Short Research Article

Self-Serving Bias in Memories

Selectively Forgetting the Connection Between Negative Information and the Self

Yanchi Zhang,[1] Zhe Pan,[1] Kai Li,[2] and Yongyu Guo[2]

[1]School of Psychology, Central China Normal University, Wuhan, PR China
[2]School of Psychology, Nanjing Normal University, Nanjing, PR China

Abstract: Protecting one's positive self-image from damage is a fundamental need of human beings. Forgetting is an effective strategy in this respect. Individuals show inferior recall of negative feedback about themselves but unimpaired recognition of self-related negative feedback. This discrepancy may imply that individuals retain negative information but forget that the information is associated with the self. In two experiments, participants judged whether two-character trait adjectives (positive or negative) described themselves or others. Subsequently, they completed old-new judgments (Experiment 2) and attribution tasks (Experiments 1 and 2). Neither old-new recognition nor source guessing bias was influenced by word valence. Participants' source memory was worse in the negative self-referenced word processing condition than in the other conditions. These results suggest there is a self-serving bias in memory for the connection between valence information and the self.

Keywords: self-image, forgetting, connection, negative information

Viewing oneself as positive is a vital human need (Leary, 2007). However, individuals in daily life frequently receive negative comments from friends, relatives, employers, and even strangers. Therefore, one's positive self-image may often be threatened. Consequentially, how to protect one's self-image from damage is a critical problem.

In this respect, the first choice is to seek affirmative information and to avoid disapproving information. Some studies have shown that individuals selectively attend to positive information about themselves (Ditto & Lopez, 1992) and report feeling better after receiving positive feedback (Kwang & Swann Jr., 2010). When asked to rate the degree to which they expected to receive four types of feedback (self-enhancing, self-effacing, self-improving, and no-feedback) from others, both Chinese and Americans were uninterested in self-effacing feedback (Gaertner, Sedikides, & Cai, 2012). In contrast, Baumeister and Ilko (1995) asked participants to furnish public or private accounts of their recent experiences of success. Compared with private accounts, public accounts contained more references to receiving aid from others. This suggests that self-presentation prompts individuals to consider possible reactions from other people and to behave modestly in public situations (Baumeister & Ilko, 1995; Gould, Brounstein, & Sigall, 1977). In short, people actively pursue information that enhances the self, but evade information that hurts the self (Gregg, Sedikides, & Gebauer, 2011; Sedikides, 1993). Unfortunately, reality does not always follow the will of human beings. In the long term, everyone will receive self-threatening feedback. In response to this frustrating situation, individuals can choose another coping strategy to maintain a positive self-image: selectively forgetting information that has disadvantageous implications for the self.

Since self-threatening information is usually upsetting, individuals try their utmost to erase such information from memories (Anderson & Green, 2001; Gagnepain, Henson, & Anderson, 2014; Kim & Yi, 2013). One of the most powerful lines of evidence is autobiographical memory research, which indicates that it is more difficult to recall negative life-events than positive life-events. For example, in one study, participants were asked to remember real-life behaviors at Time 1 and recall them at Time 2. If the target was the self, participants recalled a significantly lower proportion of negative than positive behaviors (Ritchie, Sedikides, & Skowronski, 2017).

Researchers have systematically examined the selective forgetting of self-threatening negative feedback – the mnemic neglect effect (MNE). In an initial experiment by Sedikides and Green (2000), participants read either self- or other-referent mixed-valence lists of evaluations that consisted of central and peripheral trait-relevant behaviors.

Then, participants were given a surprise free-recall task. Participants in the self-referent condition recalled fewer central-negative behaviors than those in the other-referent condition, and self-referential participants recalled fewer central-negative than central-positive behaviors. Mnemic neglect occurred both in mundane reality settings, wherein evaluations were based on a purportedly reliable and valid personality inventory, and in minimal-feedback settings, wherein feedback was hypothetical (Sedikides & Green, 2000; Sedikides, Green, Saunders, Skowronski, & Zengel, 2016).

As is commonly appreciated, feedback usually contains two aspects: the referent and the content. The referent refers to the person who is evaluated. The content refers to the valence information (positive or negative) implied in the feedback. Although there are many studies exploring the memory of self-threatening negative information (Green, Sedikides, & Gregg, 2007; Newman, Sapolsky, Tang, & Bakina, 2014; Sedikides & Green, 2004), no researchers have explored the memory of connections between the valence information and the referent (self or others) that were developed while processing feedback. How do individuals perceive the connection between valence information and themselves, especially when the information is obviously negative? Are they likely to establish a weaker connection when the information is self-threatening? Do they have a preference for generating a connection with positive information? More generally, what strategy do individuals use to protect the self-image from negative feedback, when addressing the relationship between the self and self-threatening information?

Evidence from several sources suggests that individuals weaken their connection with negative information or strengthen their connection with positive information. According to psychodynamics, to narrow the gap between the actual and ideal self, defense mechanisms (e.g., denial, projection) inhibit awareness of unfavorable events, especially when those events are highly self-threatening (Baumeister, Dale, & Sommer, 2010). In addition, studies indicate there is a self-serving bias in attribution, which is helpful for maintaining one's positive self-image and even physical well-being (Hu, Zhang, & Yang, 2015; Sanjuán & Magallares, 2014). Moreover, people make assimilation-based judgments about positive self-related information, which leads positive information to be integrated with stored self-knowledge. However, individuals make contrast-based judgments about negative self-related information, which leads threatening information to be separated from stored self-knowledge (Pinter, Green, Sedikides, & Gregg, 2011). All of these results imply that individuals prefer to sever their connection with negative information. Therefore, we expected that participants would poorly memorize the connection between negative information and the self.

The Current Study

We borrowed and modified the classic self-reference paradigm (Conway & Dewhurst, 1995; Rogers, Kuiper, & Kirker, 1977) to address the above hypothesis in two experiments. Both experiments consisted of two phases. During the learning phase, participants were asked to process trait words in both the self-reference and other-reference conditions. During the testing phase, memory performance was measured. In the first experiment, the test asked participants to attribute each word they processed to the word's referent. If participants forgot the word's referent, they could choose "not remember" as their response. In the second experiment, the test first asked participants to judge whether the words presented were "new" or "old." If a word was judged as "old," the test also required participants to attribute the "old" word to its referent. Collectively, with the two experiments we elucidated the memory of the connections generated while processing positive and negative words in both self- and other-reference conditions.

Experimental Materials

To obtain valence, arousal, familiarity, and meaningfulness measurements of the stimulus words, pretests were conducted with individuals who did not participate in the main experiments. First, 413 two-character trait words were selected from a pool of Chinese personality-trait adjectives (Wang & Cui, 2005) and randomly divided into two groups (the first group contained 210 words, while the second contained 203 words). Second, 66 college students were recruited to rate the words (33 raters for the first-group words, 33 raters for the second-group words) in terms of valence using a 7-point Likert scale (1 = *extremely bad* to 7 = *extremely good*); 66 college students rated the words (33 raters for the first-group words, 33 raters for the second group) in terms of arousal (1 = *extremely unexciting* to 7 = *extremely exciting*); 64 college students rated the words (32 raters for the first-group words, 32 raters for the second group) in terms familiarity (1 = *extremely unfamiliar* to 7 = *extremely familiar*); and 60 college students rated the words (30 raters for the first-group words, 30 raters for the second group) in terms of meaningfulness (1 = *extremely meaningless* to 7 = *extremely meaningful*). Each student only rated one dimension of the words. Finally, 40 positive and 40 negative trait words were selected as the experimental materials for the first experiment and 80 positive and 80 negative trait words were selected as the experimental materials for the second experiment. Trait words of both experiments are provided in ESM 5.

Experiment 1

Past research has established that individuals selectively forget negative information about themselves, relative to positive information. However, memory of connections between positive or negative information and the self or other entities has not been examined. The first experiment examined participants' overall attribution of words to their referents. The attribution is likely to include memory components and judgment bias.

Participants

Thirty-nine Chinese university students (27 females, 12 males; aged 18–24 years, M_{age} = 20.18 years) were recruited as paid volunteers. All participants were right-handed, native Chinese speakers with normal or corrected-to-normal vision. None of the participants had a history of, or currently reported, neurological or psychiatric conditions.

Design and Materials

The experiment used a two-factorial design that varied both the valence (positive vs. negative) and the referent (self vs. other) within subjects. The number of correct responses each participant produced during word attribution was used as the dependent variable.

The experimental stimuli were 80 two-character trait adjectives, consisting of 40 positive words and 40 negative words. Positive words were significantly more desirable than negative words, $t(78) = 43.84$, $p < .001$. These words were divided into two lists. Each list contained 20 positive words and 20 negative words. Word valence was matched between the two lists. In addition, word arousal, word familiarity, and word meaningfulness were matched both within and between lists. One list was assigned to the self-reference condition, and the other list was assigned to the other-reference condition. The two lists were exchanged between conditions across participants, so that trait words that appeared in the self-reference condition for half of the participants appeared in the other-reference condition for the other participants.

Procedure

Upon arrival in the lab, participants were given a brief overview of the study and signed the informed consent form. Then, participants completed the Beck Depression Inventory (M = 8.33, SD = 4.84) (Beck, Ward, Mendelson, Mock, & Erbaugh, 1961), the State-Trait Anxiety Inventory (M_{state} = 35.54, SD_{state} = 6.81; M_{trait} = 43.62, SD_{trait} = 9.90) (Spielberger, Gorsuch, Lushene, Vagg, & Jacobs, 1983) and the Rosenberg Self-esteem Scale (M = 27.97, SD = 4.06) (Rosenberg, 1979). Finally, the primary experimental task was started.

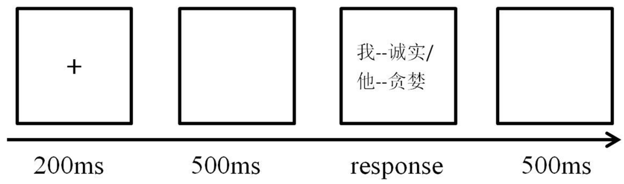

Figure 1. Each trial started with a fixation (200 ms), followed by a blank screen (500 ms). Next, a target adjective (positive/negative) was presented with a personal pronoun (I/he) until a response was given. Then, 500 ms after the response, the next trial started.

The primary experiment was conducted via E-prime 2.0 (retrieved from https://pstnet.com/welcome-to-e-prime-2-0/). During the learning phase, participants were asked to perform a two-choice judgment task of individually presented trait words paired with referent cues (Figure 1). In the self-reference condition, each trait word was presented with the first-person pronoun "I" (我), and participants were asked to indicate whether the word could be used to describe themselves. In the other-reference condition, each trait word was presented with the third-person pronoun "he" (他), and participants were asked to indicate whether the word could be used to describe others. All stimuli were randomly presented on a computer monitor in white font centered on a gray background. Participants were asked to indicate their choice by pressing the appropriate "d" or "k" key with one of their index fingers. The assignment of keys to response alternatives was balanced across participants. After completion of the study task, participants were asked to count backward from 100 to 1 aloud for 5 min to prevent maintenance rehearsal.

During the testing phase, the 80 words were randomly presented, one word at a time, on the computer screen. Participants needed to attribute the words to their referents ("I "or "he") by pressing the appropriate "d" or "k" key. If they could not remember which personal pronoun ("I "or "he") the trait word had been paired with, they were to press the space bar. It should be noted that, before the memory test was officially started, participants did not know there was *a* test.

Results and Discussion

The raw data of both experiments is provided in ESM 1–4. Attribution counts were the numbers of the trait words that were correctly attributed to their referents in each condition by each participant (Table 1). Attribution counts were subjected to a 2 (Valence: positive vs. negative) × 2 (Referent: self vs. other) ANOVA. The interaction between referent and valence was significant, $F(1, 38) = 19.25$, $p < .001$, $\eta^2 = .336$. In the self-reference condition, participants correctly attributed fewer negative words than positive

Table 1. Participants' behavioral performance in the testing phase in Experiment 1

	Participants' responses		
Actual referent during learning	Attributed to "I"	Attributed to "He"	Do not remember
Referent "I"			
Positive	14.97 (2.87)	3.33 (2.85)	1.69 (1.70)
Negative	10.92 (4.05)	6.05 (3.95)	3.03 (3.41)
Referent "He"			
Positive	7.23 (3.60)	9.21 (3.16)	3.56 (2.98)
negative	5.13 (3.32)	10.97 (3.92)	3.90 (3.37)

Note. Rows represent presentation during learning, columns denote the response of participants, and cells contain average numbers (and standard deviations).

words, $t(38) = 5.09$, $p < .001$. In the other-reference condition, participants correctly attributed more negative words than positive words, $t(38) = 2.45$, $p < .05$. For positive words, participants correctly attributed more self-referenced words than other-referenced words, $t(38) = 7.97$, $p < .001$. However, for negative words, there was no significant difference between the attribution of self-referenced words and that of other-referenced words, $t(38) = 0.05$, $p > .05$. These results suggest that (1) participants more accurately attributed negative other-referenced words than positive other-referenced words, but less accurately attributed negative self-referenced words than positive self-referenced words; (2) self-reference processing can improve the attribution of positive information rather than that of negative information.

The experimental outcome is consistent with attribution studies, which suggest that individuals are always ready to separate themselves from negative information or to link themselves to positive information (Mezulis, Abramson, Hyde, & Hankin, 2004; Seidel et al., 2010). When negative information is unavoidable, one can achieve self-protection by severing the connection between negative information and the self.

It is important to note that there are two other possible reasons for this experimental outcome, in addition to the memory of the connections between trait words and their referents. The first reason is that participants might have forgotten more negative self-referenced words than positive self-referenced words, and thus they failed to recognize whether the word had been presented. The second reason is that there might have existed judgment bias in cases where the referent was forgotten, and thus participants attributed more positive words than negative words to the self.

Experiment 2

The second experiment examined the memory effect and judgment bias simultaneously. We added some "new" words in the testing phase, and participants needed to judge whether each word was "old" or "new" and to attribute the words judged as "old" to their referents ("I" or "he").

Participants

Seventy Chinese university students (54 females, 16 males; aged 18–24 years, $M_{age} = 20.63$ years) were recruited as paid volunteers. All participants were right-handed, native Chinese speakers with normal or corrected-to-normal vision. None of the participants had a history of, or currently reported, neurological or psychiatric conditions. No participant took part in the first experiment.

Design and Materials

The experiment used a two-factor design that varied both the valence (positive vs. negative) and the referent (self vs. other) within subjects.

A set of 160 two-character trait adjectives (80 positive, 80 negative) was selected as the experimental stimulus. These words were divided into two lists, with 80 words (40 positive, 40 negative) in each list. The two lists were equated in terms of valence, arousal, familiarity, and meaningfulness. There was no significant difference between positive and negative words in terms of arousal, familiarity, and meaningfulness in each list. One of the two lists was used as study list. Half of the study list words (20 positive, 20 negative) were assigned to the self-reference condition, and half to the other-reference condition. There was no significant difference between the two halves in terms of valence, arousal, familiarity, and meaningfulness. The two halves were exchanged between conditions across participants, so that trait words that appeared in the self-reference condition for half of the participants appeared in other-reference condition for the other participants. All 160 trait words were used in the test phase.

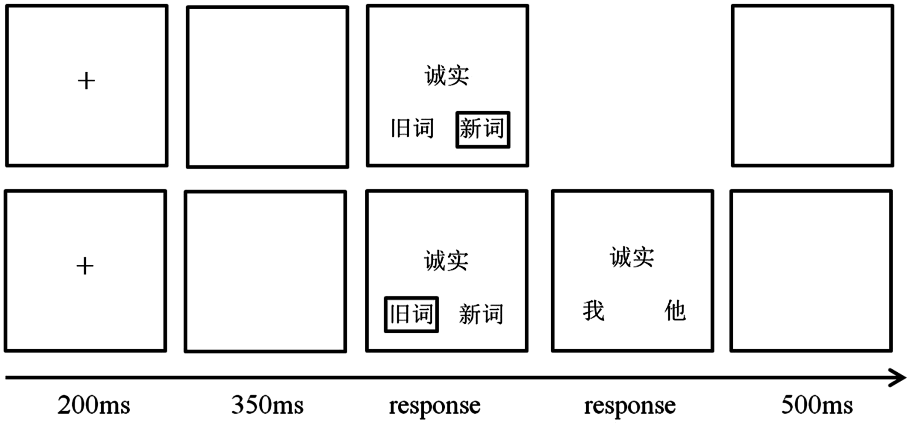

Figure 2. Each trial started with a fixation (200 ms), followed by a blank screen (350 ms). Then, an adjective (old/new) was presented, and the participant pressed keys representing "old" or "new." If the word was judged as "old," the participant needed to indicate which personal pronoun (I or he) had been paired with it. Otherwise, the word was judged as "new." The next trial started 500 ms after the response.

Procedure

Upon arrival to the lab, participants were given a brief overview of the study and signed the informed consent form. Then, participants completed the Beck Depression Inventory (M = 8.26, SD = 4.64), the State-Trait Anxiety Inventory (M_{state} = 34.99, SD_{state} = 9.02; M_{trait} = 40.54, SD_{trait} = 8.49), and the Rosenberg Self-esteem Scale (M = 27.91, SD = 2.76). Finally, the primary experimental task was started.

The primary experiment was conducted via E-prime 2.0. The learning phase was the same as in Experiment 1. For the test, the full set of 160 trait adjectives was randomly presented, one word at a time, on the computer screen (Figure 2). Participants were instructed to indicate whether the word had been presented by pressing the appropriate "d" or "k" key. When a trait adjective was judged as an "old" word, participants then needed to indicate which personal pronoun ("I" or "he") had been paired with it by pressing the appropriate "d" or "k" key. In the case of "new" judgments, 500 ms after the response, the next trial started. The assignment of keys to response alternatives was balanced across participants.

Results and Discussion

To disentangle item memory, source memory, and guessing biases, we applied the two-high threshold source memory (2HTSM) model (Bayen, Murnane, & Erdfelder, 1996; Bröder & Meiser, 2007; Vogt & Bröder, 2007). The 2HTSM model is depicted in Figure 3. There is one processing tree for each of six classes of words (i.e., Source I positive words, Source I negative words, Source He positive words, Source He negative words, new positive words, and new negative words). Source I positive (Source I negative, Source He positive, Source He negative) words will be detected as "old" with probability D_{Ip} (D_{In}, D_{Hp}, D_{Hn}). New positive (negative) words will be detected as "new" with probability D_{Np} (D_{Nn}). If old words are detected as old, their source can either be remembered (d_{Ip}, d_{In}, d_{Hp}, d_{Hn}) or not ($1 - d_{Ip}$, $1 - d_{In}$,$1 - d_{Hp}$,$1 - d_{Hn}$). In the latter case, there is a guessing probability a_p (a_n) of guessing "Source I" and $1 - a_p$ ($1 - a_n$) of guessing "Source He." If old words are undetected, they can nevertheless be guessed as "old," depending on bias b_p (b_n). Since the participant has to judge a source in this case, there is another guessing probability g_p (g_n) to guess "Source I" and $1 - g_p$ ($1 - g_n$) to guess "Source He." A "new" response only results for old words if they are neither detected nor guessed as old. If new words are not detected as new, they can be guessed as "old," and the source guessing probability g_p (g_n) also applies.

Considering that self-referential processing may promote participants' old-new recognition and the promotion may be influenced by words' valence, we let D_{Ip} and D_{In} vary freely. In addition, a fully restricted model (all D equal) did not fit the data, but leaving D_{Ip} and D_{In} free led to a very good model fit. Therefore, the final restrictions on parameters were as follows: $D_{Hp} = D_{Hn} = D_{Np} = D_{Nn}$; $a_p = g_p$; $a_n = g_n$.

The model-based analyses were conducted based on the aggregated data presented in Table 2. Parameter estimates and the goodness-of-fit index G^2 were obtained via the EM algorithm using the multiTree v046 (Moshagen, 2010). The parameter estimates and confidence intervals are shown in Table 3. The model fit the data, $G^2(1) = 0.06$, $p = .81$.

Using this model as a baseline model, model fit significantly decreased when d_{Ip}, d_{In}, d_{Hp}, and d_{Hn} were restricted to be equal, $\Delta G^2(3) = 37.91$, $p < .001$, suggesting an overall effect of learning condition on source memory. Then, we conducted pair-wise comparisons among parameters d_{Ip}, d_{In}, d_{Hp}, and d_{Hn}. The parameter d_{In} was significantly smaller than d_{Ip}, $\Delta G^2(1) = 36.68$, $p < .001$, suggesting that in the self-reference condition, participants had worse source memory for negative words than positive words. The parameter d_{Hn} is significantly larger than d_{Hp}, $\Delta G^2(1) = 7.98$,

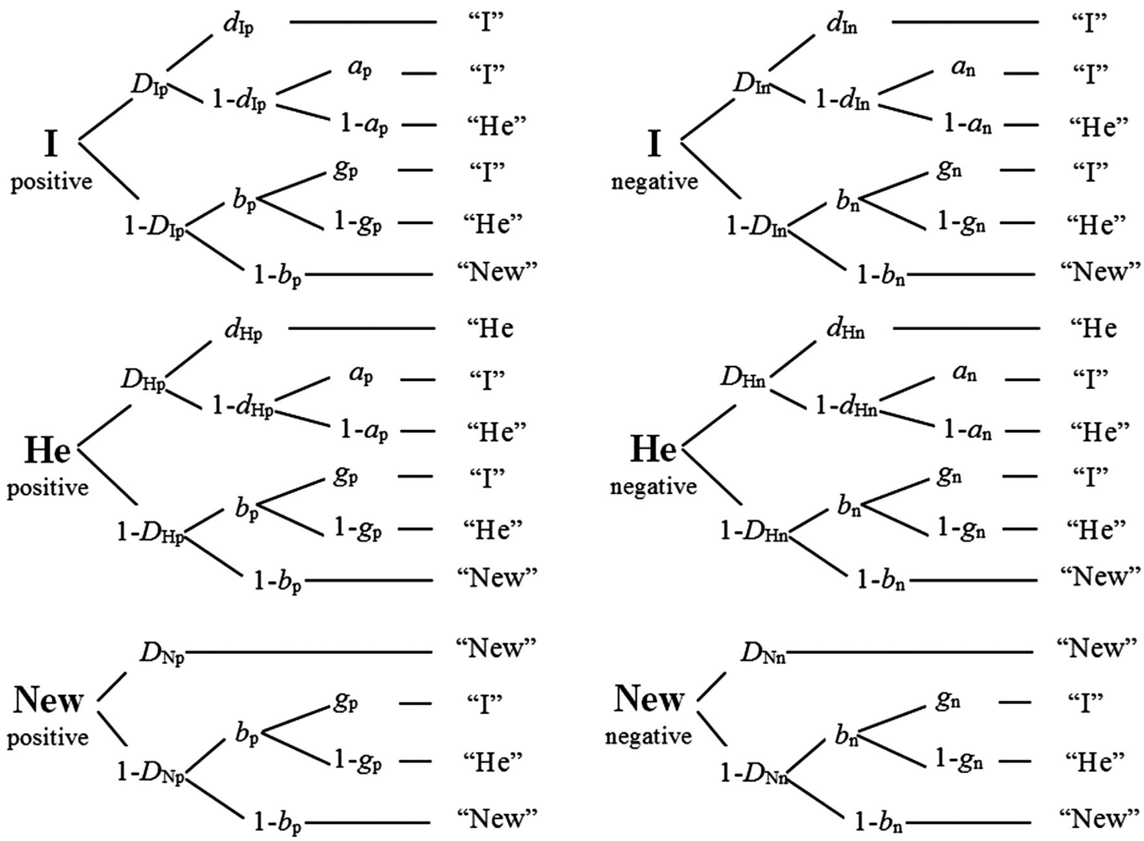

Figure 3. Processing tree depicting the 2HTSM model. The words are either from Source I, Source He, or they are new. $D_{Ip}/D_{In}/D_{Hp}/D_{Hn}$ = probabilities of detecting Source I/He words as old; $d_{Ip}/d_{In}/d_{Hp}/d_{Hn}$ = probabilities of remembering a word's Source; a_p/a_n = probabilities of guessing "Source I" when the source was forgotten; b_p/b_n = probabilities of guessing "old" if a word is not detected as old; g_p/g_n = probabilities of guessing "Source I" when a word was guessed as old; D_{Np}/D_{Nn} = probabilities of detecting a new word as new.

Table 2. Participants' behavioral performance in the testing phase in Experiment 2

	Participants' responses		
Actual source during learning	Attributed to "I"	Attributed to "He"	Do not remember
Source "I"			
Positive	922	165	313
Negative	713	310	377
Source "He"			
Positive	365	527	508
Negative	274	584	542
New words			
Positive	438	309	2,053
Negative	404	290	2,106

Note. Rows represent presentation during learning, columns denote the response of the participant, and cells contain raw frequencies.

$p < .01$, suggesting that in the other-reference condition, participants had better source memory for negative words than positive words. The parameter d_{In} is significantly smaller than d_{Hn}, $\Delta G^2(1) = 16.57$, $p < .001$, suggesting that for negative words, participants had worse source memory in the self-reference condition than in the other-reference condition.

For source guessing parameters, the parameter g_p was significantly larger than 0.5, $\Delta G^2(1) = 22.39$, $p < .001$; the parameter g_n was also significantly larger than 0.5,

Table 3. Parameter estimates and confidence intervals for the baseline model

Parameters	*M*	*SD*	95% CI
D_{Ip}	0.61	0.02	[.57–.65]
D_{In}	0.56	0.02	[.52–.60]
$D_{Nn} = D_{Np} = D_{Hp} = D_{Hn}$	0.37	0.01	[.35–.39]
d_{Ip}	0.80	0.04	[.72–.89]
d_{In}	0.36	0.06	[.24–.48]
d_{Hp}	0.52	0.06	[.40–.64]
d_{Hn}	0.75	0.06	[.64–.87]
b_p	0.42	0.01	[.40–.45]
b_n	0.39	0.01	[.37–.41]
$a_p = g_p$	0.59	0.02	[.55–.62]
$a_n = g_n$	0.58	0.02	[.55–.62]

Note. M = mean; *SD* = standard deviation; CI = confidence interval.

$\Delta G^2(1) = 18.81$, $p < .001$; and there was no significant difference between parameter g_p and g_n, $\Delta G^2(1) = 0.03$, $p = .87$, suggesting that there was a general guessing bias to attribute words (although judged as old, participants did not remember their source) to the self, but the guessing bias was not influenced by words' valence.

For old-new recognition parameters, the parameter D_{Ip} was significantly larger than D_{Hp}, $\Delta G^2(1) = 82.22$, $p < .001$; the parameter D_{In} was also significantly larger than D_{Hn}, $\Delta G^2(1) = 51.81$, $p < .001$; and there was no significant difference between parameter D_{Ip} and D_{In}, $\Delta G^2(1) = 3.39$, $p = .07$, suggesting that self-reference processing promoted participants' old-new recognition performance, but this promotion was not influenced by words' valence.

Although there was a general source guessing bias, this bias was not influenced by words' valence. Hence, source guessing bias is an unlikely candidate for the result in Experiment 1. Self-reference processing promoted old-new recognition performance, but this promotion was not influenced by words' valence. Hence, it is also unlikely that the attribution differences in Experiment 1 were caused by old-new recognition. Participants had inferior source memory in the negative self-referenced words learning condition compared to other learning conditions, suggesting that individuals selectively forget the connection between negative information and oneself.

General Discussion

The present study explored individuals' memory of the connection between positive or negative words and their referents. In the self-reference condition, the memory of the connection between negative words and the self was significantly worse than that between positive words and the self. For negative words, the memory of the connection between negative words and the self was significantly worse than that between negative words and the others. What could explain the inferior memory of the connection between negative information and oneself?

One possibility is the role of self-protection motivation in memory. Self-protection refers to the motivation for defending positive components of the self-image against harm (Sedikides, Skowronski, & Gaertner, 2004). Self-protection motivation not only guides people to avoid self-threatening information, but also influences people's memory of self-related information by processing self-threatening information in a shallow manner (Alicke & Sedikides, 2009; Sedikides & Green, 2009; Sedikides et al., 2016). As a result, the memory of self-threatening information is weakened. Autobiographical memory research indicates that unpleasant life experiences are recalled more poorly than pleasant life experiences (Ritchie et al., 2017; Skowronski & Walker, 2004; Skowronski, Betz, Thompson, & Shannon, 1991). Moreover, individuals selectively forget evaluations that have unfavorable implications for the self (Green et al., 2007; Newman et al., 2014; Pinter et al., 2011; Sedikides & Green, 2000, 2004; Sedikides et al., 2016). In the present study, to achieve the purpose of self-protection, the individual may avoid building up a connection with negative information by allocating less cognitive resources to the connection. Consequently, negative information becomes weakly connected with the self in memories.

Individuals may process self-related positive and negative information not only to different degrees, but also in different ways. Previous studies indicate that people have poor recall of self-threatening information compared to nonthreatening information (Newman, Nibert, & Winer, 2009; Newman et al., 2014; Sedikides & Green, 2000, 2004, 2009; Sedikides et al., 2004). However, they show equivalent recognition for threatening and nonthreatening information (Green et al., 2007). These findings imply that people process positive and negative self-related evaluations in different ways: they make assimilation-based judgments about positive evaluations ("they are suitable to describe me"), but contrast-based judgments about negative evaluations ("they are not suitable to describe me"). Therefore, self-related positive evaluations would be integrated into stored self-knowledge, but negative evaluations would be separated from stored self-knowledge. This claim is supported by a study in which participants received instructions designed to make them integrate some items of behavioral information, but separate other items. The integration instruction led to better information recall than did the separation instruction. However, the manipulation had no impact on levels of recognition (Pinter et al., 2011). In our experiments, there was no significant difference in recognition performance for positive and negative self-referenced words. This is consistent with the previous study that showed self-protectively neglected information can be retrieved (Green et al., 2007). Nonetheless, participants' source memory was extremely poor in the negative self-referenced words learning condition. This is probably because people integrate self-related positive information into stored self-knowledge but keep negative information separate.

When processing self-referenced information, participants' source memory is worse for negative information than positive information. When processing negative information, participants' source memory is worse for self-referenced information than other-referenced information. These results suggest that there is a self-serving bias in memory for the connection between valence information and the self.

Acknowledgment
The original materials used to conduct the research (including analysis code) are available to other researchers for purposes of replicating the procedure or reproducing the results. We thank Xiaoguang Wang for his help in data analysis.

Electronic Supplementary Materials
The electronic supplementary material is available with the online version of the article at https://doi.org/10.1027/1618-3169/a000409

ESM 1. Data (.csv)
Raw data of Experiment 1 (part 1).
ESM 2. Data (.csv)
Raw data of Experiment 1 (part 2).
ESM 3. Data (.csv)
Raw data of Experiment 2 (part 1).
ESM 4. Data (.csv)
Raw data of Experiment 2 (part 2).
ESM 5. Wordlist (.csv)
Stimulus material and word properties.

References

Alicke, M. D., & Sedikides, C. (2009). Self-enhancement and self-protection: What they are and what they do. *European Review of Social Psychology, 20*, 1–48. https://doi.org/10.1080/10463280802613866

Anderson, M. C., & Green, C. (2001). Suppressing unwanted memories by executive control. *Nature, 410*, 366–369. http://doi.org/10.1038/35066572

Baumeister, R. F., Dale, K., & Sommer, K. L. (2010). Freudian defense mechanisms and empirical findings in modern social psychology: Reaction formation, projection, displacement, undoing, isolation, sublimation, and denial. *Journal of Personality, 66*, 1081–1124. http://doi.org/10.1111/1467-6494.00043

Baumeister, R. F., & Ilko, S. A. (1995). Shallow gratitude: Public and private acknowledgement of external help in accounts of success. *Basic and Applied Social Psychology, 16*, 191–209. https://doi.org/10.1080/01973533.1995.9646109

Bayen, U. J., Murnane, K., & Erdfelder, E. (1996). Source discrimination, item detection, and multinomial models of source monitoring. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22*, 197–215. https://doi.org/10.1037/0278-7393.22.1.197

Beck, A. T., Ward, C. H., Mendelson, M., Mock, J., & Erbaugh, J. (1961). An inventory for measuring. *Archives of General Psychiatry, 4*, 561–571. https://doi.org/10.1001/archpsyc.1961.01710120031004

Bröder, A., & Meiser, T. (2007). Measuring source memory. *Zeitschrift für Psychologie/Journal of Psychology, 215*, 52–60. https://doi.org/10.1027/0044-3409.215.1.52

Conway, M. A., & Dewhurst, S. A. (1995). The self and recollective experience. *Applied Cognitive Psychology, 9*, 1–19. https://doi.org/10.1002/acp.2350090102

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

Gaertner, L., Sedikides, C., & Cai, H. (2012). Wanting to be great and better but not average: On the pancultural desire for self-enhancing and self-improving feedback. *Journal of Cross-Cultural Psychology, 43*, 521–526. https://doi.org/10.1177/0022022112438399

Gagnepain, P., Henson, R. N., & Anderson, M. C. (2014). Suppressing unwanted memories reduces their unconscious influence *via* targeted cortical inhibition. *Proceedings of the National Academy of Sciences of the United States of America, 111*, E1310–E1319. https://doi.org/10.1073/pnas.1311468111

Gould, R., Brounstein, P. J., & Sigall, H. (1977). Attributing ability to an opponent: Public aggrandizement and private denigration. *Sociometry, 40*, 254–261. https://doi.org/10.2307/3033532

Green, J. D., Sedikides, C., & Gregg, A. P. (2007). Forgotten but not gone: The recall and recognition of self-threatening memories. *Journal of Experimental Social Psychology, 44*, 547–561. https://doi.org/10.1016/j.jesp.2007.10.006

Gregg, A. P., Sedikides, C., & Gebauer, J. E. (2011). Dynamics of identity: Between self-enhancement and self-assessment. In S. J. Schwartz, K. Luyckx, & V. L. Vignoles (Eds.), *Handbook of identity theory and research* (Vol. 1, pp. 305–327). New York, NY: Springer. https://doi.org/10.1007/978-1-4419-7988-9_14

Hu, T., Zhang, D., & Yang, Z. (2015). The relationship between attributional style for negative outcomes and depression: A meta-analysis. *Journal of Social and Clinical Psychology, 34*, 304–321. https://doi.org/10.1521/jscp.2015.34.4.304

Kim, K., & Yi, D. J. (2013). Out of mind, out of sight: Perceptual consequences of memory suppression. *Psychological Science, 24*, 569–574. https://doi.org/10.1177/0956797612457577

Kwang, T., & Swann, W. B. Jr. (2010). Do people embrace praise even when they feel unworthy? A review of critical tests of self-enhancement versus self-verification. *Personality and Social Psychology Review, 14*, 263–280. https://doi.org/10.1177/1088868310365876

Leary, M. R. (2007). Motivational and emotional aspects of the self. *Annual Review of Psychology, 58*, 317–344. https://doi.org/10.1146/annurev.psych.58.110405.085658

Mezulis, A. H., Abramson, L. Y., Hyde, J. S., & Hankin, B. L. (2004). Is there a universal positivity bias in attributions? A meta-analytic review of individual, developmental, and cultural differences in the self-serving attributional bias. *Psychological Bulletin, 130*, 711–747. https://doi.org/10.1037/0033-2909.130.5.711

Moshagen, M. (2010). multiTree: A computer program for the analysis of multinomial processing tree models. *Behavior Research Methods, 42*, 42–54. https://doi.org/10.3758/BRM.42.1.42

Newman, L. S., Nibert, J. A., & Winer, E. S. (2009). Mnemic neglect is not an artifact of expectancy: The moderating role of defensive pessimism. *European Journal of Social Psychology, 39*, 477–486. https://doi.org/10.1002/ejsp.598

Newman, L. S., Sapolsky, M. S., Tang, Y., & Bakina, D. A. (2014). What's recalled depends on the nature of the recall procedure: The case of mnemic neglect. *Social Psychology, 45*, 93–102. https://doi.org/10.1027/1864-9335/a000164

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*, 612–624. https://doi.org/10.1521/soco.2011.29.5.612

Rogers, T. B., Kuiper, N. A., & Kirker, W. S. (1977). Self-reference and the encoding of personal information. *Journal of Personality and Social Psychology, 35*, 677–688. https://doi.org/10.1037/0022-3514.35.9.677

Rosenberg, M. (1979). *Rosenberg Self-Esteem Scale*. New York, NY: Basic Books.

Ritchie, T. D., Sedikides, C., & Skowronski, J. J. (2017). Does a person selectively recall the good or the bad from their personal past? It depends on the recall target and the person's favourability of self-views. *Memory, 25*, 934–944. https://doi.org/10.1080/09658211.2016.1233984

Sanjuán, P., & Magallares, A. (2014). Coping strategies as mediating variables between self-serving attributional bias and subjective well-being. *Journal of Happiness Studies, 15*, 443–453. https://doi.org/10.1007/s10902-013-9430-2

Sedikides, C. (1993). Assessment, enhancement, and verification determinants of the self-evaluation process. *Journal of Personality and Social Psychology, 65*, 317–338. https://doi.org/10.1037/0022-3514.65.2.317

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

Sedikides, C., & Green, J. D. (2004). What I don't recall can't hurt me: Information negativity versus information inconsistency as determinants of memorial self-defense. *Social Cognition, 22*, 4–29. https://doi.org/10.1521/soco.22.1.4.30987

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

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

Sedikides, C., Skowronski, J. J., & Gaertner, L. (2004). Self-enhancement and self-protection motivation: From the laboratory to an evolutionary context. *Journal of Cultural and Evolutionary Psychology, 2*, 61–79. https://doi.org/10.1556/JCEP.2.2004.1-2.4

Seidel, E., Eickhoff, S. B., Kellermann, T., Schneider, F., Gur, R. C., Habel, U., & Derntl, B. (2010). Who is to blame? Neural correlates of causal attribution in social situations. *Social Neuroscience, 5*, 335–350. https://doi.org/10.1080/17470911003615997

Skowronski, J. J., Betz, A. L., Thompson, C. P., & Shannon, L. (1991). Social memory in everyday life: Recall of self-events and other-events. *Journal of Personality and Social Psychology, 60*, 831–843. https://doi.org/10.1037/0022-3514.60.6.831

Skowronski, J. J., & Walker, W. R. (2004). How describing autobiographical events can affect autobiographical memories. *Social Cognition, 22*, 555–590. https://doi.org/10.1521/soco.22.5.555.50764

Spielberger, C. D., Gorsuch, R. L., Lushene, R., Vagg, P. R., & Jacobs, G. A. (1983). *Manual for the Stait-Trait Anxiety Inventory*. Palo Alto, CA: Consulting Psychologists Press.

Vogt, V., & Bröder, A. (2007). Independent retrieval of source dimensions: An extension of results by Starns and Hicks (2005) and a comment on the ACSIM measure. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33*, 443–450. https://doi.org/10.1037/0278-7393.33.2.443

Wang, D. F., & Cui, H. (2005). *Explorations of Chinese personality* (in Chinese). Beijing, China: Social Sciences Digest Press

Received August 15, 2017
Revision received March 7, 2018
Accepted March 7, 2018
Published online August 31, 2018

Yongyu Guo
School of Psychology
Nanjing Normal University
Nanjing, Jiangsu 210097
PR China
yyguo@njnu.edu.cn