Routledge
Taylor & Francis Group

Psychological Inquiry

An International Journal for the Advancement of Psychological Theory

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/hpli20

Moral Memories and Identity Protection

Felipe De Brigard & Matthew L. Stanley

To cite this article: Felipe De Brigard & Matthew L. Stanley (2021) Moral Memories and Identity Protection, Psychological Inquiry, 32:4, 240-246, DOI: 10.1080/1047840X.2021.2004817

To link to this article: https://doi.org/10.1080/1047840X.2021.2004817

Published online: 20 Jan 2022.

PSYCHOLOGICAL INQUIRY
2021, VOL. 32, NO. 4, 240–246
https://doi.org/10.1080/1047840X.2021.2004817

Moral Memories and Identity Protection

Felipe De Brigard and Matthew L. Stanley

Department of Psychology and Neuroscience, Center for Cognitive Neuroscience, Duke University, Durham, North Carolina, USA

In 1998, Gilbert and Wilson et al. coined the term "psychological immune system" to refer to the set of cognitive mechanisms that help individuals fend off psychological discomfort and undesirable negative affect (Gilbert et al. 1998). Although, as they themselves acknowledged, this idea had been suggested previously in the literature (Freud, 1936; Vaillant, 1993), they utilized the term to explain and understand a number of different phenomena—including, of course, biases in affective forecasting (Gilbert, 2006). Gilbert, though, did not mean for the notion of a psychological immune system to be taken literally. A few years after the publication of that seminal paper, in an interview published by *The New York Times*, Gilbert explicitly stated that he and Wilson meant for the term to be interpreted metaphorically: "We've used the metaphor of the 'psychological immune system' –it's just a metaphor, but not a bad one for that system of defenses that helps you feel better when bad things happen." (Gertner, 2003). The claim that our mind is furnished with a psychological immune system was, therefore, offered as an attractive and useful strategy for explaining and understanding diverse psychological phenomena, and the interpretation of which was meant to be merely figurative.

Gilbert's ontological hesitation does not appeal to Sedikides, who has written an intriguing piece inviting us to think of the psychological immune system in a literal sense: as an actual, evolved set of cognitive mechanisms and operations whose adaptive purpose is to protect our sense of personal identity (Sedikides, this issue). The proposal builds heavily upon a series of connections drawn from features of our biological immune system and features of our putative psychological immune system. As a result, it comprises a large number of moving parts, some of which stand on shakier ground than others, and some of which leave us with more questions than they seem to answer. For instance, some of the evidence Sedikides adduces in support of his view comes from the fact that certain psychological tendencies and biases are conducive to beneficial behaviors for the organism. Since such individual benefits are taken to be adaptive, then the conclusion that the system that brought them about must have evolved for said purpose—i.e., psychological homeostasis—seems ineluctable. Unfortunately, the jump from "beneficial to me" to "selected for" or "having the function of" is often an unwarranted line of reasoning (Garson, 2016). One can easily engage in behaviors that are beneficial for oneself, but those behaviors can simultaneously be not-adaptive for organisms like us, in the sense of conferring evolutionary advantages. When psychologists use the term 'adaptive' they normally mean something like 'non-detrimental for the organism', which is not identical to the biologists' sense of 'adaptive'—meaning the organism's propensity toward increased fitness in a local environment—which is the sense needed to underwrite functional and evolutionary claims.

These concerns aside, the idea that many of our cognitive faculties operate *as if* they were safeguarding our sense of self through time is attractive and definitively worth exploring in detail. Yet, we believe that it is worth further examining whether we should take the '*as if*' literally, as Sedikides insinuates, rather than metaphorically, as Gilbert and Wilson originally advised. Because, after all, there seems to be a rather evident and critical disanalogy between the biological immune system and the alleged psychological immune system, the existence of which Sedikides wants us to believe in. On the one hand, most of the components of our biological immune system—whether innate or adaptive—are specialized and perform specific roles in contributing to the overall function of (and lets agree with Sedikides here) preserving homeostasis. Leucocytes, innate lymphoid cells, B cells, T cells, and most other components of the biological immune system evolved to play very specific, narrow roles. On the other hand, we have no evidence whatsoever to the effect that there are specialized cognitive processes whose sole function in our cognitive economy is to help to protect our self-concept. Instead, what seems more likely is that we *re*deploy cognitive processes that are typically used for other purposes—such as perception, memory, imagination, and the like—in the service of preserving and protecting certain features of our self-concept. If so, then the conclusion we ought to draw is not that there actually is—ontologically speaking—a psychological immune system, as suggested by Sedikides, but rather that our cognitive processes often behave *as if* they were a psychological immune system—but they aren't, really, no matter how appealing analogies between the biological and psychological immune systems may seem.

In this brief commentary, we argue in favor of this alternative conclusion. We focus primarily on a specific aspect of our personal identity, namely our *moral* identity. Over the course of the past decade, there has been rapidly growing

CONTACT Felipe De Brigard felipe.debrigard@duke.edu Department of Psychology and Neuroscience, Center for Cognitive Neuroscience, Duke University, Durham, NC, USA.

interest in the moral features of personal identity across different domains of psychological research. This has produced a wealth of empirical findings—obtained using complementary methods and approaches—that help to support the position that basic cognitive processes are more likely to be redeployed to preserve and protect a desirable self-concept. As such, we first suggest that our pronounced tendency to think of ourselves as morally upstanding individuals even in the face of counterevidence is precisely the kind of psychological phenomena that an alleged psychological immune system is postulated to explain. But then we argue that such psychological phenomena could just as well be accounted for by the same memory processes by means of which we explain other ordinary mnemonic phenomena. As a result, we conclude that it is unnecessary to postulate specialized mechanisms of an alleged psychological immune system to account for at least one kind of phenomena for which such a system is currently proposed. Ultimately, we suggest that this redeployment account can inspire and guide many potentially fruitful avenues of future research.

The moral self-concept

Generally speaking, healthy adults are motivated to maintain a positive self-concept (Alicke & Sedikides, 2009; Baumeister, 1998). They go to great lengths to build up a positive self-concept (known as self-enhancement) and to avoid a negative one (known as self-protection) (Alicke & Govorun, 2005; Sedikides & Gregg, 2008). But what specifically is most important to people when people attempt to build up a positive self-concept and to avoid a negative one? Many traits and qualities are positive, but presumably, not all positive traits and qualities are equally important in constructing a positive self-concept. Several different lines of research now point to the same answer as to what is particularly important in constructing this self-concept: *moral* traits and qualities.

Many theorists have suspected that people tend to think of themselves as fundamentally and truly morally good, and that the importance they give to morality in shaping up the persons they think or like to be is predictive of good moral behavior (Bandura, 1999; Kohlberg, 1971). The pioneering work of Aquino and Reed (2002) on *moral identity* corroborated this suspicion, as they not only showed that the vast majority of their participants thought of themselves as morally good, but also that the importance they give moral features in their self-concept was actually predictive of subsequent prosocial behavior. Further developments show that the construct of moral identity, understood as an aspect of one's self-concept, is predictive of other moral behavior, including intentions to volunteer, actual volunteerism, a sense of obligation toward others, and social engagement, to name a few (see Boegershausen, Aquino, & Reed, 2015, for a recent review).

A more recent, complementary line of research strongly suggests that morality is actually quite central to our personal identity. Strohminger and Nichols (2014) set out to investigate whether all mental traits are equally important when it comes to our judgments of personal identity, or if, rather, some psychological features are more central than others for judgments of personal identity through time. Their results suggest that people perceive moral changes to be the most disruptive to personal identity, as participants were less likely to say that an individual was the same person they were before after undergoing a changing experience (e.g., brain trauma, drug ingestion, body exchange, reincarnation, aging) that modified their moral traits relative to their memories, desires, and perceptual and somatic capacities. Consistent with this finding, Strohminger and Nichols (2015) also showed that relatives of patients with neurological diseases are more likely to think that their loved ones are not quite the same persons they were before the illness, if they perceived a moral change more so than a memory or a motor change in them. Taken together, these and related pieces of evidence (e.g., Heiphetz, Strohminger, & Young, 2017) indicate that moral features are particularly important to personal identity—or, as they put it, that the true self is a *moral* self.

Further evidence has refined this claim by showing that such a true self is not only moral: it is also morally *good*. In a clever study, Tobia (2015) randomly sorted participants into one of two conditions. In one condition, participants were presented with the case of Phineas Gage, who was described as an extremely nice individual who, after suffering a brain injury while working on a railroad, became cruel and morally corrupt. In the other condition, Phineas Gage was described as a morally wicked character who, after the accident, becomes nice and morally good. Tobia found that participants were much more likely to say that Phineas pre- and post-accident were the same person if the change was in the direction of moral improvement rather than deterioration. Consistent with this finding, Molouki and Bartels (2017) report that people not only think that their moral traits are the most central of their personality characteristics for their own personal identity, but also that if their morals were to change in a negative direction, they would see that as more disruptive to their self-concept than if the change was in a positive direction. Researchers interpret these and related studies a suggesting that changes that lead to moral deterioration are perceived as more threatening to personal continuity than changes that lead to moral improvement, because the latter are judged as instances of self-discovery (i.e., as episodes in which individuals uncover their true good self; Strohminger, Knobe, & Newman, 2017). The true self is, thus, a morally *good* self.

Remarkably, the notion of a true good self seems to cross cultural boundaries. In a recent study, De Freitas et al. (2018) presented participants from Colombia, Singapore, Russia and the United States, with twelve scenarios depicting agents undergoing one of three kinds of changes: from morally good to morally bad, from morally bad to morally good, and a morally neutral change. They found that across all four cultures, personal changes that were in the direction of moral improvement were thought of as reflecting and as having been originated by the agent's true self, relative to morally neutral and morally negative changes. Moreover, the

centrality of our positive moral traits to personal identity appears to be present even among Hindu Indians and Buddhist Tibetans, despite the fact that their religious traditions explicitly deny the existence of a single self (Nichols et al., 2016; Garfield, Nichols, Rai, & Strohminger, 2015).

In sum, the available evidence strongly suggests that morality is central to people's personal identity. So important morality appears to be, that some researchers have recently suggested that being morally good may be a basic psychological need, on a par with other basic psychological needs such as autonomy, competence and the like (Prentice et al., 2019). Indeed, the extent to which people satisfied this basic moral need predicted other important psychological outcomes like positive affect and enhanced well-being. We are, in a nutshell, essentially morally good.

Moral memories

Despite the centrality of morality for our personal identity, people commit moral transgressions with exorbitant frequency. And we are not talking just about the kinds of egregious wrongdoings that make the news every day. We are talking about ordinary, everyday moral transgressions people regularly commit. In a recent study, for instance, Hofmann, Wisneski, Brandt, and Skitka (2014) used ecological momentary assessment to record the frequency with which individuals experienced moral transgressions in a three-day period, finding that participants reported 14% of the time that they had "committed, were the target of, witnessed, or learned about" a moral transgression within the past hour.

But if morality is so central to our personal identity, and we take ourselves to be essentially morally good, why is it that people commit all sorts of immoral actions with such remarkable frequency? According to Sedikides' proposal, the answer would appear to be that the psychological immune system kicks in and protects our personal identity in the face of such threatening evidence. In other words, when one commits and/or is reminded of a personal past wrongdoing, our psychological immune system is activated, as it seeks to eliminate—or, at least, mitigate—the effects of threats to our beliefs that we are fundamentally and truly morally good. But is the postulation of such a specialized system required to explain this phenomenon, or could we find a more parsimonious explanation that, in the spirit of Occam's razor, does not multiply psychological entities unnecessarily? We believe we can. Specifically, with an account in terms of good-old, widely-studied features of our autobiographical memory, one can explain how people can still maintain their belief that they are essentially morally good selves while being reminded that they have also committed moral transgressions (Stanley and De Brigard, 2019).

Consider, to begin with, a very clever study by Shu, Gino, and Bazerman (2011), in which participants were asked to read an honor code meant to bring awareness of honesty standards prior to completing a monetarily rewarded task in which it was very easy to cheat. After completing the task, participants received a surprise memory test in which they were asked about details of the honor code. Interestingly, Shu et al. found that participants who cheated remembered fewer details of the honor code relative to those who did not cheat. Or take a complementary study by Reczek, Irwin, Zane, and Ehrich (2018), in which participants were asked to read the description of several products, some of which included ethically questionable attributes (e.g., the product was manufactured by child labor), prior to rating their likelihood of purchasing them. Then, in a surprise memory test, participants were asked to recall details of the products they were presented with. Reczek et al. found that participants were more likely to forget ethically questionable information about a product if they had expressed more inclination toward purchasing it relative to those participants who felt uninclined to buy them. Or, finally, consider a study by Bell, Schain, and Echterhoff (2014), in which participants were presented with faces associated with behaviors that were either morally right or morally wrong. Critically, they also manipulated the consequences of such behaviors for the participants themselves, so that the behaviors associated with half of the faces resulted in a cost to the participant, while the other half resulted in a benefit. Interestingly, they found that faces associated with immoral behaviors were remembered well when those actions resulted in personal costs, but they were poorly remembered when they resulted in personal benefits—a result that the authors interpret as reflecting a self-serving bias.

People not only forget details of personal past wrongdoings that challenge their morally good self-concept, but they also forget negative interpersonal feedback that threatens their morally good self-concept (Green, Sedikides, & Gregg, 2008; Green & Sedikides, 2004; Sedikides, Green, Saunders, Skowronski, & Zengel, 2016). For example, when a person is presented with self-threatening feedback implying that they are unkind or untrustworthy (e.g., You would make fun of others because of their looks; You would borrow other people's belongings without their knowledge), that feedback is likely to be forgotten. However, if the person is instead presented with positive, self-affirming feedback implying that they are kind and trustworthy, then this information is more likely to be processed deeply, frequently recalled, elaborated upon, and integrated with relevant episodic self-knowledge. This process of selectively forgetting negative, self-threatening feedback is strategic. Individuals do not forget all negative self-referent feedback. Instead, the most threatening feedback assailing those particular traits and qualities of utmost importance is most likely to be forgotten (Green & Sedikides, 2004; Sedikides et al., 2016).

The underlying patterns of results across these studies suggest that people better recall information that does not threaten their morally good self-concept relative to information that does. What could explain this pattern of results? The possibility suggested by Sedikides' proposal is that, when faced with the reminder cues, dedicated mechanisms of a putative psychological immune system get activated to prevent the retrieval of information that is inconsistent—but not of information that is consistent—with the individuals' morally good personal identity. But a more parsimonious

explanation involves no such dedicated mechanisms, but rather well-documented mnemonic processes, such as *intentional forgetting*, that get deployed and redeployed to different ends in different situations.

In a seminal study, Bjork, Laberge, and Legrand (1968) used a simple laboratory task to study intentional forgetting—*list-method directed forgetting*. Participants were instructed to study a list of words, and then they were instructed either to forget or to remember that list. In a subsequent memory test, participants remembered fewer of the words in the list when they were instructed to forget the list relative to when they were instructed to remember it. In more recent work, Anderson and Green (2001) utilized a think/no-think paradigm to investigate whether participants could actively and voluntarily forget previously acquired information. In their paradigm, participants initially studied cue-target pairs and were trained to recall the second word (target) in the studied pair whenever the first word in the pair (cue) appeared on the computer screen. In the next phase of the experiment, on most trials, participants were instructed to recall the target word when presented with the cue word ("Think Trials"). But on some trials, participants were instructed to "intentionally forget" by avoiding retrieving the target word when presented with the cue word ("No-Think Trials"). Recollection for target words was better in the "Think Trials" relative to the "No-Think Trials", which suggests that inhibitory control processes during retrieval can lead to forgetting the suppressed material. Since then, many studies have helped to further understand the nature of intentional forgetting, revealing, for instance, that effect sizes tend to increase with repeated suppression of the target at retrieval (Anderson, Reinholz, Kuhl, & Mayr, 2011), as well as when the valence of the suppressed information is negative (Anderson & Huddleston, 2012; for a recent review see Anderson & Hanslmayr, 2014).

A critical take-away from the literature is that intentional forgetting is *not* exclusive to negative and/or morally questionable information that would appear to threaten a desired self-concept. For example, the stimuli used in the research conducted by Bjork et al. (1968) and by Anderson and Green (2001) did not present threats to a favorable self-concept. Further supporting this contention, Payne and Corrigan (2007) found that people are better at intentionally forgetting neutral pictures than emotional pictures, where none of the pictures presented a clear threat to a favorable self-concept. Intentional forgetting many also serve other functions unrelated to personal identity. For instance, we may try to intentionally forget previously learned information after coming to believe that the information is false, misleading, or unhelpful (e.g., Golding & Keenan, 1985). In addition, some research suggests that we intentionally forget information to put it out of mind when it could interfere with performing some subsequent task (Lehman & Malmberg, 2011). Ultimately, we can and do intentionally forget a variety of different information for a variety of different reasons.

Might it be possible that intentional forgetting is behind people's tendencies to forget past wrongdoings? As it turns out, some researchers support this explanation. In a recent study, Kouchaki and Gino (2016) asked participants to imagine taking an exam in which they have the possibility to cheat. Roughly half of the participants were asked to imagine actually cheating on the exam, while the other half were asked to imagine not cheating. A subsequent surprise memory test revealed that those participants who imagined cheating on the exam forgot more details about the imagined event than participants who imagined not cheating on the exam (Kouchaki & Gino, 2016; although, see Stanley, Yang, & De Brigard, 2018). The authors interpret their pattern of results as arising from the same intentional forgetting processes identified by Anderson and Green (2001). The thought is that even just imagining committing a moral transgression encourages people to intentionally forget the details of the experience, which results in poorer recollective performance relative to other kinds of details. More importantly, a similar explanation is available for the results of Shu et al. (2011) and Reczek et al. (2018), mentioned above. Participants likely didn't forget that they cheated on the task or that they expressed a stronger preference for the unethically manufactured product, but they did strategically forget details associated with the moral nature of their choice to protect their morally good self-concept (at least to some extent). A parallel explanation is available for the aforementioned findings by Green and Sedikides (2004), as participants who received negative feedback that threatened their morally good self-concept were more likely to forget the feedback than those who received positive feedback (although, to be fair, only further studies could provide direct evidence that this finding is actually due to *intentional* forgetting; our claim here is that such a process could in principle explain the results). And if these sorts of explanations are on the right track, then intentional forgetting, as a basic psychological process that can be deployed and redeployed for a variety of reasons, may be all is needed to be invoked to account for these effects, rendering unnecessary to postulate an additional dedicated mechanism of a putative psychological immune system.

Memory, morality and identity protection

That people tend to engage in self-enhancement and self-protection of their moral identity appears to be a rather unassailable fact of human psychology—or, at least, among the populations in which it has been studied. What is questionable, though, is whether we need to postulate the existence of a dedicated psychological system whose sole purpose is to protect the self in just such a way. In this short commentary, we explored one particular phenomenon—the selective and intentional forgetting of information that could threaten a morally good personal identity—the explanation of which, on Sedikides view, would be achieved by alleged dedicated mechanisms of this putative psychological immune system. In our view, no such dedicated mechanisms are required. Instead, we argue that ordinary memory processes, such as intentional forgetting, which are non-dedicated insofar as they can be redeployed for all sorts of

purposes in all sorts of contexts, may suffice to explain the selective loss of information that could threaten a morally good self-concept.

Indeed, similar explanations are available for other memory-related phenomena that could threaten a morally good self-concept. For instance, among the findings reported in the aforementioned paper by Kouchaki and Gino (2016) is the fact that the phenomenology—not just the accuracy—of our recollections of past wrongdoings is dampened relative to more positive memories. More precisely, our memories of past wrongdoings seem to be experienced less vividly, less clearly and with less emotional intensely than our memories for morally permissible behaviors (see also, Huang, Stanley, & De Brigard, 2020). Is this the product of a dedicated mechanism from a putative psychological immune system? Not necessarily. There is plenty of evidence that negative memories that do not threaten a morally good self-concept are also remembered less vividly, less clearly, and with less emotional intensity over time than positive memories—a mnemonic phenomenon known as *fading affect bias* (Walker & Skowronski, 2009). As such, this very same process may underwrite the phenomenological effect reported by Kouchaki and Gino (2016), rendering it unnecessary to postulate the existence of a dedicated mechanism from an alleged psychological immune system.

Or consider the recent finding that people exhibit a "knew-it-all-along effect" after cheating on a test (Stanley, Stone, & Marsh, 2021). The authors argue that cheating might not seem as distasteful (and might even be justified) when test-takers believe that they already knew the answers. Accordingly, cheating could be reconstrued as accidental or unnecessary, and thus, not a real threat to an otherwise positive self-concept. Is this also the product of a dedicated process that helps us to protect a morally good self-concept? Likely not. People often have difficulty retrospectively determining what they knew prior to acquiring all kinds of new information (e.g., Fischoff, 1975; Hasher, Attig, & Alba, 1981; Hawkins & Hastie, 1990; Roese & Vohs, 2012; Wood, 1978). Seminal research on the hindsight bias (Fischoff, 1975) exemplifies this observation: participants read historical scenarios and answered multiple-choice questions about several possible outcomes. After being informed of the actual outcomes, participants were asked to indicate which outcome they would have guessed, if they had not been provided the correct answer. Participants were consistently biased toward reporting they would have selected the correct answer all along (see also, Guilbault, Bryant, Brockway, & Posavac, 2004). Similar effects occur across a variety of domains that should have no bearing on personal identity (Christensen-Szalanski & Willham, 1991; Guilbault et al., 2004; Hawkins & Hastie, 1990). Building off early work on the hindsight bias, other research has found that, after attempting to answer obscure trivia questions and then receiving the correct answers to the questions, people often come to believe that they just knew the correct answers all along (Fischoff, 1977; Hasher et al., 1981; Jacoby & Kelley, 1987; Metcalfe & Finn, 2011; Wood, 1978). So, the effect obtained by Stanley et al. (2021) is really just an amplified variant of the classic knew-it-all-along effect, which has been documented all sorts of material, many of which in no way threaten a favorable self-concept.

Something similar may occur with other well-established effects in cognitive science, offering several potentially fruitful avenues for future research. For example, decades of research have shown that people interpret easy processing, or fluency, as a cue for truth (Alter & Oppenheimer, 2009; Boehm, 1994; Unkelbach, 2007). As a result, repetition and rehearsal tend to boost truth judgments for a variety of information (e.g., trivia facts, political positions, fake news; Brashier & Marsh, 2020; Dechêne, Stahl, Hansen, & Wänke, 2010). This may suggest that, if you frequently rehearse events exemplifying your possess positive moral traits and qualities, and if you repeatedly tell yourself that you possess positive moral traits and qualities, then you should be more likely to believe you do, in fact, possess positive moral traits and qualities (Stanley et al, 2019). It could also be the case that, since in certain contexts more positive information tends to be more deeply encoded than negative information (Kensinger, 2009), a levels-of-processing effect could account for why people tend to better recall information that impinges positively on their self-concept rather than negatively.

Sedikides' proposal draws from a number of enticing commonalities between what such a putative psychological immune system may be and what our real biological immune system actually is. Yet, behind the allure of these apparent similarities, rests a critical difference between the putative psychological immune system and its template, the biological one: we have no evidence whatsoever of the existence of dedicated psychological mechanisms that, like those of the biological immune system, have as their sole role to attack pathogens and protect the integrity of our organism. By contrast, some evidence suggests that self-enhancement and self-protection may be achieved by redeploying the same cognitive processes used for all sorts of psychological tasks. In this brief commentary, we explained how ordinary memory mechanisms may be behind our tendency to protect a morally good personal identity—rather than the kind of dedicated mechanisms suggested by Sedikides proposal. As such, we can't help but agree with Gilbert's ontological caution: the psychological immune system is just a metaphor, not a hypothesis about the structure of our cognitive architecture, for it finds no grounds in empirical evidence.

ORCID

Felipe De Brigard http://orcid.org/0000-0003-0169-1360

References

Alicke, M. D., & Govorun, O. (2005). The better-than-average effect. In M. D. Alicke, D. A. Dunning, & J. I. Krueger (Eds.), *The self in social judgment* (pp. 85–106). New York: Psychology Press.

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), 1–48. doi:10.1080/10463280802613866

Alter, A. L., & Oppenheimer, D. M. (2009). Uniting the tribes of fluency to form a metacognitive nation. *Personality and Social Psychology Review*, *13*(3), 219–235. doi:10.1177/1088868309341564

Anderson, M. C., & Green, C. (2001). Suppressing unwanted memories by executive control. *Nature*, *410*(6826), 366–369. doi:10.1038/35066572

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

Anderson, M. C., & Huddleston, E. (2012). Towards a cognitive and neurobiological model of motivated forgetting. In R. F. Belli (Ed.), *True and false recovered memories* (pp. 53–120). New York, NY: Springer

Anderson, M. C., Reinholz, J., Kuhl, B. A., & Mayr, U. (2011). Intentional suppression of unwanted memories grows more difficult as we age. *Psychology and Aging*, *26*(2), 397–405. doi:10.1037/a0022505

Aquino, K., & Reed, I. I. (2002). The self-importance of moral identity. *Journal of Personality and Social Psychology*, *83*(6), 1423–1440.

Bandura, A. (1999). Social cognitive theory of personality. In L. A. Pervin & O. P. John (Eds.), *Handbook of personality: Theory and research* (pp. 154–196). New York: Guilford Press.

Baumeister, R. F. (1998). The self. In D. T. Gilbert, S. T. Fiske, & G. Lindzey (Eds.), *Handbook of social psychology* (4th ed., Vol. 1, pp. 680–740). New York: McGraw-Hill.

Bell, R., Schain, C., & Echterhoff, G. (2014). How selfish is memory for cheaters? Evidence for moral and egoistic biases. *Cognition*, *132*(3), 437–442.

Bjork, R. A., Laberge, D., & Legrand, R. (1968). The modification of short-term memory through instructions to forget. *Psychonomic Science*, *10*(2), 55–56. doi:10.3758/BF03331404

Boegershausen, J., Aquino, K., & Reed, A. (2015). Moral Identity. *Current Opinion in Psychology*, *6*, 162–166. doi:10.1016/j.copsyc.2015.07.017

Boehm, L. E. (1994). The validity effect: A search for mediating variables. *Personality and Social Psychology Bulletin*, *20*(3), 285–293. doi:10.1177/0146167294203006

Brashier, N. M., & Marsh, E. J. (2020). Judging truth. *Annual Review of Psychology*, *71*, 499–515.

Christensen-Szalanski, J. J., & Willham, C. F. (1991). The hindsight bias: A meta-analysis. *Organizational Behavior and Human Decision Processes*, 48(1), 147–168.

De Freitas, J., Sarkissian, H., Newman, G. E., Grossmann, I., De Brigard, F., Luco, A., & Knobe, J. (2018). Consistent belief in a good true self in misanthropes and three interdependent cultures. *Cognitive Science*, *42*, 134–160. doi:10.1111/cogs.12505

Dechêne, A., Stahl, C., Hansen, J., & Wänke, M. (2010). The truth about the truth: A meta-analytic review of the truth effect. *Personality and Social Psychology Review*, *14*(2), 238–257.

Fischoff, B. (1975). Hindsight ≠ foresight: The effect of outcome knowledge on judgment under uncertainty. *Journal of Experimental Psychology*, *1*, 288–299.

Fischoff, B. (1977). Perceived informativeness of facts. *Journal of Experimental Psychology: Human Perception and Performance*, *3*, 349–358.

Freud, A. (1936). *Ego and the mechanisms of defense* (rev. ed., C. Baines, Trans.). New York: International Universities Press.

Garfield, J. L., Nichols, S., Rai, A. K., & Strohminger, N. (2015). Ego, egoism and the impact of religion on ethical experience: What a paradoxical consequence of Buddhist culture tells us about moral psychology. *The Journal of Ethics*, *19*(3–4), 293–304. doi:10.1007/s10892-015-9210-9

Garson, J. (2016). *A critical overview of biological functions*. Berlin: Springer.

Gertner, J. (2003). The futile persuit of happiness. *The New York Times*. Sept 7.

Gilbert, D. T. (2006). *Stumbling on Happiness*. New York: Vintage.

Gilbert, D. T., Pinel, E. C., Wilson, T. D., Blumberg, S. J., & Wheatley, T. P. (1998). Immune neglect: A source of durability bias in affective forecasting. *Journal of Personality and Social Psychology*, *75*(3), 617–638.

Golding, J. M., & Keenan, J. M. (1985). Directed forgetting and memory for directions to a destination. *The American Journal of Psychology*, *98*(4), 579–590. doi:10.2307/1422511

Green, J. D., & Sedikides, C. (2004). Retrieval selectivity in the processing of self-referent information: Testing the boundaries of self-protection. *Self and Identity*, *3*(1), 69–80. doi:10.1080/13576500342000059a

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. doi:10.1016/j.jesp.2007.10.006

Guilbault, R. L., Bryant, F. B., Brockway, J. H., & Posavac, E. J. (2004). A meta-analysis of research on hindsight bias. *Basic and Applied Social Psychology*, *26*(2–3), 103–117. doi:10.1080/01973533.2004.9646399

Hasher, L., Attig, M. S., & Alba, J. W. (1981). I knew it all along: Or, did I. *Journal of Verbal Learning and Verbal Behavior*, *20*(1), 86–96. doi:10.1016/S0022-5371(81)90323-6

Hawkins, S. A., & Hastie, R. (1990). Hindsight: Biased judgments of past events after the outcomes are known. *Psychological Bulletin*, *107*(3), 311–327. doi:10.1037/0033-2909.107.3.311

Heiphetz, L., Strohminger, N., & Young, L. (2017). The role of moral beliefs, memories, and preferences in representations of identity. *Cognitive Science*, *41*(3), 744–767. doi:10.1111/cogs.12354

Hofmann, W., Wisneski, D. C., Brandt, M. J., & Skitka, L. J. (2014). Morality in everyday life. *Science*, *345*(6202), 1340–1343. doi:10.1126/science.1251560

Huang, S., Stanley, M. L., & De Brigard, F. (2020). The phenomenology of remembering our moral transgressions. *Memory & Cognition*, *48*(2), 277–286. doi:10.3758/s13421-019-01009-0

Jacoby, L. L., & Kelley, C. M. (1987). Unconscious influences of memory for a prior event. *Personality and Social Psychology Bulletin*, *13*(3), 314–336. doi:10.1177/0146167287133003

Kensinger, E. A. (2009). Remembering the details: Effects of emotion. *Emotion Review: journal of the International Society for Research on Emotion*, *1*(2), 99–113. doi:10.1177/1754073908100432

Kohlberg, L. (1971). Stages of moral development as a basis for moral education. In C. M. Beck, B. S. Crittenden, & E. V. Sullivan (Eds.), *Moral education: Interdisciplinary approaches*. Toronto: University of Toronto Press.

Kouchaki, M., & Gino, F. (2016). Memories of unethical actions become obfuscated over time. *Proceedings of the National Academy of Sciences*, *113*(22), 6166–6171. doi:10.1073/pnas.1523586113

Lehman, M., & Malmberg, K. J. (2011). Overcoming the effects of intentional forgetting. *Memory & Cognition*, *39*(2), 335–347.

Metcalfe, J., & Finn, B. (2011). People's hypercorrection of high-confidence errors: Did they know it all along? *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *37*(2), 437–448.

Molouki, S., & Bartels, D. M. (2017). Personal change and the continuity of the self. *Cognitive Psychology*, *93*, 1–17. doi:10.1016/j.cogpsych.2016.11.006

Nichols, S., Strohminger, N., Rai, A., & Garfield, J. (2018). Death and the self. *Cognitive science*, 42, 314–332.

Payne, B. K., & Corrigan, E. (2007). Emotional constraints on intentional forgetting. *Journal of Experimental Social Psychology*, *43*(5), 780–786. doi:10.1016/j.jesp.2006.07.005

Prentice, M., Jayawickreme, E., Hawkins, A., Hartley, A., Furr, R. M., & Fleeson, W. (2019). Morality as a basic psychological need. *Social Psychological and Personality Science*, *10*(4), 449–460. doi:10.1177/1948550618772011

Reczek, R. W., Irwin, J. R., Zane, D. M., & Ehrich, K. R. (2018). That's not how I remember it: Willfully ignorant memory for ethical product attribute information. *Journal of Consumer Research*, *45*(1), 185–207. doi:10.1093/jcr/ucx120

Roese, N. J., & Vohs, K. D. (2012). Hindsight bias. *Perspectives on Psychological Science*, *7*(5), 411–426.

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

Sedikides, C., & Gregg, A. P. (2008). Self-enhancement: Food for thought. *Perspectives on Psychological Science*, *3*(2), 102–116.

Shu, L. L., Gino, F., & Bazerman, M. H. (2011). Dishonest deed, clear conscience: When cheating leads to moral disengagement and motivated forgetting. *Personality & Social Psychology Bulletin*, *37*(3), 330–349.

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. doi:10.1177/0963721419847990

Stanley, M. L., Henne, P., & De Brigard, F. (2019). Remembering moral and immoral actions in constructing the self. *Memory & Cognition*, *47*(3), 441–454. doi:10.3758/s13421-018-0880-y

Stanley, M. L., Stone, A. R., & Marsh, E. J. (2021). Cheaters claim they knew the answers all along. *Psychonomic Bulletin & Review*, *28*(1), 341–350. doi:10.3758/s13423-020-01812-w

Stanley, M. L., Yang, B. W., & De Brigard, F. (2018). No evidence for unethical amnesia for imagined actions: A failed replication and extension. *Memory & Cognition*, *46*(5), 787–795.

Strohminger, N., Knobe, J., & Newman, G. (2017). The true self: A psychological concept distinct from the self. *Perspectives on Psychological Science*, *12*(4), 551–560.

Strohminger, N., & Nichols, S. (2014). The essential moral self. *Cognition*, *131*(1), 159–171. doi:10.1016/j.cognition.2013.12.005

Strohminger, N., & Nichols, S. (2015). Neurodegeneration and identity. *Psychological Science*, *26* (9), 1469–1479. doi:10.1177/0956797615592381

Tobia, K. (2015). Personal identity and the Phineas Gage effect. *Analysis*, *75*(3), 396–405. doi:10.1093/analys/anv041

Unkelbach, C. (2007). Reversing the truth effect: Learning the interpretation of processing fluency in judgments of truth. *Journal of Experimental Psychology*, *33*(1), 219–230.

Vaillant, G. (1993). *The wisdom of the ego*. Cambridge, MA: Harvard University Press.

Walker, W. R., & Skowronski, J. J. (2009). The fading affect bias: But what the hell is it for? *Applied Cognitive Psychology*, *23*(8), 1122–1136. doi:10.1002/acp.1614

Wood, G. (1978). The knew-it-all-along effect. *Journal of Experimental Psychology*, *4*(2), 345–353.