Current Psychology (2024) 43:7564–7577
https://doi.org/10.1007/s12144-023-04917-2

When hope springs a leak: Aversion to positivity as a key to understanding depressed persons

Michael R. Gallagher[1] · **Taban Salem**[2] · **E. Samuel Winer**[3]

Accepted: 23 June 2023 / Published online: 13 July 2023

Abstract

Individuals with depression tend to focus on negativity and interpret neutral situations in a negative light. Cognitive theories commonly posit that depressed persons' focus on negative stimuli leaves them incapable of prioritizing positive emotional information. Reward Devaluation Theory (RDT; Winer and Salem, 2016), which stipulates that depressed persons not only focus on negativity, but also exhibit a *systematic and motivated avoidance* of positive and potentially rewarding stimuli, offers an alternative theoretical framework to understanding cognitive/affective biases in depression. We here unpack the theoretical underpinnings of RDT, reviewing the empirical evidence surrounding positivity avoidance and depression. Studies using cognitive/behavioral tasks, as well as those examining anticipatory and responsive devaluation strategies, are summarized. Future directions in the measurement of RDT, including expanded investigation of self-referential processing, are introduced for consideration. The clinical implications of RDT, which are potentially profound, are also discussed.

Keywords Reward devaluation theory · Happiness · Depression · Anhedonia

Individuals with elevated depressive symptoms tend to experience negative stimuli as more salient (Gotlib & Joormann, 2010). Positive stimuli are generally less salient for depressed individuals. Recently, the questions of how and why positivity is less available to depressed individuals have received increased theoretical and empirical scrutiny. What has emerged from these novel investigations is a complex picture that suggests that the devaluation of positivity – the opposite, not merely the lack, of salience – is also an influential explanatory mechanism in the etiology and maintenance of some core depressive symptoms such as anhedonia. Moreover, devaluing positivity, due to hope and positivity *aversion*, is potentially a definitional feature of many depressed states that is more predictive than broad negativity salience.

✉ Michael R. Gallagher
mrg463@msstate.edu

[1] Mississippi State University, P.O. Box 6161, Mississippi State, MS 39762, USA

[2] Millsaps College, Jackson, USA

[3] The New School for Social Research, New York, NY, USA

Helping to usher in this recent research impetus, Winer and Salem (2016) conducted a meta-analysis examining attentional biases related to positive and negative stimuli in depressed and anxious individuals. They found that individuals with elevated symptoms of depression demonstrated between-subjects avoidance of positive stimuli, as evidenced by slower reaction times on the dot-probe task, compared to other clinical and control groups. Furthermore, depressed individuals also exhibited within-subject avoidance of positive stimuli: when presented with a positive and a neutral cue simultaneously, they systematically directed their attention away from the positive cue, choosing instead to focus on the less positive (i.e., neutral) information. Importantly, unlike other emotion-based attentional biases, this avoidance of positive information by depressed individuals represents a reversal of normative biases, which are exaggerated toward positive information in other groups (Pool et al., 2016). This meta-analysis provided an empirical base for Reward Devaluation Theory (RDT; Winer and Salem, 2016).

Reward Devaluation Theory posits that depressed individuals often actively avoid or inhibit prospectively rewarding self-relevant stimuli. Through repeated pairings of prospectively rewarding stimuli and negative outcomes,

such as disappointment or failure, individuals may learn to automatically inhibit pursuit of positivity to evade the negative consequences they have come to expect (Frischen et al., 2012; Winer et al., 2011). Thus, positive information is perceived as prospectively meretricious in the sense that it appears rewarding when it is in fact harmful. As a result of these learned experiences, individuals with depression may not only lose hope, but become *hope-averse*. An intriguing hypothesis resulting from this framework is that this would mean many depressed individuals would view potentially positive and self-relevant information as less plausible and more potentially harmful than even genuinely negative and explicitly depressogenic information (i.e., stimuli associated with sadness or explicit dejection). Hope, in this case, would feel dangerous to experience. For many people, hope may provide positive thoughts and emotions because positive outcomes or improvements appear to be on the horizon. However, hope may appear deceitfully dangerous for some depressed individuals. Depressed persons anticipate negative information (Beck & Bredemeier, 2016). When faced with negative information such as a realistic roadblock in goal pursuit or a sad memory, the information is likely hurtful, but also anticipated and familiar based on the person's chronic expectations of their environment. Hope, on the other hand, has come to be associated with a risk of unanticipated negative outcomes (e.g., hoping for goal achievement only to be let down), which is not only hurtful but also threatens equilibrium. As a result, hope, an objectively positive construct to many, represents a signal of danger and a letting of one's guard down to some depressed individuals.

Bolstering their argument, Winer and Salem (2016) noted studies showing that positive social goals that are coactivated with negative information leads to cessation of those goals (Aarts et al., 2007; see also Joormann & Gotlib, 2010; Veling et al., 2008). Moreover, Winer and Salem noted that whereas devaluation of rewarding experiences may inhibit reward seeking behavior, repeated inhibition of goal pursuit results in further devaluation of a prospective reward and the positive emotions with which that goal had formerly been associated (Frischen et al., 2012). As an example, Veling and Aarts (2009) conducted a study with 66 undergraduate participants, in which participants were made thirsty, followed by a go/no-go task, in which some participants were instructed (via no-go cues) to withhold approach behaviors to a rewarding stimulus (i.e., bottle of water). They found that repeated inhibition of reward seeking behavior reduced the subjective reward value placed on the rewarding stimulus. Thus, it appears that not only can incongruent pairings of positive stimuli with negative outcomes lead to an avoidance of positivity, but that additional inhibition of rewarding situations can lead to further reward devaluation.

So, RDT is a framework explaining a core mechanism, counterintuitive to much of the cognitive literature, that could further conceptual understanding of depression and produce novel hypotheses. The goal of this paper is to review the influx of research (e.g., Ao et al., 2020; Ballard et al., 2017; Barkus, 2021; Bean et al., 2022; Blanco et al., 2019; Bryant et al., 2023; Elgersma et al., 2018; Ellis et al., 2019; Everaert et al., 2020; Ji et al., 2017; Jordan et al., 2021; Millgram et al., 2020; Vazquez et al., 2018; Yaroslavsky et al., 2019) relevant to the underlying mechanisms and clinical implications of RDT since its release in 2016. How has the theory fared? What changes demand accommodation in the face of this new body of evidence and what additional research is needed to further the understanding of RDT and depression?

Cognitive/behavioral tasks

The majority of the original evidence for RDT arose through the dot-probe task (Winer & Salem, 2016), and further research has expanded on the main principles of RDT through various other cognitive/behavioral tasks. Bartoszek and Winer (2015) examined the approach-avoidance component of RDT through utilization of the Approach-Avoidance Task (AAT) in concert with a novel analysis of *duration* of pulling or pushing various stimuli (as opposed to simply analyzing response time). The AAT presents emotional pictures and instructs respondents to either pull the image towards themselves or push the image away via joystick responses and a zoom-in/zoom-out effect in response to pulling and pushing respectively. There were 120 participants included in the study, pre-selected into spider-fearful, depressed, and control groups. Bartoszek and Winer (2015) found that individuals with depression exhibited a diminished approach motivation to positive stimuli compared to neutral stimuli, as well as in comparison to the control group, as evidenced by shorter durations of pulling positive stimuli towards them. Depressed participants were less likely to persist in approaching positive stimuli in comparison to neutral stimuli. This reveals that depressed persons would rather continuously engage with or savor a neutral (i.e., less positive) stimulus than encounter positivity. Although this was interpreted as an absence of approach motivation, it is possible that depressed individuals were merely incapable of staying in the moment with positivity, which is a matter of avoidance or escape.

Salem et al. (2018) sought to examine emotional biases across a variety of cognitive domains, in order to gain a more accurate picture of the biased processing that exists in the realm of depression. The Combined Cognitive Bias Hypothesis (CCBH; Hirsch et al., 2006) posits that biased

processing likely occurs at the stages of attention, interpretation, and memory. The biases across these three levels of processing can interact in a causal chain, potentially magnifying the distortion of information processing that would occur via each of these mechanisms individually (Everaert et al., 2012). Under the premises of the CCBH, Salem et al. (2018) examined attention and memory biases via the dot-probe and the two-alternative forced choice task (2AFC; Snodgrass and Shevrin, 2006) respectively, in order to assess the combined biases in relation to anhedonia. There were 152 participants included in the study, which measured levels of anhedonia on a continuum. The 2AFC is a word recognition task, in which participants are presented with a target word, typically of emotional meaning, for a brief period. Participants are then presented with two word choices and are instructed to indicate which word they previously viewed. Anhedonia is a multidimensional construct that is inherently linked to positivity avoidance. From an RDT perspective, anhedonia, loss of interest or inability to experience pleasure in formerly enjoyable experiences, may result from the devaluative process applied to previously positive experiences. The results of the study revealed that biases in processing (favoring negative and avoiding/inhibiting positive stimuli) across the two tasks were predictive of anhedonia, although neither task reached significance on its own. These results suggest that biased processing of emotional information across domains is a better predictor of anhedonia than either bias alone. So, avoidance of positivity through attentional avoidance and memory salience both appear to be influential in understanding anhedonia.

In another study looking at the impact of childhood trauma on attentional biases using the dot-probe task with faces, Günther and colleagues (2015) collected data from 45 participants, all of whom were depressed. Thus, this study did not provide between-group comparisons of depressed individuals to healthy controls, but rather within-subjects measures of bias when comparing congruent to incongruent trials, and between-subjects measures of bias in relation to levels of childhood trauma and depressive symptoms within a depressed sample. There was not a significant bias away from happy faces for the sample as a whole, perhaps reflecting that given the heterogeneity of depression (Fried & Nesse, 2015), not all depressed persons can be expected to show evidence of RDT. However, a number of interesting findings emerged with regard to happy faces, including that bias for happy faces was negatively correlated with bias for sad faces—those who were more attracted to sad faces were more repelled by happy faces (Günther et al., 2015). Further, higher levels of emotional abuse and emotional neglect in childhood were correlated with bias away from happy faces. These findings appear to partially support RDT. At least some of the depressed participants showed evidence of reward devaluation, and reward devaluation was linked to a history of negative and traumatic relationships, which would reasonably teach a person not to trust the smiles and seemingly positive advances of others.

Ao et al. (2020) further examined the processing of emotional faces through a variation of the dot-probe task in a sample that consisted of 25 participants with Major Depressive Disorder (MDD) and 25 controls. The stimuli (i.e., faces) were presented for 100 ms, as attentional engagement was the measurement of interest. Ao et al. (2020) found that the depressed group demonstrated a bias in attentional engagement of negative faces as opposed to neutral faces (i.e., vigilance to negativity). Whereas the control group was faster at identifying the probe following happy faces, the depressed group did not display such a bias, which was interpreted as a lack of initial engagement with positivity. This is an interesting finding suggesting processing-bias differences in as little as 100 ms. However, given the short duration of the trials (i.e., 100 ms), it is difficult to conclude whether the lack of attentional engagement to positivity in the depressed group was due to the positive stimuli being perceived as less emotionally salient, resulting in a bias toward the more negative stimuli (i.e., neutral faces) or rather due to an automatic avoidance of the rewarding stimuli (i.e., happy faces).

Ruhe and colleagues compared 69 recurrently, remitted depressed individuals to 43 controls on a number of tasks (Ruhe et al., 2019). First, participants were asked to categorize by valence and then later freely recall a list of personality adjectives. The instructions were designed to evoke self-referential processing. For the second task participants were shown blends of emotional and neutral faces and asked to categorize the emotion as quickly as possible. For accurate responses, after correcting for baseline symptoms of depression the formerly depressed group was slower in categorizing positive words than the control group. There was also a significant valence by group by accuracy interaction such that the remitted depressed group was slower when they inaccurately categorized positive words in comparison to controls, which may reflect greater cognitive engagement and meaning-based miscategorization, as opposed to miscategorization due to distraction. In other words, these individuals may have experienced increased ambivalence regarding self-referential positive words. These outcomes suggest that reward devaluation may persist even once symptoms have remitted. There were no differences between remitted depressed and control participants in memory for positive words. There was also no significant group difference in categorization of positive faces on the second task, possibly indicating that interpretation recovers more quickly from the effects of reward devaluation than some other cognitive processes.

Ellis et al. (2019) examined the extent to which the behavioral tenets of approach and avoidance motivation were present when examining neural reward processing (see also Messerotti Benvenuti et al., 2017). Ellis et al. (2019) examined event-related potentials, specifically reward positivity (RewP), which is a measure of neural activation in the frontocentral region in response to rewarding feedback acquired via EEG (Proudfit et al., 2015). Twenty adolescent-aged individuals were recruited with varying levels of depressive symptomatology and severity. Adolescents' RewP were examined while they completed an affective Posner task, in order to examine neural reward sensitivity. Participants also completed self-report measures to quantify approach and avoidance behaviors. Ellis et al. (2019) found that a longer RewP latency (i.e., later onset) was related to depressive symptomatology, but this relation only emerged when avoidance motivation scores were taken into account. As noted by the authors, these results may indicate that the reward avoidance behaviors outlined in RDT may be a core construct in understanding the interaction between reward sensitivity and depression; however, the limited sample size is a notable caveat.

Anhedonia

Depression is extremely heterogeneous with over 1,000 identified unique symptom profiles (Fried & Nesse, 2015). Reward devaluation likely does not occur in all cases of depression, but those who experience anhedonia as their primary symptom likely engage in reward devaluative processes to some extent, whether explicitly as a form of self-protection or via a learned automatic bias. Therefore, it is important to consider positivity avoidance from a transdiagnostic perspective and within different contexts. For example, Winer et al. (2019) review the different types of anhedonia, as well as the numerous factors that contribute to anhedonia. Anticipatory and consummatory anhedonia may influence motivation differently and influence avoidance in unique ways. Physical and social anhedonia, in which social anhedonia refers to not deriving pleasure from social activities, whereas physical anhedonia is not experiencing pleasure from physical sensations, would have different implications for RDT, with RDT more relevant to social settings. Finally, taking a more fine-grained approach, Winer et al. (2019) distinguish between trait level anhedonia and recent state level changes in anhedonic features, as these have different etiologies and may uniquely influence functioning.

Recent research has examined how anhedonia interacts with avoidance of positivity and other symptoms of depression. Jordan et al. (2018) examined the role of anhedonia in explaining the relationship between fear of positive evaluation (FPE; Weeks et al., 2012)—trepidation or uneasiness surrounding favorable evaluation or praise—and depressive symptoms. Specifically, the study analyzed differences between anticipatory and consummatory anhedonia. Jordan et al. (2018) note that anticipatory anhedonia, not anticipating or looking forward to prospective positive experiences or pleasurable events, has previously been associated with a lack of motivation to pursue reward in depressed individuals (Sherdell et al., 2012). Consummatory anhedonia (i.e., not experiencing pleasure while engaging in pleasurable events in the moment) has not typically been associated with reward motivation given that the reward has already been realized or has already occurred (Jordan et al., 2018). These differential findings in relation to reward motivation may be due to the fact that there is uncertainty surrounding potential rewards in the context of anticipatory anhedonia—that the pursuit of the hoped-for reward may ultimately result in a negative outcome. In light of this uncertainty, individuals may perceive reward cues as threatening and exhibit diminished reward motivation.

Jordan et al. (2018) included 196 participants who were recruited online, as they examined participant responses at three time points over a four month period, and found that anticipatory anhedonia mediated the relationship between FPE and depressive symptomology, whereas consummatory anhedonia did not mediate the same temporal relationship. Fear of being positively evaluated or praised may lead to inhibition of the tendency to strive for praise or to look forward to positive social events in which positive evaluation may occur. This inhibition may, in turn, lead to devaluation of those prospectively positive social events, and resultant social anhedonia. The development of anhedonia may then lead to or exacerbate other depressive symptoms.

The lack of or prospective reversal of reward motivation that characterizes some expressions of anhedonia is hypothesized to be one of the core factors resulting from reward devaluation. Therefore, recent research has specifically examined reward motivation through a variety of methodologies to better understand its complexity. For example, Bryant et al. (2017) utilized the Effort-Expenditure for Reward Task (EEfRT; Treadway et al., 2009), a task in which participants have the choice to complete simpler tasks for smaller monetary gains or more difficult tasks for the opportunity to win larger monetary gains, to assess reward motivation. The EEfRT was analyzed together with anhedonia and action-orientation, a potential protective factor against anhedonia. Action orientation is the ability to upregulate positive emotions in the face of stress (Kuhl, 1994). As described by Kuhl, action orientation is effectively an inverse motivational state to reward devaluation; individuals who are action oriented summon positivity in the face of negativity. A total of 76

participants completed the study, who were screened for either high or low action orientation, while also demonstrating some level of anticipatory anhedonia. The study found that when anhedonia was high, individuals exhibited diminished reward motivation, meaning action orientation did not act as a protective factor against reward devaluation. However, when anhedonia was low, action orientation did act as a protective factor against reward devaluation, as evidenced by the association between action orientation and the high cost/high reward option on the EEfRT. Action orientation appears to be a valid protective factor to reward devaluation at low levels of anhedonia but is unable to combat the devaluative process that occurs at high levels of anhedonia. The results reflect the impact of anhedonia and its ability to overwhelm other protective factors that might otherwise help to preserve reward motivation. Those who have lost interest in previously pleasurable experiences, perhaps through learned experiences, are less likely to seek prospectively rewarding situations in the future.

Winer et al. (2017) also sought to better understand the role of anhedonia, this time specifically within the context of anxiety and depression. This three-part longitudinal study assessed the role anhedonia plays in connecting anxiety and depression, while considering other important factors, such as relinquishment of enjoyment due to avoidance. Many individuals avoid activities due to their potential to provoke anxiety, but not all avoided activities result in a relinquishment of enjoyment. Therefore, Winer et al. (2017) examined not only how often one avoids activities due to anxiety, but also the degree of enjoyment the individual relinquishes by avoiding the activity. Study 1 included 109 participants with high anxiety scores. The results of study 1 revealed that the degree of prospective enjoyability relinquishment served as a moderator in the relationship between anxiety and depression, such that high levels of enjoyability relinquishment related to anxiety were associated with increased levels of depressive symptoms, even among participants pre-selected to have high levels of anxiety. Study 3 used a longitudinal design consisting of 196 participants in the first model. The first model of study 3 found that anhedonia temporally mediated the relationship between anxiety and depression. Anxiety predicted anhedonia, which predicted depression at both 5 and 11 month follow ups. A second mediational model consisting of 165 participants, in which anhedonia was entered as the predictor and anxiety as the mediator variable, also predicted depression. Regardless of the order of the model, it is clear that anhedonia is a core factor in understanding the relationship between anxiety and depression. Anhedonia and anxiety may develop in tandem. Increased levels of uncertainty regarding the outcomes of prospectively positive events may lead individuals to devalue and avoid positive, enjoyable experiences. This devaluation, relinquishment, or avoidance of positivity may be associated with increased feelings of anxiety surrounding loss of social support and the inability to engage in socially pleasurable activities.

Finally, given the importance of anhedonia in understanding depression specifically, and psychopathology more generally, Winer et al. (2014) examined the role of anhedonia in predicting suicidality in a psychiatric inpatient population consisting of 1,529 total participants, in which over half met criteria for a depressive disorder. Winer et al. (2014) found anhedonia to be a predictor of suicidality at both baseline and termination of treatment. Anhedonia remained a significant predictor of suicidality when accounting for other symptoms of depression (an effect size largely replicated by Ballard et al., 2017). The devaluative and avoidant tendencies that are prevalent in those who experience anhedonia may thus be important in understanding suicidality, depression, and psychopathology as a whole. However, further work parsing anhedonia to examine what types of anhedonia may relate to reward devaluation is still needed.

Fear of happiness

Fear of happiness, while similar to previously discussed constructs such as FPE and relinquishment of enjoyment due to avoidance, offers a unique perspective in understanding devaluative processes. FPE and enjoyability relinquishment both involve fear and avoidance of positive experiences (e.g., public praise, enjoyable activities), rather than the positive feeling associated with the experiences. Fear of happiness uniquely encompasses a fear of actually experiencing positive emotions, likely due to incongruent past pairings of happiness with aversive outcomes.

Research has focused on understanding the interactivity of fear of happiness with other cognitive/affective factors, as well as the overall impact fear of happiness has on one's functioning. Gilbert et al. (2014) examined the connectedness of fear of positive emotions, depression, and anxiety, as well as a number of other factors within a clinical setting of 52 participants experiencing moderate to extremely severe depression. Gilbert et al. (2014) found both fear of happiness and fear of compassion to be associated with depression, anxiety, and stress. In line with the tenets of RDT, due to past experiences happiness may be perceived as so threatening to certain individuals that they not only devalue the prospective "reward" of happiness, but rather avoid it at all costs.[1]

[1] Please see the work of Joshanloo and colleagues for a preeminent psychometrically valid and culturally considered review of fear of happiness (Joshanloo, 2013; Joshanloo & Weijers, 2014).

Further research has examined the pervasiveness of the effects of fear of happiness on individual functioning. Vanderlind et al. (2017) examined whether fear of happiness minimized the utility of positive mood enhancement strategies in individuals with depression. The sample consisted of 94 undergraduate students. Following a brief mood induction, participants were asked to recall positive autobiographical memories as part of a positive affect repairment. The results indicated that those who feared happiness recalled less positive autobiographical memories and exhibited less of an ability to repair both positive and negative affect than those who did not fear positivity (Vanderlind et al., 2017). The results of this study demonstrate the pervasiveness of fear of positive emotions, extending into the realm of emotion regulation and memory.

Recently, Jordan et al. (2021) examined major components of RDT, specifically avoidance of positivity, longitudinally in 476 participants through a variety of self-report measures. Jordan et al. (2021) examined negative affect interference (NAI; DePierro et al., 2018), negative affect related to trauma that infringes on positive experiences, as well as fear of happiness, in relation to depressive symptomatology. NAI and fear of happiness were both found to be uniquely and separately predictive of greater depression severity. These results could have emerged because NAI is a reactive form of positivity avoidance, given that it involves avoiding positive emotions and experiencing negative affect in response to positive events. Fear of happiness also encompasses avoidance of positivity, but does so by avoiding prospective happiness or positivity (see Barkus, 2021 for a theoretical comparison of RDT and NAI). Regardless, these findings illustrate the influential role positivity avoidance has on depression, but also point to the need for further understanding of distinct pathways to devaluation of positivity.

Theoretical summary

We have summarized evidence of slower, deliberate (fear of happiness and depression self-report relations) and fast-paced, implicit (e.g., perceptual sensitivity, the dot-probe at one second) effects relevant to RDT. In the past years, robust network analytic studies associated with RDT using self-reported variables have perhaps shifted the balance such that RDT might be interpreted as a theory of effortful, intentional processing. We here wish to explicitly note that RDT was in fact originally conceptualized as a theory about inhibition of positivity via attentional, perceptual, memory, and related cognitive/affective processes that occur, primarily, outside of conscious awareness. Although people evidencing reward devaluation likely do so deliberately in many instances, they likely also come to automatize devaluative responses to positivity such that the world that they experience consciously is unrealistically devoid of positive stimuli and experience.

Beyond the laboratory, the clinical implications of having a set of cognitive/affective processes working to recalibrate basic approach-related motivation toward reward to now operate *in reverse* resulting in avoidance of reward are profound. Consider that the therapeutic alliance, self-awareness, and fostering hope are the three most robust principles of change agreed upon via consensus by clinicians across therapeutic modalities (Twomey et al., 2023). Understanding that a client has an unconscious tendency to avoid potential positivity and associate it with threat (i.e., that the client is not just hopeless but hope-averse) has immediate implications for all three principles. Alliance is built by a therapist understanding more fully what the client is experiencing, the client's self-awareness of this unique implicit bias will allow them to come to gain agency over it, and hope can finally be fostered once the dangerousness and ambivalence that this person associates with hope is addressed.

Future directions

Self-reference

RDT is a self-referential framework. Some depressed individuals avoid positive information because of what it represents about themselves. However, the initial evidence base supporting RDT and the evidence summarized herein commonly does not explicitly reference the self.

Within the context of depression research, self-reference is typically operationalized by having individuals indicate whether or not traits (e.g., negative or positive) apply to them personally. Individuals with depression tend to rate more negative self-referential traits than positive (Lemoult & Gotlib, 2019), and negative self-referential thoughts are associated with depression severity (Disner et al., 2017; Phillips et al., 2010). One way these negative referential thoughts manifest is through cognitive biases to emotional stimuli. Kaiser et al. (2018) found that in a sample of 53 women, the MDD group attended to negative information more than positive, but this attentional pattern only surfaced when the stimuli were self-descriptive. These findings indicate that the emotional attention biases outlined previously may be restricted to or better explained in the context of self-relevant material.

While the majority of the literature surrounding self-referential processing has examined the effect of negative self-referential thoughts on attention and memory biases towards negative information, fewer studies have examined

the impact of referential processing on avoidance of positivity. Research has shown that individuals with depression tend to exhibit a deficit in the recall of positive information, but only when the content of the stimuli is self-relevant (Blaney, 1986; Dozois & Dobson, 2001; Gaddy & Ingram, 2014; Matt et al., 1992). Additionally, Dozois and Dobson (2001) examined the influence of self-reference within an emotional context with a clever task design. Participants were presented with a grid consisting of two axes, one of which measures self-reference and the other pertaining to valence. Participants read different adjectives and were asked to move the point on the grid to correspond with their perspective of the word regarding its valence and relation to self. Dozois and Dobson (2001) have used the distances on the grid to distinguish between depressed, anxious, and control groups.

There have also been a few studies examining the impact self-reference has on attentional avoidance of positivity. Ji et al. (2017) found that in a sample of 45 undergraduates who were screened for high and low depression symptomatology, individuals with greater depression severity displayed decreased attentional allocation to positive stimuli compared to a low depression group, but this disparity was only present when the content of the stimuli was self-relevant. Similar patterns were identified by Atchely et al. (2012) in a sample of 33 participants split into a MDD and control group, as they found that individuals with depression exhibited a deficit in the detection of positive person-referent words. These results illustrate the potential importance self-referential processing has on both avoidance of positivity and depression (see also Hsu et al., 2020).

Self-referential processing has also been examined within a neuroscience framework. Benau et al. (2019) conducted a study analyzing self-reference and the distinct brain activity associated with this type of processing in individuals with depression. Participants, consisting of 20 depressed and 20 age matched controls, completed a task similar to the Self-Referential Encoding Task (SRET; Derry and Kuiper, 1981), in which they indicated whether or not certain sentences were "true". These sentences either referred to the self or others and contained negative or positive descriptors. The late positive potential (LPP) was analyzed as this is a type of event related potential that has been associated with the processing of emotional information (Proudfit et al., 2015). Results indicated that the depressed group endorsed negative self-referential sentences at a higher rate than controls (Benau et al., 2019). The study also found individuals with depression to be more likely to *reject* positive self-referential sentences. Individuals with depression were shown to display an increased LPP when presented with negative sentences compared to positive, which could be interpreted as more processing resources for negativity. However, given that positivity was rejected more often by depressed participants, the extent to which there are differential and non-parsimonious neural processing biases for positivity deserves further empirical scrutiny.

As more evidence emerges regarding the role of self-reference across multiple levels of analysis, the way we understand and measure avoidance of positivity will evolve. The findings outlined thus far also have clinical implications, as treatments such as Attention Bias Modification may be more inclined to include tasks that specifically focus on retraining the processing of stimuli that have referential significance to the individual client. Therefore, future research surrounding Reward Devaluation Theory will likely further examine the role of self-reference regarding aspects of biased emotional responding.

Perceptual salience and dampening

Another future area of research that is likely to assist in the way we measure avoidance of positivity is perceptual salience and dampening. Past research has continually linked perceptual dampening, particularly of positive information, with depression (Raes et al., 2012; Werner-Seidler et al., 2013). Meta-analytic findings of Bean et al. (2022) found compelling evidence for the unique relationship between dampening of positivity and depression symptoms. Dampening of positivity, the extent to which one diminishes or downplays experiences of positive emotionality with negative thoughts, is typically measured via the Responses to Positive Affect questionnaire (RPA; Feldman et al., 2008). Raes et al. (2014) examined positivity dampening, via the RPA, in 187 pregnant women during their third trimester, in order to assess the extent to which dampening may influence postpartum depression. When controlling for baseline depression and depressive history, Raes et al. (2014) found that dampening of positivity at baseline was significantly associated with depression severity at 12 and 24 weeks after giving birth, suggesting a unique connection between dampening and depressive symptomatology may exist.

Given that dampening and anhedonia both involve devaluative responses to positivity, research has sought to better understand the relationship between these two constructs. For example, Werner-Seidler et al. (2013) examined dampening strategies in a variety of samples, including one study that examined 50 participants who were currently depressed, remitted depressed, or controls. The findings suggest that individuals with depression tend to dampen positivity more often than controls. Werner-Seidler et al. (2013) also found preliminary evidence that dampening may be uniquely associated with anhedonia. Similar findings have emerged within community samples (Nelis et al., 2015).

Current research is attempting to understand the role of dampening in relation to other important subfactors of psychopathology. Everaert et al. (2020) recently examined the role positivity dampening plays in maladaptive emotion regulation strategies, as well as depression and social anxiety symptoms. The initial study included 250 participants with a range of depression and social anxiety symptoms. An atemporal path model analysis revealed that dampening of positive emotions emerged as a mediator between the tendency to interpret negative information in an inflexible or rigid manner and depression, as well as social anxiety symptoms (Everaert et al., 2020). Thus, individuals with inflexible negative interpretations, even in the presence of positive information, were more likely to dampen positive emotions. Moreover, higher dampening scores, measured by way of the RPA, were positively associated with depressive and social anxiety symptomology. Thus, those who tend to exhibit negative interpretations of their environment may dampen any positive experiences they encounter. Continual diminishment of positive emotionality may give rise to depression and social anxiety symptoms.

As evidence accumulates regarding the influence dampening has on psychopathology, Li et al. (2017) examined the role of various response patterns to positive emotionality in daily life within 157 undergraduate students experiencing a range of depression symptoms. The study utilized a daily diary method over a two-week period in order to measure the impact responsivity to positive affect has on overall functioning and, more specifically, depression. Li et al. (2017) found that daily dampening of positivity was related to greater depression severity, as well as lower levels of positive emotionality after two weeks of tracking. These results once again indicate the powerful role perceptual dampening has on human depression, even in the context of daily functioning. Future research should examine dampening of positivity through behavioral measures, neuroimaging, and in the context of reward circuitry, while carefully delineating the difference between a lack of normative approach toward potential positivity, and the motivated avoidance of it.

In addition to dampening positive emotionality after it is experienced, some research has found that depressed persons may experience a blunted sensitivity when processing positive information. For example, individuals high in depression symptom severity have been shown to be less sensitive to positive compared to negative information (Atchley et al., 2012). Schlipf et al. (2013) examined the extent to which blunted processing of positivity extended to semantic information. A 23 person MDD group and a 22 person control group were presented with emotional words, which were presented verbally in a happy, angry, or neutral tone, to assess subjective appraisal of the stimuli. Schlipf et al. (2013) found that the MDD group not only rated the positive words as less emotionally salient based on the semantic meaning of the words, but also based on the intonation of the words as compared to controls. Given this reduction in perceived emotional intensity of positive words and positive prosody, the MDD group may experience a blunted perceptual saliency of positive stimuli. Further research would be helpful in teasing apart whether this group disparity is caused by a blunted sensitivity to positivity or if the positive stimuli rather lost their rewarding qualities through a devaluative process, resulting in lower comparative valence ratings.

Blunted processing of positive stimuli has also expanded into the neuroscience realm. Yang et al. (2016) utilized fMRI methodology to examine functional brain activity while viewing positive emotional stimuli. In particular, Yang et al. (2016) measured changes in brain areas responsible for reward processing and saliency processing of emotional information in 19 participants with MDD, as well as 19 healthy controls. Yang et al. (2016) found that the MDD group exhibited decreased activation in the reward circuity and saliency neural networks. This suggests that differential processing of rewarding stimuli, as well as a blunted appraisal of the saliency of potentially rewarding stimuli may be at play in individuals with depression. This study provides support for RDT in the sense that depression appears to be associated with a disparity in reward processing and blunted saliency of positive information as compared to controls. It is important to note that neural activation alone is not adequate to discern between blunting of positivity as opposed to a complete reversal in the processing of positivity, in which positivity is devalued.

Evidence supporting blunting of emotionality (even-handedness) and devaluation of reward continues to be pushed forward. Only limited work has had the power to directly examine whether reduced biases toward positivity in depression is related to a lack of approach or a reversal of normative processing (i.e., bias toward positivity) in which positivity is avoided and eventually devalued. Therefore, future mechanistic studies will benefit from disentangling these two theoretical explanations of positivity processing in depression.

Repeated pairings of positive and negative stimuli

Another call for future research focus involves concurrent pairings of positivity and negativity. Most studies that examine avoidance of positivity do so by separately comparing measures of positive and negative emotional attention biases (Winer & Salem, 2016). Very few studies have examined the effect of simultaneous positive and negative stimuli pairings for methodological reasons. The effect that coinciding positive and negative stimuli have on attention

is difficult to operationalize because positive and negative biases are confounded in one another. For example, if an individual with depression displays vigilance toward negative stimuli and avoidance away from positive stimuli when these affective stimuli are presented concurrently, how does one distinguish which bias is most prevalent? By pairing both positive and negative stimuli with a neutral stimulus, researchers are able to directly gauge negative and positive biases by comparing these emotional responses to a neutral comparator. This represents a large gap in the literature: examination of the effect positive-negative pairings have on processing, interpretation, and approach/avoidance behaviors among depressed individuals.

In a classic study, Gotlib et al. (1988) administered an emotional perceptual task to 12 individuals with depression, as well as 12 control participants. This task, the Deployment of Attention Task (DOAT; Gotlib et al., 1988), presented two words printed over one another, each in a different color. The words utilized in the study consisted of three conditions, one of which was a manic-depressed condition (i.e., positive-negative). Whereas nondepressed participants exhibited a bias towards positive words during the task, the depressed participants did not display such a bias (Gotlib et al., 1988). Although the depressed group attended to negative words during manic-depressed trials more often than the nondepressed group, Gotlib et al. (1988) revealed that the disparity between these groups likely occurred because of the nondepressed group's bias towards positivity. The lack of attentional bias towards positivity demonstrated by the depressed group may have accounted for the disparity between groups, rather than a vigilance towards negativity (Gotlib et al., 1988).

The utilization of simultaneous emotional pairings to assess for cognitive biases has recently been extended to the dot-probe task. Blanco et al. (2019) conducted a study consisting of 151 participants split across a dysphoric and non-dysphoric group, in which pairings of positive and negative faces were employed to assess for attentional biases. This study employed a free viewing task, which utilized eye tracking to measure a broad range of attentional data associated with joint emotional pairings. Results indicated that individuals with dysphoria spent less time viewing positive faces in all conditions in which positive faces were presented. Blanco et al. (2019) also found that on positive-negative trials, there was a reduction in the bias toward positive stimuli in both groups (i.e., dysphoric and control). This finding provides novel information, as it has been shown that healthy individuals tend to approach positive stimuli, both generally and in the context of attention tasks. Therefore, the findings by Blanco et al. (2019) suggest that there may be something inherently different in the way individuals process simultaneous positive-negative pairings of stimuli, in the sense that these pairings appear to inhibit reward pursuit.

Few studies have examined joint positive-negative stimulus pairings outside the realm of attentional biases. Beevers et al. (2009) examined interpretation biases in 107 participants across a currently dysphoric group and non-dysphoric group through the utilization of mixed sad and happy facial expressions. Morphing technology enabled Beevers et al. (2009) to combine faces of varying emotions into coherent, yet ambiguous images. Participants were then presented with these faces of ambiguous emotional states and asked to identify which emotion was present. The results revealed that on happy-sad trials, the dysphoric group tended to identify sadness as the primary emotion, rather than happiness, even though the faces consisted of nearly equal components of both emotions. This negativity bias was not found on trials when sadness was paired with other negative emotions (e.g., anger), suggesting that the combination of happy and sad emotional faces specifically was responsible for the inhibition of reward responsivity found throughout the study.

The limited research examining pairings of positive and negative stimuli suggests that these incongruent stimuli interfere in processing, potentially resulting in the devaluation of rewarding stimuli. These incongruent valence pairings may alter the way in which individuals perceive their environment, leading them to avoid positivity because of the potential harm associated with these stimuli from past experiences. Additional inhibition of apparently rewarding stimuli may lead to further devaluation and diminished reward pursuit. This repeated inhibition of reward may help explain how reward devaluation can expand to a broad range of social rewards, in line with RDT. Despite the interesting findings outlined above, more research is needed to understand the learning mechanisms at play during co-activation. Also, additional studies examining the way in which these incongruent pairings alter emotional processing, specifically resulting in the devaluation of positivity, would be beneficial.

The development of sensitive behavioral tasks that quantify reward devaluation can aid greatly in the understanding of positivity avoidance. For example, tasks that utilize inherently positive stimuli, such as smiling faces, may be able to be manipulated in a way to operationalize reward devaluation. Facial morphing is an innovative way to gradually combine faces of opposing valence (i.e., happy and sad faces). The combination of positive and negative emotions through facial presentations may be able to measure an individual's subjective positivity and negativity thresholds. These emotion recognition thresholds can be analyzed for disparities between the emotional intensity necessary to accurately recognize happiness versus sadness. In accordance with RDT, it would be expected that individuals with

depression would require an excessive amount of happiness to accurately recognize the emotion, given their automatic avoidance of rewarding stimuli. Studies utilizing happy and sad faces separately have found preliminary emotion recognition evidence that aligns with the tenets of RDT (Joormann & Gotlib, 2006). Further research in which facial morphs of happy and sad faces are utilized concurrently may be necessary to more precisely quantify reward devaluation.

Limitations/focus of future research

It is important to note that one limitation is the use of a narrative review and not a systematic review. Although a narrative review was utilized, critical studies detailing a variety of findings within the positivity literature were included. Additionally, while a number of studies have found support for RDT, it is important to note that many of these studies operationalize the positivity avoidance bias through cognitive tasks. Cognitive and behavioral tasks while useful, may lack ecological validity to real world settings given the nature of the tasks. Exhibiting emotional biases in a controlled lab environment provides evidence that the biases exist, but it is unclear how exactly these biases manifest in everyday life. One potential solution is to use more ambiguous stimuli in cognitive tasks. The world is filled with stimuli of mixed emotional valence that require constant interpretation. Therefore, utilizing ambiguous stimuli, such as morphed emotional faces or ambiguous emotional statements like in the emotional bias against disconfirmatory evidence (BADE; Everaert et al., 2018) task, may advance assessment of ecological validity. Fine grained data collection such as ecological momentary assessment will also contribute to more ecologically valid research.

Another limitation of relying primarily on cognitive tasks is the reliability of the tasks. Many cognitive tasks require a large number of trials to reliably identify an effect. For example, Chapman et al. (2019) recently outlined concerns surrounding the reliability of the dot-probe task. Despite this limitation, the dot-probe can still be used to measure attentional biases if employed properly, as Price et al. (2015) provide several recommendations for improving the reliability of the dot-probe. In particular, Price et al. (2019) found that using an advanced statistical method, the drift diffusion model (DDM; Ratcliff and McKoon, 2008), increased the reliability of the dot-probe task. Additionally, eye-tracking is becoming more common in attentional tasks like the dot-probe, as eye tracking methodology results in fine grained measurements of a variety of different eye movements (e.g., fixations, saliency maps). Finally, tasks that do not rely on reaction time, but rather focus on other measures such as response accuracy, may provide more reliable and stable effects going forward.

Clinical implications

Reward Devaluation Theory leads to several potentially profound clinical implications, specifically in regard to positive affect, anhedonia, and reward circuitry. Positive Affect Treatment (PAT; Craske et al., 2016) is an innovative clinical treatment given its focus on positive emotionality. PAT is designed to improve positive affect by targeting reward processing deficits through three domains: anticipation, consumption, and learning (Craske et al., 2016). Clients are taught to savor positive emotions in each of these domains. Research has begun to show the utility of PAT as a potential depression treatment (Craske et al., 2019).

Another clinical modality that encompasses several of the tenets of RDT is Augmented Depression Therapy (ADepT; Dunn et al., 2019). Whereas many psychological depression treatments focus solely on the reduction of negative symptoms (e.g., depressed mood), ADepT focuses on both the negative and positive valence systems and attempts to target anhedonia directly by increasing positive affect through a variety of techniques. Preliminary research has found ADepT to result in a reduction of anhedonia and overall depression (Dunn et al., 2019).

Taylor et al. (2017) also use a neuroscientific based approach to target the positive valence system, developing a novel treatment known as Amplification of Positivity (i.e., AMP). AMP is comprised of 10 sessions which consist of psychoeducation, generating positive thoughts and emotions, as well as exercises to increase positive emotions. In particular, a unique positive activity plan is designed for each individual to promote further engagement in positive activities and prevent future relapse. Initial research has found that AMP leads to increases in positive affect and psychological well-being, while also leading to an increase in social connectedness in those with depression and anxiety (Taylor et al., 2017, 2020). These findings provide promising evidence for the utility of AMP in upregulating the positive valence system.

Reward devaluation may be able to be addressed through these promising new anhedonia treatments. However, given their focus on savoring, approaching, and upregulating positive emotions, depressed individuals may avoid engaging in them or indicate a preference for a less explicitly positive treatment. For example, in a recent study Bryant et al. (2023) found that individuals high in fear of happiness rated PAT as having a lower treatment fit, preference, and effectiveness as a depression treatment compared to psychodynamic therapy, a less explicitly positive treatment. These results indicate that those who fear happiness may avoid or be hesitant to engage in the treatments that would most benefit their presenting problem. Therefore, clinicians should be aware of this potential resistance to positivity-based

interventions. Collaborative discussion between clinicians and clients in which psychoeducation regarding the reality of reward devaluation biases and benefits of positivity-based treatments are provided, may aid in reducing client resistance to positivity-based interventions.

Additional studies have begun to implement variations of positivity focused interventions. Geschwind et al. (2019) discussed the potential utility of positive cognitive behavioral therapy (CBT), a treatment that implements CBT based skills, while also focusing on increasing positive emotions. Positive CBT focuses on building from a strength-based approach rather than reducing undesired thoughts or behaviors. Positive CBT has shown promising initial evidence as a treatment modality (Geschwind et al., 2019, 2020).

Vazquez et al. (2018) conducted a study in which they examined the influence group therapy with a focus on positive psychology interventions, as well as a CBT group have on maladaptive attentional biases (i.e., preference for negative and avoidance of positive) in a depressed sample. Vazquez et al. (2018) found via eye-tracking that participants spent less time viewing sad faces and more time viewing happy faces after receiving either of the two group interventions as compared to prior of the training. Thus, additional research regarding the extension of positivity-based interventions into the group therapy realm may be helpful, especially in altering the cognitive biases typically exhibited in depression.

One final clinical consideration that stems from the cognitive domain is Attention Bias Modification (ABM), which may be a beneficial supplement to clinical interventions. ABM is a cognitive retraining tool, which utilizes a manipulated paradigm similar to the dot-probe to influence attentional biases. Given the impact maladaptive cognitive biases have on psychopathology, it may prove useful to target cognitive biases directly. ABM has been established as a potential treatment supplement in anxiety populations (Hakamata et al., 2010), and has begun to extend into the depression realm. Beevers et al. (2015) found that ABM was able to reduce negative attention bias in a sample with MDD. However, the majority of research surrounding ABM in depressive samples focuses on reducing negative attention biases. Future research can benefit from assessing the utility of ABM in altering the avoidance of positivity bias outlined throughout this paper. Given that cognitive training takes place in a controlled lab environment, in addition to the reliability concerns with tasks such as the dot-probe, further research regarding the generalizability of attention bias modification to other realms of cognition and depressive symptomology will be valuable. Finally, given the importance of self-referential positive information established previously, increased research using self-referential stimuli in ABM is recommended.

Conclusion

The goal of this paper was to provide (i) a summary of the theoretical import of Reward Devaluation Theory, (ii) a synopsis of the empirical literature surrounding RDT that has accumulated since its introduction, and (iii) a set of future directions that will likely elucidate the parameters of avoidance of positivity in relation to depression and incorporate further clinical translation. We initially summarized the origins and literature base of RDT, which posits that individuals with depression not only exhibit vigilance in the face of negativity, but also demonstrate an avoidance or inhibition of positivity. Previous pairings of positive stimuli with negative outcomes may result in learned behaviors in which prospectively rewarding stimuli appear dangerous and are thus avoided. Coactivation of positive stimuli with negative outcomes can lead to a decrease in goal seeking behavior. This inhibition to approach rewarding stimuli also leads to devaluation of positivity, thus reinforcing the inhibitory behaviors. This feedback loop alters one's view of what is rewarding or positive, as what was previously interpreted as rewarding stimuli/situations may appear more dangerous than neutrality, given the lack of warning signs inherently associated with positivity.

Thus far, positivity avoidance has been examined through a variety of cognitive and behavioral tasks, as our understanding of positivity avoidance, both anticipatory and reactive, continues to develop. Future research across levels of analysis examining the domains of self-referential processing and repeated positive and negative stimulus pairings will likely aid in the understanding of devaluation. Additionally, the way in which reward devaluation is operationalized will likely continue to be refined, as cognitive/behavioral tasks that incorporate advanced methodology have the prospect of mapping devaluative components onto specific person- and symptom-based factors. As such, Reward Devaluation Theory will hopefully continue to lead to novel hypotheses that increase knowledge of the complexity of depression as well as aiding in understanding of the parameters of emotional information processing in general.

Role of funding sources Research reported in this publication was supported by the National Institute of Mental Health of the National Institutes of Health under Award Number R15MH101573. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Data availability Data sharing is not applicable to this article as no datasets were generated or analyzed during the current study.

Declarations

Ethical approval This is a review paper. This project did not involve human participants and/or animals.

Conflict of interest On behalf of all authors, the corresponding author states that there is no conflict of interest.

Informed consent Informed consent was not required, given the nature of the manuscript.

References

Aarts, H., Custers, R., & Holland, R. W. (2007). The nonconscious cessation of goal pursuit: When goals and negative affect are coactivated. *Journal of Personality and Social Psychology*, *92*(2), 165–178. https://doi.org/10.1037/0022-3514.92.2.165.

Ao, X., Mo, L., Wei, Z., Yu, W., Zhou, F., & Zhang, D. (2020). Negative bias during early attentional engagement in major depressive disorder as examined using a two-stage model: High sensitivity to sad but bluntness to happy cues. *Frontiers in Human Neuroscience*, *14*, https://doi.org/10.3389/fnhum.2020.593010.

Atchley, R. A., Ilardi, S. S., Young, K. M., Stroupe, N. N., O'Hare, A. J., Bistricky, S. L., & Lepping, R. J. (2012). Depression reduces perceptual sensitivity for positive words and pictures. *Cognition and Emotion*, *26*(8), 1359–1370. https://doi.org/10.1080/02699931.2012.660134.

Ballard, E. D., Wills, K., Lally, N., Richards, E. M., Luckenbaugh, D. A., Walls, T., & Zarate, C. A. (2017). Anhedonia as a clinical correlate of suicidal thoughts in clinical ketamine trials. *Journal of Affective Disorders*, *218*, 195–200. https://doi.org/10.1016/j.jad.2017.04.057.

Barkus, E. (2021). The effects of anhedonia in social context. *Current Behavioral Neuroscience Reports*, *8*(3), 77–89. https://doi.org/10.1007/s40473-021-00232-x.

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.

Bean, C. A. L., Summers, C. B., & Ciesla, J. A. (2022). Dampening of positive affect and depression: A meta-analysis of cross-sectional and longitudinal relationships. *Behaviour Research and Therapy*, *156*, 104153. https://doi.org/10.1016/j.brat.2022.104153.

Beck, A. T., & Bredemeier, K. (2016). A unified model of depression. *Clinical Psychological Science*, *4*(4), 596–619. https://doi.org/10.1177/2167702616628523.

Beevers, C. G., Clasen, P. C., Enock, P. M., & Schnyer, D. M. (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. https://doi.org/10.1037/abn0000049.

Beevers, C. G., Wells, T. T., Ellis, A. J., & Fischer, K. (2009). Identification of emotionally ambiguous interpersonal stimuli among dysphoric and nondysphoric individuals. *Cognitive Therapy and Research*, *33*(3), 283–290. https://doi.org/10.1007/s10608-008-9198-6.

Benau, E. M., Hill, K. E., Atchley, R. A., O'Hare, A. J., Gibson, L. J., Hajcak, G., & Foti, D. (2019). Increased neural sensitivity to self-relevant stimuli in major depressive disorder. *Psychophysiology*, *56*(7), e13345. https://doi.org/10.1111/psyp.13345.

Blanco, I., Poyato, N., Nieto, I., Boemo, T., Pascual, T., Roca, P., & Vazquez, C. (2019). Attentional biases in dysphoria when happy and sad faces are simultaneously presented. *Journal of Behavior Therapy and Experimental Psychiatry*, *65*, 101499. https://doi.org/10.1016/j.jbtep.2019.101499.

Blaney, P. H. (1986). Affect and memory: A review. *Psychological Bulletin*, *99*(2), 229–246. https://doi.org/10.1037/0033-2909.99.2.229.

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*, 101830. https://doi.org/10.1016/j.jbtep.2022.101830.

Bryant, J., Winer, E. S., Salem, T., & Nadorff, M. R. (2017). Struggling toward reward: Recent experience of anhedonia interacts with motivation to predict reward pursuit in the face of a stressful manipulation. *PLOS ONE*, *12*(3), e0173439. https://doi.org/10.1371/journal.pone.0173439.

Chapman, A., Devue, C., & Grimshaw, G. M. (2019). Fleeting reliability in the dot-probe task. *Psychological Research Psychologische Forschung*, *83*(2), 308–320. https://doi.org/10.1007/s00426-017-0947-6.

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.

Craske, M. G., Meuret, A. E., Ritz, T., Treanor, M., Dour, H., & Rosenfield, D. (2019). Positive affect treatment for depression and anxiety: A randomized clinical trial for a core feature of anhedonia. *Journal of Consulting and Clinical Psychology*, *87*(5), 457–471. https://doi.org/10.1037/ccp0000396.

DePierro, J., D'Andrea, W., Frewen, P., & Todman, M. (2018). Alterations in positive affect: Relationship to symptoms, traumatic experiences, and affect ratings. *Psychological Trauma: Theory Research Practice and Policy*, *10*(5), 585–593. https://doi.org/10.1037/tra0000317.

Derry, P. A., & Kuiper, N. A. (1981). Schematic processing and self-reference in clinical depression. *Journal of Abnormal Psychology*, *90*(4), 286–297. https://doi.org/10.1037//0021-843x.90.4.286.

Disner, S. G., Shumake, J. D., & Beevers, C. G. (2017). Self-referential schemas and attentional bias predict severity and naturalistic course of depression symptoms. *Cognition and Emotion*, *31*(4), 632–644. https://doi.org/10.1080/02699931.2016.1146123.

Dozois, D. J. A., & Dobson, K. S. (2001). Information processing and cognitive organization in unipolar depression: Specificity and comorbidity issues. *Journal of Abnormal Psychology*, *110*(2), 236–246. https://doi.org/10.1037/0021-843X.110.2.236.

Dunn, B. D., Widnall, E., Reed, N., Owens, C., Campbell, J., & Kuyken, W. (2019). Bringing light into darkness: A multiple baseline mixed methods case series evaluation of Augmented Depression Therapy (ADepT). *Behaviour Research and Therapy*, *120*, 103418. https://doi.org/10.1016/j.brat.2019.103418.

Elgersma, H. J., Koster, E. H. W., van Tuijl, L. A., Hoekzema, A., Penninx, B. W. J. H., Bockting, C. L. H., & de Jong, P. J. (2018). Attentional bias for negative, positive, and threat words in current and remitted depression. *PLOS ONE*, *13*(10), e0205154. https://doi.org/10.1371/journal.pone.0205154.

Ellis, A. J., Salgari, G., Miklowitz, D. J., & Loo, S. K. (2019). The role of avoidance motivation in the relationship between reward sensitivity and depression symptoms in adolescents: An ERP study. *Psychiatry Research*, *279*, 345–349. https://doi.org/10.1016/j.psychres.2019.03.011.

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., Bronstein, M. V., Castro, A. A., Cannon, T. D., & Joormann, J. (2020). When negative interpretations persist, positive emotions don't! Inflexible negative interpretations encourage depression and social anxiety by dampening positive emotions. *Behaviour Research and Therapy*, *124*, 103510. https://doi.org/10.1016/j.brat.2019.103510.

Everaert, J., Koster, E. H. W., & Derakshan, N. (2012). The combined cognitive bias hypothesis in depression. *Clinical Psychology Review*, *32*(5), 413–424. https://doi.org/10.1016/j.cpr.2012.04.003.

Feldman, G. C., Joormann, J., & Johnson, S. L. (2008). Responses to positive affect: A self-report measure of rumination and dampening. *Cognitive Therapy and Research*, *32*(4), 507–525. https://doi.org/10.1007/s10608-006-9083-0.

Fried, E. I., & Nesse, R. M. (2015). Depression is not a consistent syndrome: An investigation of unique symptom patterns in the STAR*D study. *Journal of Affective Disorders, 17*, 296–102. https://doi.org/10.1016/j.jad.2014.10.010.

Frischen, A., Ferrey, A. E., Burt, D. H. R., Pistchik, M., & Fenske, M. J. (2012). The affective consequences of cognitive inhibition: Devaluation or neutralization? *Journal of Experimental Psychology: Human Perception and Performance*, *38*(1), 169–179. https://doi.org/10.1037/a0025981.

Gaddy, M. A., & Ingram, R. E. (2014). A meta-analytic review of mood-congruent implicit memory in depressed mood. *Clinical Psychology Review*, *34*(5), 402–416. https://doi.org/10.1016/j.cpr.2014.06.001.

Geschwind, N., Arntz, A., Bannink, F., & Peeters, F. (2019). Positive cognitive behavior therapy in the treatment of depression: A randomized order within-subject comparison with traditional cognitive behavior therapy. *Behaviour Research and Therapy*, *116*, 119–130. https://doi.org/10.1016/j.brat.2019.03.005.

Geschwind, N., Bosgraaf, E., Bannink, F., & Peeters, F. (2020). Positivity pays off: Clients' perspectives on positive compared with traditional cognitive behavioral therapy for depression. *Psychotherapy*, *57*(3), 366–378. https://doi.org/10.1037/pst0000288.

Gilbert, P., McEwan, K., Catarino, F., Baião, R., & Palmeira, L. (2014). Fears of happiness and compassion in relationship with depression, alexithymia, and attachment security in a depressed sample. *British Journal of Clinical Psychology*, *53*(2), 228–244. https://doi.org/10.1111/bjc.12037.

Günther, V., Dannlowski, U., Kersting, A., & Suslow, T. (2015). Associations between childhood maltreatment and emotion processing biases in major depression: Results from a dot-probe task. *Bmc Psychiatry*, *15*(1), https://doi.org/10.1186/s12888-015-0501-2.

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., McLachlan, A. L., & Katz, A. N. (1988). Biases in visual attention in depressed and nondepressed individuals. *Cognition and Emotion*, *2*(3), 185–200. https://doi.org/10.1080/02699938808410923.

Hakamata, Y., Lissek, S., Bar-Haim, Y., Britton, J. C., Fox, N. A., Leibenluft, E., & Pine, D. S. (2010). Attention bias modification treatment: A meta-analysis toward the establishment of novel treatment for anxiety. *Biological Psychiatry*, *68*(11), 982–990. https://doi.org/10.1016/j.biopsych.2010.07.021.

Hirsch, C. R., Clark, D. M., & Mathews, A. (2006). Imagery and interpretations in social phobia: Support for the combined cognitive biases hypothesis. *Behavior Therapy*, *37*(3), 223–236. https://doi.org/10.1016/j.beth.2006.02.001.

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

Ji, J. L., Grafton, B., & Macleod, C. (2017). Referential focus moderates depression-linked attentional avoidance of positive information. *Behaviour Research and Therapy*, *93*, 47–54. https://doi.org/10.1016/j.brat.2017.03.004.

Joormann, J., & Gotlib, I. H. (2006). Is this happiness I see? Biases in the identification of emotional facial expressions in depression and social phobia. *Journal of Abnormal Psychology*, *115*(4), 705–714. https://doi.org/10.1037/0021-843x.115.4.705.

Joormann, J., & Gotlib, I. H. (2010). Emotion regulation in depression: Relation to cognitive inhibition. *Cognition and Emotion*, *24*(2), 281–298. https://doi.org/10.1080/02699930903407948.

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.

Jordan, D. G., Winer, E. S., Salem, T., & Kilgore, J. (2018). Longitudinal evaluation of anhedonia as a mediator of fear of positive evaluation and other depressive symptoms. *Cognition and Emotion*, *32*(7), 1437–1447. https://doi.org/10.1080/02699931.2017.1289895.

Joshanloo, M. (2013). The influence of fear of happiness beliefs on responses to the satisfaction with life scale. *Personality and Individual Differences*, *54*(5), 647–651. https://doi.org/10.1016/j.paid.2012.11.011.

Joshanloo, M., & Weijers, D. (2014). Aversion to happiness across cultures: A review of where and why people are averse to happiness. *Journal of Happiness Studies*, *15*(3), 717–735. https://doi.org/10.1007/s10902-013-9489-9.

Kaiser, R. H., Snyder, H. R., Goer, F., Clegg, R., Ironside, M., & Pizzagalli, D. A. (2018). Attention bias in rumination and depression: Cognitive mechanisms and brain networks. *Clinical Psychological Science*, *6*(6), 765–782. https://doi.org/10.1177/2167702618797935.

Kuhl, J. (1994). Action versus state orientation: Psychometric properties of the Action Control Scale (ACS-90). *Volition and personality: Action versus state orientation*, *47*(56).

Lemoult, J., & Gotlib, I. H. (2019). Depression: A cognitive perspective. *Clinical Psychology Review*, *69*, 51–66. https://doi.org/10.1016/j.cpr.2018.06.008.

Li, Y. I., Starr, L. R., & Hershenberg, R. (2017). Responses to positive affect in daily life: Positive rumination and dampening moderate the association between daily events and depressive symptoms. *Journal of Psychopathology and Behavioral Assessment*, *39*(3), 412–425. https://doi.org/10.1007/s10862-017-9593-y.

Matt, G. E., Vázquez, C., & Campbell, W. K. (1992). Mood-congruent recall of affectively toned stimuli: A meta-analytic review. *Clinical Psychology Review*, *12*(2), 227–255. https://doi.org/10.1016/0272-7358(92)90116-p.

Messerotti Benvenuti, S., Buodo, G., & Palomba, D. (2017). Appetitive and aversive motivation in dysphoria: A time-domain and time-frequency study of response inhibition. *Biological Psychology*, *125*, 12–27. https://doi.org/10.1016/j.biopsycho.2017.02.007.

Millgram, Y., Huppert, J. D., & Tamir, M. (2020). Emotion goals in psychopathology: A new perspective on dysfunctional emotion regulation. *Current Directions in Psychological Science*, *29*(3), 242–247. https://doi.org/10.1177/0963721420917713.

Nelis, S., Holmes, E. A., & Raes, F. (2015). Response styles to positive affect and depression: Concurrent and prospective associations in a community sample. *Cognitive Therapy and Research*, *39*(4), 480–491. https://doi.org/10.1007/s10608-015-9671-y.

Phillips, W. J., Hine, D. W., & Thorsteinsson, E. B. (2010). Implicit cognition and depression: A meta-analysis. *Clinical Psychology Review*, *30*(6), 691–709. https://doi.org/10.1016/j.cpr.2010.05.002.

Pool, E., Brosch, T., Delplanque, S., & Sander, D. (2016). Attentional bias for positive emotional stimuli: A meta-analytic investigation. *Psychological Bulletin*, *142*(1), 79–106. https://doi.org/10.1037/bul0000026.

Price, R. B., Brown, V., & Siegle, G. J. (2019). Computational modeling applied to the dot-probe task yields improved reliability and mechanistic insights. *Biological Psychiatry*, *85*(7), 606–612. https://doi.org/10.1016/j.biopsych.2018.09.022.

Price, R. B., Kuckertz, J. M., Siegle, G. J., Ladouceur, C. D., Silk, J. S., Ryan, N. D., & Amir, N. (2015). Empirical recommendations for improving the stability of the dot-probe task in clinical research. *Psychological Assessment*, *27*(2), 365–376. https://doi.org/10.1037/pas0000036.

Proudfit, G. H., Bress, J. N., Foti, D., Kujawa, A., & Klein, D. N. (2015). Depression and event-related potentials: Emotional disengagement and reward insensitivity. *Current Opinion in Psychology*, *4*, 110–113. https://doi.org/10.1016/j.copsyc.2014.12.018.

Raes, F., Smets, J., Nelis, S., & Schoofs, H. (2012). Dampening of positive affect prospectively predicts depressive symptoms in non-clinical samples. *Cognition and Emotion*, *26*(1), 75–82. https://doi.org/10.1080/02699931.2011.555474.

Raes, F., Smets, J., Wessel, I., Van Den Eede, F., Nelis, S., Franck, E., & Hanssens, M. (2014). Turning the pink cloud grey: Dampening of positive affect predicts postpartum depressive symptoms. *Journal of Psychosomatic Research*, *77*(1), 64–69. https://doi.org/10.1016/j.jpsychores.2014.04.003.

Ratcliff, R., & McKoon, G. (2008). The diffusion decision model: Theory and data for two-choice decision tasks. *Neural Computation*, *20*(4), 873–922. https://doi.org/10.1162/neco.2008.12-06-420.

Ruhe, H. G., Mocking, R. J. T., Figueroa, C. A., Seeverens, P. W. J., Ikani, N., Tyborowska, A., & Schene, A. H. (2019). Emotional biases and recurrence in major depressive disorder. Results of 2.5 years follow-up of drug-free cohort vulnerable for recurrence. *Frontiers in Psychiatry*, *10*, https://doi.org/10.3389/fpsyt.2019.00145.

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.

Schlipf, S., Batra, A., Walter, G., Zeep, C., Wildgruber, D., Fallgatter, A., & Ethofer, T. (2013). Judgment of emotional information expressed by prosody and semantics in patients with unipolar depression. *Frontiers in Psychology*, *4*, https://doi.org/10.3389/fpsyg.2013.00461.

Sherdell, L., Waugh, C. E., & Gotlib, I. H. (2012). Anticipatory pleasure predicts motivation for reward in major depression. *Journal of Abnormal Psychology*, *121*(1), 51–60. https://doi.org/10.1037/a0024945.

Snodgrass, M., & Shevrin, H. (2006). Unconscious inhibition and facilitation at the objective detection threshold: Replicable and qualitatively different unconscious perceptual effects. *Cognition*, *101*(1), 43–79. https://doi.org/10.1016/j.cognition.2005.06.006.

Taylor, C. T., Lyubomirsky, S., & Stein, M. B. (2017). Upregulating the positive affect system in anxiety and depression: Outcomes of a positive activity intervention. *Depression and Anxiety*, *34*(3), 267–280. https://doi.org/10.1002/da.22593.

Taylor, C. T., Pearlstein, S. L., Kakaria, S., Lyubomirsky, S., & Stein, M. B. (2020). Enhancing social connectedness in anxiety and depression through amplification of positivity: Preliminary treatment outcomes and process of change. *Cognitive Therapy and Research*, *44*(4), 788–800. https://doi.org/10.1007/s10608-020-10102-7.

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), e6598. https://doi.org/10.1371/journal.pone.0006598.

Twomey, C., O'Reilly, G., & Goldfried, M. R. (2023). Consensus on the perceived presence of transtheoretical principles of change in routine psychotherapy practice: A survey of clinicians and researchers. *Psychotherapy*. https://doi.org/10.1037/pst0000489.

Vanderlind, W. M., Stanton, C. H., Weinbrecht, A., Velkoff, E. A., & Joormann, J. (2017). Remembering the good ole days: Fear of positive emotion relates to affect repair using positive memories. *Cognitive Therapy and Research*, *41*(3), 362–368. https://doi.org/10.1007/s10608-016-9775-z.

Vazquez, C., Duque, A., Blanco, I., Pascual, T., Poyato, N., Lopez-Gomez, I., & Chaves, C. (2018). CBT and positive psychology interventions for clinical depression promote healthy attentional biases: An eye-tracking study. *Depression and Anxiety*, *35*(10), 966–973. https://doi.org/10.1002/da.22786.

Veling, H., & Aarts, H. (2009). Putting behavior on hold decreases reward value of need-instrumental objects outside of awareness. *Journal of Experimental Social Psychology*, *45*(4), 1020–1023. https://doi.org/10.1016/j.jesp.2009.04.020.

Veling, H., Holland, R. W., & Van Knippenberg, A. (2008). When approach motivation and behavioral inhibition collide: Behavior regulation through stimulus devaluation. *Journal of Experimental Social Psychology*, *44*(4), 1013–1019. https://doi.org/10.1016/j.jesp.2008.03.004.

Weeks, J. W., Heimberg, R. G., Rodebaugh, T. L., Goldin, P. R., & Gross, J. J. (2012). Psychometric evaluation of the fear of positive evaluation scale in patients with social anxiety disorder. *Psychological Assessment*, *24*(2), 301–312. https://doi.org/10.1037/a0025723.

Werner-Seidler, A., Banks, R., Dunn, B. D., & Moulds, M. L. (2013). An investigation of the relationship between positive affect regulation and depression. *Behaviour Research and Therapy*, *51*(1), 46–56. https://doi.org/10.1016/j.brat.2012.11.001.

Winer, E. S., Bryant, J., Bartoszek, G., Rojas, E., Nadorff, M. R., & Kilgore, J. (2017). Mapping the relationship between anxiety, anhedonia, and depression. *Journal of Affective Disorders*, *221*, 289–296. https://doi.org/10.1016/j.jad.2017.06.006.

Winer, E. S., Cervone, D., Newman, L. S., & Snodgrass, M. (2011). Subchance perception: Anxious, non-defensive individuals identify subliminally-presented positive words at below-chance levels. *Personality and Individual Differences*, *51*(8), 996–1001. https://doi.org/10.1016/j.paid.2011.08.010.

Winer, E. S., Jordan, D. G., & Collins, A. C. (2019). Conceptualizing anhedonias and implications for depression treatments. *Psychology Research and Behavior Management, 12*, 325–335. https://doi.org/10.2147/prbm.s159260.

Winer, E. S., Nadorff, M. R., Ellis, T. E., Allen, J. G., Herrera, S., & Salem, T. (2014). Anhedonia predicts suicidal ideation in a large psychiatric inpatient sample. *Psychiatry Research*, *218*(1–2), 124–128. https://doi.org/10.1016/j.psychres.2014.04.016.

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.

Yang, Y., Zhong, N., Imamura, K., Lu, S., Li, M., Zhou, H., & Li, K. (2016). Task and resting-state fMRI reveal altered salience responses to positive stimuli in patients with major depressive disorder. *PLOS ONE*, *11*(5), e0155092. https://doi.org/10.1371/journal.pone.0155092.

Yaroslavsky, I., Allard, E. S., & Sanchez-Lopez, A. (2019). Can't look away: Attention control deficits predict rumination, depression symptoms and depressive affect in daily life. *Journal of Affective Disorders*, *245*, 1061–1069. https://doi.org/10.1016/j.jad.2018.11.036.