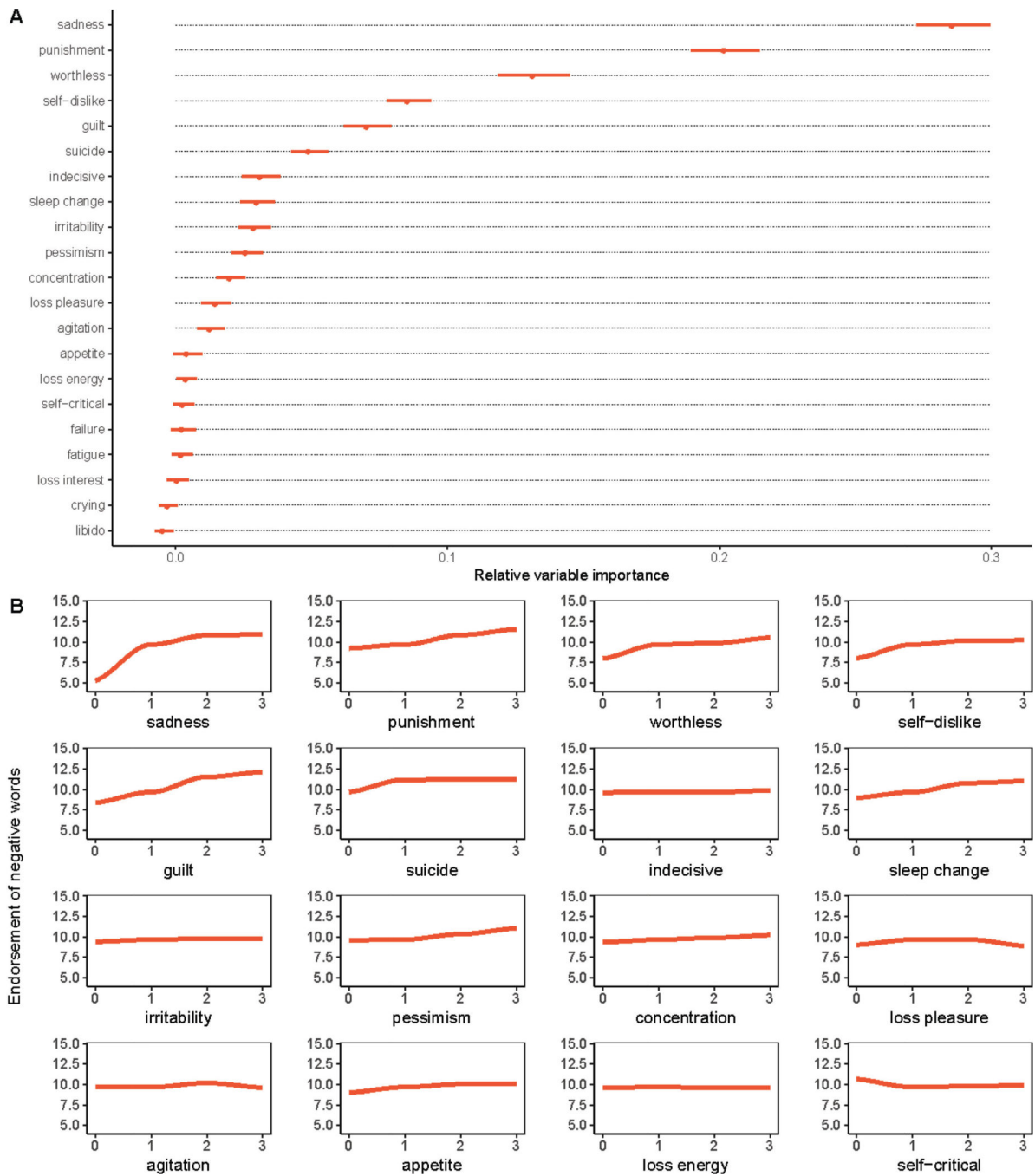
Figure 2.
Symptom importance (A) and partial dependence (B) plots for the random forest model for endorsement of negative words on the SRET.

Table 1:

Demographic characteristics of study sample.

Characteristic	N = 218
Age in years, mean (*SD*)	23.3 (4.88)
Female gender (%)	144 (66.1%)
Hispanic ethnicity (%)	67 (30.7%)
Caucasian (%)	218 (100%)
Never married (%)	185 (84.9%)
Years in school (*SD*)	14.4 (2.39)
Private health insurance (%)	126 (57.8%)
Household income (%)	
$0 – $24,999	84 (38.5%)
$25,000 – $49,999	45 (20.6%)
$50,000 – $74,999	26 (11.9%)
$75,000 – $99,999	19 (8.7%)
$100,000 +	44 (20.2%)
Current MDD	60 (27.5%)
Lifetime MDD	132 (60.6%)
Any current anxiety disorder	62 (28.4%)
Any current psychiatric disorder	115 (52.8%)
Beck Depression Inventory – II (*SD*)	18.00 (11.76)

Table 2.

Means, standard deviations, and correlations with confidence intervals for attention bias metrics and depression sum score.

Variable	*M*	*SD*	1	2	3	4
1. RT bias	−0.00	0.03				
2. TLBS toward sad stimuli	0.11	0.03	.31**			
			[.17, .44]			
3. gaze bias	0.01	0.04	.32**	.10		
			[.18, .45]	[−.05, .25]		
4. % of trials toward sad stimuli	0.50	0.06	.16*	.03	.72**	
			[.01, .31]	[−.12, .18]	[.64, .79]	
5. BDI-II	18.21	11.73	−.00	.22**	.04	−.11
			[−.15, .15]	[.07, .36]	[−.19, .11]	[−.26, .04]

Note. M and *SD* are used to represent mean and standard deviation, respectively. Values in square brackets indicate the 95% confidence interval for each correlation.

*
indicates *p* < .05.

**
indicates *p* < .01. RT = reaction time.

TLBS = trial level bias score. BDI-II = Beck Depression Inventory.