Journal of Experimental Psychology: General

Avoiding Positivity at a Cost: Evidence of Reward Devaluation in the Novel Valence Selection Task

Mya Urena, E. Samuel Winer, and Caitlin Mills

Online First Publication, January 13, 2025. https://dx.doi.org/10.1037/xge0001702

AMERICAN PSYCHOLOGICAL ASSOCIATION

Journal of Experimental Psychology: General

ISSN: 0096-3445

https://doi.org/10.1037/xge0001702

Avoiding Positivity at a Cost: Evidence of Reward Devaluation in the Novel Valence Selection Task

Mya Urena[1], E. Samuel Winer[2], and Caitlin Mills[1]

[1] Department of Educational Psychology, College of Education and Human Development, University of Minnesota, Twin Cities

[2] Department of Psychology, The New School for Social Research

Reward devaluation theory (RDT) posits that some depressed individuals may not only be biased toward negative material but also actively avoid positive material (i.e., devaluing reward). Although there are intuitive, everyday life consequences for individuals who "devalue reward" or positivity, prior work has not established if (and how) reward devaluation manifests in tasks that encompass aspects of our daily lives, such as reading. The current research thus assessed if devaluation presents in a novel Valence Selection Task, akin to a reading assignment. In three studies, participants read incomplete reading prompts and were instructed to choose between a positively valenced, negatively valenced, or neutral sentence ending—all of which were viable sentence endings. Study 1 demonstrated that participants exhibiting depressive symptoms (assessed via fear of happiness) were less likely to select the positive endings, in line with RDT predictions. Study 2 replicated these findings, regardless of who the "subject" of the reading prompt was (self vs. other). Finally, results from Study 3 suggest that participants who displayed depressive symptoms were less likely to choose the positively valenced response, even when it was manipulated to be the objectively correct answer. These findings underscore the relevance of RDT in novel contexts and highlight potential clinical and educational applications.

Public Significance Statement

Although the negativity bias in depression has been well-established, the corresponding idea of positivity avoidance (or reward devaluation) is still much less understood. That is, people with depression may not only be biased toward negativity, they may also go so far as to actively avoid positivity. Our goal was to test whether this effect translates to situations that are more akin to daily life, such as processing text, by developing a novel task that was meant to mimic sentence completion tasks found in educational settings. Across three studies, results suggest that individuals with depressive symptoms were indeed more likely to avoid selecting positive endings to a short story, even when it meant selecting the objectively "wrong" answer. These findings not only provide necessary support for the reward devaluation theory but also highlight the extent to which depressive symptoms may shape the way we encounter and respond to affectively charged materials throughout our daily lives, particularly at work and in classrooms.

Keywords: reward devaluation theory, fear of happiness, depression, education, attentional bias

Supplemental materials: https://doi.org/10.1037/xge0001702.supp

Jessica D. Payne served as action editor.

Mya Urena https://orcid.org/0009-0003-7471-2449

E. Samuel Winer https://orcid.org/0000-0002-3817-874X

Caitlin Mills https://orcid.org/0000-0003-4498-0496

Data are available at https://osf.io/whdpb/?view_only=3adf4170914b4baf8b85b5b5fa20c509. This study was not preregistered. Prior to publication, this research was presented at the Psychonomic Society's Annual Meeting in San Francisco, California, in 2023 and the Society for Text and Discourse's annual meeting in Chicago, Illinois, in 2024. The authors have no conflicts of interest to declare that are relevant to the content of this article.

Mya Urena served as lead for conceptualization, investigation, data curation, formal analysis, methodology, software, visualization, writing-original draft, writing-review and editing. E. Samuel Winer served as lead in conceptualization, a supporting role in methodology, supervision, validation, writing-original draft, writing-review and editing, and project administration. Caitlin Mills served as lead for conceptualization, funding acquisition, project administration, supervision, validation, and writing–review and editing.

Correspondence concerning this article should be addressed to Mya Urena, Department of Educational Psychology, College of Education and Human Development, University of Minnesota, Twin Cities, 104 Burton Hall, 178 Pillsbury Drive SE, Minneapolis, MN 55455, United States. Email: urena014@umn.edu

Depression is one of the most commonly diagnosed mental disorders in children; about 2.7 million diagnoses were recorded between 2016 and 2019 (*Data and Statistics on Children's Mental Health | CDC*, 2022), with rates only rising further after the peak of the COVID-19 pandemic (Śniadach et al., 2021). The rates of depression in children are particularly concerning in an educational context, given that roughly 20.9% of adolescents (ages 12–17) have experienced major depressive disorder (Bitsko, 2022). Indeed, depression has been consistently and negatively linked to students' classroom engagement, educational achievement, and prospects for higher education (Fergusson & Woodward, 2002; Fernandes et al., 2021; Quiroga et al., 2013; Wickersham et al., 2021). At the same time, depression presents heterogeneously across individuals (Fried & Nesse, 2015); a student experiencing depression following the loss of a close friend may differ significantly from another experiencing anhedonic symptoms, such as a loss of interest in social interaction. However, these differences are not always built into the diagnostic process or corresponding treatments—particularly from an educational standpoint (Ciuhan & Iliescu, 2021; Humensky et al., 2010).

One of the well-known ways that depression can manifest is through altering and biasing reward systems. This is particularly relevant from a learning standpoint given that (a) our daily lives are rife with affectively charged content and (b) traditional education systems are often rooted in rewards and positive feedback. Historically, the attentional bias literature has focused on establishing a consistently negative bias in depression, whereby depressed individuals tend to prioritize negative stimuli over positive or neutral stimuli (Gotlib & Joormann, 2010; Teachman et al., 2012; Yiend, 2010). More recently, however, theoretical and meta-analytic evidence has suggested that it may not be as simple as a bias toward negativity or a lack of a bias toward positivity, but rather a *de*valuation of positivity and reward (Gallagher et al., 2024; Salem et al., 2018; Winer & Salem, 2016).

Reward devaluation theory (RDT) posits that some depressed individuals respond negatively to positive material to such an extent as to actively avoid it. The genesis of RDT stems mainly from meta-analytic evidence compiled from a well-known sustained attention task, the dot-probe (DP) task (MacLeod et al., 1986). The crux of the DP task is the measurement of reaction times to observe whether anxious or depressed participants have an attentional bias toward valenced stimuli and will respond faster to the valenced stimuli. Winer and Salem (2016) synthesized findings from multiple DP studies to establish the concept of RDT after observing a consistent trend among depressed individuals—ultimately indicating a greater tendency to direct attention away from positively valenced cues.

Some of these RDT predictions converge with existing frameworks in the depression literature, such as those centered around belief-updating, interpretation bias, and reward insensitivity (Beck & Clark, 1991; Deng et al., 2022; Eshel & Roiser, 2010; Everaert et al., 2017, 2018; Kube, 2023; Whitton et al., 2015). For example, biased belief-updating frameworks, as outlined recently by Kube (2023), emphasize the idea that people are likely to have a set of "belief" priors that are iteratively updated when people are presented with novel positive information; however, individuals with depression may be resistant to updating their negative beliefs regarding their future performance and relationships (Kube, 2023). Similarly, interpretation bias, or the tendency of depressed individuals to interpret novel, ambiguous information as negative (Beck & Clark, 1991), provides another complementary set of predictions. Prior work in this area has demonstrated a clear relationship between depression and a bias toward negativity (Everaert et al., 2017, 2018) across multiple tasks (Deng et al., 2022), but does not capture the possible *avoidance* of positivity. Perhaps the closest framework to RDT, as seen in Gotlib's evenhandedness theory (Gotlib & McCabe, 1992; McCabe & Gotlib, 1995), is the idea of reward insensitivity, which highlights the tendency for depressed individuals to experience an apathetic response to reward (Eshel & Roiser, 2010; Whitton et al., 2015).

Although all three of these frameworks map neatly onto RDT's acknowledgment that depressed individuals' attention will likely be biased toward negativity and negative information, RDT provides a more specific mechanistic prediction that those same individuals would also bias away from reward and positivity through a systematic devaluation of reward. The current research is thus focused on elucidating how depressed individuals actively avoid positivity and reward, as described by RDT.

The original predictions made by RDT were almost entirely founded on meta-analytic evidence of the DP task—that is, on reaction times alone; however, more recent work suggests it may also extend outside of this lab-based task. For example, Bryant et al. (2023) demonstrated that individuals who scored high on the Fear of Happiness Scale (FHS) reported that they would avoid choosing therapies that specifically target positivity (e.g., positive affective treatment). Fear of happiness, in particular, is thought to be associated with reward devaluation; as Jordan et al. (2021) summarized, individuals who are depressed, specifically those who experience a fear of happiness, will not only feel negative when presented with positive outcomes but encounter negative consequences instead; subsequently, these individuals begin to experience fear in response to them. However, despite this link with participants' anticipatory behavior, there is no clear current evidence that people change their concurrent behavior to avoid positivity in other contexts, particularly ones that mimic the types of information processing one may see in educational contexts (i.e., building a mental model of a story, completing a sentence with the "best" ending).

Although prior work establishes a helpful foundation for the basic function of reward devaluation, its potential application to real-world settings like education (and beyond) has not yet been empirically tested. We, therefore, developed a novel paradigm, which we refer to as the Valence Selection Task (VST), to address the following questions across three different studies. First, we assessed whether predictions from RDT still hold in a novel reading task—that is, do participants who "devalue" reward (or positivity) alter their selection behaviors beyond reaction times (as originally observed in the DP task)? Second, does this finding replicate regardless of how the positivity is framed, either in the context of positive outcomes for themselves versus others? Finally, we tested how far the avoidance of positivity extends, assessing one of the most novel tenets of RDT: Are people who experience a greater fear of happiness willing to avoid positivity at the expense of being incorrect?

Study 1

Method

Participants

Participants had to be 18 years old and fluent in English to participate in the study. Given that the meta-analytic group differences ($g = .399$) from Winer and Salem (2016) did not translate to a correlation effect

size, we instead used the smallest medium effect size ($r = .3$; Cohen, 1988) as a basis for our power calculations in Study 1. G*Power was utilized to determine that 82 participants were required to detect an effect size of 0.3 with 0.80 power and α set to .05.

A total of 82 participants were recruited through Prolific, an online data collection platform. One participant did not adequately complete the survey and was therefore excluded, resulting in a final sample of 81 participants ($N = 81$). Approximately 60% identified as women, 39% identified as men, and 1% identified as nonbinary. The age range of participants was 18–73 years ($M = 28.04$; $SD = 9.72$). Based on internal piloting, we determined that the study would take approximately 15 min to complete. Using this estimation and Prolific's cost calculator tool, participants received $3 for their involvement in the study.

Materials

Valence Selection Task. We developed a novel paradigm to test the RDT predictions, mimicking educationally relevant scenarios where people have to fill in the blank with the "best" ending according to their judgment. In the VST, participants were given a set of 30 self-focused incomplete sentences (e.g., "You are walking to the office. You only have about 5 minutes left of your walk. You are debating whether to stop and buy a coffee when…"). The task required participants to finish the prompt using one of three sentence endings provided for each prompt. Each item of the VST was displayed on its own page. On each page, both the incomplete prompts and the sentence endings were randomized and presented simultaneously in a multiple-choice format. Of the three endings provided, one was positively valenced (e.g., "you see a $10 bill on the sidewalk"), one was negatively valenced (e.g., "you trip and fall on the sidewalk"), and one was neutral (e.g., "you see a coffee shop"). Participants were assured that there were no correct answers to any of the prompts and that they were only required to pick the ending that they thought best fit the sentence (see Supplemental Materials for the full list).

Fear of Happiness. The FHS is a five-item self-report measure that assesses a participant's belief that happiness or good fortune is a sign of impending unhappiness or bad incidents (Joshanloo et al., 2014). Items on the FHS were rated on a scale from 1 (*strongly disagree*) to 7 (*strongly agree*) and were summed for a total score (5–35). Higher scores indicate a greater fear of happiness.

Convergent Measures of Depression Over Time. We included two additional measures as indicators of convergence (see Supplemental Materials for full results on these measures). First, general depressive symptoms were assessed via the Quick Inventory of Depressive Symptomatology Self-Report (QIDS-SR). The QIDS-SR (https://ids-qids.org/about.html) is a 16-item self-report measure that is used widely in the clinical literature, having been developed as a free-for-use alternative to the Beck Depression Inventory–II (Beck et al., 1996) that closely coincides with the symptoms of depression that make up *Diagnostic and Statistical Manual of Mental Disorders*-based mood disorder diagnoses (Rush et al., 1996, 2003, 2005).

Items from the QIDS-SR were scored on a scale from 0 to 4. Item 12 from the QIDS-SR concerning suicidal ideation was removed from the present study because the study was conducted online with multiple participants taking the survey at once. This would not allow researchers to be directly available to individuals who demonstrated acute distress in response to Item 12. The remaining 15 items were categorized based on the eight foremost symptoms of depression: sleep disturbance (the highest value from Items 1 through 4), sad mood (Item 5), changes in appetite and weight (the highest value from Items 6 through 9), concentration difficulties (Item 10), self-criticism (Item 11), loss of interest (Item 13), energy/fatigue (Item 14), and psychomotor functioning (the highest value from either Item 15 or 16). Higher scores indicated a more significant presence of depressive symptoms. The thresholds for symptom severity ranged from 0 to 5 ("none"), 6 to 10 ("mild"), 11 to 15 ("moderate"), 16 to 20 ("severe"), and 21 to 27 ("very severe"; Rush et al., 2003).

Second, anhedonic symptoms were assessed using the Specific Loss of Interest in Pleasure Scale (SLIPS), a 23-item self-report measure that focuses on one's loss of interest in pleasure (Winer et al., 2014). For each item of the SLIPS, participants are presented with four options scored from 0 to 3. Choosing Option 1 (coded as 0) suggested that the participant had experienced no loss of interest or pleasure (e.g., "I still enjoy going with friends"). Choosing Option 2 (coded as 1) suggested that the participant had some loss of interest or pleasure (e.g., "I don't enjoy going out with friends as much as I used to"). Choosing Option 3 (coded as 2) suggested that the participant had experienced a loss of most interest or pleasure (e.g., "I no longer enjoy going out with friends"). Choosing Option 4 (coded as 3) suggested that the participant had never experienced interest or pleasure in that scenario (e.g., "I have never enjoyed going out with anyone"). After data collection was completed, any item coded as 3 was recoded as 0 to demonstrate the change in anhedonic symptoms over time, per standard scoring of the SLIPS. Scores ranged from 0 to 46. In prior work, researchers have categorized participants into three subgroups using two cutoff points within their SLIPS results: (a) "no loss of pleasure" with a SLIPS score of 0; (b) "some loss of pleasure" with SLIPS scores ranging from 1 to 8; and (c) "marked diminishment of pleasure" with SLIPS scores of 8 or higher (Ritsner & Ratner, 2019). Note again that the full results for the QIDS-SR and SLIPS can be found in the Supplemental Materials.

Procedure

Participants underwent an informed consent procedure before completing the three aforementioned self-report measures: (a) FHS, (b) QIDS-SR, and (c) SLIPS. Participants then completed the VST, followed by a demographic information survey before ending the study.

Analytical Approach

All analyses were performed using R Statistical Software (v4.2.2; R Core Team, 2021). For valence bias, the proportion of positive, negative, and neutral answers selected in the valence selection task was calculated for analyses (i.e., how often did someone choose the respective valenced endings?). These were then correlated with the self-reported depression scales. The data for the VST were not normally distributed, so Spearman's ρ was calculated instead, though we note that the pattern of results does not change if Pearson's r is used.

Results

Sample Characteristics

The mean scores of the self-report measures for the final sample were as follows: FHS, 14.2 ($SD = 6.94$); QIDS-SR, 8.78 ($SD = 4.57$);

SLIPS, 10.5 (SD = 8.34). When referencing QIDS-SR scores, our sample represented a mild range of depressive symptoms (M = 8.78, SD = 4.57). However, sum scores varied from no depressive symptoms to very severe depressive symptoms (ranging from 0 to 22). Scores from both the SLIPS and the FHS, when compared to norm values and results from prior studies, also suggest that participants in this sample demonstrated mild to moderately depressive symptoms. On average, it took participants 6.79 min (SD = 3.30) to complete the VST. Additionally, participants selected positive responses 43% of the time (SD = .207), negative responses 17% of the time (SD = .135), and neutral responses 40% of the time (SD = .163).

VST Results

Negative. Attendance to negative information is prevalent in depression literature and is a significant component of reward devaluation. Replicating previous work (Gotlib & Joormann, 2010; Teachman et al., 2012; Yiend, 2010), individuals with high scores on the FHS were more inclined to select the negatively valenced responses (ρ = 0.40, p < .001; Figure 1). A similar relationship was found with the QIDS-SR (ρ = 0.35, p = .001) and SLIPS (ρ = 0.38, p < .001).

Positive. In line with predictions from RDT, the proportion of positive responses was significantly and negatively correlated with scores on the FHS (ρ = −0.31, p = .005; Figure 1). This pattern was similar for both the QIDS-SR (ρ = −0.29, p = .009) and SLIPS (ρ = −0.28, p = .011) as well. Taken together, results indicate that participants experiencing depressive symptoms avoided choosing the positive response, suggesting the presence of reward devaluation.

Neutral. No significant correlations were found between the self-report measures and the proportions of neutral response types across all three measures: FHS (ρ = 0.04, p = .748; Figure 1); QIDS-SR (ρ = 0.07, p = .558); and SLIPS (ρ = 0.00, p = .991). In other words, there was no relationship between participants experiencing depressive symptoms and their inclination to choose the neutral response.

Discussion

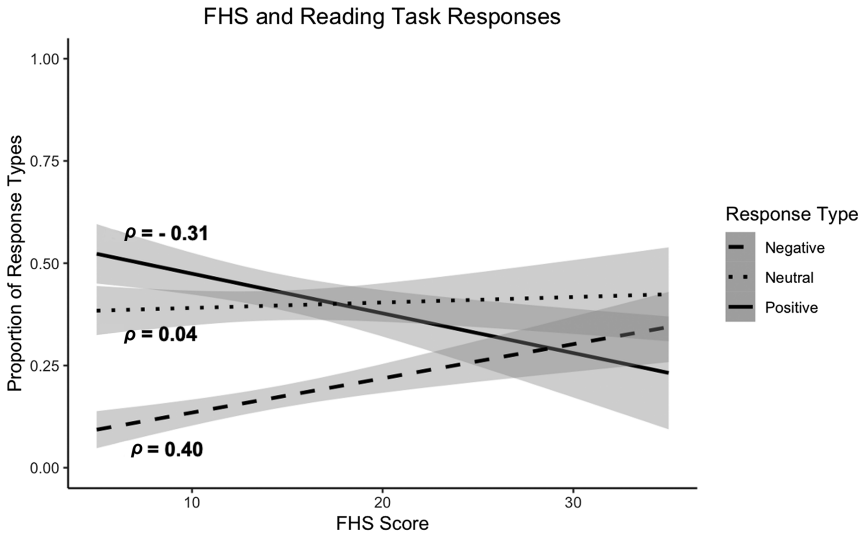
Study 1 extends support of RDT in the context of a novel task that is more akin to daily life compared to the DP task. That is, millisecond changes in reaction times demonstrate a key phenomenon but may not have cascading effects in everyday life; however, avoiding the selection of positive endings is much more similar to tasks that are often seen in education (i.e., sentence completion tasks).

Previous work by Winer and Salem (2016) demonstrated a discriminant signal for avoidance of positive information by depressed persons that emerged from meta-analyzing a large number of studies that used positive stimuli. However, the dot-probe is neither reliable for use in a single study nor is its meaning conceptually clear to nonexperts. The VST provides an accessible and perhaps more practically relevant metric for assessing reward devaluation.

At the same time, a limitation of Study 1 is that all of the reading prompts in the VST were self-referential, that is, participants were the "subject" of each prompt (i.e., "*You* were walking down the street"). Given that prior research has highlighted that depressed individuals tend to harbor negative self-images (Collins et al., 2023; Collins & Winer, 2024), it may be possible that we only observe support for RDT when participants view the positive scenario as applying to themselves. Another limitation of Study 1 is that we did not reach our desired sample size based on a priori power calculations (82 required; 81 people with usable data for analyses). We originally completed the study and analyses without consideration of exclusion (e.g., not for people who did not complete the entire task); because of this, we avoided adding data retrospectively. However, underpowered results such as these can pose real issues in the scientific community, and therefore it was important for us to test the replicability of these findings to ensure they were not a result of our underpowered sample size. Study 2 thus aims to both replicate and extend upon the original study design by incorporating both self-referential and other-referential reading prompts within the VST with a larger sample size.

Figure 1

Correlations Between FHS Score and Proportion of Response Types



Note. FHS = Fear of Happiness Scale.

Study 2

The objectives of Study 2 were to (a) replicate the findings of Study 1 while (b) assessing whether the subject reference group (self-referential vs. other-referential) moderated observations of reward devaluation.

Method

Participants

In total, 285 ($N = 285$) participants (47% identified as women, 51% identified as men, 1% identified as nonbinary, 0.4% identified as other; $M_{age} = 40.36$, $SD_{age} = 14.82$) who were 18 years or older and fluent English speakers were recruited through Prolific. We chose to collect a larger sample size for Study 2 in order to have enough power to detect a potential moderation, as well as conduct the main correlational analyses. We did not have a prior effect size to determine sample size for a moderation of this nature, given the novelty of the research question; we, therefore, chose to collect up to 285, as this was the maximum amount in our budget to minimize the possibility of Type II errors. Similar to Study 1, participants received $3 for their participation.

Procedure

The methods and procedure were kept identical to Study 1, with a few key changes: We included a within-subjects manipulation, where participants were exposed to both self-referential and other-referential prompts as part of the VST. The 30 prompts from Study 1 were adapted by altering the subject of the sentence from "you" to another person's name, creating a new set of 30 prompts that were other-referential (see Supplemental Materials for a full list). For example, if a self-referential prompt read, "You sign up to run a marathon. On the day of the marathon you…," the corresponding mirrored other-referential prompt would read, "Valerie signed up to run a marathon. On the day of the marathon, she…" Participants in the study received a total of 30 reading prompts, 15 of the self-referential and 15 of the other-referential prompts, and they did not see the same prompt twice regardless of reference condition. The prompts were counterbalanced, such that a given prompt could show up in either condition to avoid item effects.

Analytical Approach

In Study 1, we used Spearman's ρ given that we had a single score on the FHS for each person and their corresponding proportions for negative, positive, and neutral options. In Study 2, there is a repeated measures data structure, such that each participant has proportion scores for both self-referential and other-referential prompts. Given that a primary research question is whether the referential group moderates the relationship between FHS and valence selection, we used linear mixed-effects regression models using the *lme4* (Bates et al., 2015) R package. Three models were constructed for negative, positive, and neutral proportions, respectively, with the following model structure: Response Type Proportion (negative/positive/neutral) ~ Fear of Happiness Score + Referential Condition + Fear of Happiness Score × Referential Condition + (1 | Participant ID). This model structure allowed us to test for (a) the relationship between FHS and valence selection (replicating Study 1); (b) whether the referential group was a moderator of this relationship (extension of Study 1); and (c) the main effect of referential group, which is not germane to the primary research questions, but is reported for completeness.

Results

Sample Characteristics

The sample for Study 2 showcased a comprehensive array of depressive symptoms, spanning from 0 (*no symptoms*) to 21 (*severe symptoms*). Overall, scores on the QIDS-SR demonstrated that our sample represented a mildly depressed population, with a mean score of 6.47 ($SD = 4.79$). The average FHS score was 11.69 ($SD = 6.91$), and the average SLIPS score stood at 8.32 ($SD = 9.68$). On average, participants took approximately 5.53 min ($SD = 2.75$) to complete the VST.

VST Results: Does the Relationship Between FHS and Valence Selection Replicate?

Negative. As in Study 1, we found that participants who scored higher on the FHS were more inclined to select the negatively valenced responses ($\beta = 0.364$, $SE = 0.056$, $p < .001$).

Positive. Participants who scored higher on the FHS were again less likely to select the positively valenced responses ($\beta = -0.370$, $SE = 0.056$, $p < .001$), replicating findings from Study 1 and providing further support for RDT.

Neutral. Unlike Study 1, there was a significant relationship between FHS score and the proportion of neutral responses selected ($\beta = 0.128$, $SE = 0.059$, $p = .030$), such that higher FHS scores were related to more neutral selections. Although not predicted, this result is still in line with RDT, as participants' avoidance of selecting the positively valenced response may have resulted in them choosing the neutral response more frequently as well (though the relationship with selecting negative is stronger; Figure 2).

VST Results: Does the Reference Group Moderate the Relationship Between FHS and Valence Selection?

Negative. The relationship between FHS and the proportion of negative responses depended on whether the participant completed a self- or other-referential prompt (interaction: $\beta = -0.091$, $SE = 0.043$, $p = .033$). A simple slopes analysis was performed using the *interactions* package (v1.1.0; Long, 2019), which indicated that there was a significant relationship between FHS score and the proportion of negative responses in both the self-referential ($B = 0.004$, $SE = 0.001$, $t = 4.85$, $p < .001$) and other-referential conditions ($B = 0.005$, $SE = 0.001$, $t = 6.46$, $p < .001$).

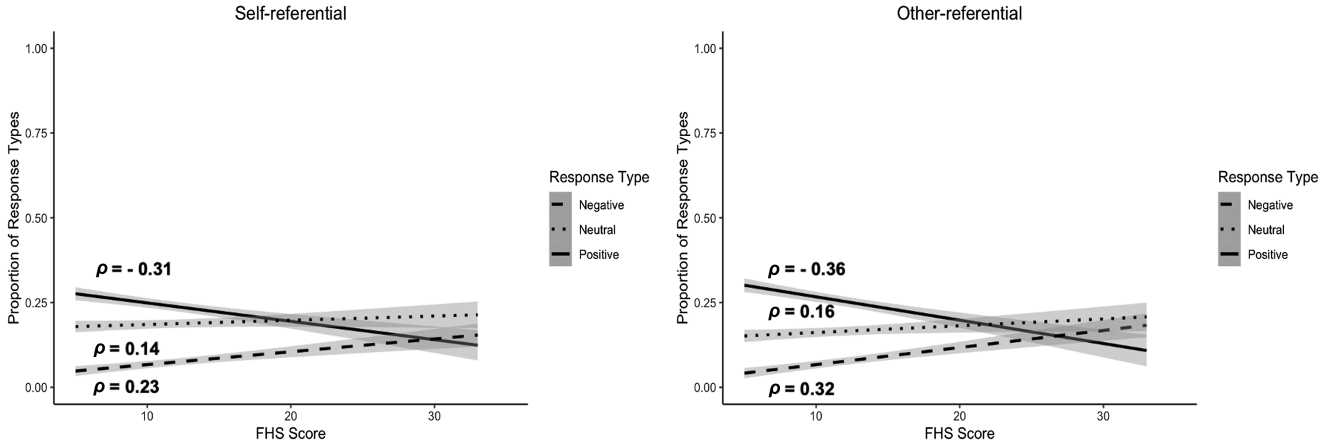
Positive. There was no significant interaction between the FHS score's association with the proportion of positive responses and the self/other-referential condition.

Neutral. There was no significant interaction between the FHS score's association with the proportion of neutral responses and the self/other-referential condition.

Overall, the results indicate that participants who scored higher on the FHS were more inclined to choose the negatively valenced sentence ending when reading other-referential prompts. For positively valenced and neutral responses, the self/other-referential condition had no impact.

Figure 2

Correlations Between FHS Score and Proportion of Response Types From Self-Referential and Other-Referential Prompts



Note. FHS = Fear of Happiness Scale.

Discussion

Results from Study 1 replicated in Study 2: Individuals who reported depressive symptoms (specifically a fear of happiness) were more likely to choose the negatively valenced response and less inclined to choose the positively valenced response. The VST thus seems reliable across studies and provides robust evidence that reward devaluation impacts participant responses in a way that may be likely to impact their daily lives (i.e., as they process affectively charged information).

It is worth noting a few possibilities for why we did not observe any moderation effects that may be fruitful for future research. The other-referential prompts featured random individuals generated by the researchers. There were no references to individuals that were specifically associated with the participant. For example, the prompt "Valerie signed up to run a marathon. On the day of the marathon, she…" might elicit a different response if "Valerie" was the name of one's school bully compared to the name of a best friend. In future studies, it may be useful to ask participants the names of their close friends or favorite coworkers at the beginning of the study and then implement that information in the VST, which may be more in line with other self-referential memory paradigms (Collins & Winer, 2024). The lack of significant interactions for the positive responses could also be attributed to the differences in tense between the self- and other-referential prompts. When generating the prompts for Study 1 and Study 2, the self-referential prompts were written in the present tense (e.g., "You *sign* up to run a marathon. On the day of the marathon you…"), while the other-referential prompts were written in the past tense (e.g., "Valerie *signed* up to run a marathon. On the day of the marathon, she…"). Future work should focus on how changing the settings of the unfinished prompt may impact the proportion of response types that are chosen.

These possible limitations notwithstanding, the main conclusion is that we did not find evidence that the self/other-referential condition impacted how individuals with a fear of happiness completed the reading prompts, suggesting that RDT may be a somewhat general phenomenon that is not selectively applied to self-referential situations; it may simply be positivity in general. This idea of the possible "generalized" nature of RDT brings up another interesting question: How far are people willing to go to avoid positivity? Are they willing to avoid the positively valenced answer, even if it means being incorrect?

Study 3

In Studies 1 and 2, sentence endings for the reading prompts were equally plausible. No matter what sentence ending was chosen, the sentence ultimately made sense. In Study 3, we opted to change the positively valenced answers so that they *were* the best fit for the prompt, rather than all answers being equally plausible. We generated word-problem-like prompts that included intentionally easy (one step) math problems (e.g., 4 green balloons + 6 red balloons = 10 total balloons), so that participants could readily distinguish the correct answer from the incorrect. This allowed us to assess whether participants who exhibited a fear of happiness were willing to be wrong to avoid selecting the positive outcome rather than getting the math problem wrong due to the problem being difficult.

Method

The following changes were made to Study 3. We created a new set of 16 prompts for the VST (see Supplemental Materials for a full list), so that the positively valenced sentence ending was the most correct of the three endings, while the negatively valenced and neutral endings were incorrect. This was done by utilizing a similar format to the Study 1 VST, where a scenario was proposed to the participant and the participant had to select the ending that they believed fit the best. For example, a prompt would read "Gary had $30 in his wallet to spend at the mall. He bought a T-shirt for $12, a pair of socks for $5, and a hat for $8. When he reached the cashier, he found that…" The participants were given three choices: (a) Positive: "he had $5 remaining in his wallet to buy a small accessory"; (b) Negative: "he didn't have enough money and had to return one of the items"; and (c) Neutral: "the t-shirt was $15." Again, all questions were inherently straightforward, so we were not measuring math ability or fluency but rather whether people would

avoid selecting the positive answers, despite being able to easily calculate the correct one.

In Study 3, we also made the decision to only have participants complete the FHS given that it was our main variable of interest. All prompts were other-referential given that we found no evidence of moderation in Study 2 and conservatively assumed that participants may be even more likely to select positive answers for other-referential prompts based on previous results. In this way, evidence for support of RDT would be strongest if we set up the condition when they are most likely to select the positive answer. No other changes were made to the procedure.

As in Study 1 and Study 2, before beginning the VST, participants were informed that there were no incorrect answers when completing the prompts. This consistency was maintained throughout to ensure that we could accurately compare results across the three studies and that participants had the freedom to select any response that they thought was appropriate.

Participants

For Study 3, we collected a total of 100 participants (52% identifed as women, 45% as men, 2% as nonbinary, and 1% as other). Using the correlation values from Studies 1 and 2, detecting an effect of $\rho = 0.33$, we determined that Study 3 would require about 67 participants (with α at .05 and $1 - \beta$ at .8, two-tailed). However, we also expected our correlation in Study 3 to be a bit lower given the restricted range that would come along with making the answers obviously correct. Given that we did not have an exact effect sizer to use for an a priori estimate, we therefore estimated an effect in line with $\rho = 0.25$, which required a sample size of 100 using a one-tailed test. Internal piloting suggested the study would take approximately 10 min to complete; based on this time estimate, we used Prolific's cost calculator tool to determine that participants should receive $2 for their involvement in the study.

Analytical Approach

Spearman's ρ was used (as in Study 1) given that we did not have a within-subjects manipulation.

Results

Sample Characteristics

The mean FHS score was 13.3 ($SD = 7.19$), similar to other work that has utilized the FHS (Calafiore et al., 2024; Jordan et al., 2021). On average, participants selected positively valenced responses 95% of the time ($SD = .077$), selected negatively valenced responses approximately 2.6% of the time ($SD = .050$), and selected neutral responses approximately 2.8% of the time ($SD = .049$). We note that across the three studies, the selection of neutral and negative was numerically lower. Overall it took participants 4.71 min ($SD = 2.35$) to complete the VST.

VST Results

Negative. Similar to both Studies 1 and 2, participants who scored higher on the FHS were more inclined to select the negatively valenced response ($\rho = 0.23$, $p = .023$), even though the negatively valenced response was technically incorrect.

Positive. The proportion of positive responses was significantly negatively correlated with higher scores on the FHS ($\rho = -0.23$, $p = .022$). This finding and general effect size were replicated across Studies 1, 2, and 3. Even though the positively valenced response was technically the most correct of the three, those who experience a fear of happiness were still willing to be incorrect to avoid selecting the positive response.

Neutral. There was no significant correlation found between scores on the FHS and the proportion of neutral responses; FHS ($\rho = 0.15$, $p = .131$). This result is not surprising given that the neutral response was also technically not the best selection in terms of correctness (Figure 3).

Discussion

The purpose of Study 3 was to further validate the concept of reward devaluation in the face of incorrectness. Participants who demonstrated a greater fear of happiness were more inclined to select the negatively valenced response, despite the response being technically incorrect. This finding suggests that people are not only willing to change their behaviors to avoid positivity, but they are willing to do so even when it means being incorrect.

It is worth noting that we observed a very high rate of positive responses and a very low rate of negative responses. For the purpose of this study, the math problems chosen were very straightforward, with the intention that participants could complete the problems in their heads using one operation. This allowed us to imply correctness without placing the emphasis on the performance of the participant. Results suggested that this implicit sense of correctness was well-received by participants, such that we observed a ceiling effect for the positive, correct answers. On the one hand, this can be interpreted as people generally selecting the "correct" answer. On the other, taken together with Studies 1 and 2 (when there were no correct/incorrect answers), it suggests that even when the distribution is skewed very much in favor of selecting the best answer, those with higher FHS scores are still willing to give up their accuracy compared to their "typical" FHS counterparts. It is quite easy to imagine how this could manifest in real-world situations, particularly when students are faced with affectively charged materials (that are often positive) at school every day.

Transparency and Openness

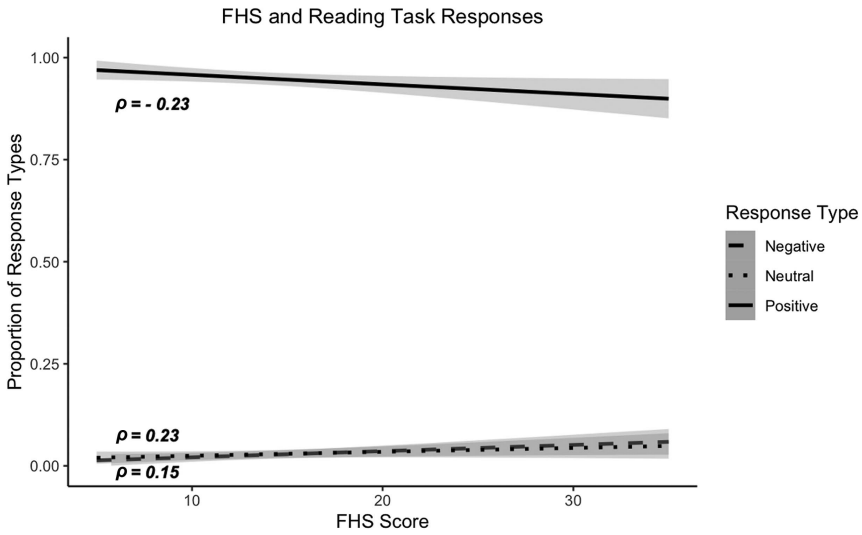
All research practices described here were approved by the Institutional Review Board at the University of Minnesota Twin Cities. This study was not preregistered before data analysis. We report how we determined all data exclusions, manipulations, and measures in the study. Data are accessible to the public on the Open Science Framework repository found here: https://osf.io/whdpb/?view_only=3adf4170914b4baf8b85b5b5fa20c509.

General Discussion

Across three studies, we have made several key contributions. Theoretically, while the VST could still be classified as a lab-based task, these studies are testing reward devaluation in a novel way that more closely aligns with activities seen in other domains such as education (i.e., sentence completion tasks). People who are experiencing depressive symptoms, and particularly those experiencing

Figure 3

Study 3: Correlations Between FHS Score and Proportion of Response Types



Note. FHS = Fear of Happiness Scale.

fear of happiness, are willing to avoid positivity in general—even when it means they will be incorrect. In other words, individuals who are happiness-averse do not merely disregard potential positivity. The presence of positivity *inhibits* their ability to correctly solve problems. Importantly, these are individuals who performed fine when responding to neutral prompts; it is the interaction of positively valenced prompting and positivity-fearing individuals that causes this stunning effect, which RDT predicts.

Many frameworks suggest differential processing of positive information in depression (for reviews, see Everaert et al., 2017; Kube, 2023; Winer & Salem, 2016). Frameworks that somewhat dovetail with RDT are those that highlight attenuated belief-updating in depression or insensitivity to reward. Indeed, substantial evidence suggests that depressed individuals interpret ambiguous information more negatively than do nondepressed persons and that depressed persons have difficulty updating negative interpretations even when they are presented with increasing evidence that their beliefs are more negative than they are justified to be (see Everaert, 2018). We interpret these as complimentary frameworks to RDT that stop short of predicting the findings presented here (especially those of Study 3).

This systematic avoidance of positive responses can not be explained merely by a general lack of belief-updating or reward insensitivity, despite there being clear theoretical overlap. Belief-updating frameworks do not discriminately focus on reward devaluation (i.e., reward will result in dejection) as a principle underlying unyielding belief systems in depression. Moreover, these three studies outline a mechanism by which avoidance of positivity explains how priors do not get updated. Persons with symptoms of depression or high fear of happiness are systematically getting information incorrect through a pattern of avoidance of positivity and positivity alone. Devaluation of reward is of course governed by belief systems, and disconfirmation of positive information is part of a devaluative process, but getting questions wrong in the effort of positivity avoidance is not merely disregarding evidence; it is the point at which potential evidence is ephemerally shifted by the perceiver to be processed as confirmatory. In other words, this process is incompletely described as a bias that disconfirms, disregards, or undervalues reward; the process systematically eliminates or *de*values the rewarding option.

Our interpretation is similar for reward insensitivity, such as what is been described in Gotlib's evenhandedness theory (Gotlib & McCabe, 1992; McCabe & Gotlib, 1995). Just as with belief-updating approaches, RDT does not so much disagree with reward insensitivity approaches but instead provides additional theoretical import on *de*valuation that lead to independent hypotheses, such as those of Study 3 of this article. In reward-insensitivity approaches, emphasis is paid to an *absence* of normative approach-related biases toward positivity in depressed individuals. We agree with this (see Bartoszek & Winer, 2015), but go further to note that there is also a *presence* of an avoidance-related bias away from positivity in depressed individuals (Winer & Salem, 2016). This is, from our view and as it relates to the current set of studies, the more interesting aspect of RDT. This is also the reason that we hypothesized Study 3's results.

So, referring to the systematic avoidance of positive responses that are the correct answer choices as a general lack of belief-updating or as "reward insensitivity" may be incomplete, despite the overarching similarity and agreement among these theoretical approaches. What is being demonstrated in the previous dot-probe findings (Winer & Salem, 2016), and in Study 3 here, suggests a particular type of emotional expertise that allows the perceiver to continue to perceive the world in line with negative expectations by eradicating implausible positive stimuli. That rejection can only happen from a hypersensitive inhibition of positive information, just as systematic avoidance can only occur after initial identification (even if that identification is ephemeral in nature).

Methodologically, we introduced a novel task (the VST) that can be used in future work to help assess the boundaries of negativity bias and reward devaluation and may eventually be used to help identify people who are experiencing such devaluation in settings when reward may have unintended consequences. Whether buying a

house, evaluating a therapeutic relationship, motivating oneself in school, or trusting a doctor's prognosis, if one is guided by the otherwise irrelevant devaluation of reward, one is more likely to choose incorrectly. This paradigm, especially with the modification made in Study 3, allows inaccurate reward devaluation to be identified. Providing this information to someone who is happy- or hope-averse will, over time, help make devaluative biases more salient and less virulent for them.

From a clinical standpoint, there are three key contributions of this work. First, taken together, the three studies translate objective measurement of reward devaluation in depression from millisecond differences on an antiquated, attentional bias task to performance-based differences on an easily administered and conceptually clear valence selection test. This greatly increases the clinical utility of reward devaluation and focuses on more precise cognitive/affective mechanisms than the conscious endorsement of happiness aversion.

Second, the implications for related clinical treatments that focus on patient expectations (e.g., Kube et al., 2019, 2020; Rief & Joormann, 2019) are potentially critical. During cognitive treatments that focus on expectation change, clinicians attempt to alter prior adverse expectations by presenting disconfirmatory evidence. As an example, if an individual who is depressed holds the expectation that "the world is an unforgiving place," the clinician would work with the patient to interrogate and hopefully adopt the evidence that the world was not, singularly, unforgiving but a far more chaotic environment that can yield good things ("I may be forgetting something, but didn't you just tell me that someone stopped to change your tire for no pay when you were broken down in a storm?"). Nondepressed individuals would likely adopt this information and alter their prior expectations to some extent. Depressed individuals who are particularly hard to treat, however, will exhibit immunization strategies. That is, they have established strategies that discount potentially positive information so as to maintain their prior beliefs in a negative environment. The novel VST and set of findings show the extent to which a person might go to maintain such prior beliefs and outline a precise mechanism by which such immunization takes place. Honing in on the specific mechanism by which one maintains a depressive state allows for more personalized interventions.

Third, the implications for clinical treatment are potentially profound for individuals exhibiting reward devaluation, as identified in Study 3. If a person is at the extreme point of choosing incorrect options in the service of maintaining their devaluative environment, that person is similarly likely to avoid positivity-focused treatment options that might otherwise work best for them. For example, some recent work has demonstrated this by showing that the FHS predicts an aversion to positivity-focused treatments in comparison to other treatment options (Bryant et al., 2023). Individuals ranging from high to low on fear of happiness were given two treatment descriptions, one describing positive affect treatment (Craske et al., 2016) and one describing psychodynamic treatment, a comparison control without as much explicit positivity. In two studies, individuals with higher fear of happiness were more likely to choose a psychodynamic treatment in comparison to positive affect treatment despite positivity-focused treatments being constructed with them in mind. Incorporating inoculative psychoeducation about reward devaluation in concert with the valence selection task early in treatment might provide ways to better fit positivity treatments to those who avoid positivity.

Finally, from an educational perspective, although the effect sizes are small, they are highly relevant for educational practice where effect sizes of this magnitude are the norm and not the exception (Kraft, 2020). Sentence completion tasks are often used as a way to practice reading comprehension skills as well as assess them (Hutt et al., 2021, 2024; McCray & Brunfaut, 2018). Here, we mimicked their structure for the VST, as we felt this had a direct implication for education as it is rife with reward-based systems and affectively charged materials. Students who are currently experiencing a fear of happiness may be at a great disadvantage if they are willing to be incorrect or withhold their participation to avoid positivity. Perhaps adaptive and personalized educational technologies can help mitigate some of these issues, but more work is also needed in this space to understand the boundaries and consequences of reward devaluation over time. The fact that depression is linked to negative educational outcomes (Fergusson & Woodward, 2002; Fernandes et al., 2021; Quiroga et al., 2013; Wickersham et al., 2021) may be partly related to some of these behavioral changes and avoidance tendencies.

Limitations and Next Steps

It is important to bring attention to a few limitations of this work. We utilized Prolific to generate our sample. To improve the generalizability of our results, future studies should prioritize larger sample sizes and incorporate more diverse populations, particularly clinical populations.

There was also a lack of actual reward incorporated into the VST. RDT suggests that depressed individuals devalue positive information and *reward*. Future studies could incorporate rewards utilizing a gambling task. In this task, participants would be presented with rewards and gambling options that are either associated with a higher likelihood of earning more rewards or a higher likelihood of losing rewards (e.g., Pizzagalli et al., 2005; Treadway et al., 2009). Would participants who experience reward devaluation be more inclined to make riskier decisions to avoid the positive outcome of obtaining more rewards?

Additionally, we observed a ceiling effect in Study 3, such that participants were highly likely to select the correct answers given that the task was designed to be extremely easy. Although we observed a replicated effect of reward devaluation using this set of questions, it would be valuable for future research to investigate whether this effect intensifies when participants must invest more effort in solving a problem, such as extracting numerical details from a word problem while grappling with the inherent positivity of the context. Finally, our hypotheses and analysis plans were not preregistered before conducting these studies. However, our findings were replicated across three studies, reducing the possibility of Type 1 and 2 errors.

Conclusion

The current studies provide strong evidence suggesting that individuals exhibiting depressive symptoms, specifically those who experience a fear of happiness, are more likely to avoid positive information. Moreover, they do so even when it means being incorrect. Looking ahead, the idea of reward devaluation seems much more important in practical spaces than ever before; future work focusing on how devaluation manifests and ways to support patients, workers, and students who avoid positivity could be tremendously

helpful for the large portion of the population who experiences depression.

Constraints on Generality

Our findings provide evidence of reward devaluation as outlined by RDT (Winer & Salem, 2016) in participants who report higher rates of depressive symptomatology. Given that this attendance to negativity and avoidance of positivity was replicated across three studies and prevailed even when the avoidance of positivity meant that participants had to be blatantly incorrect, we expect our results to generalize across other populations with similar depressive symptomatology. However, as stated in our limitations, these studies were conducted online with the only major constraints being that participants had to be 18 years or older and fluent English speakers. We do not yet have evidence that findings could be replicated in more diverse samples or outside of laboratory-like settings.

References

Bartoszek, G., & Winer, E. S. (2015). Spider-fearful individuals hesitantly approach threat, whereas depressed individuals do not persistently approach reward. *Journal of Behavior Therapy and Experimental Psychiatry*, *46*, 1–7. https://doi.org/10.1016/j.jbtep.2014.07.012

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Beck, A. T., & Clark, D. A. (1991). Anxiety and depression: An information processing perspective. In R. Schwarzer & R. A. Wicklund (Eds.), *Anxiety and self-focused attention* (pp. 41–54). Harwood Academic Publishers.

Beck, A. T., Steer, R. A., & Brown, G. (1996). *Beck Depression Inventory–II (BDI-II)* [Database record]. APA PsycTests. https://doi.org/10.1037/t00742-000

Bitsko, R. H. (2022). Mental health surveillance among children—United States, 2013–2019. *MMWR Supplements*, *71*(2), 1–42. https://doi.org/10.15585/mmwr.su7102a1

Bryant, J. S., Gallagher, M. R., Collins, A. C., & Winer, E. S. (2023). Individuals fearing positivity do not perceive positive affect treatments as strong fits: A novel experimental finding and replication. *Journal of Behavior Therapy and Experimental Psychiatry*, *79*, Article 101830. https://doi.org/10.1016/j.jbtep.2022.101830

Calafiore, C., Collins, A. C., Miller, J. A. M., Watson, J. C., & Winer, E. S. (2024). Examining the unique and interactive impacts of anhedonia and fear of happiness on depressive symptoms. *Journal of Affective Disorders Reports*, *15*, Article 100702. https://doi.org/10.1016/j.jadr.2023.100702

Ciuhan, G. C., & Iliescu, D. (2021). Depression and learning problems in children: Executive function impairments and inattention as mediators. *Acta Psychologica*, *220*, Article 103420. https://doi.org/10.1016/j.actpsy.2021.103420

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Routledge. https://doi.org/10.4324/9780203771587

Collins, A. C., Lass, A. N. S., & Winer, E. S. (2023). Negative self-schemas and devaluation of positivity in depressed individuals: A moderated network analysis. *Current Psychology*, *42*(36), 32566–32575. https://doi.org/10.1007/s12144-023-04262-4

Collins, A. C., & Winer, E. S. (2024). Self-Referential processing and depression: A systematic review and meta-analysis. *Clinical Psychological Science*, *12*(4), 721–750. https://doi.org/10.1177/21677026231190390

Craske, M. G., Meuret, A. E., Ritz, T., Treanor, M., & Dour, H. J. (2016). Treatment for anhedonia: A neuroscience driven approach. *Depression and Anxiety*, *33*(10), 927–938. https://doi.org/10.1002/da.22490

Data and Statistics on Children's Mental Health | CDC. (2022). Centers for Disease Control and Prevention. https://www.cdc.gov/children-mental-health/data-research/?CDC_AAref_Val=https://www.cdc.gov/childrensmentalhealth/data.html

Deng, W., Everaert, J., Creighton, M., Bronstein, M. V., Cannon, T., & Joormann, J. (2022). Developing a novel assessment of interpretation flexibility: Reliability, validity and clinical implications. *Personality and Individual Differences*, *190*, Article 111548. https://doi.org/10.1016/j.paid.2022.111548

Eshel, N., & Roiser, J. P. (2010). Reward and punishment processing in depression. *Biological Psychiatry*, *68*(2), 118–124. https://doi.org/10.1016/j.biopsych.2010.01.027

Everaert, J., Bronstein, M. V., Cannon, T. D., & Joormann, J. (2018). Looking through tinted glasses: Depression and social anxiety are related to both interpretation biases and inflexible negative interpretations. *Clinical Psychological Science*, *6*(4), 517–528. https://doi.org/10.1177/2167702617747968

Everaert, J., Podina, I. R., & Koster, E. H. W. (2017). A comprehensive meta-analysis of interpretation biases in depression. *Clinical Psychology Review*, *58*, 33–48. https://doi.org/10.1016/j.cpr.2017.09.005

Fergusson, D. M., & Woodward, L. J. (2002). Mental health, educational, and social role outcomes of adolescents with depression. *Archives of General Psychiatry*, *59*(3), 225–231. https://doi.org/10.1001/archpsyc.59.3.225

Fernandes, M. D. S. V., Mendonça, C. R., da Silva, T. M. V., & Noll, M. (2021). The relationship between depression and quality of life in students and the academic consequences: Protocol for a systematic review with meta-analysis. *International Journal of Educational Research*, *109*, Article 101812. https://doi.org/10.1016/j.ijer.2021.101812

Fried, E. I., & Nesse, R. M. (2015). Depression sum-scores don't add up: Why analyzing specific depression symptoms is essential. *BMC Medicine*, *13*(1), Article 72. https://doi.org/10.1186/s12916-015-0325-4

Gallagher, M. R., Salem, T., & Winer, E. S. (2024). When hope springs a leak: Aversion to positivity as a key to understanding depressed persons. *Current Psychology*, *43*(8), 7564–7577. https://doi.org/10.1007/s12144-023-04917-2

Gotlib, I. H., & Joormann, J. (2010). Cognition and depression: Current status and future directions. *Annual Review of Clinical Psychology*, *6*(1), 285–312. https://doi.org/10.1146/annurev.clinpsy.121208.131305

Gotlib, I. H., & McCabe, S. B. (1992). An information-processing approach to the study of cognitive functioning in depression. *Progress in Experimental Personality & Psychopathology Research*, *15*, 131–161.

Humensky, J., Kuwabara, S. A., Fogel, J., Wells, C., Goodwin, B., & Voorhees, B. W. V. (2010). Adolescents with depressive symptoms and their challenges with learning in school. *The Journal of School Nursing*, *26*(5), 377–392. https://doi.org/10.1177/1059840510376515

Hutt, S., Krasich, K., Brockmole, R. J., & D'Mello, S. K. (2021). Breaking out of the lab: Mitigating mind wandering with gaze-based attention-aware technology in classrooms. *Proceedings of the 2021 CHI conference on human factors in computing systems* (pp. 1–14). Association for Computing Machinery. https://doi.org/10.1145/3411764.3445269

Hutt, S., Wong, A., Papoutsaki, A., Baker, R. S., Gold, J. I., & Mills, C. (2024). Webcam-based eye tracking to detect mind wandering and comprehension errors. *Behavior Research Methods*, *56*(1), 1–17. https://doi.org/10.3758/s13428-022-02040-x

Jordan, D. G., Collins, A. C., Dunaway, M. G., Kilgore, J., & Winer, E. S. (2021). Negative affect interference and fear of happiness are independently associated with depressive symptoms. *Journal of Clinical Psychology*, *77*(3), 646–660. https://doi.org/10.1002/jclp.23066

Joshanloo, M., Lepshokova, Z. K., Panyusheva, T., Natalia, A., Poon, W.-C., Yeung, V. W., Sundaram, S., Achoui, M., Asano, R., Igarashi, T., Tsukamoto, S., Rizwan, M., Khilji, I. A., Ferreira, M. C., Pang, J. S., Ho, L. S., Han, G., Bae, J., & Jiang, D.-Y. (2014). Cross-Cultural Validation of Fear of Happiness Scale across 14 national groups. *Journal of Cross-Cultural Psychology*, *45*(2), 246–264. https://doi.org/10.1177/0022022113505357

Kraft, M. A. (2020). Interpreting effect sizes of education interventions. *Educational Researcher*, *49*(4), 241–253. https://doi.org/10.3102/0013189X20912798

Kube, T. (2023). Biased belief updating in depression. *Clinical Psychology Review*, *103*, Article 102298. https://doi.org/10.1016/j.cpr.2023.102298

Kube, T., Rief, W., Gollwitzer, M., Gärtner, T., & Glombiewski, J. A. (2019). Why dysfunctional expectations in depression persist—Results from two experimental studies investigating cognitive immunization. *Psychological Medicine*, *49*(9), 1532–1544. https://doi.org/10.1017/S0033291718002106

Kube, T., Schwarting, R., Rozenkrantz, L., Glombiewski, J. A., & Rief, W. (2020). Distorted cognitive processes in major depression: A predictive processing perspective. *Biological Psychiatry*, *87*(5), 388–398. https://doi.org/10.1016/j.biopsych.2019.07.017

Long, J. A. (2019). *Interactions: Comprehensive, user-friendly toolkit for probing interactions* (R package Version 1.1.0) [Computrer software]. https://cran.r-project.org/package=interactions

MacLeod, C., Mathews, A., & Tata, P. (1986). Attentional bias in emotional disorders. *Journal of Abnormal Psychology*, *95*(1), 15–20. https://doi.org/10.1037/0021-843X.95.1.15

McCabe, S. B., & Gotlib, I. H. (1995). Selective attention and clinical depression: Performance on a deployment-of-attention task. *Journal of Abnormal Psychology*, *104*(1), 241–245. https://doi.org/10.1037/0021-843X.104.1.241

McCray, G., & Brunfaut, T. (2018). Investigating the construct measured by banked gap-fill items: Evidence from eye-tracking. *Language Testing*, *35*(1), 51–73. https://doi.org/10.1177/0265532216677105

Pizzagalli, D. A., Sherwood, R. J., Henriques, J. B., & Davidson, R. J. (2005). Frontal brain asymmetry and reward responsiveness: A source-localization study. *Psychological Science*, *16*(10), 805–813. https://doi.org/10.1111/j.1467-9280.2005.01618.x

Quiroga, C. V., Janosz, M., Bisset, S., & Morin, A. J. S. (2013). Early adolescent depression symptoms and school dropout: Mediating processes involving self-reported academic competence and achievement. *Journal of Educational Psychology*, *105*(2), 552–560. https://doi.org/10.1037/a0031524

R Core Team. (2021). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Rief, W., & Joormann, J. (2019). Revisiting the cognitive model of depression: The role of expectations. *Clinical Psychology in Europe*, *1*(1), Article e32605. https://doi.org/10.32872/cpe.v1i1.32605

Ritsner, M. S., & Ratner, Y. (2019). Predicting predischarge anhedonia among inpatients with schizophrenia and schizoaffective disorders: A large-scale analysis. *Journal of Nervous and Mental Disease*, *207*(1), 12–21. https://doi.org/10.1097/NMD.0000000000000923

Rush, A. J., Gullion, C. M., Basco, M. R., Jarrett, R. B., & Trivedi, M. H. (1996). The Inventory of Depressive Symptomatology (IDS): Psychometric properties. *Psychological Medicine*, *26*(3), 477–486. https://doi.org/10.1017/S0033291700035558

Rush, A. J., Trivedi, M. H., Carmody, T. J., Ibrahim, H. M., Markowitz, J. C., Keitner, G. I., Kornstein, S. G., Arnow, B., Klein, D. N., Manber, R., Dunner, D. L., Gelenberg, A. J., Kocsis, J. H., Nemeroff, C. B., Fawcett, J., Thase, M. E., Russell, J. M., Jody, D. N., Borian, F. E., & Keller, M. B. (2005). Self-reported depressive symptom measures: Sensitivity to detecting change in a randomized, controlled trial of chronically depressed, nonpsychotic outpatients. *Neuropsychopharmacology*, *30*(2), 405–416. https://doi.org/10.1038/sj.npp.1300614

Rush, A. J., Trivedi, M. H., Ibrahim, H. M., Carmody, T. J., Arnow, B., Klein, D. N., Markowitz, J. C., Ninan, P. T., Kornstein, S., Manber, R., Thase, M. E., Kocsis, J. H., & Keller, M. B. (2003). The 16-Item Quick Inventory of Depressive Symptomatology (QIDS), clinician rating (QIDS-C), and self-report (QIDS-SR): A psychometric evaluation in patients with chronic major depression. *Biological Psychiatry*, *54*(5), 573–583. https://doi.org/10.1016/S0006-3223(02)01866-8

Salem, T., Winer, E. S., & Nadorff, M. R. (2018). Combined behavioural markers of cognitive biases are associated with anhedonia. *Cognition and Emotion*, *32*(2), 422–430. https://doi.org/10.1080/02699931.2017.1307808

Śniadach, J., Szymkowiak, S., Osip, P., & Waszkiewicz, N. (2021). Increased depression and anxiety disorders during the COVID-19 pandemic in children and adolescents: A literature review. *Life*, *11*(11), Article 1188. https://doi.org/10.3390/life11111188

Teachman, B. A., Joormann, J., Steinman, S. A., & Gotlib, I. H. (2012). Automaticity in anxiety disorders and major depressive disorder. *Clinical Psychology Review*, *32*(6), 575–603. https://doi.org/10.1016/j.cpr.2012.06.004

Treadway, M. T., Buckholtz, J. W., Schwartzman, A. N., Lambert, W. E., & Zald, D. H. (2009). Worth the 'EEfRT'? The effort expenditure for rewards task as an objective measure of motivation and anhedonia. *PLOS ONE*, *4*(8), Article e6598. https://doi.org/10.1371/journal.pone.0006598

Whitton, A. E., Treadway, M. T., & Pizzagalli, D. A. (2015). Reward processing dysfunction in major depression, bipolar disorder and schizophrenia. *Current Opinion in Psychiatry*, *28*(1), 7–12. https://doi.org/10.1097/YCO.0000000000000122

Wickersham, A., Sugg, H. V. R., Epstein, S., Stewart, R., Ford, T., & Downs, J. (2021). Systematic review and meta-analysis: The association between child and adolescent depression and later educational attainment. *Journal of the American Academy of Child & Adolescent Psychiatry*, *60*(1), 105–118. https://doi.org/10.1016/j.jaac.2020.10.008

Winer, E. S., & Salem, T. (2016). Reward devaluation: Dot-probe meta-analytic evidence of avoidance of positive information in depressed persons. *Psychological Bulletin*, *142*(1), 18–78. https://doi.org/10.1037/bul0000022

Winer, E. S., Veilleux, J. C., & Ginger, E. J. (2014). Development and validation of the Specific Loss of Interest and Pleasure Scale (SLIPS). *Journal of Affective Disorders*, *152–154*, 193–201. https://doi.org/10.1016/j.jad.2013.09.010

Yiend, J. (2010). The effects of emotion on attention: A review of attentional processing of emotional information. *Cognition and Emotion*, *24*(1), 3–47. https://doi.org/10.1080/02699930903205698

Received October 25, 2023
Revision received September 27, 2024
Accepted October 16, 2024 ■