Coping with Moral Self-Threat: A Moral Memory Bias Perspective

Authors: Wang Xiuxin, Shen Yifan, Wang Xiuxin

Date: 2022-01-14T11:32:48+00:00

Abstract

People sometimes engage in immoral behaviors, which may threaten their positive moral self-concept. To cope with such moral self-threat, individuals exhibit moral memory bias, characterized by impaired recall of immoral events or information that threaten the moral self. In recent years, researchers have provided empirical evidence for moral memory bias through autobiographical memory paradigms, game paradigms, perspective-taking paradigms, and self-reference paradigms. Further research has demonstrated that this moral memory bias may be motivated by the need to defend against moral self-threat. It is noteworthy that the manifestation of moral memory bias may be contingent on certain conditions. Future studies should broaden the research scope of moral memory bias, uncover its underlying cognitive mechanisms, and investigate its interplay with other strategies for coping with moral self-threat.

Full Text

How to Cope with Threats to the Moral Self? A Perspective on Moral Memory Bias

WANG Xiuxin, SHEN Yifan
School of Psychology, Qufu Normal University, Qufu 273165, China

Abstract

People sometimes behave unethically, which may threaten their positive moral self-concept. To cope with this moral self-threat, individuals exhibit moral memory bias, characterized by forgetting unethical events or information that threatens their moral self. Recent research using autobiographical memory paradigms, game paradigms, imagination paradigms, and self-reference paradigms has provided empirical support for moral memory bias. Studies further suggest that

this bias stems from the need to cope with moral self-threat. Notably, the existence of moral memory bias may require certain conditions. Future research should expand the scope of moral memory bias studies, reveal its cognitive mechanisms, and explore its relationship with other strategies for coping with moral self-threat.

Keywords: moral self, self-protective motivation, ethical dissonance, motivated forgetting

People sometimes behave unethically, which may threaten their positive moral self-concept (Li & Yu, 2013). How do individuals cope with such moral self-threat? Existing research has addressed this question from perspectives such as moral disengagement and self-serving justifications (see reviews by Wang et al., 2016; Wu & Bai, 2019). Recent studies have approached the issue from a memory perspective, finding that people tend to forget relevant events or information that threaten their moral self. This moral memory bias allows individuals to maintain a positive moral self-concept even after engaging in unethical behavior. This paper focuses on moral memory bias, reviewing research evidence and explanations, discussing the conditions for its occurrence, and proposing future research directions.

1.1 Evidence from the Autobiographical Memory Paradigm

Autobiographical memory refers to individuals' memories of their past experiences or events (Rubin, 1986). The autobiographical memory paradigm investigates participants' recall of autobiographical memories with different characteristics. For instance, Kouchaki and Gino (2016) asked participants to recall and write about moral, immoral, or neutral actions performed by themselves or others, then evaluate the clarity and vividness of these memories. Results showed that compared to moral or neutral events they had performed themselves, participants recalled their own immoral events as less clear and vivid. However, no significant differences were found in recall of events performed by others. Similarly, Ritchie et al. (2017) had participants write about numerous positive or negative events performed by themselves or others (e.g., being kind to others vs. gossiping), then recall them after a delay. Participants recalled more positive than negative events they had performed themselves, but showed no significant difference in recall of events performed by others.

The autobiographical memory paradigm uses participants' real-life experiences, offering high ecological validity, but it has several limitations. First, differences between memories are uncontrollable. Although participants recall moral or immoral events, these events may differ in vividness, recency, or frequency of rehearsal, which reduces internal validity. Second, researchers cannot typically participate in participants' autobiographical memories and thus cannot objectively assess memory accuracy. Consequently, when defining and measuring moral memory bias, researchers must rely on indicators such as vividness or recall frequency rather than the more common measure of memory accuracy.

1.2 Evidence from the Game Paradigm

In the game paradigm, researchers typically have participants complete a "game-like" task, then after an interval, ask them to recall relevant information from the "game." Compared to the autobiographical memory paradigm, the game paradigm creates observable real memories that allow researchers to assess memory accuracy. For example, Shu and Gino (2012) had participants complete a problem-solving task where better performance yielded greater rewards, but participants could lie about their performance to obtain extra payment. Notably, researchers presented participants with moral rules (e.g., "do not lie") before the task. Results showed that participants who lied were more likely to forget these moral rules, and this forgetting persisted even when monetary rewards were offered for correct recall.

Similarly, Kouchaki and Gino (2016) had participants play a coin-tossing game where they could lie for extra profit, then measured their memory of the game two weeks later, using their memory of dinner on the experimental day as a control. Participants who lied recalled the game as less vivid and with fewer associated thoughts or emotions compared to non-lying participants, while no significant difference was found between groups in recalling the dinner experience. Other researchers have used the dictator game to investigate this phenomenon. Carlson et al. (2020) had participants act as allocators deciding how much money to give to recipients, then after a distractor task, asked them to recall the amount given. Stingy participants (who gave less) recalled giving more than they actually had, whereas generous participants showed no such bias (see also Tasimi & Johnson, 2015). This effect persisted even with monetary incentives for accurate recall. Saucet and Villeval (2019) presented participants with two allocation options: a selfish option where the allocator received more than the recipient, and an altruistic option where the allocator received less. After participants made their choice and completed a distractor task, they were asked to recall the amount given to the recipient. Notably, the original allocation schemes were presented but the amount given to the recipient in the chosen option was hidden. Results showed that recall accuracy for selfish allocations was significantly lower than for altruistic allocations.

The game paradigm creates observable real memories that allow researchers to assess memory accuracy, making it superior to the autobiographical memory paradigm in this regard. However, in this paradigm, researchers classify participants as moral or immoral based on their performance in the game. "Moral" or "immoral" status is based on participants' choices rather than being manipulated by researchers, which somewhat reduces internal validity.

1.3 Evidence from the Imagination Paradigm

In the imagination paradigm, researchers typically describe a moral or immoral action to participants and ask them to imagine it from first-person or third-person perspectives. For example, Kouchaki and Gino (2016) described scenar-

ios of cheating or not cheating on an exam and asked participants to imagine them from either perspective. Four days later, participants evaluated these memories. Results showed that when imagining from a first-person perspective, participants recalled cheating scenarios as less clear and vivid than non-cheating scenarios, but no significant difference was found when imagining from a third-person perspective.

Kouchaki and Gino (2016) also had participants imagine dishonest or honest scenarios, then measured memory accuracy one week later. They found that participants who imagined dishonest scenarios showed poorer memory accuracy. However, Stanley et al. (2018) replicated this study and found no significant differences in memory accuracy between conditions, a result that persisted even when considering other scenarios and controlling for differences in participants' imagination vividness. Stanley et al. (2018) argued for distinguishing between phenomenological features of memory (such as vividness and clarity) and accuracy. Phenomenological features are not perfectly correlated with accuracy —vivid or clear memories are not necessarily accurate. In the imagination paradigm, moral memory bias may exist only at the phenomenological level, not at the accuracy level.

Although the imagination paradigm offers high internal validity, inconsistent findings warrant further investigation. Additionally, the imagination paradigm can be considered "virtual autobiographical memory," while the game paradigm represents "laboratory autobiographical memory." In other words, whether using autobiographical memory, game, or imagination paradigms, the focus is on episodic memory for events happening to oneself. Does moral memory bias also occur in semantic memory? Researchers have investigated this question using the self-reference paradigm.

1.4 Evidence from the Self-Reference Paradigm

The self-reference paradigm typically consists of encoding and test phases (Liu & Zhu, 2002). In the encoding phase, participants are presented with words and asked to process them either self-referentially (e.g., "To what extent does this word describe you?") or other-referentially (e.g., "To what extent does this word describe others?") or semantically (e.g., "Does this word mean the same as XX?"). After encoding and a distractor task, participants enter the test phase where they are presented with both old and new words and must judge whether each word is new, self-referential, other-referential, or semantic-referential. Data analysis typically examines two indicators: recognition memory (assessing whether participants can distinguish new from old words) and source memory (assessing whether participants can correctly identify whether a word was self-, other-, or semantic-referential among non-new words).

Zhang et al. (2018) used positive and negative words as encoding materials, asking participants to make self- and other-referential judgments about whether the words could describe themselves or others. These words included many moral

terms (e.g., selfish vs. unselfish). Results showed that in the self-reference condition, source memory for negative words was significantly worse than for positive words, whereas no significant difference was found in the other-reference condition. Moreover, compared to other-reference, source memory for negative words was worse in self-reference, while source memory for positive words showed no significant difference between self- and other-reference. Rowell and Jaswal (2021) presented participants with three types of morally relevant action words (friendly/polite, mean/impolite, and neutral; e.g., "opening a door for someone," "pushing someone," "catching a ball") and obtained similar results.

These studies suggest that people more easily forget or less frequently recall associations between themselves and immoral words, demonstrating moral memory bias. Notably, these studies found moral memory bias in source memory but only found better memory for self-referential words than other-referential words in recognition memory. This may be because recognition memory more reflects depth of processing, whereas source memory more reflects organizational processing of relationships between words (Durbin et al., 2017; Rowell & Jaswal, 2021). Self-referential processing promotes deep processing of words (Liu & Zhu, 2002; Turk et al., 2008), which may lead to better recognition memory for self-referential words. Self-referential processing also promotes organizational processing (Liu & Zhu, 2002; Klein & Loftus, 1988), but organizational processing is more susceptible to the compatibility between words and self-schema. Immoral words may be less compatible with self-schema, resulting in poorer source memory for negative words (Durbin et al., 2017).

2 Explanations for Moral Memory Bias

Researchers primarily explain moral memory bias from the perspective of moral self-threat. People generally hold positive moral self-concepts, while occasional immoral events may threaten these concepts. To cope with moral self-threat, people may exhibit moral memory bias. This bias is mainly manifested as forgetting events or information that threaten the moral self, without corresponding memory enhancement for moral events that strengthen the moral self (Rowell & Jaswal, 2021). This pattern aligns with self-perceptions: Klein and Epley (2016, 2017) found that people do not necessarily see themselves as better than others, just less evil. This suggests that moral memory bias primarily serves self-protective motivation, providing indirect evidence for the moral self-threat explanation.

In this sense, moral memory bias can be subsumed under the mnemic neglect effect, which posits that self-protective motivation leads people to ignore self-threatening negative information (see review by Sedikides et al., 2016). Moral memory bias is similarly based on self-protective motivation but focuses specifically on the moral self, highlighting selective forgetting of events or information that threaten moral self-concept.

Direct evidence also supports the moral self-threat explanation. Research

shows that lying participants have lower moral self-evaluations, and this moral self-threat mediates the relationship between lying and moral memory bias (Kouchaki & Gino, 2016). Moral self-threat can also be understood through the lens of ethical dissonance—the discrepancy between a positive moral self and unethical behavior (Barkan et al., 2015). Studies indicate that lying participants indeed experience this ethical dissonance, which mediates the relationship between lying and moral memory bias (Kouchaki & Gino, 2016).

Research further demonstrates that moral memory bias disappears when moral self-threat is absent. For example, this bias only occurs in self-relevant contexts, with no memory differences for other-relevant events or information (Kouchaki & Gino, 2016; Rowell & Jaswal, 2021). Moreover, moral memory bias primarily appears in individuals who have committed immoral acts, such as liars (Shu & Gino, 2012), allocators who make more selfish choices (Saucet & Villeval, 2019), or allocators who violate their own fairness principles (Carlson et al., 2020). In dictator games, when allocations are randomly assigned rather than chosen by the allocator, the allocator bears no responsibility and experiences no moral self-threat, eliminating moral memory bias (Saucet & Villeval, 2019; Carlson et al., 2020).

Furthermore, moral memory bias may stem not only from moral self-threat experienced after committing immoral acts but also from anticipated moral self-threat before such acts. This corresponds to Shalvi et al.' s (2015) distinction between experienced threat to the moral self and anticipated threat to the moral self. Kouchaki and Gino (2016) suggested that moral memory bias may alleviate anticipated moral self-threat and subsequently lead to further unethical behavior. Galeotti et al. (2020) also noted that moral memory bias can help individuals achieve self-forgiveness to rationalize future unethical actions.

Beyond motivational mechanisms, researchers have explored cognitive mechanisms of moral memory bias. First, does moral memory bias truly involve forgetting? One possibility is genuine forgetting, but another is that participants deliberately misreport despite remembering. Researchers can reduce potential misreporting by rewarding accurate recall. Studies show that memory bias persists even when rewards encourage accurate recall (Carlson et al., 2020; Saucet & Villeval, 2019), suggesting genuine forgetting.

Second, if genuine forgetting occurs, what is the process? Researchers propose two possibilities: biased encoding, where individuals encode moral events more deeply and immoral events more shallowly; and retrieval suppression, where individuals actively suppress retrieval of immoral events or information (Anderson & Hanslmayr, 2014). Shu and Gino (2012) found that moral memory bias appeared after but not before unethical behavior, suggesting no difference in encoding depth. They further found that lying reduced the accessibility of moral words, implying that moral memory bias may result from active retrieval suppression. However, some research suggests that motivated forgetting may occur during encoding (e.g., Rigney et al., 2021), warranting further investigation. Neuroimaging studies can provide more reliable dependent measures; for

example, retrieval suppression reduces memory-related brain activity (Hu et al., 2015; 2017), and suppression-induced forgetting increases activation in the right dorsolateral prefrontal cortex (dlPFC) and middle frontal gyrus (MFG) (Guan & Wang, 2021).

Moral memory bias may also occur during storage. The constructive view of memory suggests that memory changes during storage under the influence of individual experience and psychological schemas (Mo & Gao, 2011). Individuals generally hold relatively positive moral self-concepts (Tappin & Mckay, 2017). Influenced by this self-schema, individuals may forget immoral events or information. This constructive nature of memory may require time; longer intervals provide more "room for manipulation," leading to greater memory bias. Tasimi and Johnson (2015) found that moral memory bias indeed increases over time.

3 Conditions for the Existence of Moral Memory Bias

First, this phenomenon may only occur when people intentionally commit immoral acts. Researchers distinguish between intentional and unintentional unethical behavior: the former refers to deliberate, conscious immoral actions, while the latter refers to unconscious immoral actions (e.g., implicit biases; Gino, 2015). Intentional unethical behavior is often profit-driven, such as lying for greater rewards (Mazar et al., 2008) or making unfair allocations (Otto & Bolle, 2015; Rode & Menestrel, 2011). When committing intentional unethical acts, people should be aware that their behavior violates moral standards, experience moral self-threat, and consequently exhibit moral memory bias. When committing unintentional unethical acts, people may not experience moral self-threat and thus show no moral memory bias. Unless otherwise specified, all immoral behavior discussed in this paper refers to intentional unethical behavior.

Second, this phenomenon may only occur when people commit relatively minor immoral acts. Self-concept maintenance theory posits that people engage in some degree of unethical behavior for self-interest while employing psychological mechanisms that allow them to maintain a positive moral self-concept (Mazar et al., 2008). These psychological mechanisms represent strategies for coping with moral self-threat. Notably, these unethical acts are not completely immoral but only moderately so. For example, in allocation decisions, most participants do not keep all the money for themselves but give some to others (Otto & Bolle, 2015). These less severe immoral acts make successful coping with moral self-threat possible. One might consider these less severe immoral acts as behavioral strategies for coping with moral self-threat, while moral memory bias represents a psychological strategy. People may understand the subtle relationship between behavioral and psychological strategies and limit their behavioral strategies to what their psychological strategies can handle. However, when moral violations are severe, these psychological strategies may be "overwhelmed," and moral memory bias may disappear. Existing research provides some evidence: for instance, 46% of violent offenders experience intrusive memories, and their recall of these criminal experiences is more detailed and vivid (Evans et al., 2007a;

2007b). Similarly, people recall severe moral violations more frequently and with greater detail and vividness (Huang et al., 2020).

What constitutes a "not severe" immoral act? First, social norms serve as an important criterion for judging severity. Social norms convey expectations for appropriate or moral behavior in specific social contexts, and people often use them to evaluate individual behavior (Mcdonald & Crandall, 2015). Not severe immoral acts are those that do not seriously violate social norms, such as lying about one' s performance for economic reward (Mazar et al., 2008). Severe immoral acts seriously violate social norms, such as violent crimes mentioned above. Second, conforming to a moral threshold may provide a specific definition of "not severe." People generally do not pursue saint-like morality but adhere to "ordinary person morality" where they do some good and some bad (Gan, 2017). Building on this, the moral threshold model further proposes that people' s behavior has a bottom line—the moral threshold. Behavior typically does not cross this threshold, and acts conforming to the moral threshold allow individuals to maintain a positive moral self-concept (Zlatev et al., 2020). When immoral acts do not cross the moral threshold, individuals can use psychological strategies like moral memory bias to maintain their moral self-concept; once crossed, moral memory bias disappears.

4 Discussion and Future Directions

Existing research has accumulated evidence for moral memory bias and its moral self-threat mechanism, but inconsistent findings exist, and other mechanisms may be involved. Future research should attend to the conditions under which moral memory bias occurs and further expand studies on moral memory bias in both breadth and depth.

4.1 Expanding Research on Moral Memory Bias

First, researchers should identify potential moderating variables to integrate inconsistent findings. For example, using the dictator game paradigm, Saucet and Villeval (2019) found that recall accuracy for amounts given to recipients was higher in altruistic than selfish choices, with no bias in the amounts recalled. In contrast, Carlson et al. (2020) found that selfish participants recalled giving more to recipients than they actually had. This inconsistency may result from different memory retrieval cues. Saucet and Villeval (2019) presented participants with both selfish and altruistic options, then later presented the original allocation schemes as cues, providing strong retrieval support that may have reduced memory bias in amount recall while only showing higher recall accuracy for altruistic choices. Carlson et al. (2020), however, asked participants to recall amounts without cues, potentially leading to greater memory bias. Notably, the influence of retrieval cues on memory may operate not only at the cognitive level but also at the motivational level. More retrieval cues leave less "room for manipulation" in subjective memory construction, whereas fewer cues

in ambiguous settings allow participants to fully exercise subjective construction, showing more motivated forgetting. Investigating these issues can reveal situational differences in how people use moral memory bias to cope with moral self-threat.

Second, researchers should examine cultural differences in moral memory bias and its generalization. Current researchers generally believe that moral memory bias stems from self-protective motivation to cope with moral self-threat. Cross-cultural research shows that individuals in Eastern cultures have stronger self-protective motivation than those in Western cultures (Sedikides et al., 2015). Does moral memory bias manifest more strongly in Eastern cultural groups? Additionally, moral memory bias generalizes to others; for example, Ritchie et al. (2017) found that individuals forget negative events performed by people they like. To whom does this generalization extend? From the moral self-threat perspective, answering this requires understanding for which others' unethical behavior individuals feel responsible—that is, which others' behaviors are incorporated into one' s own moral self-concept. Cultural differences exist in the extent to which people incorporate others into their self-concept; for example, Zhu et al. (2007) found that Chinese incorporate their mothers into their self-concept while Westerners do not. Investigating these issues can reveal cultural differences in using moral memory bias to cope with moral self-threat.

Third, researchers should further expand the scope of moral memory bias research. Current studies primarily focus on lying and allocation contexts, but unethical behavior also exists in other domains such as harm and loyalty (Hofmann et al., 2014). For example, in moral dilemmas, people may harm a few to save many (Liu & Liao, 2021). In conflict-of-interest advice-giving situations, advisors may give biased advice for self-interest (Barneron & Yaniv, 2020). Does moral memory bias exist in these domains? Moreover, does the degree of moral memory bias differ across moral domains? Investigating these questions can better answer how people use moral memory bias to cope with moral self-threat.

4.2 Deepening Investigation of Moral Memory Bias Mechanisms

Researchers often view moral self-threat as an internal motivation, believing that people want to maintain their moral image in their own eyes (Mazar et al., 2008). Similarly, some researchers call moral memory bias a "self-impression management strategy," where people do not want to see themselves as immoral (Saucet & Villeval, 2019). However, people want to maintain not only their self-image but also their image in others' eyes (Dana et al., 2007). This impression management motivation to maintain one' s image before others also influences psychological and behavioral performance; for example, people make more utilitarian/deontological choices in moral dilemmas to appear competent/warm (Rom & Conway, 2018). People may exhibit moral memory bias not only to cope with moral self-threat but also to maintain their moral image before others. Although difficult to completely separate, researchers could use public/private settings to explore their relative contributions.

4.3 Exploring Relationships Between Moral Memory Bias and Other Strategies for Coping with Moral Self-Threat

To some extent, moral memory bias represents a functional approach to memory research (Li & Guo, 2009), focusing on its role in coping with moral self-threat. Notably, beyond moral memory bias, people have many other strategies for coping with moral self-threat, such as moral disengagement and self-justification mentioned earlier. What is the relationship between moral memory bias and these strategies? This is a difficult but valuable question. One possibility is that moral memory bias and these strategies complement each other and work together. For example, Stanley and Brigard (2019) suggested that moral memory bias and psychological distancing are complementary: people selectively forget immoral events that threaten their moral self, and when these events cannot or are not forgotten, people may psychologically distance themselves by believing they have changed substantially since those events. Another possibility is that other strategies may create conditions for moral memory bias to emerge. After committing immoral acts, people may use moral disengagement and self-justification to lower their moral standards and reframe their evaluation of the event, making it easier to blur the immoral behavior and exhibit moral memory bias. This warrants future research attention for a more comprehensive understanding of how people cope with moral self-threat.

Acknowledgments: We thank the anonymous reviewers for their constructive comments, Dr. Zhang Yikang from Maastricht University and Dr. Sun Qian from East China Normal University for reviewing the English abstract, and Ge Shuchun and Li Fang for their critical reading of the manuscript.

References

Gan, S. P. (2017). The scale of ordinary person morality. *Morality and Civilization*, (3), 30–36.

Guan, X. X., & Wang, H. B. (2021). Suppression-induced forgetting and its neural mechanisms. *Advances in Psychological Science*, 29(4), 665–676.

Li, H. H., & Yu, J. (2013). Moral self-threat and its coping mechanisms. *Journal of Guangxi Normal University (Philosophy and Social Sciences Edition)*, 49(1), 147–153.

Liu, C. J., & Liao, J. Q. (2021). Paradigm evolution and theoretical value of moral dilemma research. *Advances in Psychological Science*, 29(8), 1–13.

Liu, X. M., & Zhu, Y. (2002). The self-reference effect in memory. *Advances in Psychological Science*, 10(2), 121–126.

Li, J. G., & Guo, X. Y. (2009). The functional approach to memory research. *Advances in Psychological Science*, 17(5), 923–930.

Mo, Y. T., & Gao, S. C. (2011). A social constructionist interpretation of memory. *Psychological Research*, 4(3), 21–27.

Wang, J., Wu, M. Z., & Sun, X. L. (2016). Self-regulation strategies for ethical dissonance. *Psychological Science*, 39(6), 1473-1478.

Wu, L., & Bai, Z. Q. (2019). Why does unethical behavior easily repeat? A psychological explanation. *Journal of Central China Normal University (Humanities and Social Sciences Edition)*, 58(2), 184-192.

Anderson, M. C., & Hanslmayr, S. (2014). Neural mechanisms of motivated forgetting. *Trends in Cognitive Sciences*, 18(6), 279-292.

Barkan, R., Ayal, S., & Dan, A. (2015). Ethical dissonance, justifications, and moral behavior. *Current Opinion in Psychology*, 6, 157-161.

Barneron, M., & Yaniv, I. (2020). Advice-giving under conflict of interest: Context enhances self-serving behavior. *Journal of Experimental Social Psychology*, 91, 104046.

Carlson, R. W., MA Maréchal, Oud, B., Fehr, E., & Crockett, M. J. (2020). Motivated misremembering of selfish decisions. *Nature Communications*, 11, 2100.

Dana, J., Weber, R. A., & Kuang, J. X. (2007). Exploiting moral wiggle room: experiments demonstrating an illusory preference for fairness. *Economic Theory*, 33(1), 67-80.

Durbin, K. A., Mitchell, K. J., & Johnson, M. K. (2017). Source memory that encoding was self-referential: The influence of stimulus characteristics. *Memory*, 25(9), 1191-1200.

Evans, C., Ehlers, A., Mezey, G., & Clark, D. M. (2007a). Intrusive memories and ruminations related to violent crime among young offenders: Phenomenological characteristics. *Journal of Traumatic Stress*, 20(2), 183-196.

Evans, C., Ehlers, A., Mezey, G., & Clark, D. M. (2007b). Intrusive memories in perpetrators of violent crime: Emotions and cognitions. *Journal of Consulting and Clinical Psychology*, 75(1), 134-144.

Galeotti, F., Saucet, C., & Villeval, M. C. (2020). Unethical amnesia responds more to instrumental than to hedonic motives. *Proceedings of the National Academy of Sciences*, 117(41), 25423-25428.

Gino, F. (2015). Understanding ordinary unethical behavior: Why people who value morality act immorally. *Current Opinion in Behavioral Sciences*, 3, 107-111.

Hu, X., Bergstrom, Z. M., Bodenhausen, G. V., & Rosenfeld, J. P. (2015). Suppressing unwanted autobiographical memories reduces their automatic influences. *Psychological Science*, 26(7), 1098-1106.

Hu, X., Bergstrm, Z. M., Gagnepain, P., & Anderson, M. C. (2017). Suppressing unwanted memories reduces their unintended influences. *Current Directions in Psychological Science*, 26(2), 197-206.

Huang, S., Stanley, M. L., & De Brigard, F. (2020). The phenomenology of remembering our moral transgressions. *Memory and Cognition*, 48(2), 277–286.

Hofmann, W., Wisneski, D. C., Brandt, M. J., & Skitka, L. J. (2014). Morality in everyday life. *Science*, 345(6202), 1340–1343.

Klein, N., & Epley, N. (2016). Maybe holier, but definitely less evil, than you: Bounded self-righteousness in social judgment. *Journal of Personality and Social Psychology*, 110(5), 660–674.

Klein, N., & Epley, N. (2017). Less evil than you: Bounded self-righteousness in character inferences, emotional reactions, and behavioral extremes. *Personality and Social Psychology Bulletin*, 43(8), 1202–1212.

Klein, S. B., & Loftus, J. (1988). The nature of self-referent encoding: The contributions of elaborative and organizational processes. *Journal of Personality and Social Psychology*, 55(1), 5–11.

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

Mazar, N., Amir, O., & Ariely, D. (2008). The dishonesty of honest people: A theory of self-concept maintenance. *Journal of Marketing Research*, 45(6), 633–644.

Mcdonald, R. I., & Crandall, C. S. (2015). Social norms and social influence. *Current Opinion in Behavioral Sciences*, 3, 147–151.

Otto, P. E., & Bolle, F. (2015). Exploiting one' s power with a guilty conscience: An experimental investigation of self-serving biases. *Journal of Economic Psychology*, 51, 79–89.

Rigney, A.E., Schnyer, D.M., Hu, X., & Beer, J.S. (2021). Mechanisms of a spotless self-image: Navigating negative, self-relevant feedback. *Self and Identity*, 20(8), 1057–1076.

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(8), 1047–1058.

Rode, J., & Menestrel, M. L. (2011). The influence of decision power on distributive fairness. *Journal of Economic Behavior and Organization*, 79(3), 246–255.

Rom, S. C., & Conway, P. (2018). The strategic moral self: Self-presentation shapes moral dilemma judgments. *Journal of Experimental Social Psychology*, 74, 24–37.

Rowell, S. F., & Jaswal, V. K. (2021). I remember being nice: Self-enhancement memory bias in middle childhood. *Memory*, 29(2), 261–269.

Rubin, D. C. (1986). *Autobiographical memory.* New York, NY: Cambridge University Press.

Saucet, C., & Villeval, M. C. (2019). Motivated memory in dictator games. *Games and Economic Behavior*, 117, 250–279.

Sedikides, C., Gaertner, L., & Cai, H. (2015). On the panculturality of self-enhancement and self-protection motivation: The case for the universality of self-esteem. *Advances in Motivation Science*, 2, 185–241.

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.

Shalvi, S., Gino, F., Barkan, R., & Ayal, S. (2015). Self-serving justifications: Doing wrong and feeling moral. *Current Directions in Psychological Science*, 24(2), 125–130.

Shu, L. L., & Gino, F. (2012). Sweeping dishonesty under the rug: How unethical actions lead to forgetting of moral rules. *Journal of Personality and Social Psychology*, 102(6), 1164–1177.

Stanley, M. L., & De Brigard, F. (2019). Moral memories and the belief in the good self. *Current Directions in Psychological Science*, 28(4), 387–391.

Stanley, M. L., Yang, B. W., & De Brigard, F. (2018). No evidence for unethical amnesia for imagined actions: A failed replication and extension. *Memory and Cognition*, 46, 787–795.

Tappin, B. M., & McKay, R. T. (2017). The illusion of moral superiority. *Social Psychological and Personality Science*, 8(6), 623–631.

Tasimi, A., & Johnson, M. K. (2015). A self-serving bias in children's memories? *Journal of Experimental Psychology: General*, 144(3), 528–533.

Turk, D. J., Cunningham, S. J., & Macrae, C. N. (2008). Self-memory biases in explicit and incidental encoding of trait adjectives. *Consciousness and Cognition*, 17(3), 1040–1045.

Zlatev, J. J., Kupor, D. M., Laurin, K., & Miller, D. T. (2020). Being "Good" or "Good Enough" : Prosocial Risk and the Structure of Moral Self-Regard. *Journal of Personality and Social Psychology*, 118(2), 242–253.

Zhang, Y., Pan, Z., Li, K., & Guo, Y. (2018). Self-serving bias in memories: Selectively forgetting the connection between negative information and the self. *Experimental Psychology*, 65(4), 236–244.

Zhu, Y., Zhang, L., Fan, J., & Han, S. H. (2007). Neural basis of cultural influence on self-representation. *NeuroImage*, 34, 1310–1316.

Note: Figure translations are in progress. See original paper for figures.