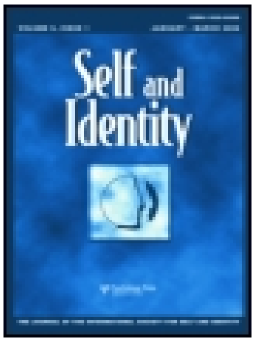

Self and Identity

Routledge
Taylor & Francis Group

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/psai20

Mechanisms of a spotless self-image: Navigating negative, self-relevant feedback

Anastasia E. Rigney , David M. Schnyer , Xiaoqing Hu & Jennifer S. Beer

To cite this article: Anastasia E. Rigney , David M. Schnyer , Xiaoqing Hu & Jennifer S. Beer (2020): Mechanisms of a spotless self-image: Navigating negative, self-relevant feedback, Self and Identity, DOI: 10.1080/15298868.2020.1806918

To link to this article: https://doi.org/10.1080/15298868.2020.1806918

SELF AND IDENTITY
https://doi.org/10.1080/15298868.2020.1806918

Routledge
Taylor & Francis Group

ARTICLE

Mechanisms of a spotless self-image: Navigating negative, self-relevant feedback

Anastasia E. Rigney[a], David M. Schnyer[a], Xiaoqing Hu[b] and Jennifer S. Beer[a]

[a]Department of Psychology, University of Texas at Austin, Austin, TX, USA; [b]Department of Psychology, The University of Hong Kong, Pok Fu Lam, Hong Kong

ABSTRACT

The present research investigated whether the differential recognition thresholds associated with memory for self-relevant negative feedback stem from processes occurring at encoding and/or suppression at retrieval. Socioemotional and monetary incentives offered before and after encoding did not significantly affect recognition thresholds for negative, self-relevant personality feedback (Studies 1–2). However, when presented before encoding took place, the combination of socioemotional and monetary incentives did impact recognition thresholds for negative personality feedback (about another person: Study 2). Differences in memory (rather than concealed knowledge) predicted ERP patterns associated with forgotten negative, self-relevant feedback as early as the encoding stage (Study 3). Results suggest that disrupted processes during encoding may play a role in the differential recognition thresholds associated with memory for negative, self-relevant feedback.

ARTICLE HISTORY
Received 17 July 2019
Accepted 1 August 2020

KEYWORDS
Self; social cognition; self-enhancement; motivation

People tend to apply higher standards before acknowledging the self-relevance of negative feedback and this higher bar for recognition is considered a fundamental and normative aspect of self-evaluation (e.g., Alicke et al., 2013; Ditto & Lopez, 1992; Greenwald, 1980; Taylor & Brown, 1988). Yet it remains unclear when the difference in recognition standards arises in the information processing stream (D'Argembeau & Van der Linden, 2008; Sedikides & Green, 2009; Walker et al., 2003). Some research suggests disrupted processing happens as early as the point of encoding (Sedikides & Green, 2000), while other research has investigated the possibility of disrupted processing at retrieval (Green, Sedikides et al., 2008; Djikic et al., 2007, 2005). However, it has been difficult to draw strong conclusions from existing research. The results from the encoding stage are ambiguous; it is possible that observed recognition impairments are merely an artifact of experimental disruption to encoding processes rather than a reflection of the encoding impairments that arise in the absence of experimental interference. Additionally, the existing research on retrieval processes has yielded mixed results. The present research addresses existing issues in two ways: (a) testing whether people can be incentivized at

CONTACT Jennifer S. Beer beerutexas@gmail.com Department of Psychology, The University of Texas at Austin, Austin, TX 78712-1043, USA
Supplemental data for this article can be accessed here.

the encoding stage to lower their standards for how much negative, self-relevant feedback must feel familiar before it is recognized (presented after encoding: Study 1; presented before encoding: Study 2) and (b) investigating whether neural signals associated with processing self-relevant, negative feedback are distinguishable in relation to memory differences (i.e., remembered versus forgotten) or are consistent with markers of concealed knowledge at retrieval (Study 3).

Impaired memory for negative, self-relevant feedback has been robustly established in the extant literature (D'Argembeau & Van der Linden, 2008; Djikic et al., 2005; Sedikides & Green, 2009). Memory for negative, self-relevant feedback is diminished whether it be for bogus feedback about personality traits (Djikic et al., 2005; Sedikides & Green, 2009), past unethical behaviors (Kouchaki & Gino, 2016), or past academic performance (Gramzow & Willard, 2006). It is still not clear why memory for negative, self-relevant feedback tends to be poor; processes at both encoding and retrieval have been considered as the culprit. In other words, we do not yet know whether people begin to distort their ability to recognize negative feedback as self-relevant when they learn the feedback (i.e., encoding) or when they are faced with trying to remember it later (i.e., retrieval).

The possibility of encoding has been studied by examining whether experimental manipulations known to undermine memory in general affect memory for self-relevant feedback. That is, when participants have limited time to encode feedback (Sedikides & Green, 2000) or perform a simultaneous task while encoding feedback (Zengel et al., 2018), memory tends to be equally poor for positive and negative self-relevant information. Therefore, the existing research demonstrates that experimentally-introduced factors known to encourage shallow encoding also tend to produce poor memory for negative self-relevant information. However, a stronger conclusion about the role of encoding could be made if it was the case that people cannot improve their memory for negative, self-relevant feedback after encoding has taken its natural course. That is, if people are incentivized to accurately identify negative, self-relevant feedback after they have already been presented with the feedback, can they lower their standards for recognition of negative, self-relevant feedback to reap those rewards? Or would they need to know about the rewards before encoding takes place?

Alternatively (or additionally), research has raised the possibility that retrieval is a critical point at which memory for negative, self-relevant feedback tends to fail. Researchers have theorized that the poorer memory arises because people use different recognition thresholds when remembering negative versus positive self-relevant feedback. The concept of a recognition threshold is rooted in Signal Detection Theory which posits that people partly assess their memory for information based on how strongly the information elicits feelings of familiarity (typically captured by the parameter of location (c) in Signal Detection calculations: e.g., Stanislaw & Todorov, 1999). Consistent with the broader hypothesis that people are more skeptical of the self-relevance of negative information (Ditto & Lopez, 1992), people should require stronger feelings of familiarity (i.e., a more conservative threshold) with negative feedback before they will acknowledge their recognition of that feedback as self-relevant.

If negative, self-relevant feedback faces a larger burden at retrieval but is not altered during encoding, then it should be possible to improve recognition with retrieval aids. Therefore, researchers have tested whether experimental intervention to aid retrieval (i.e., not asking participants to freely recall information but rather to distinguish "old" feedback from "new" information) helps reduce the disparities in recognition thresholds across negative and positive self-relevant feedback. However, research has not consistently found that aiding retrieval reduces disparities in recognition thresholds. Whereas one study found that participants exhibited fairly similar recognition thresholds across negative and positive self-relevant feedback when asked to pick the feedback out of a list of feedback and lures (as measured by location c: Study 2: Green, Sedikides et al., 2008), this finding was not replicated in another study where differences in recognition threshold for negative, self-relevant feedback persisted (as measured by location c: Study 1: Green, Sedikides et al., 2008). Furthermore, studies employing customized indices of recognition thresholds also find persistent differences for negative feedback in a recognition task (Djikic et al., 2007, 2005). Procedural differences are unlikely to account for the inconsistent results across studies. Although the findings were inconsistent, many studies used a similar design in which participants were presented with lists of statements about personality behaviors, asked to freely recall as much of the feedback task as they could, and then performed a recognition task (i.e., Djikic et al., 2007, 2005; Study 1 and Study 2: Green, Sedikides et al., 2008).

Taken together, it has been challenging to integrate the research on encoding and retrieval to form a picture of whether people's higher standards of recognition for negative, self-relevant feedback begin during encoding or arise at the time of retrieval, or both. For example, the retrieval studies and the encoding studies have not delved into memory processing at the same level (i.e., recognition threshold versus the broader measure of accuracy). Additionally, despite the procedural parallels, the retrieval studies do not use a consistent operationalization of the primary variable of concern (i.e., recognition threshold). Finally, very little empirical attention has been given to an alternative hypothesis, that is, that the conservative recognition threshold reflects people's suppression of knowledge about negative, self-relevant feedback.

Therefore, the present research examined whether feedback encoded in relation to the self is subject to differential recognition thresholds (i.e., extent to which feedback must feel familiar before it is recognized) when the desire to cast oneself in a positive light is irrelevant or further incentivized with financial reward after encoding has taken place (Study 1) or before encoding has taken place (Study 2). Furthermore, Study 3 examines whether the neural markers of forgotten negative, self-relevant feedback are better characterized by patterns previously associated with memory differences or with suppressed knowledge.

Study 1

Study 1 tested how recognition thresholds for negative self-relevant feedback are affected by socioemotional incentive or a combination of socioemotional and financial incentives. That is, the two levels of incentives permit the investigation of whether recognition thresholds show significant change in relation to the removal of threat (i.e., socioemotional concerns) or the combination of the removal of threat and the possibility of reward (i.e., financial gain). Participants were presented with bogus feedback about their personality traits and then given a surprise recognition test. If the differential recognition thresholds for negative feedback are not dependent on processes that occur as early as encoding, then participants should be capable of lowering their recognition thresholds to be more liberal (i.e., act on lower signals of familiarity) once they become aware that feedback is non-threatening or the stakes become even more desirable with the addition of financial reward. However, if recognition thresholds stem from processes that occur during encoding, then recognition thresholds for negative feedback should continue to be more conservative despite the removal of self-threat or the further addition of financial reward because those incentives cannot undo processes that transpired during encoding.

Methods

Participants

Study 1 included two samples who performed a direct replication of the experimental procedure and were analyzed according to the recommendations of Integrative Data Analysis which advocates for the pooling of data sets to optimize statistical power and assess replication when the original data are available (rather than meta-analyses when only effect sizes are available: Curran & Hussong, 2009). 367 total participants were included in the analyses (252 females, M_{age} = 19.08 years, SD = 1.21; demographic information, and results from the individual samples are included in the Supplement). A priori exclusion criteria ensured analyses were based on meaningful trials and participants who were engaged in the task (see Supplement for additional information on distribution across conditions). That is, additional participants were excluded because they responded on fewer than 80% of the trials (Sample 1a: 5 participants; Sample 1b: 13 participants) and one additional participant was excluded due to confusion about the task (Sample 1a). Participants received course credit for participation and gave informed consent in compliance with the human subject regulations of the University of Texas at Austin.

Procedure overview

Study 1 modified a bogus personality feedback procedure used in previous research investigating memory for self-relevant feedback (Djikic et al., 2005). Participants completed a personality assessment task, received bogus feedback about their personality, and then completed a surprise recognition test for the feedback which introduced a feedback manipulation at the halfway point (Figure 1).

Feedback conditions were randomly assigned and manipulated the extent to which memory for negative feedback was non-threatening or non-threatening and financially rewarding (Non Self-Relevant, Non Self-Relevant + Financial Reward, Control) for the second part of the surprise recognition test. The task was presented using E-Prime 2 (Psychology Software Tools, INC., Sharpsburg, PA). Each component is described below.

Personality assessment. All participants completed personality questionnaires: the Narcissistic Personality Inventory (Raskin et al., 1991), the Rosenberg Self-Esteem Scale (Rosenberg, 1965), and the Big Five Inventory (John et al., 2008). Two subjective tasks were also included to increase the believability of subsequent feedback (see Supplement).

Bogus personality feedback. Participants then received feedback ostensibly calculated from the personality assessment. However, the feedback was actually the same for all

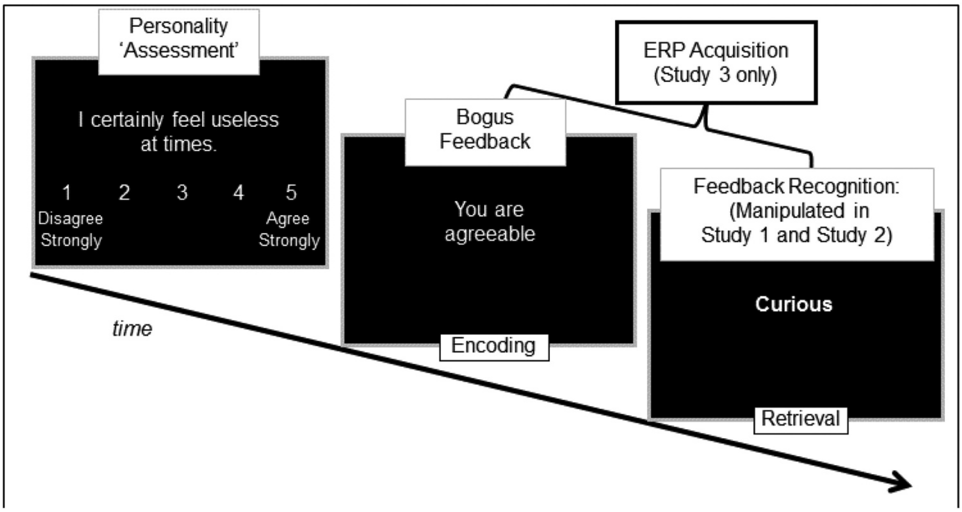

Figure 1. Study design. Participants completed a series of personality questions. Next, they received bogus personality feedback (50% positive trait words, 50% negative trait words). Finally, participants completed a recognition test for the feedback. In Study 1, a manipulation was introduced halfway through the surprise Feedback Recognition task to vary the extent to which memory for negative feedback was non-threatening or non-threatening and financially rewarded. None of the participants were aware of the incentives until after encoding had occurred. In Study 2, we manipulated the relevance of the Personality Assessment, Bogus Feedback, and Feedback Recognition to the participant's personality or the personality of another person. Additionally, we manipulated whether a financial incentive for memory was offered before encoding had begun. In Study 3, ERP data was acquired during the Bogus Feedback and a surprise Feedback Recognition task.

participants (80 positive and 80 negative traits: Anderson, 1968, see https://osf.io/q26xu/). Participants first saw a screen which said "You are" (1000 ms). The "You are" stem was then completed with one trait randomly drawn from a list of 160 traits (2000 ms). To ensure that participants were attending to the feedback, they were asked to press a key when the trait appeared on screen. Trials were separated by screens with a fixation cross (1000 ms).

Surprise recognition test of feedback. The surprise recognition test included 320 trait words (https://osf.io/q26xu/): the 160 traits presented in the experiments and 160 lures (80 positive, 80 negative). Trait words were presented (1000 ms) and trials were separated by a screen with a fixation cross (1500 ms). Participants used the keyboard to indicate whether they had previously seen the trait in their feedback or if it was a new word (responses collected during the trait word presentation and the following fixation screen: 2500 ms total to respond).

For all three recognition conditions, the random assignment affected the instructions that participants received after completing the first half of the recognition test (40 positive old, 40 positive new, 40 negative old, and 40 negative new). In the first half, participants reported their recognition for the trait words with the understanding that the trait feedback had come from their personality assessment. After the first half (i.e. Part 1), participants were interrupted by the experimenter and told one of three things. In the Non Self-Relevant Feedback condition, participants were told that the feedback they received was actually meant for someone else and given to them by mistake. They were told that despite the error, we were still interested in their memory performance and asked to finish the task. This manipulation ensured that the negative feedback actually had no bearing on the self and, therefore, was not threatening to retrieve during the recognition task. In the Non Self-Relevant + Financial Incentive condition, participants were also told that the feedback was actually meant for someone else and were further instructed that they would receive a cash bonus for correct identification of feedback as being old or new (e.g., a bonus of up to 10 USD based on two randomly selected trials from the remaining recognition test). This manipulation added a financial incentive to retrieve memories of negative feedback. In the Control condition, participants were told that the interruption was to prevent fatigue.

Behavioral analysis. The present research focused on a standardized measure of recognition threshold used in previous research on self-enhancement of memory (Green, Sedikides et al., 2008; Paulhus et al., 2003). Specifically, criterion location (c) from Signal Detection Theory (SDT) assessed the extent to which participants claimed recognition based on little feeling of familiarity (e.g., Stanislaw & Todorov, 1999). From the perspective of SDT, criterion location(c) indicates the strength of an internal feeling of familiarity that a participant needs to claim recognition. Criterion location(c) is calculated by considering hits and false alarms:

$$C = (\text{Hits} + \text{FalseAlarms})/2$$

Higher numbers reflect a more liberal threshold, that is, lower levels of internal familiarity are needed before claiming recognition. Data were analyzed with a two between-subjects factors (Recognition condition: Non Self-Relevant Feedback, Non Self-Relevant Feedback + Financial Incentive, Control; Study: Sample 1a, Sample 1b) and two within-subject factors (Valence: Positive, Negative; Time: Part 1, Part 2) ANOVA. The factor of time was analyzed because it was part of the experimental procedure and we planned to test whether any significant interactions may be accounted for by fatigue. Note that the results are similar if only considered in relation to proportion of correctly recognized traits or after the recognition manipulation (see Supplement).

Results

Participants had a more conservative threshold for reporting that they remembered the negative traits from their feedback (Main effect of Valence, $F(1,361) = 101.74$, $p < 0.001$, $\eta^2 = 0.22$; Control Condition: $F(1,361) = 20.34$, $\eta^2 = .053$ (Part 1), $F(1,361) = 21.58$, $\eta^2 = .056$ (Part 2), ps < .001; Non Self-Relevant: $F(1,361) = 28.66$, $\eta^2 = .074$ (Part 1), $F(1, 361) = 26.82$,, $\eta^2 = .069$ (Part 2), ps < .001; Non Self-Relevant + Financial Reward: $F(1, 361) = 34.67$, $\eta^2 = .088$ (Part 1), $F(1,361) = 42.13$, $\eta^2 = .105$ (Part 2), ps < .001) which persisted even after threat was removed and after financial reward was additionally possible (interaction between Valence, Time, and Recognition condition, $F(2,361) = 0.214$, $p = 0.807$, $\eta^2 = 0.001$; Figure 2; this pattern was consistent even when data from Part 2 was considered in isolation, see Supplement). Data sample did not significantly affect results (interaction of Study, Valence, Recognition, and Time, $F(2,361) = 1.571$, $p = 0.209$, $\eta^2 = 0.009$). One interpretation of Study 1's results is that encoding may operate as

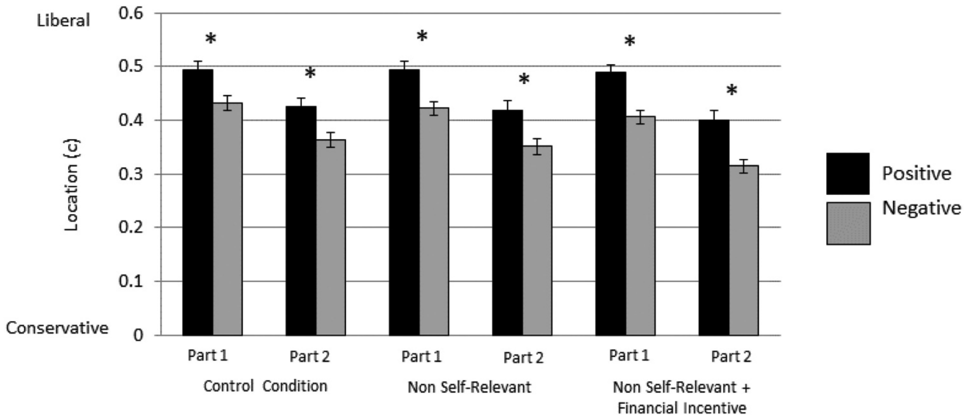

Figure 2. Study 1 Results. Participants used more liberal thresholds to claim recognition of positive feedback as compared to negative feedback regardless of condition (Non Self-Relevant, Non Self-Relevant + Financial Incentive, or Control). In all conditions, Part 1 required participants to complete the recognition test while they believed the feedback to be self-relevant. Additional manipulation was introduced in Part 2 (or not in the case of the Control Condition). Providing psychological or additional financial incentives for memory did not result in a significant shift of threshold for claiming negative feedback. (Bars indicate standard error).

a "point of no return" for some of the underlying processes that people subsequently use to assess whether they recognize feedback. If recognition thresholds are shaped as early as encoding, then incentives presented after the fact would be unlikely to be effective. That is, once encoding has taken place, people may be unable to adjust how much negative, self-relevant feedback needs to feel familiar before they recognize it. Despite the enticement of the incentives, participants were unable to capitalize on the incentives because they could not adjust their recognition thresholds for negative, self-relevant feedback (to a more liberal state where they are comparable to the recognition thresholds used for positive, self-relevant feedback) once encoding had taken place. (Study 2 will address the alternative possibility that incentives do not have a significant effect on recognition thresholds regardless of timing).

Equivalence testing of valence effects

Equivalence testing (i.e., two one-sided tests: Lakens, 2017) was conducted to contextualize the valence effects found in the main analysis. Upper and lower bounds were selected (raw difference of .06 to −.06, see Lakens, 2017) to test whether observed valence effects might be considered equivalent to an effect size that is too small to consider as a meaningful difference between conditions (i.e., fell significantly within upper and lower bounds) or an effect that may be of interest (i.e., fell outside upper and lower bounds, that is, not significantly within the equivalence bounds). In the main analyses, significant differences were found between positive and negative valence in all conditions and interpreted as consistent with previous research which shows that people need to have a stronger feeling of familiarity before they recognize negative feedback as self-relevant (when compared to positive feedback). Therefore, we expected that the equivalence testing would find that the observed effects did not significantly fall within a range of differences that were equal to or close to zero. As expected, for all conditions, the observed valence effect sizes were outside of the equivalence bounds: Control Condition: t(120) = −0.3, p = .62 (Part 1), −0.2, p = .58 (Part 2); Non Self-Relevant: t (126) = −1.04, p = .84 (Part 1), −.76, p = .77 (Part 2); Non Self-Relevant + Financial Reward: t(118) = −0.13, p = .55 (Part 1), −0.44, p = .66 (Part 2). Therefore, the equivalence test results were consistent with the interpretation of the main analyses. The equivalence test results do not support the concern that some of the conditions yielded valence effects (which happen to be statistically significant from a null hypothesis testing approach) that are small enough to overlap with non-meaningful differences.

Study 2

Study 2 built on Study 1 by further investigating the effects of socioemotional and financial incentives on recognition thresholds for personality feedback. Study 1 suggests that negative, self-relevant feedback is processed in a manner in which recognition thresholds cannot be adjusted if incentives are presented after encoding has occurred. However, an alternate explanation is that the incentives were simply incapable of changing recognition thresholds. Therefore, Study 2 examined how escalating degrees of socioemotional and financial incentives presented before encoding affected differences in recognition thresholds. Participants performed a similar task to Study 1 (Figure 1). Study 2 tested how differences in recognition thresholds were affected by knowing about a financial reward, non self-relevance, or the combination before encoding occurred. If people can be incentivized to engage in spontaneous encoding of negative feedback that does not skew recognition thresholds, then knowing about incentives (socioemotional, financial, both) before encoding should significantly reduce differences in recognition thresholds between negative and positive personality feedback.

Methods

Participants

As in Study 1, Study 2 included two samples analyzed according to the recommendations of Integrative Data Analysis (Curran & Hussong, 2009). 586 participants were included in the analyses (416 females, M_{age} = 19.04 years, SD = 2.36; demographic information, power analysis, and results from the individual samples are included in the Supplement, preregistration at osf.io/wy9v6, osf.io/w7hb8). Additional participants were excluded if they responded on fewer than 80% of recognition test trials (Sample 2a = 9; Sample 2b = 11), asked that their data not be used (Sample 2a = 10; Sample 2b = 7), or answered manipulation check questions incorrectly (Sample 2a = 2; Sample 2b = 14; see Supplement for more information on distribution across conditions). Participants received course credit for participation and gave informed consent in compliance with the human subject regulations of the University of Texas at Austin. All measures, manipulations and exclusions are reported.

Procedure

he procedure for Study 2 was very similar to Study 1 with three exceptions: (1) there was no break during the recognition task, (2) non self-relevance was manipulated by asking participants to perform all tasks about a peer and (3) the incentive conditions were designed to reflect increasing degrees of incentive (Self-Relevant, Self-Relevant + Financial Reward, Non Self-Relevant, and Non Self-Relevant + Financial Reward) (see Figure 1). The Self-Relevant condition paralleled the control condition in Study 1: participants completed the personality assessment about themselves, the recognition task was a surprise, and no financial incentives were offered. In the Self-Relevant + Financial Reward condition, participants were informed about the possibility of receiving a 10 USD bonus in a subsequent recognition task before they completed the personality assessment about themselves. In the Non Self-Relevant condition, participants completed the personality assessment, bogus feedback, and surprise feedback recognition task about a peer. Participants were shown a picture of a peer ostensibly named Chris (see Green, Sedikides et al., 2008) and gender matched to the participant. Chris was presented as a student who had previously participated in the study and agreed to share their personality feedback (study materials available at https://osf.io/9rbhj/). No financial incentive was offered. In the Non Self-Relevant + Financial Reward condition, participants were instructed about the possibility of a receiving a 10 USD bonus in a subsequent recognition task before they completed the personality assessment task about Chris.

Behavioral analysis. As in Study 1, Study 2 examined location(c) from Signal Detection Theory. An ANOVA with one within-subject factor (Valence: Positive, Negative) and two between-subject factors (Incentive condition: Self-Relevant, Self-Relevant + Financial Reward, Non Self-Relevant, Non Self-Relevant + Financial Reward; Study: Sample 2a, Sample 2b) tested for significant effects on recognition thresholds (i.e., location (c); see Supplement for results from proportion of correctly recognized items and for comparisons between the Self-Relevant condition in Study 2 and Part 2 of the Control condition from Study 1).

Results

When financial reward was combined with non self-relevance, participants were significantly less likely to use differential recognition thresholds for positive versus negative personality feedback (interaction between Valence and Incentive condition, $F(3,578) = 2.99$, $p = 0.031$, $\eta^2 = 0.015$; Figure 3). Participants had a significantly more conservative threshold for remembering negative feedback compared to positive feedback (Main effect of Valence, $F(1,578) = 43.07$, $p < 0.001$, $\eta^2 = 0.069$; Self-Relevant: $F(1, 578) = 19.86$, $p < .001$, $\eta^2 = .033$; Self-Relevant + Financial Reward: $F(1, 578) = 14.28$, $p < .001$, $\eta^2 = .024$; Non Self-Relevant: $F(1, 578) = 18.19$, $p < .001$, $\eta^2 = .031$.) except in the Non Self-Relevant + Financial Reward condition ($F(1, 578) = .48$, $p = .48$, $\eta^2 = .001$) where thresholds for negative feedback tended to rise upwards to the thresholds used for positive feedback. Data sample did not have a significant effect on the results (interaction between Valence, Incentive condition, and Study, $F(3,578) = 0.686$, $p = 0.561$, $\eta^2 = 0.004$). These results suggest that the combination of psychological (i.e., removing concerns about self-enhancement) and financial incentive presented before encoding can significantly influence the extent to which people draw on similar recognition thresholds for personality feedback across valence. In other words, if sufficiently motivated before encoding takes place, people can lower the extent to which negative, self-relevant

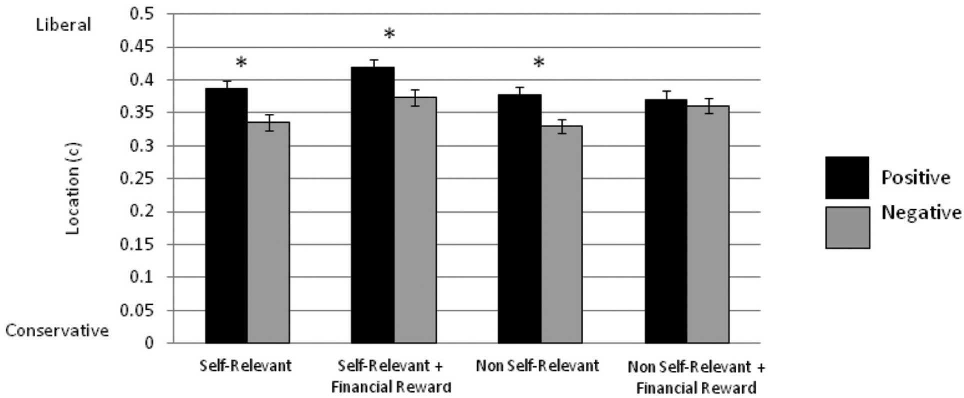

Figure 3. Study 2 Results. In Study 2, participants knew about the possibility of financial reward, non self-relevance, or their combination before encoding occurred. Participants were more likely to use more conservative thresholds to claim recognition of negative feedback compared to positive feedback except when they were incentivized by both the elimination of self-esteem threat and possibility of financial reward.

feedback needs to feel familiar such that it is comparable to their standards for recognizing positive self-relevant feedback.

Equivalence testing of valence effects

As in Study 1, equivalence testing (Lakens, 2017) with the same upper and lower bounds was conducted to contextualize the valence effects found in the main analysis. Consistent with the main analyses, observed valence effects fell outside of the equivalence bounds for the Self-Relevant condition ($t(144) = 0.75$, $p = .22$), Self-Relevant + Financial Reward condition ($t(131) = 1.37$, $p = .09$), and Non Self-Relevant condition ($t(166) = 1.15$, $p = .13$) yet fell within the equivalence bounds for the Non Self-Relevant + Financial Reward condition ($t(141) = 5.2$ $p < .001$). Therefore, the equivalence test results were consistent with the interpretation of the main analyses: the Non Self-Relevant + Financial Reward condition alone yielded an observed effect that was suggestive of a lack of difference in recognition thresholds for negative and positive feedback. When they knew about the possibility of financial reward and the feedback was not personally threatening before they began to learn the feedback, people tended to use similar standards for how much feedback had to feel familiar across the negative and positive conditions.

Study 3

Study 3 builds on Study 1 and 2 by testing other markers, that is, neurophysiological differences associated with forgotten negative, self-relevant feedback at encoding and retrieval. The results from Study 1–2 are consistent with the hypothesis that recognition thresholds may be rooted in processing that transpires as early as the encoding of negative, self-relevant feedback; incentives can affect recognition thresholds but only when people are aware of them before encoding and the incentives are large enough (i.e., need for self-defense has been removed and there is the opportunity for financial reward). Therefore, Study 1 and Study 2 suggest that negative, forgotten self-relevant feedback should be associated with neurophysiological markers that distinguish it from negative, remembered self-relevant feedback during encoding and retrieval. Furthermore, the implications of Study 1 and Study 2 suggest that there should not be support for the alternative hypothesis that has received little empirical attention so far in the literature, that is, that people are suppressing their recognition of the negative, self-relevant feedback at retrieval. The examination of neurophysiological differences between forgotten negative, self-relevant feedback and new negative feedback is one way to test whether participants may be failing to explicitly express recognition of negative, self-relevant feedback.

Therefore, the central question becomes how do the neurophysiological markers of forgotten negative, self-relevant feedback compare to the neurophysiological markers of remembered negative, self-relevant feedback and novel negative information? One hypothesis is that self-reported memory is meaningful, that is, event-related potentials (ERPs) from ongoing EEG activity can significantly distinguish between negative self-relevant feedback that is forgotten versus remembered. Previous research suggests that ERPs associated with forgotten information should be distinguishable from remembered information at the time of encoding and retrieval (Neville et al., 1986; Paller et al., 1987). An alternative hypothesis is that ERPs associated with forgotten, negative self-relevant feedback suggest suppressed knowledge of that feedback. Previous research finds that ERPs show significant differences for information that has been encoded but suppressed at the time of retrieval when compared to novel information (i.e., the late posterior negativity, Hu et al., 2015). In other words, if information is genuinely forgotten, then processing it should be similar to the processing of new information. However, when people simply conceal their recognition of information, their concealment is reflected in neurophysiological distinctions between the concealed information and new information.

Therefore, Study 3 examines two possible neurophysiological patterns to characterize the processing of negative, self-relevant feedback: (1) a pattern of difference associated with memory differences (i.e., a significant difference between remembered negative versus forgotten negative feedback at the time of encoding and retrieval) and (2) a pattern of difference associated with suppression of recognition (i.e., a significant difference between forgotten negative, self-relevant feedback compared to correctly identified new information, that is, correct negative rejections during a recognition task). We draw on a permutation approach for analyzing ERPs that has been used previously in our lab and others (Griffin & Schnyer, 2020; Trujillo, Allen, Schnyer, & Peterson, 2010; Sanguinetti et al., 2016). A permutation approach addresses the issues commonly associated with ERP analytic approaches that allow for experimenter flexibility in selecting time windows and electrode locations as well as inappropriate correction of multiple comparisons (see Luck & Gaspelin, 2017). For example, prior ERP research on distinguishing remembered from forgotten information does not use a consistent time window (e.g., Neville et al., 1986; Paller et al., 1987).

Methods

Participants

Analysis included 36 participants (28 females, M_{age} = 19.53 years, SD = 2.40) (see Supplement for power analysis). As in Studies 1–2, three additional participants were excluded due to responses on less than 80% of either encoding or recognition trials. Participants were right-handed, native English speakers, and were screened for medications, neurological, or psychological conditions that might affect the neural responses or psychological effects being tested (i.e., clinical depression, head trauma, epilepsy, etc.). All participants gave informed consent in compliance with the human subject regulations of the University of Texas at Austin.

Procedure

The behavioral procedure for Study 3 was similar to Studies 1 and 2 with three exceptions: (1) there was no recognition incentive manipulation, (2) the bogus feedback consisted of 85 positive traits and 85 negative traits, and (3) the feedback recognition task included 85 positive lures and 85 negative lures. The increased numbers of traits ensured there would be sufficient power (i.e., trials per condition) to conduct the planned ERP analyses as a function of remembered and forgotten feedback. All measures, manipulations and exclusions are reported.

Behavioral analysis. As in Studies 1–2, Study 3 tested for differences in location(c) for positive and negative personality feedback.

ERP acquisition and processing. Sixty-four channels of continuous EEG data were recorded using BrainVision PyCorder and processed with the Analyzer 2 software (BrainVision LLC, Morrisville, NC) and custom Matlab scripts. Four additional electrodes, in and outside of the cap, were used to record horizontal and vertical eye movements. Impedances were kept below 5 k Ω. Caps were constructed and positioned on each participant to conform to the extended 10–20 International System.

Offline, data were band-pass filtered (0.1–30 Hz, respectively) and re-referenced to the linked mastoids (TP9 and TP10). Continuous EEG was then epoched starting at 200 ms before to 2000 ms after the onset of the stimulus for each condition. Ocular artifacts were removed by deriving bipolar eye channels and employing the Gratton & Coles method of ocular correction. Finally, trials were averaged into individual conditions (Encoding and Retrieval: negative later remembered ($M_{\text{trial count}}$ = 38), negative later forgotten ($M_{\text{trial count}}$ = 46); Retrieval: correctly identified as new negative feedback ($M_{\text{trial count}}$ = 60)) and all epochs were baselined to an average of the peristimulus period of −200 to 0 ms.

Epoched data were analyzed using non-parametric randomized permutation pairwise comparison approach and were cluster corrected for multiple comparisons across time and electrode site ($p < 0.05$, 20,000 permutations; Nichols & Holmes, 2002; Trujillo, Allen, Schnyer, & Peterson, 2010; Sanguinetti et al., 2016). This method of analysis is advantageous because it utilizes recorded data from across the whole scalp, thereby avoiding subjective decisions about regions of interest and time windows as in past methods of ERP analysis (see Trujillo et al., 2010; Sanguinetti et al., 2016; Nichols & Holmes, 2002). By applying cluster correction algorithms for multiple comparisons, it also avoids the problems of inflated alpha levels associated with traditional t-tests.

To perform these tests, independent statistical significance thresholds for each data point were determined by estimating a *t*-distribution from the data for each electrode and time-point, computing *t*-statistics from each of 20,000 random between condition permutations of data across conditions under the null hypothesis. For each of these permutations, a random subset of conditions were swapped before *t*-values were computed. Under the null hypothesis, these *t*-values are elements of the null distribution. Thus, 20,000 *t*-values are created to form a data driven distribution, and a two-tailed $p = .05$ primary threshold was determined for each data point. These thresholds form a three-dimensional matrix where two dimensions preserve the topographic organization of the electrodes, and the third dimension is time.

In a second step, these significance thresholds were used to determine contiguous locations where clusters of data exceeded the significance thresholds. A second round of 20,000 permutations were computed. During each permutation, the $p = .05$ thresholds achieved in the first step were applied at each data point, thus determining which points exceed this threshold. Contiguous clusters were formed from points that have *t*-values above these thresholds; a maximal cluster size is determined for each permutation step, yielding a distribution of 20,000 maximal cluster values under the null hypothesis. Lastly, in a third step, this distribution of maximal cluster sizes is used to test *t*-statistic cluster sizes from the true dataset. Clusters in the actual dataset with *t*-statistics greater than the maximal cluster distribution's $p = .05$ criterion cluster size are considered significant at the two-tailed level, thus providing strong control for type-I errors.

Results

Behavioral results: More conservative thresholds for recognizing negative, self-relevant feedback compared to positive, self-relevant feedback

Consistent with Studies 1 and 2, participants had more conservative thresholds for claiming familiarity with negative feedback than positive feedback, $t(35) = -4.58$, $p < 0.001$, $d = 0.84$ (Figure 4). That is, participants needed a stronger feeling of familiarity with negative feedback before they recognized it from the self-relevant feedback task (when compared to the feeling of familiarity needed to claim recognition of positive, self-relevant feedback).

Self-reported memory is associated with neurophysiological distinctions for negative self-relevant feedback at encoding and retrieval

Consistent with the implications raised by Study 1 and Study 2 results, the ERP analysis suggested that self-reports of memory were associated with differences at the neurophysiological level of analysis. ERPs revealed differences associated with forgotten negative feedback at the time of encoding and retrieval when compared to remembered negative feedback. During the encoding phase of the task, a late time period of the ERP was generally smaller for negative feedback that would later be forgotten (in the surprise recognition task) than for feedback that would subsequently be remembered (i.e., a cluster spanning frontal to posterior sites between 700–800 ms after stimulus onset, see Figure 5(a)). During the retrieval phase of the task, it was a later portion of the ERP that was associated with forgotten negative feedback, where there was a larger, more positive response for forgotten negative feedback (i.e., a cluster on the central scalp between 600–1000 ms after stimulus onset, see Figure 5(b)). In other words, participants' neurophysiological responses to negative feedback were consistent with their claims that they remembered some of the feedback and forgot other parts of it. Negative, self-relevant

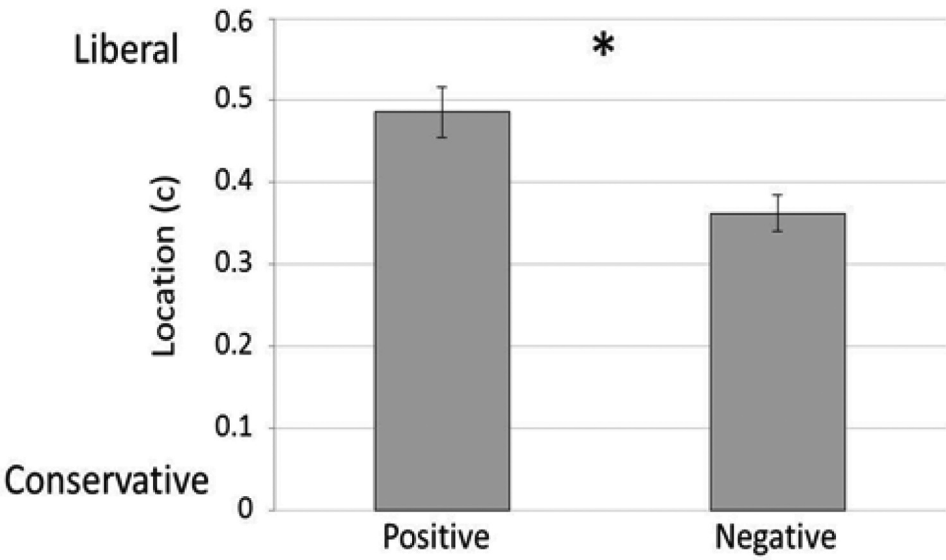

Figure 4. Study 3 Behavioral Results. As in Studies 1 and 2, participants had a more conservative threshold for claiming recognition of negative feedback compared to positive feedback.

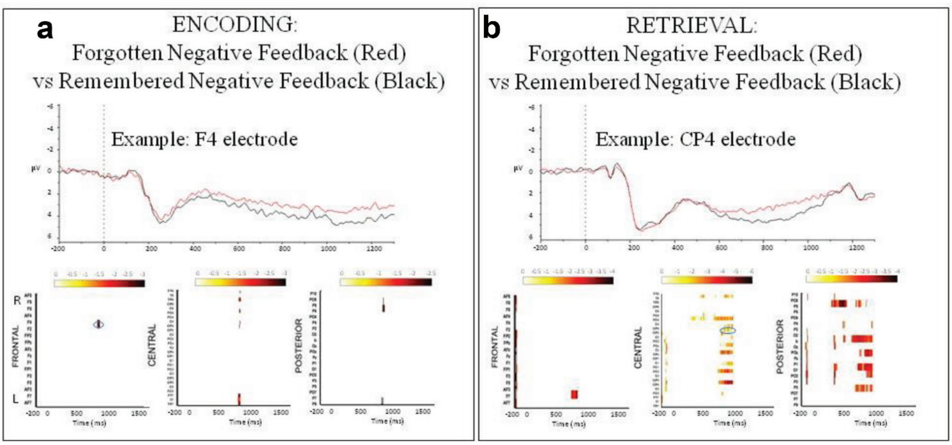

Figure 5. ERP analysis found significant distinctions as a function of self-reported memory for negative self-relevant feedback at both encoding (Panel A) and retrieval (Panel B). In each panel, the bottom graphs indicate clusters where there are significant differences (color gradient scale indicates effect size) between conditions as a function of time (x-axis) and electrode location (y-axis). Top graphs show representative waveforms of each condition at one electrode for visualization purposes (electrode site circled in blue on bottom graphs).

feedback that was remembered was neurophysiologically distinct from negative, self-relevant feedback that was forgotten and differences were observed both during encoding and retrieval.

ERPs associated with negative, self-feedback that is self-reported as forgotten are not significantly distinct from ERPs associated with correctly identified novel feedback

Also consistent with the implications raised by Study 1 and Study 2 results, participants did not significantly exhibit neurophysiological patterns that would suggest that they were suppressing their recognition of the negative, self-relevant feedback. The cluster corrected threshold analyses did not identify any statistically significant ERP differences at the permutation threshold of .05 (two-tailed) for the forgotten negative feedback compared to feedback words that were correctly identified as novel. In other words, participants' neurophysiological response to negative feedback they claimed to not remember was not statistically distinguishable from the ERP response to information they were seeing for the first time.

General discussion

The present research provides greater insight into the processes that contribute to differential recognition thresholds for negative, self-relevant feedback. The mixed findings in the existing literature are difficult to integrate because they have not addressed naturally occurring encoding and have employed inconsistent memory indices to examine encoding versus retrieval (e.g., Green, Sedikides et al., 2008; Djikic et al., 2007, 2005; Sedikides & Green, 2000; Zengel et al., 2018). The current research focused on a previously studied and relatively subjective aspect of memory: recognition threshold which is the strength of internal feelings of familiarity needed to claim familiarity (i.e., location c). In Study 1, recognition thresholds for negative feedback remained relatively more conservative even when incentivized through decreased self-relevance or decreased self-relevance and opportunity for financial gain. Study 2 established that it is possible to incentivize more equivalent recognition thresholds across positive and negative feedback but only when self-relevance is decreased and a financial incentive is presented before encoding takes place. Study 3 investigated neurophysiological signatures to more fully characterize processing within the negative, self-relevant feedback condition itself. Consistent with the implications raised by Study 1 and Study 2 findings, Study 3 suggested that neurophysiological associations with self-reported forgotten negative feedback are better characterized by memory differences than suppressed knowledge. Taken together, the results from the present research support the hypothesis that impaired recognition thresholds for negative feedback stem from processing that occurs as early as the stage of encoding and that future research will benefit from understanding these effects in relation to self-perception motives as well as perceptions of other people.

The current research builds on previous research by addressing how recognition thresholds are related to encoding processes for negative, self-relevant feedback. Previous research focused on the encoding stage had shown that variables known to impair encoding such as limited time or mental load (i.e., factors known to make encoding difficult) introduced in relation to encoding personality feedback can undermine memory (Sedikides & Green, 2000; Zengel et al., 2018). However, it was unknown whether these experimental interventions exacerbated impaired processing which typically operates during encoding of personal feedback or created experimental artifacts. It was also unclear how to integrate findings from the encoding stage with research focused on the retrieval of personality feedback which has more often used a different marker of poor memory (i.e., recognition threshold). The findings from the current research build on previous research by suggesting that differences in recognition thresholds for negative, self-relevant feedback are rooted in processing that occurs at the encoding stage rather than suppression at the retrieval stage. The combined incentive of financial reward and reduced self-enhancement concerns was associated with a change in recognition threshold only when the incentive was known before encoding took place. That is, it took the greatest level of incentive to improve recognition thresholds for negative, self-relevant feedback and it was only effective when participants knew about them before encoding took place. Future research should investigate whether even more incentive would be needed if the negative feedback was tailored to be particularly threatening to each individual. Furthermore, the ERP findings favored a memory difference explanation rather than significantly suggesting that participants successfully encoded feedback but later suppressed their recognition. Therefore, the current research advances our understanding of impaired memory for negative self-relevant feedback by suggesting that compromised processing at the encoding stage contributes to differences in subsequent recognition thresholds.

The current research also raises new hypotheses about how various self-perception motivations may impact recognition thresholds for self-relevant information. Whereas previous research has failed to find robust support for the role of self-verification in free recall of negative, self-relevant feedback (e.g., Study 2: Sedikides & Green, 2004; Study 3 failed to find statistical significance for individuals with negative self-image: Swann & Read, 1981), future research might examine the role of self-verification in recognition of self-relevant feedback. Do people with negative self-image seek to confirm that self-image by using a more liberal recognition threshold for negative self-relevant feedback? Additionally, people are not always motivated toward self-enhancement or self-protection. Sometimes they are interested in self-improvement which motivates them to seek and remember diagnostic feedback including negative feedback (Green, Pinter et al., 2005; Green, Sedikides et al., 2009). Our research suggests two new hypotheses about the influence of self-improvement on feedback processing: self-improvement motivations should be associated with (1) more conservative thresholds for recognizing all self-relevant feedback and (2) more equivocal thresholds across negative and positive self-relevant feedback.

The current research also suggests a new hypothesis about the role of memory impairment in the long term consequences of self-enhancement. For example, previous research has suggested that self-enhancement in the academic domain (defined as self-perceptions that are more favorable than an objective measure of the self's qualities) can be associated with poor long-term outcomes such as lowered self-esteem and reduced interest in academic environments (Robins & Beer, 2001; but see Dufner, Reitz, & Zander, 2014 for evidence of cultural and age differences). A prevalent explanation for the negative long-term consequences is that people eventually find themselves unable to suppress the retrieval of negative feedback. In contrast, the current findings suggest that people's tendency to recognize negative feedback as self-relevant may be weaker than their tendency to recognize the self-relevance of positive feedback, so there might be little to suppress. Therefore, it may be that the long-term consequences arise from other factors associated with self-enhancement. However, it is also possible that people may not be able to maintain preferential standards of recognition for self-relevant feedback over the long term. The current research points to the need for future research which investigates the ability to engage in sustained preferential internalization of self-relevant feedback as a function of valence and its relation to the long-term consequences associated with motivated self-processing.

The current research also suggests that recasting our understanding of memory biases for personality feedback as a uniquely self process may be warranted. While much of the previous work on recognition thresholds for personality feedback has focused on the self (e.g., Djikic et al., 2007, 2005), some studies have included both self and others (Green, Sedikides et al., 2008). As in previous research (Study 1: Green, Sedikides, et al., 2008), the current study has also found that recognition thresholds can be more conservative for negative information about another person. The present research builds on existing research by replicating previous findings and raising the possibility that recognition thresholds for feedback about another person may respond to financial incentive so long as it is presented before encoding of feedback takes place. If differential recognition thresholds for valenced personality feedback is a method of self-protection, then why have multiple studies found that this effect also occurs for recognizing personality feedback about another person? Future research might more deeply investigate whether the same or different underlying processes contribute to differential recognition thresholds found for personality feedback about self and other. For example, perhaps factors such as relationships, situations, and centrality of personality domains can predict similar movement for recognition thresholds for personality feedback about self and other. This hypotheses stems from research suggesting that poor memory for negative, self-relevant feedback may be most pronounced in certain conditions and for certain kinds of traits (e.g., Sedikides et al., 2016). Alternatively, it may be possible that the underlying mechanisms of recognition thresholds are different when processing feedback about the self versus others. For example, even the phenomenon of unrealistically positive social comparisons about the self arises from distinct underlying motivations: cognitive tendencies and self-defense (e.g., Beer, 2014; Chambers & Windschitl, 2004; Taylor & Brown, 1988). Future research is needed to more systematically understand how valence effects on recognition thresholds for self-relevant feedback relates to the valence effects on recognition thresholds for other feedback.

In summary, the current research builds on our understanding of why people exhibit a “higher bar” for recognizing negative, self-relevant feedback compared to positive, self-relevant feedback. It has been unclear as to whether people were simply claiming they did not recognize the negative, self-relevant feedback because they did not want to admit they have received it or whether the negative feedback is just processed in a more impoverished manner which makes it difficult to retrieve. Taken together the current studies favor the “impoverished processing account” over the “suppression account.” That is, negative, self-relevant feedback may be encoded in a weaker manner than positive information and the studies show that there is room to improve its encoding but only when people are sufficiently incentivized before encoding has taken place. Future research may benefit from a focus on how the timepoint at which incentives are offered interacts with their influence on recognition thresholds.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This research was supported in part by NSF-BCS [1147776] to J.S.B.

Data availability statement

Data for all of the reported samples and experiments are available at https://osf.io/k3c69/

Trait stimuli words are available at https://osf.io/q26xu/

Screen shots of experimental protocol available at https://osf.io/9rbhj/

Study 2 pre-registration available at https://osf.io/wy9v6/(Sample 2a) and https://osf.io/w7hb8/ (Sample 2b)

References

Alicke, M. D., Zell, E., & Guenther, C. L. (2013). Social self-analysis. constructing, protecting, and enhancing the self. In *Advances in Experimental Social Psychology* (Vol. 48) (pp. 173–234). Cambridge, MA: Academic Press. https://doi.org/10.1016/B978-0-12-407188-9.00004-1

Anderson, N. H. (1968). Likableness ratings of 555 personality-trait words. *Journal of Personality and Social Psychology*, *9*(3), 272–279. https://doi.org/10.1037/h0025907

Beer, J. S. (2014). Exaggerated positivity in self-evaluation: A social neuroscience approach to reconciling the role of self-esteem protection and cognitive bias. *Social and Personality Psychology Compass*, *8*(10), 583–594. https://doi.org/10.1111/spc3.12133

Chambers, J. R., & Windschitl, P. D. (2004). Biases in social comparative judgments: The role of nonmotivated factors in above-average and comparative-optimism effects. *Psychological Bulletin*, *130*(5), 813–838. https://doi.org/10.1037/0033-2909.130.5.813

Curran, P. J., & Hussong, A. M. (2009). Integrative data analysis: The simultaneous analysis of multiple data sets. *Psychological Methods*, *14*(2), 81. https://doi.org/10.1037/a0015914

D'Argembeau, A., & Van der Linden, M. (2008). Remembering pride and shame: Self-enhancement and the phenomenology of autobiographical memory. *Memory (Hove, England)*, *16*(5), 538–547. https://doi.org/10.1080/09658210802010463

Ditto, P. H., & Lopez, D. F. (1992). Motivated skepticism: Use of differential decision criteria for preferred and nonpreferred conclusions. *Journal of Personality and Social Psychology*, *63*(4), 568–584. https://doi.org/10.1037/0022-3514.63.4.568

Djikic, M., Chan, I., & Peterson, J. B. (2007). Reducing memory distortions in egoistic self-enhancers: Effects of indirect social facilitation. *Personality and Individual Differences*, *42*(4), 723–731. https://doi.org/10.1016/j.paid.2006.08.012

Djikic, M., Peterson, J. B., & Zelazo, P. D. (2005). Attentional biases and memory distortions in self-enhancers. *Personality and Individual Differences*, *38*(3), 559–568. https://doi.org/10.1016/j.paid.2004.05.010

Dufner, M., Reitz, A., & Zander, L. (2014). Antecedents, Consequences, and Mechanisms: On the Longitudinal Interplay Between Academic Self-Enhancement and Psychological Adjustment. *Journal of personality, 83*.

Gramzow, R. H., & Willard, G. (2006). Exaggerating current and past performance: Motivated self-enhancement versus reconstructive memory. *Personality & Social Psychology Bulletin*, *32*(8), 1114–1125. https://doi.org/10.1177/0146167206288600

Green, J. D., Pinter, B., & Sedikides, C. (2005). Mnemic neglect and self-threat: Trait modifiability moderates self-protection. *European Journal of Social Psychology*, *35*(2), 225–235. https://doi.org/10.1002/ejsp.242

Green, J. D., Sedikides, C., & Gregg, A. P. (2008). Forgotten but not gone: The recall and recognition of self-threatening memories. *Journal of Experimental Social Psychology*, *44*(3), 547–561. https://doi.org/10.1016/j.jesp.2007.10.006

Green, J. D., Sedikides, C., Pinter, B., & Van Tongeren, D. R. (2009). Two sides to self-protection: Self-improvement strivings and feedback from close relationships eliminate mnemic neglect. *Self and Identity*, *8*(2–3), 233–250. https://doi.org/10.1080/15298860802505145

Greenwald, A. G. (1980). The totalitarian ego. Fabrication and revision of personal history. *American Psychologist*, *35*(7), 603–618. https://doi.org/10.1037/0003-066X.35.7.603

Griffin, N. R., & Schnyer, D. M. (2020). Memory distortion for orthographically associated words in individuals with depressive symptoms. *Cognition*. [published online ahead of print, 2020 Jun 18].

Hu, X., Bergström, Z. M., Bodenhausen, G. V., & Rosenfeld, J. P. (2015). Suppressing unwanted autobiographical memories reduces their automatic influences: Evidence from electrophysiology and an implicit autobiographical memory test. *Psychological Science*, *26*(7), 1098–1106. https://doi.org/10.1177/0956797615575734

John, O. P., Naumann, L. P., & Soto, C. J. (2008). Paradigm shift to the integrative big five trait taxonomy. In *Handbook of Personality: Theory and Research* (pp. 114–158). New York, NY: Guilford Press. https://doi.org/10.1016/S0191-8869(97)81000-8

Kouchaki, M., & Gino, F. (2016). Memories of unethical actions become obfuscated over time. *Proceedings of the National Academy of Sciences*, *113*(22), 1–6. https://doi.org/10.1073/pnas.1523586113

Lakens, D. (2017). Equivalence tests: A practical primer for t tests, correlations, and meta-analyses. *Social Psychological And Personality Science*, *8*(4), 355–362. https://doi.org/10.1177/1948550617697177

Luck, S. J., & Gaspelin, N. (2017). How to get statistically significant effects in any ERP experiment (and why you shouldn't). *Psychophysiology.*, *54*(1), 146–157. https://doi.org/10.1111/psyp.12639

Neville, H. J., Kutas, M., Chesney, G., & Schmidt, A. L. (1986). Event-related brain potentials during initial encoding and recognition memory of congruous and incongruous words. *Journal of Memory and Language*, *25*(1), 75–92. https://doi.org/10.1016/0749-596X(86)90022-7

Paller, K. A., Kutas, M., & Mayes, A. R. (1987). Neural correlates of encoding in an incidental learning paradigm. *Electroencephalography and Clinical Neurophysiology*, *67*(4), 360–371. https://doi.org/10.1016/0013-4694(87)90124-6

Paulhus, D. L., Harms, P. D., Bruce, M. N., & Lysy, D. C. (2003). The over-claiming technique: Measuring self-enhancement independent of ability. *Journal of Personality and Social Psychology*, *84*(4), 890–904. http://doi.10.1037/0022-3514.84.4.890

Raskin, R., Novacek, J., & Hogan, R. (1991, March). Narcissism, self-esteem, and defensive self-enhancement. *Journal of Personality*, *59*(1), 19–38. https://doi.org/10.1111/j.1467-6494.1991.tb00766.x

Robins, R. W., & Beer, J. S. (2001). Positive illusions about the self: Short-term benefits and long-term costs. *Journal of Personality and Social Psychology*, *80*(2), 340–352. http://doi.10.1037//0O22-3514.80.2.340

Rosenberg, M. (1965). *Society and the adolescent self-image*. Princeton University Press.

Sanguinetti, J. L., Trujillo, L. T., Schnyer, D. M., Allen, J. B., & Peterson, M. A. (2016). Increased alpha band activity indexes inhibitory competition across a border during figure assignment. *Vision Research*, *126*, 120–130. https://doi.org/10.1016/j.visres.2015.06.008

Sedikides, C., & Green, J. D. (2000). On the self-protective nature of inconsistency-negativity management: Using the person memory paradigm to examine self-referent memory. *Journal of Personality and Social Psychology*, *79*(6), 906–922. https://doi.org/10.1037/0022-3514.79.6.906

Sedikides, C., & Green, J. D. (2009). Memory as a self-protective mechanism. *Social and Personality Psychology Compass*, *3*(6), 1055–1068. https://doi.org/10.1111/j.1751-9004.2009.00220.x

Sedikides, C., Green, J. D., Saunders, J., Skowronski, J. J., & Zengel, B. (2016). Mnemic neglect: Selective amnesia of one's faults. *European Review of Social Psychology*, *27*(1), 1–62. https://doi.org/10.1080/10463283.2016.1183913

Stanislaw, H., & Todorov, N. (1999). Calculation of signal detection theory measures. *Behavior Research Methods, Instruments and Computers*, *31*(1), 137–149. https://doi.org/10.3758/BF03207704

Swann, W. B., & Read, S. J. (1981). Acquiring self-knowledge: The search for feedback that fits. *Journal of Personality and Social Psychology*, *41*(6), 1119–1128.
Taylor, S. E., & Brown, J. D. (1988). Illusion and well-being: A social psychological perspective on mental health. *Psychological Bulletin*, *103*(2), 193–210. https://doi.org/10.1037/0033-2909.103.2.193
Taylor, S. E., Neter, E., & Wayment, H. A. (1995). Self-evaluation processes. *Personality & Social Psychology Bulletin*, *21*(12), 1278–1287. https://doi.org/10.1177/01461672952112005
Trujillo, L.T., Allen, J.B., Schnyer, D.M., & Peterson, M.A. (2010). Neurophysiological evidence for the influence of past experience on figure-ground perception. *Journal of Vision*, *20,* 1–21.
Walker, W. R., Skowronski, J. J., & Thompson, C. P. (2003). Life is pleasant–and memory helps to keep it that way! *Review of General Psychology*, *7*(2), 203–210. https://doi.org/10.1037/1089-2680.7.2.203
Zengel, B., Wells, B. M., & Skowronski, J. J. (2018). The waxing and waning of mnemic neglect. *Journal Of Personality And Social Psychology*, *114*(5), 719–734. https://doi.org/10.1037/pspa0000124