The Dissertation Committee for Anastasia Elena Rigney Certifies that this is the approved version of the following Dissertation:

THE ROLE OF BIASED SEARCHING THROUGH MEMORY IN MOTIVATED SOCIAL EVALUATION

Committee:

Jennifer S. Beer, Supervisor

Bertram Gawronski

Lisa A. Neff

David M. Schnyer

THE ROLE OF BIASED SEARCHING THROUGH MEMORY IN MOTIVATED SOCIAL EVALUATION

by

Anastasia Elena Rigney

Dissertation

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

Doctor of Philosophy

The University of Texas at Austin

May, 2019

Dedication

To my grandfather, Dr. Carl Jennings Rigney, for always being so eager to share his love of math and science with me.

Acknowledgements

There are many people I would like to acknowledge for their support and hard work in helping me get to where I am today. First, I would like to express my deepest gratitude to my mentor, Dr. Jennifer Beer, for training, encouraging, and inspiring me to be a better scientist. I genuinely do not have the words to express how much I have enjoyed working with you for the past five years. Second, I would like to thank my committee members, Dr. Bertram Gawronski, Dr. Lisa Neff, and Dr. David Schnyer, for their incredibly helpful feedback and support in the process of writing my dissertation and throughout my graduate training. Third, I am very grateful to the amazing lab mates I have been privileged to work with. Thank you to Dr. Gili Freedman, Dr. Taru Flagan, Dr. Jessica Koski, Dr. Michelle Harris, (soon to be Dr.) Skylar Brannon, and Serena Brandler. I would not have made it this far if it had not been for my friendship with each of you. Fourth, I would like to thank my family for encouraging and supporting me in my pursuit of higher education. And finally, I would like to thank my partner, Douglas Laustsen, for supporting me and changing the trajectory of his life to help me achieve my goals. More importantly, thank you, Doug, for always making life exciting and introducing me to new ideas and experiences every day.

Abstract

THE ROLE OF BIASED SEARCHING THROUGH MEMORY IN MOTIVATED SOCIAL EVALUATION

Anastasia Elena Rigney, Ph.D.
The University of Texas at Austin, 2019

Supervisor: Jennifer S. Beer

People do not always perceive their social world dispassionately; they often engage in motivated social evaluation. That is, people often do not evaluate themselves or other people objectively, but rather in a way that conforms to how they want to see the social target (i.e., a desired directional conclusion). For example, research shows that people tend to see themselves and liked others in an unrealistically positive light (Kruger, 1999; Tajfel & Turner, 2004; Taylor & Brown, 1988). Several researchers have posited biased searches through memory as an underlying mechanism supporting the phenomenon of seeing people in a certain light (Showers & Cantor, 1985; Kunda, 1990, Dunning, 2015). That is, it has been suggested that when aiming to paint a social target in a certain light people search through their memories or beliefs in ways that help them find information to support their desired directional conclusion. However, the methods used in existing research have made it difficult to understand if social evaluations that have been labeled as motivated actually reflect people striving for desired directional conclusions and what role biased memory searches may play. The proposed dissertation research addresses two overarching questions to understand the role of biased searches through memory in social evaluation. Research Question 1: What is the role of a) biased searches through memory and b) directional conclusions in the greater reported memory for positive self-relevant feedback (compared to negative self-relevant feedback; Studies 1, 2, & 3)? Research Question 2: Does biased searching through memory operate similarly when aiming to paint someone in a particular light (regardless of the directional conclusion) or only in a flattering light (Study 4)? A combination of experimental, neuroimaging (i.e., Event Related Potential), and computational modeling (i.e., Signal Detection Theory and Drift Diffusion Model) methods are used to address these questions.

Table of Contents

List of Figures

OVERVIEW AND BACKGROUND

People often do not see their social world as it is, but rather, the way they want to see it. Prior beliefs or expectations can shape the way information is processed about a social target. One prevalent example is the extent to which people sometimes have rose-colored glasses about the self and liked others (Taylor & Brown, 1988; Tajfel & Turner, 2004). The processes by which people evaluate social targets in overly positive ways manifest in many different domains including memory distortions (Sedikides & Green, 2009) and statistically unlikely personality judgments (Alicke, 1985). Despite many years of research exploring the phenomenon of motivated social evaluation, there is limited evidence supporting some of the underlying mechanisms that have been posited.

Motivated social evaluation: In what ways do people paint social targets in a positive or negative light?

Motivated social evaluation occurs when people's social evaluations are driven by a specific goal of how they want to view a social target (Showers & Cantor, 1985; Kunda, 1990, Dunning, 2015). For example, motivated social evaluations are theorized to arise when people have already made up their minds about their conclusion (i.e., have a desired directional conclusion) before they start to evaluate a social target. The literature finds robust effects for social evaluations that are theorized to reflect the posited role of desired directional conclusions in social evaluation. For example, people's desire to see themselves in a positive light is theorized to explain why they tend to remember positive, personal feedback at a greater rate than negative, personal feedback (Sedikides & Green, 2009), they are more likely to recall memories that support the claim that they possess desirable traits (Sanitioso, Kunda, & Fong, 1990), and they more readily forget past unethical behaviors they have perpetrated (Kouchaki & Gino, 2016). Additionally, people tend to disproportionately report self-evaluations and evaluations of liked others that are flattering when compared to evaluations of other people or objective markers (Kruger, 1999; Tajfel & Turner, 2004; Taylor & Brown, 1988). For example, evaluations of Obama and Trump depend on the evaluator's political values (Pew Research Center, 2016). More specifically, personality characterizations are consistent with the idea that people seek to confirm positive views to the extent that the political figure shares a political affiliation with the evaluator (and vice versa).

Underlying mechanisms: The role of biased searching through memory in motivated social evaluations

How are people able to paint social targets in a positive or negative light? Several theories have posited that a biased search through memory is one cognitive process that may support motivated social evaluation (Showers & Cantor, 1985; Kunda, 1990, Dunning, 2015). That is, people may search their existing memories when they are making social evaluations and this cognitive process can become biased when they apply different rules or standards for searching through their memories to support their desired directional conclusions. From this perspective, people's tendency to remember flattering personal feedback at a greater rate may reflect their motivation or desire to see themselves in a positive light (i.e., self-enhancement). Further, people may accomplish this difference in memory by having different standards for an internal sense of familiarity with positive, self-relevant memories before claiming recognition as compared to negative self-relevant memories. The role of biased searches through memory may also account for people's tendency to make flattering evaluations of themselves and people they like. That is, when evaluating a liked social target they may search more deeply through memory to find information that supports their desired directional conclusion of a positive evaluation. While several theories have posited a role for biased memory searches in motivated social evaluation, there are several limitations of extant research that have made it difficult to draw conclusions about its role (and, in some cases, even the role of desired directional conclusions).

Barriers to understanding the role of biased memory searches and desired directional conclusions in social evaluation

Why has it been challenging to understand the role of a biased search through memory in motivated social evaluations? First, previous research has operationalized memory in ways that do not shed light on underlying memory processes. Second, there are reasons to call into question whether some of the memory effects in the literature actually reflected motivated evaluations, that is, a desire to reach a directional conclusion. Finally, previous research has focused on evaluations in which people are aiming to make flattering evaluations rather than any motivated evaluation (e.g., unflattering, self-verifying, etc.), so it is unclear if biased memory searches operate similarly when reaching any desired directional conclusion as when aiming to reach flattering ones. The following sections outline in more detail these limitations of past research.

WHAT ROLE DOES BIASED MEMORY SEARCHING PLAY IN ASYMMETRIES IN MEMORY FOR POSITIVE AND NEGATIVE SELF-RELEVANT INFORMATION?

It is unclear what role biased searching through memory may play in motivated social evaluation because previous research has not actually measured biased searches through memory. Typically, memory has been operationalized as self-reported recall or recognition (Sedikides & Green, 2009). Researchers have rarely employed memory indices that allow for understanding the underlying role of biased memory searching. When researchers have used memory indices that operationalize underlying mechanisms they were either custom ones or superficially reported but not discussed or interpreted. The use of custom indices makes it difficult to interpret the findings in any psychologically meaningful way due to the lack of theoretical backing. For example, one custom index that has been used is the ratio between positive and negative stimuli that are recognized (i.e., the ‘Positive Ratio’; Djikic, Chan, & Peterson, 2007; Djikic, Peterson, & Zelazo, 2005). The use of custom indices such as the ‘Positive Ratio’ has the same limitation as using raw recognition rates. A greater ‘Positive Ratio’ could be occurring due to stimulus properties of positive and negative words rather than motivation to reach a directional conclusion. Failing to use standardized memory indices has made it difficult to understand underlying processes such as the role of biased searching through memories in motivated social evaluation.

ARE DIFFERENCES IN MEMORY FOR POSITIVE AND NEGATIVE SELF-RELEVANT INFORMATION DUE TO DIRECTIONAL CONCLUSIONS?

Research has typically conflated asymmetries in self-relevant memory with the desire to reach a specific directional conclusion (e.g., self-enhancement accounts). Past research on valence asymmetries in self-relevant memory has not considered that there may be other processes that can lead to the difference between positive and negative memory besides a motivation toward positive self-evaluations. It may be the case that there are nonmotivated explanations for phenomena that are typically assumed to be motivated (e.g., Chambers & Windschitl, 2004). As one example of how nonmotivated processes may lead to the same effects as motivated processes, research has shown that people find it easier to process positive compared to negative stimuli and therefore positive stimuli are easier to remember or feel more familiar (e.g., fluency accounts; Alves, Koch, & Unkelbach, 2017). In fact, words that have greater perceptual fluency are more frequently judged as old in a recognition task (Johnston, Hawley, & Elliott, 1991). So, positive words might be remembered at greater rates simply due to more fluent processing of positive compared to negative words. If nonmotivated processes, such as valence differences in fluency, could account for self-relevant memory differences then it would suggest that this robust effect of self-relevant memory asymmetries may not actually be an instance of motivated social evaluation (i.e., must entail a desired directional conclusion). Past research conflating asymmetries in reported memory with motivation to reach a particular conclusion has made it unclear if we have really looked at instances of motivated social evaluation or if these are instances of nonmotivated processes (e.g., inherent properties of the stimuli).

DOES BIASED SEARCHING THROUGH BELIEFS OPERATE SIMILARLY WHEN ONE HAS A DESIRE TO SEE A SOCIAL TARGET IN ANY DESIRED LIGHT (E.G., UNFLATTERING) AND IN A FLATTERING LIGHT?

Past research examining biased memory searching has focused on the motivation to see a social target in a flattering light. For example, experiments involving motivation and memory typically look at how people remember more positive than negative information about the self (Sedikides & Green, 2009) or what qualities about the self people recall after experimentally manipulating which trait is desirable (Sanitioso, Kunda, & Fong, 1990). So, past research has shown that people exhibit differences in reported memory, but only in situations where people are likely aiming to make flattering social evaluations. To more deeply understand the role of biased memory searches in motivated social evaluation it is important to expand measures of biased searching through memories to cases where people are not just aiming to paint someone in a flattering light. Can the same mechanism of biased memory searching operate similarly for both flattering and unflattering desired directional conclusions?

Study Overview and Hypotheses

Identifying limitations of past research reveals questions and avenues for a deeper understanding of the role of biased searching through memories in social evaluation. The proposed research addresses limitations in two overarching aims: (RQ1) to understand the role of a) biased memory searches and b) desired directional conclusions in asymmetric reported memory for self-relevant information (Studies 1a, 1b, 2a, 2b, and 3); and (RQ2) to understand if similar biased searches through memory may operate when people are motivated to paint someone in a particular light (given any directional conclusion) or are specific to instances where the goal is to paint someone in a flattering light (Studies 4). The current research aims to answer these questions by employing experimental, neurophysiological, and computational modeling methods.

Studies 1a & 1b. Can incentives offered after encoding reduce asymmetries in standards for claiming recognition (i.e., biased memory searching) across valence? These studies utilized an incentive paradigm to understand if valence asymmetries in memory are due to a strong drive toward a positive conclusion that is relatively resistant to competing incentives (as might be predicted from a self-enhancement perspective), or, alternatively, if valence asymmetries in memory are relatively responsive to alternate incentives (as could be predicted based on the logic of nonmotivated accounts such as the fluency account). If there is a strong underlying motivation, such as self-enhancement, then psychological or financial incentives would likely be unable to shift such a motivation. However, if there is no underlying motivation, then there is nothing to prevent incentives from reducing valence asymmetries in standards for claiming recognition. In study 1a participants received feedback ostensibly based on a personality questionnaire they completed (paradigm based on Djikic et al., 2005). Participants were offered psychological and financial incentives to recognize negative feedback they previously might claim to forget half way through a surprise recognition task. Differential standards for recognition between positive and negative were used to test for the role of biased memory searches (i.e., location (c) of Signal Detection Theory; Paulhus, Bruce, Harms, & Lysy, 2003). Further, if there is an effect of incentives on standards of recognition, then this would be more consistent with predictions based on nonmotivated accounts than the self-enhancement account suggesting valence asymmetries in memory for self-relevant information may not actually be motivated. However, if there is no effect of incentives on thresholds, then that would be more consistent with a self-enhancement account. Study 1b was a direct replication of Study 1a.

Study 2a & 2b. Can incentives offered before encoding reduce asymmetries in biased searches through memory (as measured by standards for claiming recognition) across valence? Study 2a is a replication and extension of Studies 1a and 1b. Study 1 found no effect of incentives on thresholds, which may reflect support for the self-enhancement account. Although Study 1 found support for the self-enhancement perspective, it is possible that memory asymmetries for self-relevant information begin at the time of encoding. As a more stringent test of how self-enhancement motivations may affect valence asymmetries in self-relevant memory, Study 2 aimed to test the role of self-enhancement motivations at the time of encoding. Differential standards for recognition between positive and negative were used to test for the role of biased memory searches. Further, if the self-enhancement account is again supported, then only higher levels of incentives (e.g., financially incentivized to recognize feedback about an other) could affect the asymmetries between positive and negative feedback because self-enhancement motivations are so strong that high levels of incentives would be needed. Study 2b was a direct replication of Study 2a.

Study 3. Is the conservative threshold for negative feedback at retrieval associated with concealed knowledge? Study 3 utilized Event Related Potentials (ERP) to determine if self-reports of lower recognition of negative feedback are consistent with neural signatures of concealed knowledge. Results from Studies 1 and 2 suggest that incentives only had an impact when offered before encoding and only in the most highly incentivized condition. Is it possible that part of the selective searching of negative feedback is suppression of memories for negative feedback at the time of retrieval? To understand how the negative feedback is uniquely processed, participants again saw bogus feedback (i.e., encoding phase) and then completed a surprise recognition task (i.e., retrieval phase). ERP data were collected during both the encoding and retrieval phases in order to assess if the 'forgotten' negative feedback was encoded and suppressed or not encoded in the first place. If there are differences between 'forgotten' negative, self-relevant feedback and correctly identified new information during the recognition task, then that would suggest suppression of negative feedback at the time of retrieval. This may suggest one possible mechanism for valence differences in standards for claiming recognition. Alternatively, if there are neurophysiological differences between forgotten and remembered negative, self-relevant feedback at encoding and retrieval, then that would suggest that there are multiple ways in which biased memory processes support self-enhancement motivations.

Study 4. Does a similar biased search through memory operate when aiming to paint someone in an unflattering light as well as a flattering light? The first three studies aimed to understand the role of biased memory searches in motivated social evaluation given a flattering desired directional conclusion. Building on those findings, Study 4 aimed to understand if biased searches through memory operate similarly given an unflattering directional conclusion. It is difficult to generalize about psychological processes when studying participants who have negative directional conclusions about the self (e.g., those prone to depression), therefore, Study 4 utilizes a paradigm in which the same person is likely to rate other social targets in a flattering and an unflattering light. Specifically, Democratic participants rate Republican and Democratic politicians on positive and negative traits. This 2x2 design allows for focusing on situations in which people want to make unflattering as well as flattering attributions as participants have the opportunity to rate disliked social targets negatively. There is abundant evidence that people more favorably rate in-group members over out-group members (Bartels, 2002; Brandt et al., 2014; Duarte et al., 2015; Munro et al., 2002; Tajfel & Turner, 2004) and specifically this differential rating occurs in the political domain (Pew Research Center, 2016). Drift-diffusion modeling (DDM; Ratcliff, 1978) was utilized to understand if belief searching functions similarly when the directional conclusion is unflattering as when it is flattering. If biased searching through memory functions the same given unflattering directional conclusions then there should be differential rates of evidence accumulation (i.e., drift rates from DDM) between positive and negative within each political affiliation (i.e., when given opportunities to affirm positive vs. negative traits of Democratic politicians and negative vs. positive traits of Republican politicians). Alternatively, if biased memory searching is more associated with a desire to make flattering evaluations then there should be differential rates of evidence accumulation when only when rating in-group politicians (i.e., only when given opportunities to affirm positive vs. negative traits of Democratic politicians).

STUDY 1A & 1B: CAN INCENTIVES OFFERED AFTER ENCODING REDUCE ASYMMETRIES IN STANDARDS FOR CLAIMING RECOGNITION (I.E., BIASED MEMORY SEARCHING) ACROSS VALENCE?

Study 1 tested whether biased searching through memories supports valence asymmetries in memory for self-relevant information. Extant research has conflated valence asymmetries in memory with motivation to reach a desired directional conclusion. To address this limitation of extant research, Study 1 further examined if valence asymmetries in memory are the result of motivation to reach a desired directional conclusion by testing whether incentives other than self-flattery affect biased memory searches when presented after encoding. Participants were presented with bogus feedback about their personality traits and then given a surprise recognition test. Half way through the surprise recognition test, participants in some conditions were offered incentives (i.e., either psychological incentives only or a combination of psychological and financial incentives) for accurate recognition of both positive and negative feedback. Location (c) from Signal Detection Theory (SDT) is used to operationalize the underlying memory mechanism of biased memory searching (Green & Swets, 1966). If location (c) distinguishes positive from negative, then this suggests that biased searching through memories may contribute to this form of motivated social evaluation because it represents differential standards for claiming recognition depending on valence. Further, introducing incentives allows for testing if desired directional goals account for the asymmetries in memory. If incentives result in decreased asymmetries in reported memory (i.e., reported memory for positive and negative become more similar), then this would be consistent with non-motivated accounts (e.g., fluency). However, if the asymmetries in reported memory are resistant to competing incentives, then this suggests a self-enhancement account. Incentives failing to shift asymmetric memory supports the self-enhancement account because self-enhancement is thought to be a strong, automatic motivation akin to motivations such as hunger drives. If this motivation is strong and automatic, then incentives should not be sufficient to diminish the self-enhancement drive and the asymmetry in memory for positive and negative self-relevant information should persist. Alternatively, incentives shifting asymmetric memory suggests that there is no underlying competing motivation (i.e., self-enhancement) that is resilient to a new, competing motivation such as financial gain.

Study 1 Method

Participants

Analysis of Study 1a focused on 187 participants (127 females, M_{age} = 19.07 years, *SD* = 1.16)[1]. Eight additional participants were excluded due to subject error (seven responded on less than 80% of recognition test trials and one expressed confusion about the task). Analysis of Study 1b focused on 180 participants (125 females, M_{age} = 19.10 years, *SD* = 1.28). Thirteen additional participants were excluded due to subject error (responded on less than 80% of recognition test trials). Trials that fell below two standard deviations below the mean reaction time for each experimental condition were excluded. These exclusion criteria were determined in advance for all studies to ensure analyses were based on meaningful trials and participants who were engaged in the task. Participants received course credit for their participation. All participants gave informed

[1] Sample size was determined using G*Power and based on a power level of 0.90. Effect size was determined using a small effect size (η2 = 0.15) of interactions within subjects. Studies of affirmation and financial incentives yield large effects, so the main effect of valence should be qualified by part if the retrieval bias were true (Cohen, Aronson, & Steele, 2000; Sherman, Nelson, & Steele, 2000). The recommended total sample size is 144.

consent in compliance with the human subject regulations of the University of Texas at Austin.

PROCEDURE

Study 1 modified a bogus personality feedback procedure used in previous research investigating memory for self-relevant feedback (Djikic, Peterson, & Zelazo, 2005). Participants completed a personality assessment task, received bogus feedback about their personality, and then completed a surprise recognition test for the feedback (see Figure 3). In order to manipulate the extent to which memory for negative feedback was rewarding or non-threatening, participants were randomly assigned to one of three conditions for the second part of the surprise recognition test. The task was presented using E-Prime 2 (Psychology Software Tools, INC., Sharpsburg, PA).

Personality 'Assessment'

At the beginning of the experiment, all participants completed a set of personality assessments: the Narcissistic Personality Inventory (Raskin & Terry, 1988), the Rosenberg Self-Esteem Scale (Rosenberg, 1965), the Big Five Inventory (John, Naumann, & Soto, 2008). In addition, participants completed two subjective tasks to increase the believability of subsequent feedback. First, participants were asked to pick one of four emotion words that best described a still picture (i.e., a modified form of the thematic apperception task). Second, participants completed a word association task in which they were asked to classify a given word as negative, neutral, or positive (e.g., nature).

Bogus Personality Feedback

Participants then received feedback ostensibly calculated from their responses in the personality assessment. However, the feedback was not calculated from their responses and the content was the same for all participants. Specifically, all participants were presented with 80 positive and 80 negative traits (Anderson, 1968). In order to ensure that any recognition differences were not due to differences in familiarity, positive and negative traits were matched for meaningfulness (i.e., how well participants felt they understood the meaning of the word: Positive: *M* = 3.56 *SD* = 0.18; Negative: *M* = 3.56, *SD* = 0.20; *t*(159) = -0.34, *p* = 0.73, *d* = 0.027). Participants first saw a screen which said ‘You are’ (1000 ms). The 'You are' stem was then randomly completed with one of the 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

Finally, participants were randomly assigned to one of three conditions where they completed a surprise recognition test of the feedback they received. The surprise recognition test included all 320 trait words: 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 completely new word. Their responses were collected during the trait word presentation and the following fixation screen (2500 ms total to respond for each trait).

For all three recognition conditions, participants performed the first half of the surprise recognition test (40 positive old, 40 positive new, 40 negative old, and 40 negative new). The random assignment affected the instructions that participants received after completing the first half of the recognition test. After the first half of the test was complete, 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, their memory performance was still important and they should 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 Financial Incentive condition: participants were also told that the feedback was actually meant for someone else and further instructed that they would receive a cash bonus for correct identification of feedback as being old or new. Specifically, participants were instructed that they would receive a bonus of up to $10 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 provide a break so they wouldn't feel fatigued for the last portion of the experiment.

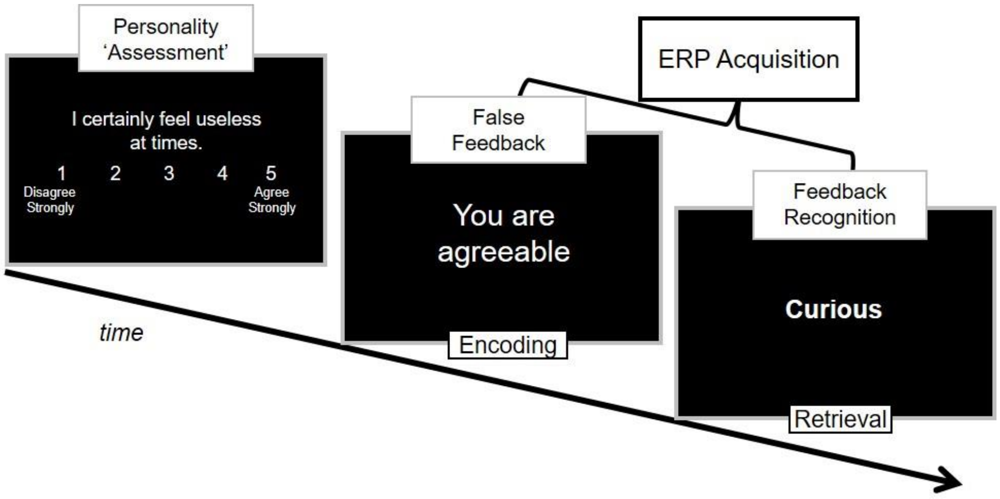


Figure 1: Study design. Participants first completed a series of personality questions. They then received bogus personality feedback which was 50% positive and 50% negative. Finally, participants completed a surprise recognition test for the feedback. In Study 1, participants completed the recognition task in two parts to manipulate the extent to which recognizing negative feedback was rewarding or non-threatening. No EEG data was collected. In Study 2, the motivational manipulation occurred prior to encoding and participants completed the recognition task uninterrupted, as depicted. In Study 3, participants completed the recognition task uninterrupted and EEG data was acquired while they received feedback and while they completed the surprise recognition task.

Behavioral Analysis

Memory for feedback was analyzed in two ways: Proportion recognized and Signal Detection Theory (SDT). Proportion recognized was calculated by dividing remembered words by the number of words that could have been recognized within a given condition (Green, Sedikides, & Gregg, 2008; Pinter, Green, Sedikides, & Gregg, 2011). Proportion recognized gives a raw rate of memory. SDT was used to calculate thresholds for recognition of feedback (i.e., criterion location (c): Paulhus, Bruce, Harms, & Lysy, 2003) and accuracy (i.e., d': Green, Sedikides, & Gregg, 2008). Standardized memory indices such as Signal Detection Theory allows for operationalizing underlying memory mechanisms rather than making assumptions about the role of memory in motivated social evaluation. 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:

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

Higher numbers reflect a more liberal threshold which indicates that lower levels of internal familiarity are needed before claiming recognition. Location (c) yields a measure of how much participants were willing to claim recognition based on very little feeling of familiarity. Further, d' indicates the ability to discriminate old stimuli from new stimuli. It is also calculated by considering hits and false alarms:

$$D' = \text{Hits} - \text{False Alarms}$$

Higher numbers reflect greater accuracy, which indicates a greater ability to distinguish old from new stimuli. Data were then analyzed using a 3-way ANOVA with two within-subject factors (Valence: Positive and Negative; Time: Part 1 and Part 2) and one between-subjects factor (Manipulation Condition: Non Self-Relevant Feedback, Non Self-Relevant Feedback + Financial Incentive, and Control) for each of these three memory measures. All three measures are reported and discussed, but special emphasis is placed on location (c) as it relates to the central research questions about the role of biased searching through memories in motivated social evaluation.

Pooling Data

Study 1 included two samples (1a and 1b) which 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). As the two samples were identical in procedure and small effect sizes were expected, IDA offers several benefits. Study was included as a factor to ensure that there were no significant differences between the two samples and whether the results replicated in each independent sample is discussed.

STUDY 1 RESULTS

More liberal thresholds for claiming recognition of positive as compared to negative trait feedback across incentive conditions and part of recognition task. Location (c) indicated different standards for claiming recognition of negative, self-relevant feedback. Participants had a more liberal threshold for remembering positive feedback compared to negative feedback (Main effect of Valence, $F(1,361) = 101.74$, $p < 0.001$, $\eta_p^2 = 0.22$) which persisted even after financial incentives were presented and/or threat was removed (interaction between Valence, Time, and Recognition condition, $F(2,361) = 0.214$, $p = 0.807$, $\eta_p^2 = 0.001$; see Figure 2). There was a main effect of Time ($F(1,361) = 190.05$, $p < 0.001$, $\eta_p^2 = 0.345$), but there was no main effect of Recognition condition ($F(1,361) = 1.22$, $p = 0.298$, $\eta_p^2 = 0.007$). These results were not affected by data sample (interaction between Valence, Time, Recognition condition, and Study, $F(2,361) = 1.571$, $p = 0.209$, $\eta_p^2 = 0.009$). The persistence of more conservative standards for claiming recognition of negative feedback in the face of financial incentive plus threat to self-esteem is eliminated or when only threat to self-esteem is eliminated is consistent with the self-enhancement hypothesis.

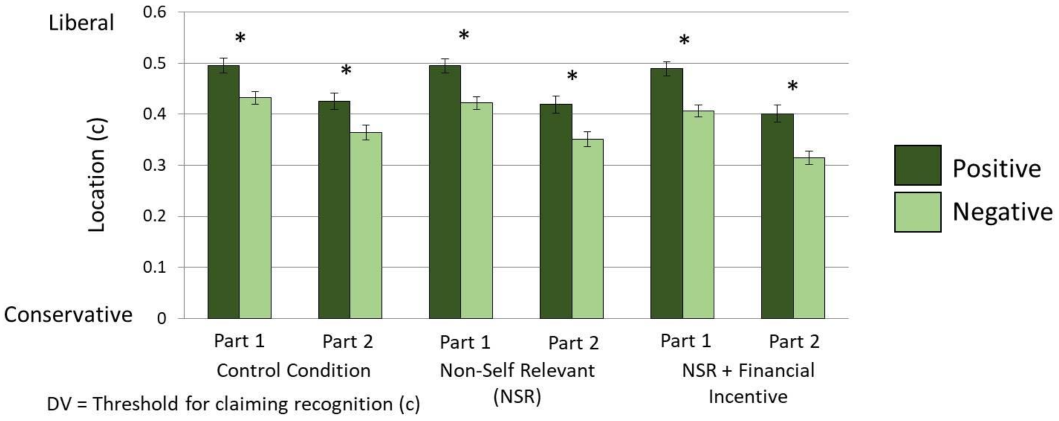


Figure 2: Study 1 Results: location c. Participants used more liberal thresholds to claim recognition of positive feedback as compared to negative feedback regardless of condition (Non-Self Relevant, Financial Incentive, or Control). Providing psychological or financial incentives for memory did not result in a significant shift of threshold for claiming negative feedback.

This pattern of results replicated within each individual sample. Participants in Study 1a had a more liberal thresholds for remembering positive feedback compared to negative feedback (Main effect of Valence, $F(1,184) = 41.33$, $p < 0.001$, $\eta_p^2 = 0.183$) which persisted even after financial incentives were presented and/or threat was removed (interaction between Valence, Time, and Recognition condition, $F(2,184) = 0.77$, $p = 0.462$, $\eta_p^2 = 0.008$). Participants in Study 1b had a more liberal threshold for remembering positive feedback compared to negative feedback (Main effect of Valence, $F(1,177) = 64.40$, $p < 0.001$, $\eta_p^2 = 0.267$) which persisted even after financial incentives were presented and/or threat was removed (interaction between Valence, Time, and Recognition condition, $F(2,177) = 0.985$, $p = 0.376$, $\eta_p^2 = 0.01$).

Greater accuracy for negative as compared to positive trait feedback across incentive conditions and part of recognition task. D' indicated greater accuracy for negative trait words than for positive trait words (Main effect of Valence, $F(1,361) = 135.92$, $p < 0.001$, $\eta_p^2 = 0.274$) which persisted even after financial incentives were presented and/or threat was removed (interaction between Valence, Time, and Recognition condition, $F(2,361) = 0.11$, $p = 0.895$, $\eta_p^2 = 0.001$; see Figure 3). There was a main effect of Time ($F(1,361) = 99.12$, $p < 0.001$, $\eta_p^2 = 0.215$), but there was no main effect of Recognition condition ($F(1,361) = 0.73$, $p = 0.483$, $\eta_p^2 = 0.004$). These results were not affected by data sample (interaction between Valence, Time, Recognition condition, and Study, $F(2,361) = 1.79$, $p = 0.168$, $\eta_p^2 = 0.01$).

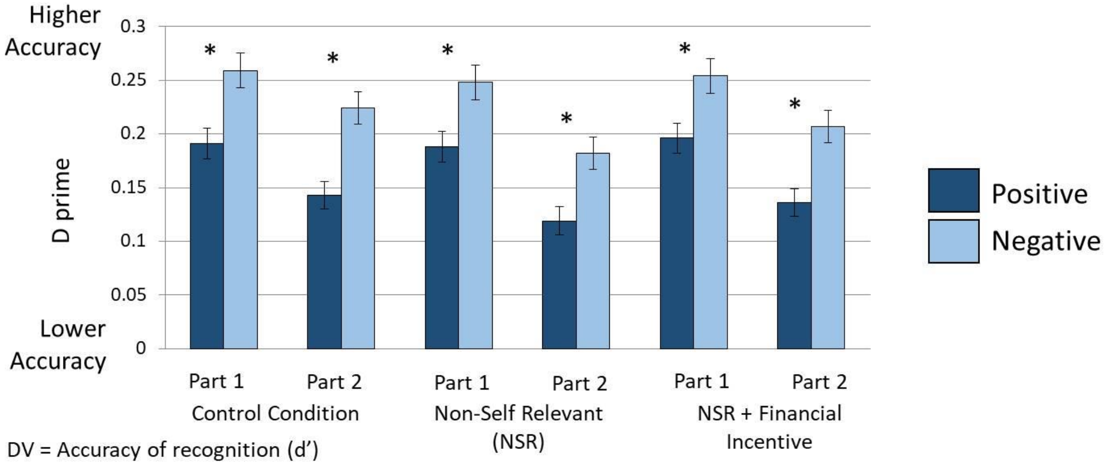


Figure 3: Study 1 Results: d prime. Participants had higher accuracy for negative as compared to positive traits across incentive condition and parts of the recognition task.

This pattern of results was replicated within each individual sample. Participants in Study 1a had greater accuracy for negative feedback compared to positive feedback (Main effect of Valence, $F(1,184) = 42.60$, $p < 0.001$, $\eta_p^2 = 0.188$) which persisted even after financial incentives were presented and/or threat was removed (interaction between Valence, Time, and Recognition condition, $F(2,184) = 0.50$, $p = 0.605$, $\eta_p^2 = 0.005$). Participants in Study 1b had greater accuracy for negative feedback compared to positive feedback (Main effect of Valence, $F(1,177) = 102.33$, $p < 0.001$, $\eta_p^2 = 0.366$) which persisted even after financial incentives were presented and/or threat was removed (interaction between Valence, Time, and Recognition condition, $F(2,177) = 1.41$, $p = 0.248$, $\eta_p^2 = 0.016$).

Greater recognition of positive as compared to negative trait feedback. An analysis of how many words were remembered found similar results to the analyses with location (c). Participants remembered more positive compared to negative traits from their personality feedback (Main effect of Valence, $F(1,361) = 23.48$, $p < 0.001$, $\eta_p^2 = 0.061$; see Figure 4) which persisted even after financial incentives were presented and/or threat was removed (interaction between Valence, Time, and Recognition condition, $F(2,361) = 0.035$, $p = 0.965$, $\eta_p^2 = 0$). There was a main effect of Time ($F(1,361) = 256.07$, $p < 0.001$, $\eta_p^2 = 0.415$), but there was no main effect of Recognition condition ($F(1,361) = 1.01$, $p = 0.364$, $\eta_p^2 = 0.006$). These results were not significantly affected by data sample (interaction between Valence, Time, Recognition condition, and Study, $F(2,361) = 2.24$, $p = 0.108$, $\eta_p^2 = 0.012$).

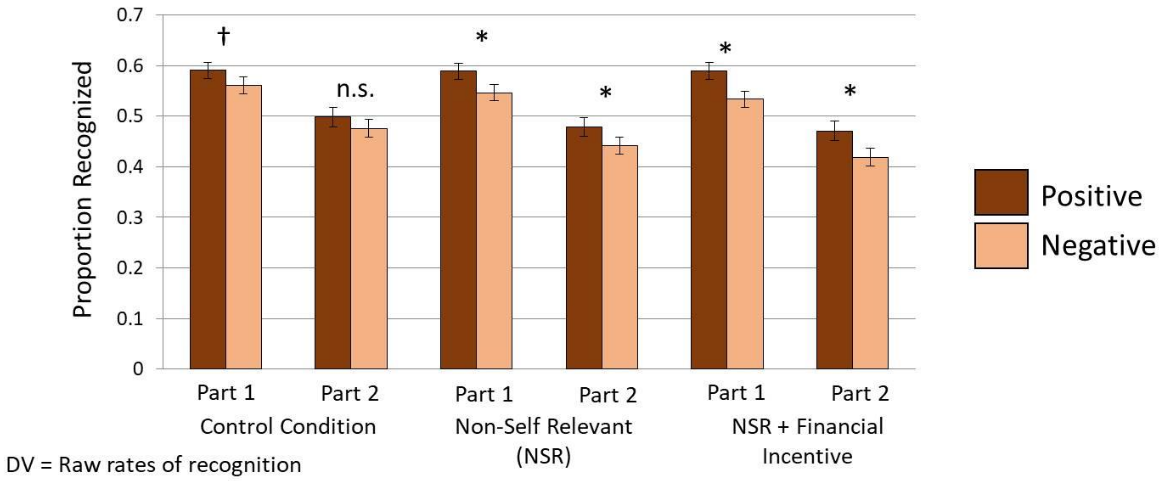


Figure 4: Study 1 Results: Recognition Rates. Participants recognized more positive traits than negative traits across incentive conditions and across parts of the recognition task. † = 0.07

This pattern of results replicated within each individual sample. Participants in Study 1a remembered more positive compared to negative traits (Main effect of Valence, $F(1,184) = 11.76$, $p = 0.001$, $\eta_p^2 = 0.060$) which persisted even after financial incentives were presented and/or threat was removed (interaction between Valence, Time, and Recognition condition, $F(2,184) = 1.05$, $p = 0.351$, $\eta_p^2 = 0.011$). Participants in Study 1b remembered more positive compared to negative traits (Main effect of Valence, $F(1,177) = 12.10$, $p = 0.001$, $\eta_p^2 = 0.064$) which persisted even after financial incentives were presented and/or threat was removed (interaction between Valence, Time, and Recognition condition, $F(2,177) = 1.22$, $p = 0.298$, $\eta_p^2 = 0.014$).

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 (i.e., location (c)). 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. 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 do not support the concern that some of the conditions yielded valence effects that happen to be statistically significant from a null hypothesis testing approach yet are small enough to overlap with non-meaningful differences.

STUDY 1 DISCUSSION

Study 1 findings are more consistent with a self-enhancement account in that asymmetries in memory are not easily responsive to alternate incentives. Further, the main effect of valence on location (c) suggests that people have different standards for claiming recognition of positive compared to negative feedback (i.e., one form of biased searching through memories). Study 1 findings suggest that asymmetries in proportion from previous research may have arisen from more liberal thresholds and lower accuracy for positive compared to negative information. The incentives provided after encoding took place failed to shift this memory distortion, which lends support to a self-enhancement account. It seems that reported memory asymmetries are likely arising from a motivational drive to claim recognition of positive feedback rather than inherent differences in positive and negative stimuli as suggested by the logic of nonmotivated accounts such as the fluency account. However, the feedback was offered after encoding took place. It could be that disruptions in memory occur at the time of encoding making any incentives unable to shift memory and thresholds for negative feedback. Would incentives influence differences in location (c) for positive and negative if presented before encoding?

STUDY 2A & 2B: CAN INCENTIVES OFFERED BEFORE ENCODING REDUCE ASYMMETRIES IN STANDARDS FOR CLAIMING RECOGNITION (I.E., BIASED MEMORY SEARCHING) ACROSS VALENCE?

While the results of Study 1 provide support for the self-enhancement account and the role of biased memory searching, it is unclear if disruptions at the time of encoding could lead to differences in recognition thresholds at the time of retrieval. Study 2 builds on Study 1 by introducing incentives prior to encoding rather than after. Participants were presented with bogus feedback about their personality traits or a peer's personality traits and then given a known or a surprise recognition test. Importantly, participants in the incentive conditions were given information about incentives (again, either psychological or psychological and financial) prior to receiving any feedback (i.e., prior to the encoding phase). This creates four distinct incentive condition levels. The first condition is similar to the control condition in Study 1 as the feedback is about the self and there is no financial reward. From there participants received increasing levels of incentives. The second condition offered financial incentives, but the feedback was about the self. The third condition offered no financial incentives, but the feedback was about an other which makes it less threatening to the self. Finally, the fourth condition offered financial incentives and the feedback was about an other. If biased memory searching does play a role in motivated social evaluation then location (c) should again distinguish positive from negative. Further, if self-enhancement accounts rather than fluency accounts are supported then incentives should again have no effect on self-relevant feedback, but may impact memory indices in the highest incentive condition (i.e., non self-relevant and financially rewarded).

Study 2 Method

Participants

Analysis of Study 2a focused on 254 participants (183 females, M_{age} = 18.70 years, *SD* = 1.39). 21 additional participants were excluded due to subject error (9 responded on fewer than 80% of recognition test trials, 10 asked not to use data, 2 answered manipulation check questions incorrectly). Analysis of Study 2b focused on 332 participants (233 females, M_{age} = 19.31 years, *SD* = 2.87). 32 additional participants were excluded due to subject error (11 responded on fewer than 80% of recognition test trials, 7 asked not to use data, 14 answered manipulation check questions incorrectly). Trials were excluded which fell below two standard deviations below the mean reaction time for each experimental condition. These exclusion criteria were determined in advance for all studies to ensure analyses were based on participants who were engaged in the task. Participants received course credit for their participation. All participants gave informed consent in compliance with the human subject regulations of the University of Texas at Austin.

Procedure

The procedure for Study 2 was almost identical to the procedure for Studies 1a and 1b. The only exceptions being (1) that half of the participants were informed of the memory test and offered financial incentives for their memory performance prior to receiving personality feedback (2) half of each incentive group saw feedback about a peer instead of the self and (3) there was no break during the recognition task. Participants were given one of four different instructions. In the Self, Surprise Memory Test condition participants completed the personality assessment about themselves and were not offered bonus money nor warned of the subsequent memory test. This manipulation served as the control condition and is similar to the control condition in the previous studies. In the Self, Warned and Incentivized condition, participants completed the personality assessment about themselves, but were additionally offered bonus money and warned of the subsequent memory test. Specifically, participants were instructed that they would receive a bonus of up to $10 based on two randomly selected trials from an upcoming recognition test. This manipulation added a financial incentive to encode memories of negative feedback. In the Other, Surprise Memory Test condition, participants completed the personality assessment about another student (displayed picture was gender matched) and were not offered bonus money nor warned of the subsequent memory test. This manipulation served to motivationally incentivize the encoding of negative feedback as it was not self-relevant. In the Other, Warned and Incentivized condition, participants completed the personality assessment about a peer, but were additionally offered bonus money and warned of the subsequent memory test. Specifically, participants were instructed that they would receive a bonus of up to $10 based on two randomly selected trials from an upcoming recognition test. This manipulation added a financial and motivational incentive to encode memories of negative feedback.

Behavioral Analysis

As in Studies 1a and 1b, Study 2 examined location (c) and d' from Signal Detection Theory (SDT) as well as raw recognition rates. Further, as in Study 1, data were collapsed across Studies 2a and 2b to maximize the power to detect small effects and provide the most accurate estimate of effect sizes (Curran & Hussong, 2009).

Study 2 Results

More liberal thresholds for claiming recognition of positive as compared to negative trait feedback except given the greatest level of incentives. Location (c) indicated more conservative thresholds for negative, self-relevant feedback in all conditions except the Other, Warned and Incentivized condition. Participants had a more liberal threshold for remembering positive feedback compared to negative feedback (Main effect of Valence, $F(1,578) = 43.07$, $p < 0.001$, $\eta_p^2 = 0.069$; see Figure 5). However, participants did not have a more liberal threshold for remembering positive feedback in all four conditions (interaction between Valence and Incentive condition, $F(3,578) = 2.99$, $p = 0.031$, $\eta_p^2 = 0.015$; see Figure 5). There was also a main effect of Incentive condition ($F(1,578) = 3.09$, $p = 0.027$, $\eta_p^2 = 0.016$). These results were not affected by data sample (interaction between Valence, Incentive condition, and Study, $F(3,578) = 0.686$, $p = 0.561$, $\eta_p^2 = 0.004$). The acceptance of negative feedback only when that feedback is no longer self-threatening and financially incentivized is consistent with the self-enhancement hypothesis. Further, the more liberal thresholds for positive as compared to negative further supports the role of biased memory searching in valence asymmetries for self-relevant feedback.

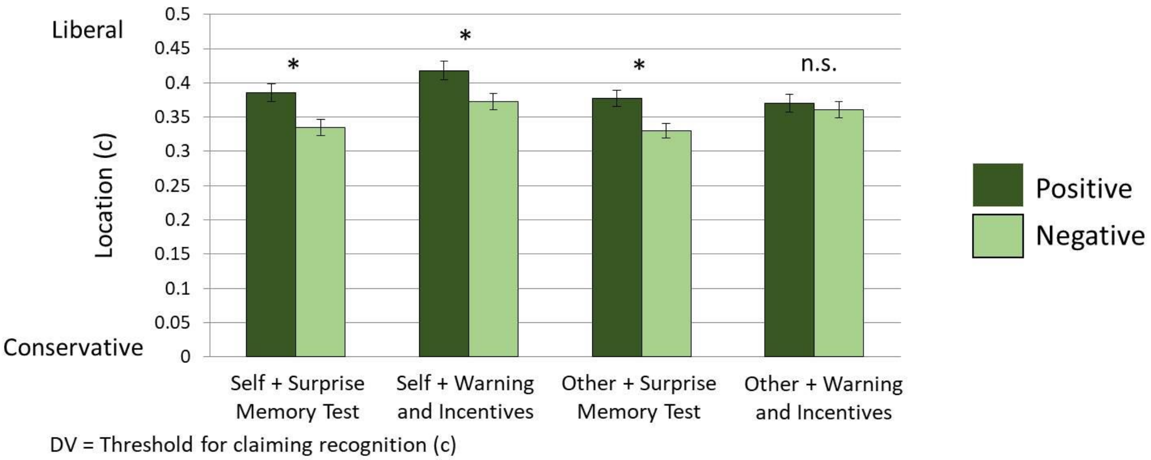


Figure 5: Study 2 Results: location c. Participants used more liberal thresholds to claim recognition of positive feedback as compared to negative feedback except when they had the most incentives (Other, Warned and Incentivized condition).

Post hoc t-tests between the Non Self-Relevant conditions revealed that the highest level of reward was related to a shift in location (c) within the negative condition, but not within the positive condition. There was no significant difference between Non Self-Relevant positive and Non Self-Relevant + Financial Reward positive ($t(307) = -0.37$, $p = 0.711$, $d = 0.044$), but there was a marginally significant difference between Non Self-Relevant negative and Non Self-Relevant + Financial Reward negative ($t(307) = 1.83$, $p = 0.068$, $d = 0.209$). The difference between the two negative conditions, but not the two positive conditions suggests that the effect of the increasing financial reward is associated with a shift in negative rather than positive.

This pattern of results replicated within each individual sample. Participants in Study 2a had a more liberal threshold for remembering positive feedback compared to negative feedback (Main effect of Valence, $F(1,250) = 16.29$, $p < 0.001$, $\eta_p^2 = 0.061$) except this difference showed a trend to be less pronounced in the Non Self-Relevant + Financial Reward condition (interaction between Valence and Incentive condition,

$F(3,250) = 1.921$, $p = 0.127$, $\eta_p^2 = 0.023$). Pairwise t-tests found that participants in the other three conditions showed significant differences in location (c) for positive and negative feedback (Self-Relevant: $t(58) = 2.85$, $p = 0.006$, $d = 0.372$; Self-Relevant + Financial Reward: $t(54) = 2.68$, $p = 0.01$, $d = 0.379$; Non Self-Relevant: $t(76) = 2.22$, $p = 0.029$, $d = 0.257$; Non Self-Relevant + Financial Reward: $t(62) = 0.08$, $p = 0.935$, $d = 0.014$). As in Study 2a, pairwise t-tests conducted on Study 2b data found that participants showed significant differences in location (c) for all conditions except the Non Self-Relevant + Financial Reward condition (Self-Relevant: $t(85) = 2.49$, $p = 0.015$, $d = 0.271$; Self-Relevant + Financial Reward: $t(76) = 3.26$, $p = 0.002$, $d = 0.378$; Non Self-Relevant: $t(89) = 3.61$, $p = 0.001$, $d = 0.310$; Non Self-Relevant + Financial Reward: $t(78) = 1.45$, $p = 0.15$, $d = 0.162$). An ANOVA found that participants had a more liberal threshold for remembering positive feedback compared to negative feedback (Main effect of Valence, $F(1,328) = 28.53$, $p < 0.001$, $\eta_p^2 = 0.080$; interaction between Valence and Incentive condition was not significant, $F(3,328) = 1.38$, $p = 0.249$, $\eta_p^2 = 0.012$).

Greater accuracy for negative as compared to positive trait feedback across incentive conditions. D’ indicated greater accuracy for negative trait words than for positive trait words (Main effect of Valence, $F(1,578) = 249.70$, $p < 0.001$, $\eta_p^2 = 0.301$). This pattern was not significantly affected by increasing levels of financial and psychological incentives (interaction between Valence and Incentive condition, $F(2,578) = 2.25$, $p = 0.082$, $\eta_p^2 = 0.011$; see Figure 6). There was also a main effect of Incentive condition ($F(1,578) = 6.85$, $p < 0.001$, $\eta_p^2 = 0.034$). These results were not affected by data sample (interaction between Valence, Incentive condition, and Study, $F(2,578) = 0.48$, $p = 0.7$, $\eta_p^2 = 0.002$).

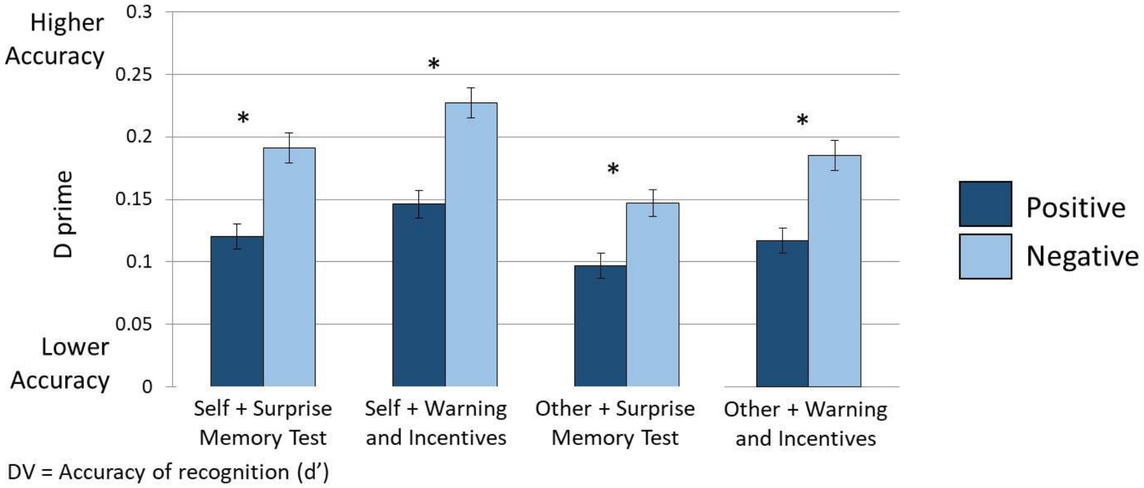


Figure 6: Study 2 Results: d prime. Participants had higher accuracy for negative as compared to positive traits across incentive condition and parts of the recognition task.

This pattern of results replicated within each individual sample. Participants in Study 2a had greater accuracy for negative trait words than for positive trait words (Main effect of Valence, $F(1,250) = 138.60$, $p < 0.001$, $\eta_p^2 = 0.358$). The interaction between Valence and Incentive condition was not significant ($F(3,250) = 0.95$, $p = 0.415$, $\eta_p^2 = 0.011$). Participants in Study 2b had greater accuracy for negative trait words than for positive trait words (Main effect of Valence, $F(1,328) = 124.21$, $p < 0.001$, $\eta_p^2 = 0.274$). The interaction between Valence and Incentive condition was not significant ($F(3,328) = 1.88$, $p = 0.134$, $\eta_p^2 = 0.017$).

No difference in recognition of positive and negative trait feedback. As in previous research, (Green et al., 2008), when proportion of words remembered was considered, participants did not remember significantly more positive as compared to negative traits from their personality feedback (Main effect of Valence, $F(1,578) = 1.24$, $p = 0.258$, $\eta_p^2 = 0.002$; see Figure 7). This pattern was not significantly affected by increasing levels of incentives (interaction between Valence and Incentive condition,

F(2,578) = 0.71, p = 0.55, η_p^2 = 0.004). However, there was a main effect of Incentive condition (F(1,578) = 2.61, p = 0.050, η_p^2 = 0.013). This pattern was not significantly affected by increasing levels of financial and psychological incentives (interaction between Valence and Incentive condition, F(3,578) = 0.711, p = 0.546, η_p^2 = 0.004).

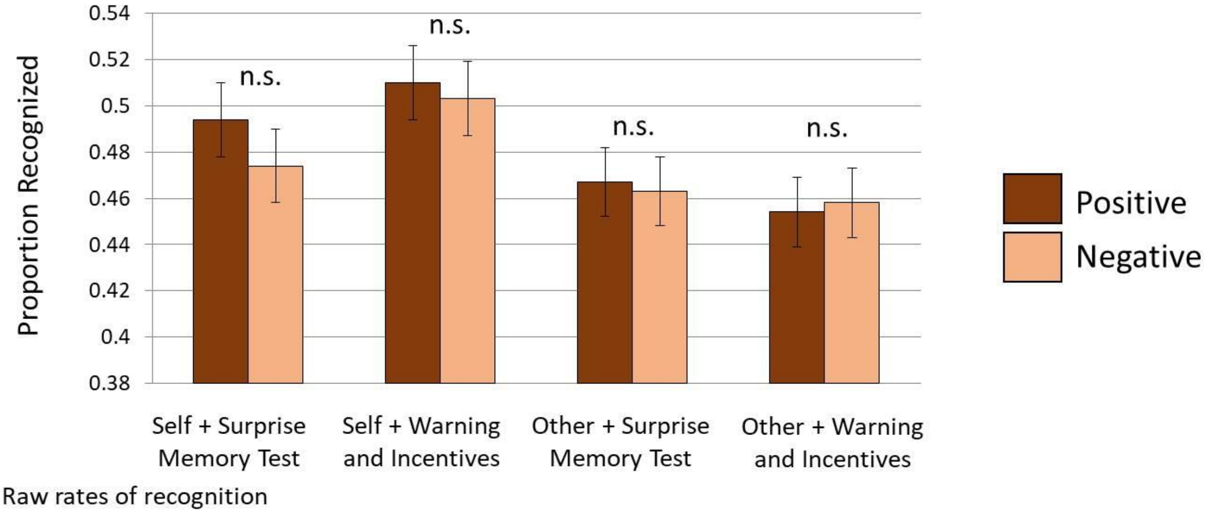


Figure 7: Study 2 Results: Recognition Rates. Participants did not recognize significantly more positive traits than negative traits across incentive conditions.

This pattern of results replicated within each individual sample. Participants in Study 2a did not remember significantly more positive compared to negative traits (Main effect of Valence, F(1,250) = 0.69, p = 0.41, η_p^2 = 0.003). The interaction between Valence and Incentive condition was not significant (F(3,250) = 1.68, p = 0.172, η_p^2 = 0.020). Participants in Study 2b did not remember significantly more positive compared to negative traits (Main effect of Valence, F(1,328) = 0.37, p = 0.545, η_p^2 = 0.001). The interaction between Valence and Incentive condition was not significant (F(3,328) = 0.91, p = 0.435, η_p^2 = 0.008).

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 (i.e., location (c)). 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.

Study 2 Discussion

Study 2 findings, like Study 1 findings, are consistent with a self-enhancement account and support the role of biased memory searching in motivated social evaluation. Asymmetries in memory searching are only responsive to the highest measured level of alternate incentives (i.e., non self-relevant and financially incentivized). This supports the self-enhancement account in that it shows how difficult it is to incentivize people to shift their thresholds for claiming negative feedback. Even when offered prior to encoding, threshold shifts did not occur except with the highest measured level of incentives. Further, as in Study 1, there is a significant main effect of valence when analyzing location (c) suggesting that motivated social evaluation is at least in part supported by biased searching through memories. Therefore, results from studies 1 & 2 suggest that people have more conservative thresholds for negative feedback and this is likely due to self-enhancement motivations. Is the conservative threshold for negative feedback at retrieval associated with concealed knowledge?

STUDY 3: ARE THE MORE CONSERVATIVE THRESHOLDS ASSOCIATED WITH REMEMBERING NEGATIVE FEEDBACK ASSOCIATED WITH CONCEALED KNOWLEDGE OF THAT FEEDBACK?

Study 3 builds on Studies 1 and 2 by examining the neural markers of self-relevant feedback at encoding and retrieval as a function of valence and memory. More specifically, what best characterizes the forgotten, negative self-relevant feedback? One 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: Hu et al., 2015). One alternative hypothesis is that self-reported memory reflects truly forgotten feedback: 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 (Paller, Kutas, & Mayes, 1987; Neville et al., 1986). Therefore, Study 3 examines two possible neurophysiological patterns to characterize the processing of negative, self-relevant feedback: (1) a pattern of difference associated with suppression (i.e., a significant difference between forgotten negative, self-relevant feedback compared to correctly identified new information, that is, correct rejections during a recognition task) and (2) 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). We draw on a permutation approach for analyzing ERPs (Nichols & Holmes, 2002; Trujillo, Allen, Schnyer, & Peterson, 2010; Sanguinetti, Trujillo, Schnyer, Allen, & Peterson, 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).

Study 3 Method

PARTICIPANTS

Analysis focused on 36 participants (28 females, M_{age} = 19.53 years, *SD* = 2.40)[2]. Three additional participants were excluded due to subject error (responded 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 1a, 1b, 2a, and 2b with a few exceptions. First, there was no manipulation presented halfway through the recognition task or prior to feedback receipt. Second, participants were presented with bogus feedback that consisted of 85 positive traits and 85 negative traits. The increase in trait feedback ensured there would be sufficient power (i.e., trials per condition) to conduct the planned ERP analyses. As in Study 1, positive and negative traits were

[2] Sample size was determined using G*Power and based on a power level of 0.90. Effect size was determined using a conservative effect size (d = 0.59) of the difference between memory for positive and negative feedback, which is consistent with effect sizes from previous research of this memory bias (d = 0.60 to 1.95; Zengel et al., 2016; Green & Sedikides, 2004; Green et al., 2008). The recommended total sample size is 33..

matched for meaningfulness (Positive: M = 3.55, SD = 0.18; Negative: M = 3.56, SD = 0.20; $t(169)$ = -0.52, p = 0.60).

Behavioral Analysis

As in Studies 1 and 2, Study 3 examined the location (c) from Signal Detection Theory.

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). Four additional electrodes were placed in and outside of the cap 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 EEGs were then epoched starting at 200 ms before to 2000 ms after the onset of the stimulus. 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 prestimulus period of -200 to 0ms.

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, Trujillo, Schnyer, Allen, & Peterson, 2016). This method of analysis is advantageous because it utilizes all of the recorded data 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.

Study 3 Results

Behavioral Results: Greater Recognition for Positive Self-Relevant Feedback Compared to Negative Self-Relevant Feedback. Consistent with the control conditions in Studies 1 and 2, participants had more liberal thresholds for claiming familiarity with positive feedback than negative feedback, $t(35) = 4.58$, $p < 0.001$, $d = 0.84$ (see Figure 8).

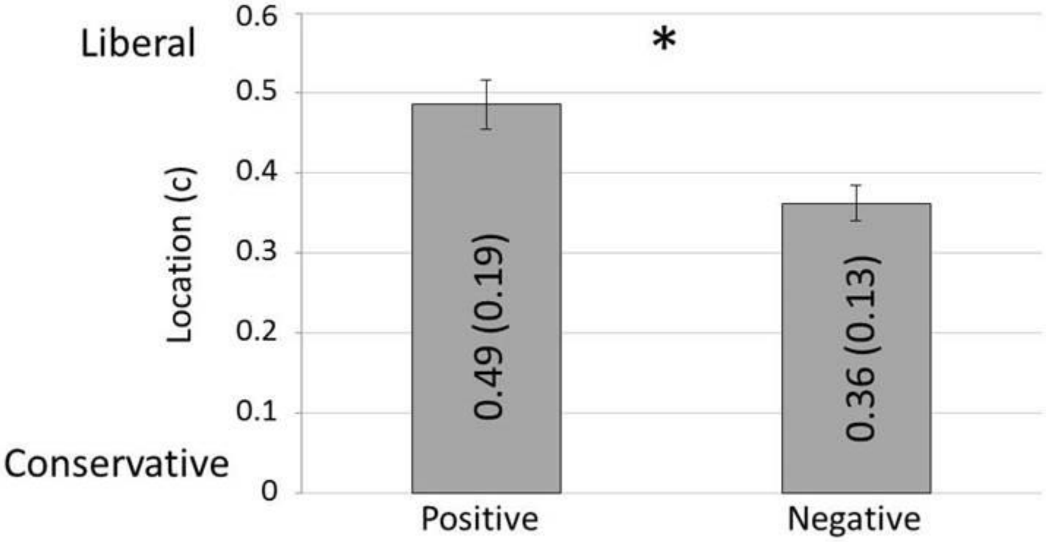
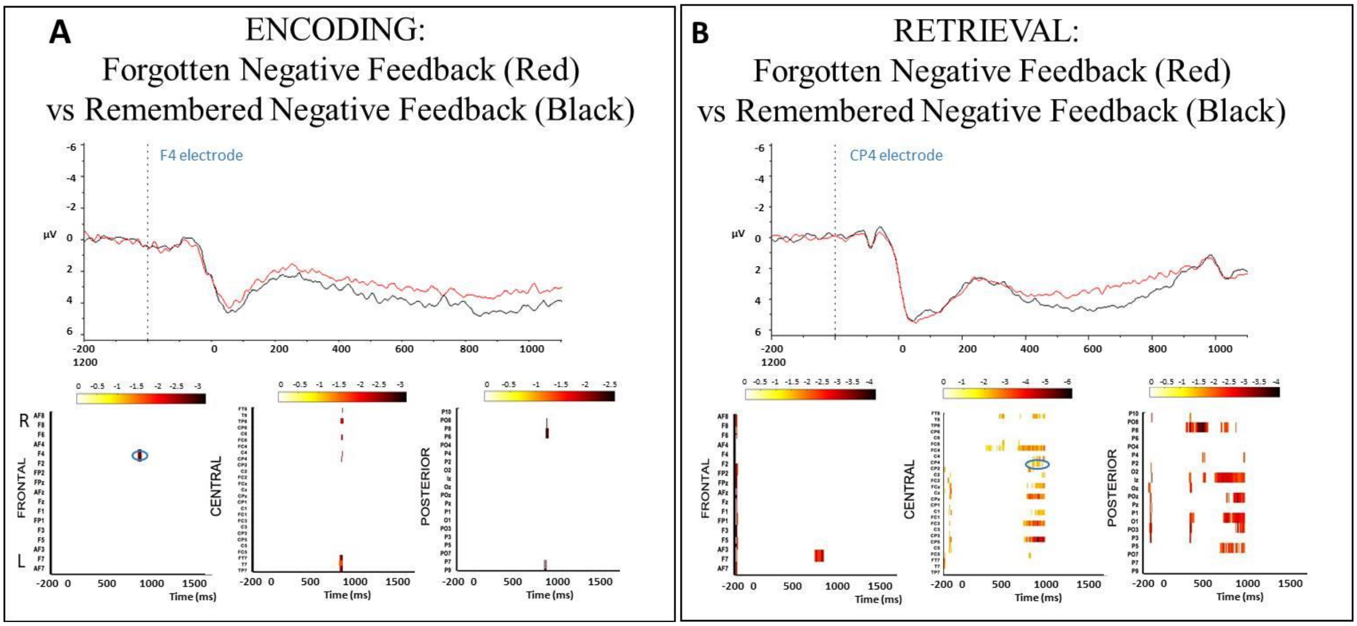


Figure 8: Study 3 Behavioral Results. As in Studies 1 and 2, participants had a more liberal threshold for claiming recognition of positive feedback as compared to negative feedback. Numbers indicate means and (standard deviations).

Self-Reported Memory is Associated with Meaningful Distinctions For Negative Self-Relevant Feedback at Encoding and Retrieval. The ERP analysis suggested that self-reports of memory were associated with differences at the neurophysiological level of analysis. There were significantly different ERPs 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, ERPs were 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 9A). During the retrieval phase of the task, ERPs associated with forgotten negative feedback were also smaller than responses for correctly remembered negative feedback (i.e., a cluster on the central scalp between 600-1000 ms after stimulus onset, see Figure 9B).

ERPs Associated with Negative Self-Feedback that is Self-Reported as Forgotten are not Significantly Distinct from ERPs Associated with Correctly Identified Novel Feedback. Previous research suggests that ERPs measured during retrieval can distinguish between suppressed information and novel information (Hu et al., 2015) yet 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.


A ENCODING: Forgotten Negative Feedback (Red) vs Remembered Negative Feedback (Black)

B RETRIEVAL: Forgotten Negative Feedback (Red) vs Remembered Negative Feedback (Black)

Figure 9: Study 3 ERP Results. ERP results were significantly able to distinguish self-reported memorydifferences at both encoding and retrieval. Bottom graphs indicate clusters where there are significant differences between the two conditions. Top graphs show representative waveforms of each condition at one electrode for visualization purposes (electrode site circled in blue on bottom graphs).

Study 3 Discussion

Study 3 results support the ‘forgotten memories’ hypothesis rather than the ‘concealed knowledge’ hypothesis. More specifically, there was a significant difference between neural patterns associated with remembered negative versus forgotten negative feedback at the time of encoding and retrieval and there was no difference between forgotten negative, self-relevant feedback compared to correctly identified new information (i.e., correct rejections) during the recognition task. These results suggest that negative, self-relevant feedback that is ‘forgotten’ is not encoded in the first place. This lower rate of encoding negative self-relevant feedback occurs in conjunction with biased searching through memories as measured by differential standards for claiming recognition of positive and negative feedback. Further, results from Studies 1, 2, & 3 all support a self-enhancement account. It is challenging to find incentives that encourage deviations from the self-flattering standards of recognition typically applied to self-relevant feedback (rather than differences arising from properties of the stimuli) and self-enhancement motivations are further supported by disruptions at the time of encoding for negative, self-relevant feedback.

STUDY 4: DOES BIASED SEARCHING THROUGH MEMORY OPERATE SIMILARLY WHEN AIMING TO PAINT SOMEONE IN AN UNFLATTERING LIGHT AS WELL AS IN A FLATTERING LIGHT?

Studies 1, 2, and 3 addressed the limitations in the extant research that were related to operationalizing the underlying mechanisms of biased searches through memory and conflation of valence asymmetries in memory with motivation toward a specific conclusion. However, there is still the limitation of only focusing on biased memory searches when the desired directional conclusion is flattering. Study 4 builds on previous research and Studies 1-3 by testing whether biased memory searching operates in cases where flattering and unflattering conclusions are desired by the same person about different social targets. Participants rated liked and disliked politicians on positive and negative traits, which created opportunities to evaluate liked others in a flattering light and disliked others in an unflattering light. Drift Diffusion Modeling (DDM), which allows for independent calculation of two parameters (i.e., starting point and drift rate), was used to understand if biased memory searching supports painting people in unflattering lights as well as flattering lights. Unlike in Studies 1-3, in which location (c) of SDT was used, Study 4 used drift rate of DDM as a measure of biased memory searches. If biased memory searching operates similarly given flattering and unflattering directional conclusions, then we would expect drift rates to differ as a function of political affiliation and rating positive and negative traits. In other words, positive and negative trait ratings should be associated with different drift rates depending on whether the evaluation is about an in-group target or an out-group target (i.e., when desired directional conclusions are likely unflattering). However, if biased memory searching supports motivated social evaluation only when aiming to paint a social target in a flattering light, then we would expect drift rates to differ for positive and negative trait ratings only within the in-group political party condition. Starting point is calculated, but only discussed to determine if this paradigm could actually be an instance of people exhibiting desired directional conclusions.

Study 4 Method

Participants

Analysis was conducted for 75 participants (50 females, 24 males, and 1 other; M_{age} = 18.59 years, *SD* = 0.89). Participants were analyzed based on political identity. Those who indicated Democratic affiliation were included in analyses. Those who indicated Republican affiliation (*N* = 25) or marked other (*N* = 26) were excluded. Analysis focused on self-identified Democratic participants because liberal participants were easier to recruit on a college campus, which ensured a large enough sample size and made comparisons between groups difficult (as there was a dramatically uneven number per group). Further, the focus on one political party reduced the chance for noise in the data because non-Democratic participants did not behave consistently in a pilot sample. All participants gave informed consent in compliance with the human subject regulations of the University of Texas at Austin.

PROCEDURE

Participants evaluated the personality traits of well-known politicians. In each trial, participants saw a picture of a politician, the politician's political party affiliation, and a prompt to if they believed each politician possessed a particular trait (Figure 10). Participants were given a two-alternative forced choice task (i.e., 'yes' or 'no') to rate all six politicians (three per political party: Republican and Democratic) on each of 60 trait words (30 per valence: positive and negative). Trait words were taken from a list of words standardized for valence (Anderson, 1968). Politicians were matched for age and gender across party to control for visual and social features of the stimuli. Democratic politicians consisted of Barack Obama, Bernie Sanders, and Wendy Davis. Republican politicians consisted of Donald Trump, Ted Cruz, and Sarah Palin. In a pilot test of this evaluation procedure, Democratic and Republican participants reported significantly different evaluations of each of these politicians in all trait.

Prior to entering the task, participants completed 10 practice trials that were identical to the full task with the exception that they were asked to rate different politicians. The practice trials ensured that participants understood how to complete the task before beginning the experiment.

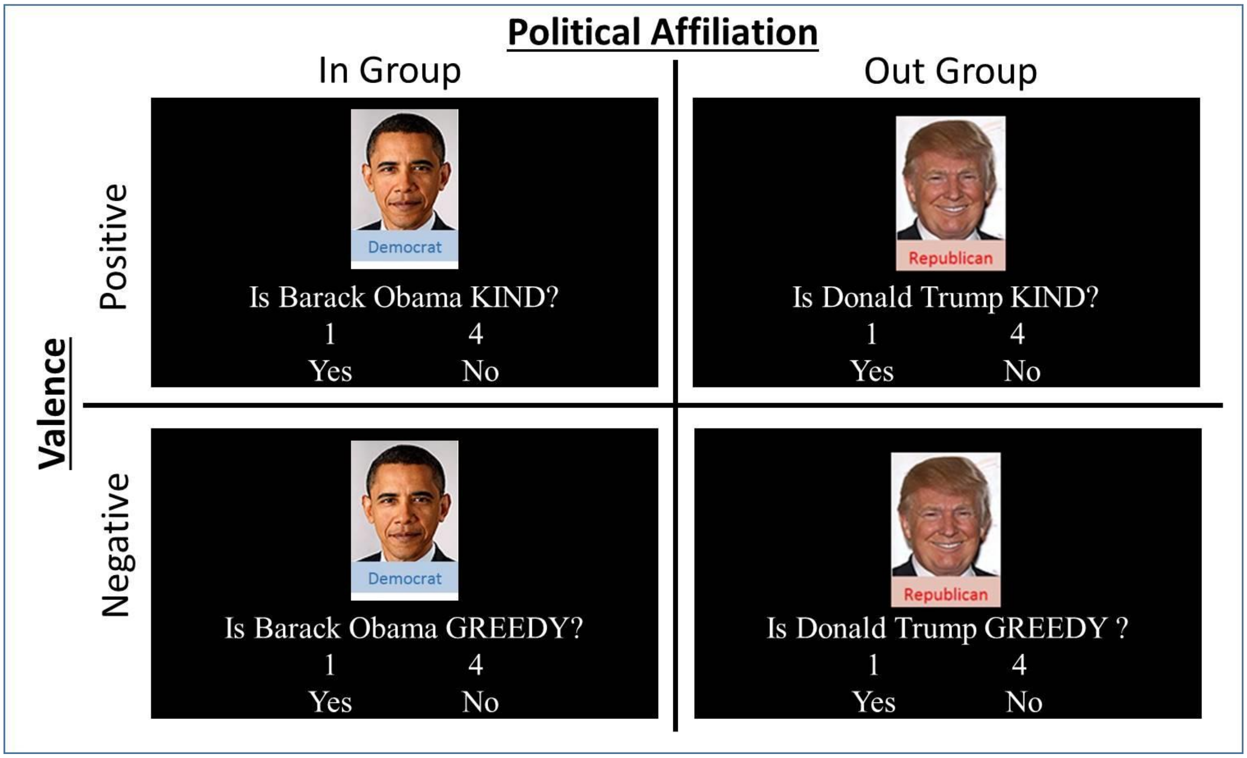


Figure 10: Social Evaluation Task. Participants (prescreened for political affiliation) evaluated the positive and negative traits for 3 politicians from each of their in-group political party and out-group political party. The social evaluation task crossed Valence (positive, negative) with Party Affiliation (in group, out group).

Drift Diffusion Modeling

To determine the roles of prior expectations and preferential evidence accumulation in evaluating a liked or disliked other, Drift Diffusion Modeling (DDM) was employed. DDM is beneficial in this context because it allows for understanding of underlying mechanisms that would otherwise be difficult to assess with self-report or reaction time data alone. Self-report is problematic because people are largely unable to introspect about the internal mechanisms that lead to their decisions (Nisbett & Wilson, 1977). Especially given that this research pertains to motivational biases and people are blind to their own biases, it would be difficult to trust any introspection about their decision making processes (Pronin, Lin, & Ross, 2002; Ehrlinger, Gilovich, & Ross, 2005). Reaction time data is also problematic because reaction times could be fast of slow due to prior expectations or preferential evidence accumulation. For example, if asked to indicate if Obama is intelligent a participant may answer quickly because they have a prior expectation that he is or they could answer quickly because they have a rapid rate of evidence accumulation. While DDM uses both self-report and reaction time, it further utilizes distributions of reaction times to calculate the underlying processes. By understanding when certain decisions are more frequently made within an RT distribution, decisions resulting from prior expectations can be teased apart from decisions resulting from preferential evidence accumulation.

DDM data analysis was conducted using fast-dm to calculate starting point (z) and drift rate (v) (Voss & Voss, 2007). Starting point indicates the prior expectations that participants have and thus how likely they are to endorse a rating as true or false. The drift rate indicates how much participants are engaging in preferential evidence accumulation before endorsing the trait as true or false. While DDM was originally used in paradigms that had correct and incorrect decisions, some recent research has shown that it can be applied to paradigms in which decisions are a matter of subjective preference with no right or wrong answer (Flagan, Mumford, & Beer, 2017; Krajbich, Lu, Camerer, & Rangel, 2012; Milosavljevic, Malmaud, Huth, Koch, & Rangel, 2010). In this paradigm starting point is a proxy for desired directional conclusion. In other words, before they have evaluated the politician they already have a direction in which they desire for their evaluation to go. It is important to note that starting point is a necessary, but not sufficient marker of possessing directional conclusions as starting point may reflect prior cognitive appraisals rather than motivation, per se. Further, drift rate indicates the rate or depth of searching through memory prior to making a decision. Differences between conditions in this paradigm relate to participants' subjective feelings and participants only have their own memories to process, so differences in drift rate must reflect differences in depth of processing their own memories.

Behavioral Analysis

Data analysis included calculating differences in raw ratings, starting points, and evidence accumulation. All three indices were estimated for each of the four conditions (Democrat-Positive, Republican-Negative, Democrat-Negative, Republican-Positive). We analyzed raw ratings in a 2 (Valence: Positive and Negative) by 2 (Politician: Democrat and Republican) within-subjects ANOVA to test if participants behaved consistently with their self-expressed political views.

Drift rates were analyzed in a 2 (Valence: Positive and Negative) by 2 (Politician: Democrat and Republican) within-subjects ANOVA to test differences in rates of accumulating evidence before giving a response.

Starting points were analyzed in two separate ways. First, a 2 (Valence: Positive and Negative) by 2 (Politician: Democrat and Republican) within-subjects ANOVA was conducted to test if there were differences between conditions in prior expectations. Further, a one-sample t-test was used to compare starting points in each condition to the center point. Starting points at the center point ($z = 0.5$) indicate that the participant has no prior expectations toward affirming or denying. Starting points closer to 0 indicate that the participant has a prior expectation of affirming the trait as true of that politician and starting points closer to 1 indicate that the participant has a prior expectation of denying the trait as true of that politician.

Study 4 Results

Raw Rating Scores. As hypothesized, liberal participants rated Democratic politicians and Republican politicians consistently with their political views (Valence*Political Affiliation: $F(1,74) = 707.28$, $p < 0.001$, $\eta_p^2 = 0.905$; see Figure 11). There was no main effect of Political Affiliation ($F(1,74) = 2.774$, $p = 0.10$, $\eta_p^2 = 0.036$), but there was a main effect of Valence ($F(1,74) = 74.625$, $p < 0.001$, $\eta_p^2 = 0.502$).

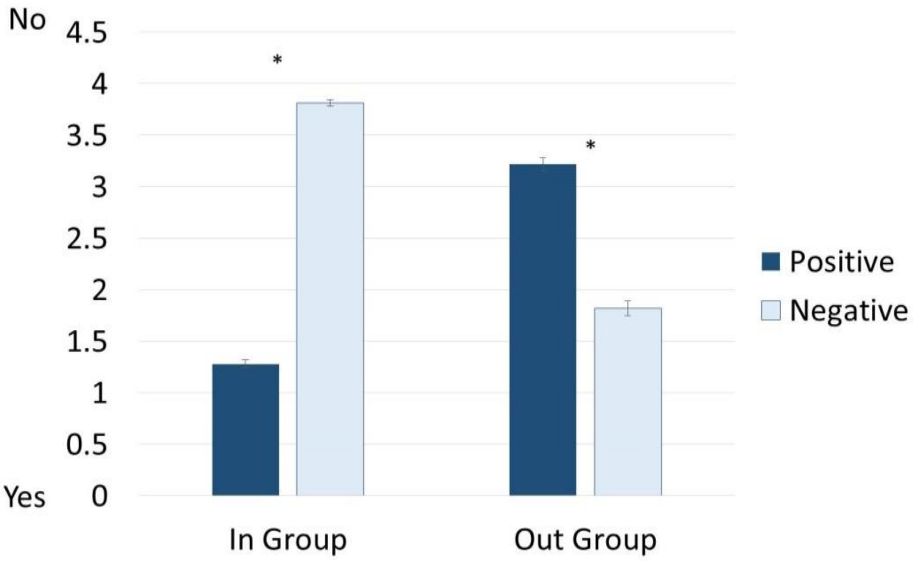


Figure 11: Raw Rating Scores. Participants rated politicians in accordance with their political leanings. Liberal participants rated Democratic politicians more positively and less negatively and Republican politicians less positively and more negatively.

Evidence Accumulation. Drift rates differed for positive and negative trait ratings, but only within the in-group condition. There was an interaction of political affiliation and trait valence on drift rates (Valence*Political Affiliation: $F(1,74) = 9.81$, $p =0.002$, $\eta_p^2 = 0.117$; see Figure 12). Participants had lower drift rates, indicating more deep processing, when rating Democrats positively ($M = 1.19$, $SD = 0.55$) than when rating Democrats negatively ($M = 1.35$, $SD = 0.49$; $t(74) = -3.698$, $p < 0.001$, $d = 0.427$). However, participants had similar drift rates when rating Republicans negatively ($M = 0.67$, $SD = 0.52$) and when rating Republicans positively ($M = 0.75$, $SD = 0.50$; $t(74) = 1.581$, $p = 0.118$, $d = 0.183$). There was a main effect of Political Affiliation ($F(1,74) = 82.17$, $p < 0.001$, $\eta_p^2 = 0.526$), but no main effect of Valence ($F(1,74) = 2.369$, $p = 0.128$, $\eta_p^2 = 0.031$).

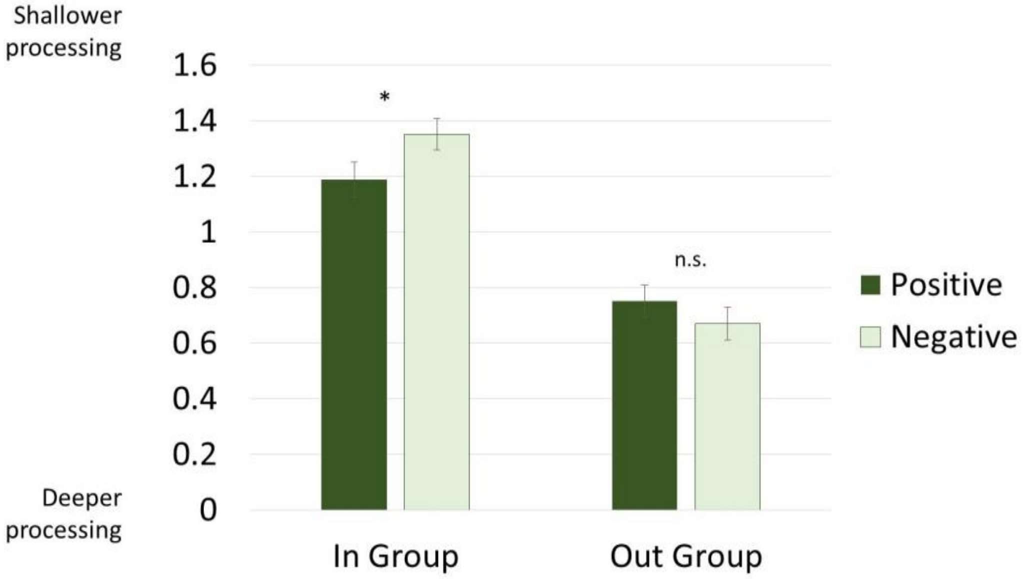


Figure 12: Drift Rates. Liberal participants had larger drift rates (more shallow processing) when evaluating Democratic politicians than when evaluating Republican politicians. There was also an interaction effect such that they had smaller drift rates when evaluating Democratic politicians positively as compared to negatively.

Starting Point. Participants started with prior expectations for rating Democrats with positive traits and Republicans with negative traits, but they did not have prior expectations for rating Democrats with negative traits and Republicans positive traits (Valence*Political Affiliation: $F(1,74) = 13.45$, $p < 0.001$, $\eta_p^2 = 0.154$; see Figure 13).

There was no main effect of Valence (F(1,74) = 0.40, p = 0.53, η_p^2 = 0.005) or of Political Affiliation (F(1,74) = 2.02, p = 0.16, η_p^2 = 0.027). Starting point for rating Democrats positively (M = 0.44, SD = 0.11) and Republicans negatively (M = 0.44, SD = 0.12) were different than the middle point (t(74) = -4.86, $p < 0.001$, d = -0.80; t(74) = -4.14, $p < 0.001$, d = -0.68). However, starting point for rating Democrats negatively (M = 0.50, SD = 0.09) and Republicans positively (M = 0.52, SD = 0.12) were not different than the middle point (t(74) = -0.13, p = 0.898, d = -0.02; t(74) = 1.23, p = 0.225, d = 0.20).

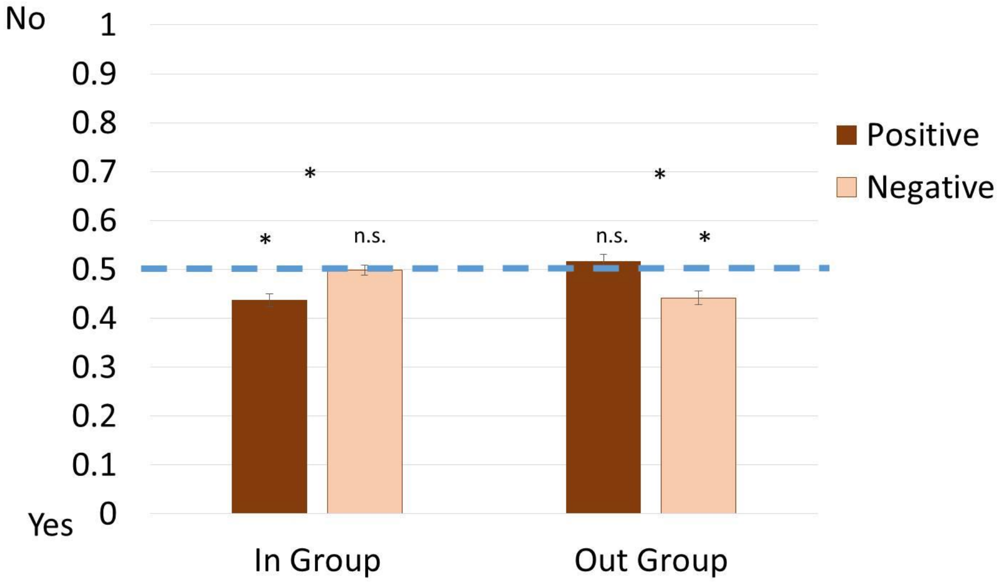


Figure 13: Starting Points. Liberal participants had starting points that were significantly different than the midpoint (blue, dashed line) when they evaluated Democrats positively and Republicans negatively.

Study 4 Discussion

Study 4 results suggest that biased memory searches occur when aiming to paint a social target in a flattering but not in an unflattering light. Participants had different drift rates for positive and negative evaluations in the in-group condition, but not in the out-group condition. This suggests that, given a desire to paint someone in an unflattering light (i.e., negative trait ratings of Trump), people did not employ biased memory searching to help them achieve making that evaluation. Starting point data suggests that rating out-group members on negative traits and in-group members on positive traits was consistent with participants' desired directional conclusions. Participants only expressed prior expectations (i.e., starting points different than the midpoint) when the evaluation question was consistent with their political views. This is consistent with the interpretation that liberals do have a desired directional conclusion toward identifying Obama as intelligent and Trump as greedy. However, reaching the desired directional conclusion that Trump is greedy did not seem to be supported by biased memory searching.

GENERAL DISCUSSION

The current research offers support for the role of biased searching through memories as one cognitive mechanism underlying motivated social evaluation. Previous literature has been unable to illuminate the role of biased searching through memory because it has failed to operationalize the underlying mechanisms, it has conflated asymmetrical memory reporting with desired directional conclusions, and it has focused on the role of biased memory searching only in situations where people's directional conclusions are flattering. The current research utilized computational modeling methods (SDT and DDM) to operationalize underlying memory mechanisms, used financial and psychological incentives to test the role of desired directional conclusions, and employed a novel paradigm to understand the role of biased memory searching when aiming to paint a social target in an unflattering as well as a flattering light. In Study 1, recognition thresholds for negative feedback remained relatively more conservative even when memory was incentivized through decreased self-relevance or decreased self-relevance and opportunity for financial gain. Study 2 suggested 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 feedback condition. Results suggest that neurophysiological associations with self-reported forgotten negative feedback are better characterized by 'forgotten memories' than 'suppressed knowledge.' Moreover, study 4 showed that biased memory searches may only support the desire to paint a social target in a flattering light, but not when the desired directional conclusion is to paint a social target in an unflattering light. Taken together, results from Studies 1-4 suggest that biased searching through memory is one cognitive process that supports flattering social evaluations about the self and others and that social evaluations that have previously been assumed to be motivated are in fact likely the result of aiming to reach a desired directional conclusion. Further, the results have implications for thinking about self-processing as well as future directions to more fully understand the role of biased memory searching in motivated social evaluations.

Implications for the role of biased searches through memory in social evaluation

The current research builds on prior research by utilizing computational modeling to more fully understand the role of biased memory searching in motivated social evaluations. Previous research often did not operationalize the underlying mechanisms involved in motivated social evaluation, so it has been unclear if biased searches through memory could explain motivated social evaluation phenomena as some researchers have suggested (Showers & Cantor, 1985; Kunda, 1990, Dunning, 2015). While there are certainly many other cognitive processes which may support motivated social evaluation, the current research suggests the critical role that biased memory searching plays.

Studies 1-3 show that differences in memory for positive and negative information are driven by different standards for claiming recognition. The use of Signal Detection Theory (SDT) allowed for understanding the underlying components of thresholds for claiming recognition and ability to discern previously seen words from completely novel words. Further, it seems that previous research utilizing proportion recognized or recalled was capturing less conservative thresholds for claiming recognition and less accuracy of positive compared to negative feedback. Results from Study 4 suggest that people engage more deeply in biased memory searches when they want to make flattering, but not unflattering, social evaluations. However, it should be noted that Study 4 analyses focused on self-identified Democrats which warrants caution in generalizing these results. Research does suggest that Democrats and Republicans exhibit similar biases when rating in-group and out-group members (Pew, 2016), but generalization should be met with caution. The results from computational modeling (i.e., location (c) from SDT and drift rate from DDM) in all four studies suggest that biased searching through memory is one cognitive process that may support motivated social evaluation. However, Study 4 results suggest that the role for biased memory searching may only support motivated social evaluation when aiming to paint a social target in a flattering light.

The current research also finds that the role of biased memory searches at the time of recognition may be affected by or co-occurring with processes that occur as early as encoding. Study 2 shows that alternate incentives only work when incentives are presented prior to encoding and Study 3 further finds that negative feedback that is ‘forgotten’ is not encoded in the first place rather than encoded but suppressed at the time of retrieval. Therefore, it seems that differences in standards for claiming recognition of positive and negative self-relevant information at the time of retrieval is occurring in conjunction with lower rates of encoding negative feedback. Taken together, the results of these studies show that biased searching through memories is involved in motivated social evaluation, and its involvement may supported by differences in encoding.

Did the current research really look at instances of motivated social evaluation?

One criticism of research on motivated social evaluation is that it fails to discount alternate, nonmotivated accounts for findings that can be interpreted from a motivated perspective (Chambers & Windschitl, 2004). Conversely, the self-enhancement account suggests that people possess a strong motivation to see the self in a positive light and that this self positivity motivation (and the resulting asymmetries in recognition thresholds) would be extremely difficult to override with other incentives. Results from Studies 1 and 2 showing that asymmetries in recognition thresholds are resilient to competing incentives suggests these findings are consistent with a self-enhancement explanation of why people remember more flattering self-relevant information. This support for the self-enhancement explanation suggests that the paradigm used in Studies 1-3 is appropriate for research aiming to understand the role of biased searches through memory. Further, the starting point findings from Study 4 also suggest that the evaluation of political figures on positive and negative traits is appropriate for studying the role of biased searches through memory in motivated social evaluation. Starting point findings suggest that people may be expressing desired directional conclusions to rate Democrats with positive traits and Republicans with negative traits (though, as noted earlier, starting point findings are not sufficient to conclusively suggest motivation).

Implications for the long-term consequences of overly positive self-evaluations

Beyond speaking to the specific questions outlined in this dissertation, the current research has more broad implications for understanding self-processing. While it has been suggested that there are positive consequences of self-enhancement (Dufner, Reitz, & Zander, 2014), there have also been many negative consequences found. 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). A prevalent explanation for the negative long-term consequences is that people eventually find themselves unable to suppress the retrieval of negative feedback, which leads to negative self-esteem. However, the research here suggests that people are not suppressing negative feedback because biased searching through memories partially stems from differences at encoding. ERP results suggest that which beliefs are available to access are constrained by which information is encoded in the first place. Further, it seems that shifting thresholds requires high levels of incentives suggesting that any negative feedback that is encoded would still be less likely to be identified as self-relevant. Incentives were only able to influence people to shift their thresholds for claiming recognition of positive and negative feedback when they were both psychological and financial, so it seems unlikely that over time the strong drive to see the self positively will diminish. Taken together, the results of Studies 1-3 suggest that long-term consequences are unlikely due to an inability to continuously suppress negative self-relevant information as suppression does not seem to be a mechanism supporting biased memory searches.

Future directions and considerations

While the current findings shed light on motivated social evaluations and the role of biased memory searches in such evaluations, there are still avenues for a deeper understanding in self-processing. Specifically, it is unclear if results from Study 4 would replicate given a self-evaluative paradigm. While study 4 utilized a person perception paradigm, the focus was on understanding the underlying role of biased memory searches in motivation to reach any desired directional conclusion and not just a flattering desired directional conclusion. It is difficult to draw conclusions or generalize about psychological processes when studying people who desire unflattering feedback about the self because self-enhancement is such a strong and prevalent motivation and the desire for unflattering feedback can indicate psychological dysfunction. Therefore, the rating of political figures paradigm was used as a situation in which people may typically have strong motivations toward unflattering evaluations. Now that there is a foundation for understanding, we could explore more deeply how biased memory searches would work within self-evaluation. One possible future direction that could build on the current research would be to manipulate people's mindsets regarding self-relevant feedback. While self-enhancement is a strong and prevalent self-perception motivation, there are other motivations people may have at different times or may be experimentally induced to have (Taylor, Netter, & Wayment, 1995). One motivation that may make people less motivated toward positive self-relevant feedback is self-improvement. In a self-improvement mindset a person has a goal to become better in a certain domain. For example, someone who would like to improve their grade in a class may be more interested in receiving any feedback (positive or negative) about their performance on a paper or exam in order to perform better on future assignments. In such a mindset, all feedback is valuable because positive feedback allows you to know what you are doing well and negative feedback allows you to understand where changes are needed. Future research could manipulate people's mindsets about the value of negative feedback to understand how depth of processing may vary given a self-processing paradigm. If people have no strong preference toward flattering feedback, would they exhibit equal depths of memory searching for flattering and unflattering self-relevant information?

While the discussion thus far has focused on neurotypical populations, the current findings may also have implications for people affected by psychological pathologies. Specifically, we know that people with depression and low self-esteem tend to have negative self-views and rather than seeking positive feedback they aim to verify those self-views by seeking negative feedback (Swann, 1983; Swann, Wenzlaff, & Tafarodi, 1992). The current and extant research has largely focused on the positive end of the spectrum. How would the specific indices measured here apply to a paradigm looking at self-evaluation in those affected by depression? Study 4 results suggest that biased memory searching does not operate the same way when desired directional conclusions are toward unflattering evaluations. However, research on those with depression does show differences in how positive and negative stimuli are processed and remembered (Coyne & Gotlib, 1983). It could be that not all motivations toward making unflattering evaluations are the same and this difference between wanting to see an other negatively and wanting to see the self- negatively needs more exploration. Or, it could be that depth of processing has been defined differently when exploring the role of depression in memory biases. In that case, utilizing DDM in a paradigm examining self-evaluation tendencies in those with depression might reveal new insights about the cognitive processes associated with this psychological disorder.

Conclusions

The roles of biased searches through memory and desired directional conclusions have been posited as supporting social evaluation. However, the support for their roles has been limited due to a lack of operationalizing underlying mechanisms of memory, conflation of memory asymmetries for positive and negative feedback with desire to see someone in a positive light, and an emphasis or focus on flattering evaluations rather than any desired directional conclusion. The current research utilized novel paradigms as well as computational modeling to allow for a deeper understanding of the roles of biased memory searches and desired directional conclusions in social evaluations. Biased belief searching is one cognitive mechanism that supports motivated social evaluation.

Computational modeling revealed two ways that biased memory searches could support motivated social evaluation: differential standards for claiming recognition and depth of processing when searching through memory. Further, theories that suggest a role for directional conclusions (rather than nonmotivated perspectives such as fluency) can account for differences in people's propensity to claim recognition of positive information at a greater rate than negative information about the self. Psychological and financial incentives were unable to diminish the difference between thresholds for claiming recognition of positive and negative self-relevant feedback except when the highest measured level of incentives were offered before encoding to recognize feedback about an other. Further, the results from these four studies shed light on the mechanisms that support motivated social evaluations and offer future avenues to explore with regards to self-evaluation.

References

Alicke, M. D. (1985). Global self-evaluation as determined by the desirability and controllability of trait adjectives. *Journal of Personality and Social Psychology*, *49*(6), 1621–1630.

Alves, H., Koch, A., & Unkelbach, C. (2017). Why good is more alike than bad: Processing implications. *Trends in Cognitive Sciences, 21*(2), 69-79.

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

Bartels L. M. (2002). Beyond the running tally: Partisan bias in political perceptions. *Political Behavior, 24,* 117–150.

Brandt, M. J., Reyna, C., Chambers, J. R., Crawford, J. T., & Wetherell, G. (2014). The ideological-conflict hypothesis: Intolerance among both liberals and conservatives. *Current Directions In Psychological Science*, *23*(1), 27-34. doi:10.1177/0963721413510932

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.

Cohen, G. L., Aronson, J., & Steele, C. M. (2000). When beliefs yield to evidence: Reducing biased evaluation by affirming the self. *Personality and Social Psychology Bulletin*, *26*(9), 1151-1164.

Coyne, J. C., & Gotlib, I. H. (1983). The role of cognition in depression: a critical appraisal. *Psychological bulletin, 94*(3), 472.

Curran, P. J., & Hussong, A. M. (2009). Integrative data analysis: the simultaneous analysis of multiple data sets. *Psychological methods*, *14*(2), 81.

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. http://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. http://doi.org/10.1016/j.paid.2004.05.010

Duarte, J. L., Crawford, J. T., Stern, C., Haidt, J., Jussim, L., & Tetlock, P. E. (2015). Political diversity will improve social psychological science. *Behavioral And Brain Sciences, e130.*

Dufner, M., Gebauer, J. E., Sedikides, C., & Denissen, J. J. (2018). Self-enhancement and psychological adjustment: A meta-analytic review. *Personality and Social Psychology Review*, 1088868318756467.

Dunning, D. (2015). Motivated cognition in self and social thought.

Ehrlinger, J., Gilovich, T., & Ross, L. (2005). Peering into the bias blind spot: People’s assessments of bias in themselves and others. *Personality and Social Psychology Bulletin, 31*(5), 680-692.

Flagan, T., Mumford, J. A., & Beer, J. S. (2017). How do you see me? the neural basis of motivated meta-perception. *Journal of cognitive neuroscience, 29*(11), 1908-1917.

Green, J. D., & Sedikides, C. (2004). Retrieval selectivity in the processing of self-referent information: Testing the boundaries of self-protection. *Self and Identity*, i(1), 69-80.

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. http://doi.org/10.1016/j.jesp.2007.10.006

Green, D. M., & Swets, J. A. (1966). Signal detection theory and psychophysics (Vol. 1). New York: Wiley.

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,* 1098-1106. doi: 10.1177/0956797615575734

John, O. P., Naumann, L. P., & Soto, C. J. (2008). Paradigm shift to the integrative Big Five Trait taxonomy. *Handbook of Personality: Theory and Research*, 114–158. http://doi.org/10.1016/S0191-8869(97)81000-8

Johnston, W. A., Hawley, K. J., & Elliott, J. M. (1991). Contribution of perceptual fluency to recognition judgments. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 17*(2), 210.

Kouchaki, M., & Gino, F. (2016). Memories of unethical actions become obfuscated over time. *Proceedings of the National Academy of Sciences, 113*(22), 6166-6171.

Krajbich, I., Lu, D., Camerer, C., & Rangel, A. (2012). The attentional drift-diffusion model extends to simple purchasing decisions. *Frontiers in psychology, 3,* 193.

Kruger, J. (1999). Lake Wobegon be gone! The" below-average effect" and the egocentric nature of comparative ability judgments. *Journal of personality and social psychology*, *77*(2), 221.

Kunda, Z. (1990). The case for motivated reasoning. *Psychological Bulletin*, *108*(3), 480–498.

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

Milosavljevic, M., Malmaud, J., Huth, A., Koch, C., & Rangel, A. (2010). The drift diffusion model can account for the accuracy and reaction time of value-based choices under high and low time pressure.

Munro, G. D., Ditto, P. H., Lockhart, L. K., Fagerlin, A., Gready, M., & Peterson, E. (2002). Biased assimilation of sociopolitical arguments: Evaluating the 1996 U.S. presidential debate. *Basic And Applied Social Psychology*, *24*(1), 15-26. doi:10.1207/153248302753439038

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. http://doi.org/10.1016/0749-596X(86)90022-7

Nichols, T. E., & Holmes, A. P. (2002). Nonparametric permutation tests for functional neuroimaging: a primer with examples. *Human brain mapping*, *15*(1), 1-25.

Nisbett, R. E., & Wilson, T. D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological review*, *84*(3), 231.

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.

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.org/10.1037/0022-3514.84.4.890

Pew Research Center, June, 2016, "Partisanship and Political Animosity in 2016"

Pinter, B., Green, J. D., Sedikides, C., & Gregg, A. P. (2011). Self-protective memory: Separation/integration as a mechanism for mnemic neglect. *Social Cognition*, *29*(5), 612-624.

Pronin, E., Lin, D. Y., & Ross, L. (2002). The bias blind spot: Perceptions of bias in self versus others. *Personality and Social Psychology Bulletin*, *28*(3), 369-381.

Raskin, R., & Terry, H. (1988). A principal-components analysis of the Narcissistic Personality Inventory and further evidence of its construct validity. *Journal of personality and social psychology*, *54*(5), 890.

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

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.

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

Sanguinetti, J. L., Trujillo, L. T., Schnyer, D. M., Allen, J. J., & Peterson, M. A. (2016). Increased alpha band activity indexes inhibitory competition across a border during figure assignment. *Vision research*, *126*, 120-130.

Sanitioso, R., Kunda, Z., & Fong, G. T. (1990). Motivated recruitment of autobiographical memories. *Journal of Personality and Social psychology*, *59*(2), 229.

Sedikides, C., & Green, J. D. (2009). Memory as a Self-Protective Mechanism. *Social and Personality Psychology Compass*, *3*, 1055–1068. http://doi.org/10.1111/j.1751-9004.2009.00220.x

Sherman, D. A., Nelson, L. D., & Steele, C. M. (2000). Do messages about health risks threaten the self? Increasing the acceptance of threatening health messages via self-affirmation. *Personality and Social Psychology Bulletin*, *26*(9), 1046-1058.

Showers, C., & Cantor, N. (1985). Social cognition: A look at motivated strategies. *Annual review of psychology*, *36*(1), 275-305.

Swann, W. B. (1983). Self-verification: Bringing social reality into harmony with the self. *Social psychological perspectives on the self*, *2*, 33-66.

Swann, W. B., Wenzlaff, R. M., & Tafarodi, R. W. (1992). Depression and the search for negative evaluations: More evidence of the role of self-verification strivings.

Tajfel, H., & Turner, J. C. (2004). The Social Identity Theory of Intergroup Behavior.

Taylor, S. E., & Brown, J. D. (1988). Illusion and well-being: a social psychological perspective on mental health. *Psychological Bulletin*, *103*(2), 193–210. http://doi.org/10.1037/0033-2909.103.2.193

Taylor, S. E., Neter, E., & Wayment, H. A. (1995). Self-evaluation processes. *Personality and Social Psychology Bulletin, 21*(12), 1278-1287.

Trujillo, L. T., Allen, J. J., Schnyer, D. M., & Peterson, M. A. (2010). Neurophysiological evidence for the influence of past experience on figure–ground perception. *Journal of Vision*, *10*(2), 5-5.

Voss, A., Nagler, M., & Lerche, V. (2013). Diffusion models in experimental psychology. *Experimental psychology*.

Voss, A., & Voss, J. (2007). Fast-dm: A free program for efficient diffusion model analysis. *Behavior Research Methods*, *39*(4), 767-775.

Voss, A., Voss, J., & Lerche, V. (2015). Assessing cognitive processes with diffusion model analyses: a tutorial based on fast-dm-30. *Frontiers in psychology*, *6*, 336.

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. doi:10.1037/pspa0000124