CrossMark

Remembering moral and immoral actions in constructing the self

Matthew L. Stanley [1] · Paul Henne [2] · Felipe De Brigard [1,2]

Abstract

Having positive moral traits is central to one's sense of self, and people generally are motivated to maintain a positive view of the self in the present. But it remains unclear *how* people foster a positive, morally good view of the self in the present. We suggest that recollecting and reflecting on moral and immoral actions from the personal past jointly help to construct a morally good view of the current self in complementary ways. More specifically, across four studies we investigated the extent to which people believe they have changed over time after recollecting their own moral or immoral behaviors from the personal past. Our results indicate that recollecting past immoral actions is associated with stronger impressions of dissimilarity and change in the sense of self over time than recollecting past moral actions. These effects held for diverse domains of morality (i.e., honesty/dishonesty, helping/harming, fairness/unfairness, and loyalty/disloyalty), and they remained even after accounting for objective, calendar time. Further supporting a motivational explanation, these effects held when people recollected their own past actions but not when they recollected the actions of other people.

Keywords Moral psychology · Autobiographical memory · Temporal self-appraisal theory · Identity · Self

Introduction

A critical function of remembering the personal past is to construct a sense of self: remembering events from the past informs who we are and who we wish to be (Bluck, 2003; Conway, 2005; Fivush, Habermas, Waters, & Zaman, 2011; McAdams, 2013; Wilson, Gunn, & Ross, 2009). But the person we believe ourselves to be in the present influences which events we readily recall, how we recall those events, and how we come to interpret the significance of those events (Conway, 2005; Greenwald, 1980; Ross, 1989; Wilson & Ross, 2003). At least in Western cultures, most people are motivated to maintain a positive view of themselves in the present (Alicke & Sedikides, 2009; Baumeister, 1998; Sedikides, 1993). Systematic biases and distortions in recollecting specific past events and in

✉ Matthew L. Stanley
matthew.stanley@duke.edu

[1] Department of Psychology and Neuroscience, Center for Cognitive Neuroscience, Duke University, Durham, NC 27708-0743, USA

[2] Department of Philosophy, Duke University, Durham, NC, USA

constructing a life narrative help to maintain, enhance, and protect this positive view of the self.

In certain contexts, people construct a positive view of the self in the present by readily and selectively recalling past behaviors that showcase their positive traits and qualities (Markus & Wurf, 1987; Pasupathi, Mansour, & Brubaker, 2007; Ross, 1989; Sanitioso, Kunda, & Fong, 1990). In one line of research, experimenters led participants to believe that either extroverts or introverts tend to be more successful (Kunda, Fong, Sanitioso, & Reber, 1993; Kunda & Sanitioso, 1989; Sanitioso, Kunda, Fong, 1990). Those led to believe that extroversion is conducive to success more readily recalled past events showcasing their extroversion relative to those led to believe that introversion is conducive to success. Other research has indicated that people tend to preferentially forget negative feedback about themselves, particularly when that negative feedback is about something of personal importance (Sanitioso & Wlodarski, 2004; Sedikides & Green, 2000). Importantly, those with the strongest motivations to view themselves positively are most likely to interpret their positive experiences as characteristic of a stable sense of self that persists over time (Kunda, 1990; Wilson & Ross, 2003). For instance, a person strongly motivated to believe she is intelligent might selectively seek out, readily accept, and successfully retain information substantiating this positive view of herself as intelligent. She might then easily and

Springer

frequently recall events from her personal past that support this view of herself, and those recollections would strengthen the belief that she currently is, and steadfastly has been, an intelligent person. In this way, autobiographical memories facilitate impressions of similarity (or consistency) over time in service of fostering a favorable view of the self.

The readiness with which we recall events showcasing our positive traits does not imply that we are more likely to forget experiences that portray the self negatively (Stanley, Yang, & De Brigard, 2018). When confronted by our own past shortcomings and failures, it is still possible to achieve a positive view of the self in the present by constructing a life narrative in which negative experiences represent turning points or indications of self-improvement over time (Conway, 2005; Libby & Eibach, 2002; Wilson & Ross, 2003). Individuals compare their current selves to their past selves to perceive positive change and improvement over time, which, in turn, helps to maintain, or even to enhance, a favorable view of the current self (Demiray & Janssen, 2015; Gebauer, Broemer, Haddock, & von Hecker, 2008; Ross & Wilson, 2003). For example, Wilson and Ross (2000) manipulated the objectives of participants when instructing them to describe themselves: some were encouraged to adopt the goal of evaluating themselves favorably, and others were encouraged to evaluate themselves accurately. Participants instructed to evaluate themselves favorably were more likely to describe inferior past selves than were participants pursuing accuracy goals. In another study, Ross and Wilson (2002) found that participants felt more psychologically distant from past negative experiences, but they felt psychologically nearer to past positive experiences. By perceiving a past achievement as psychologically close, people can continue to relish their success and establish a positive self-image; by perceiving a past failure as psychologically distant, people can dismiss that failure as belonging to a very different past self. In this way, a positive view of the self can be cultivated in the present by strategically interpreting negative past experiences within a life narrative.

When constructing a positive sense of self from autobiographical memories, memories showcasing certain traits and qualities are presumably more important than others. Recent evidence suggests that perceiving alterations in another person's moral traits leads to attributions of greater personal change for that person than perceiving alterations in other mental features such as perception, desires, or emotions (Heiphetz, Strohminger, & Young, 2017). Critically, people also treat moral traits as defining features of their own selves (Molouki & Bartels, 2017) – just as they do when making judgments about the identities of other people. More than any other feature of our mental lives, positive moral traits seem to be the most deeply rooted and causally central components of the self (Strohminger & Nichols, 2014, 2015; Strohminger, Newman, & Knobe, 2017; De Freitas et al., 2018).

Despite the now considerable evidence indicating that positive moral traits are most central to personal identity, there is little research investigating how remembering moral and immoral actions from the personal past creates a morally good self-image. Most research on morality and the self has utilized vignettes and hypothetical thought experiments, not actual memories of personal past experiences. Nevertheless, many argue that autobiographical memories of one's own personal history play a critical role in constructing the self (Conway, Singer, & Tagini, 2004; Wilson, Gunn, & Ross, 2009). Some evidence does suggest that autobiographical memories of immoral actions are particularly susceptible to biases and distortions in ways that look favorably on the remember (Escobedo & Adolphs, 2010; Kouchaki & Gino, 2016; Stanley et al., 2017). In particular, Stanley et al. (2017) found that participants judged their own moral transgressions from the distant past to be more morally wrong than their more recent transgressions. Participants advantageously utilized time to buttress a belief in personal moral improvement. This finding is consistent with the notion that people frequently compare their current selves to their past selves to perceive personal improvement over time, regardless of whether that perceived improvement is accurate (D'argembeau & Van der Linden, 2008; Demiray & Janssen, 2015; Ross & Wilson, 2000; Ryff, 1991; Wilson & Ross, 2003).

If people are especially motivated to view themselves as morally good, then recollecting their past moral and immoral behaviors might produce differential perceptions of change in the self since those events occurred. People might perceive greater change in the self after recalling their own past immoral behaviors relative to moral behaviors, even if objective, calendar time cannot be readily utilized to produce an impression of moral improvement over time. By perceiving greater change and dissimilarity in the self since committing a moral transgression, people can form an impression of personal moral improvement over time. This, in turn, might foster a positive, morally good view of the self in the present. Furthermore, these perceptions of similarity or change in the self after reflecting on past immoral relative to moral behaviors might manifest in general and specific ways. More generally, people might report perceiving greater global change in their selves after recalling their past moral transgressions relative to their past morally praiseworthy deeds. More specifically, people might report perceiving greater positive change in the extent to which they exemplify particular moral traits (e.g., honesty, loyalty) over time after recalling their past immoral behaviors relative to moral behaviors. In four studies, we investigate the extent to which people perceive change in the self after recalling moral and immoral behaviors from the personal past. We hypothesize that people strategically come to perceive meaningful change in the self after reflecting on their own immoral behaviors from the past, but that people tend to perceive similarity (lack of change) in the self over time after

reflecting on their own moral behaviors from the past. In this way, people use their autobiographical memories of moral and immoral actions in complementary ways to forge a morally good sense of self in the present.

Study 1

In Study 1, we investigated whether people perceive greater dissimilarity and change in the self after recollecting their own immoral relative to moral behaviors. We further investigated whether these perceptions of dissimilarity and change persist after accounting for when participants report that the events actually occurred in the past (i.e., objective, calendar time).

Materials and method

Participants One-hundred and ten individuals voluntarily participated in this study via Amazon's Mechanical Turk (AMT) for monetary compensation. Participant recruitment was restricted to individuals in the USA with a prior approval rating above 85%. Twenty participants were excluded for failing to provide at least one memory, for recalling an event that occurred more than 10 years ago, or for providing the incorrect type of memory based on the cue (i.e., a memory of a morally wrong behavior was provided when the participant was cued to recall a morally right behavior, or vice versa). As such, data were analyzed with the remaining 90 participants (M_{age} = 35.61 years, SD = 10.30, age range = 20–71 years, 33 females, 56 males). All participants reported being fluent English speakers. Informed consent was obtained from each participant in accordance with the protocol approved by the Duke University Campus Institutional Review Board.

Procedure The study was self-paced. Participants were asked to recall a total of six distinct behaviors, one at a time, from their personal pasts that occurred within the past 10 years. Three of these memories involved morally wrong actions committed by the participant; the other three memories involved morally right actions committed by the participant. Participants were told that these remembered actions must have occurred on a particular day in a particular place.

For each remembered behavior, participants described the event in two to five sentences. They then typed in the month and year that it occurred, and they selected one of the following options to best describe when it occurred: within the past day, within the past week, within the past two weeks, within the past month, within the past 2 months, within the past 6 months, within the past year, within the past 2 years, within the past 5 years, within the past 10 years. As a manipulation check, participants answered the following: how morally wrong or morally right was your behavior in this instance? (1 = *very morally wrong*, 7 = *very morally right*). Participants

then answered the following three questions in random order: as you think about this memory now, do you feel like you are a *different* person now than you were then? (1 = *definitely no*, 7 = *definitely yes*); as you think about this memory now, do you feel like you are the *same* person now that you were then? (1 = *definitely no*, 7 = *definitely yes*); how much have you *changed* as a person since this event occurred? (1 = *not at all*, 7 = *a lot*). It is worth noting that these last three questions assess perceptions of psychological change over time, not whether the former self ceased to exist after the creation of a new self (Starmans & Bloom, 2018). Upon completion of the study, participants were monetarily compensated for their time.

Data analyses Data were analyzed using R (R Development Core Team, 2009) with the lme4 software package (Bates, Maechler, Bolker, & Walker, 2015). Data were fit to linear mixed-effects models, and subject was included as a random effect (random intercepts only) in all models. Significance for fixed effects was assessed using Satterthwaite approximations to degrees of freedom, and 95% confidence intervals (CIs) around beta-values were computed using parametric bootstrapping. Objective time was assessed in two complementary ways. One objective time variable (hereafter referred to as $time_A$) was coded as follows: 0 = *within the past day*; 1 = *within the past week*; 2 = *within the past 2 weeks*; 3 = *within the past month*; 4 = *within the past 2 months*; 5 = *within the past 6 months*; 6 = *within the past year*; 7 = *within the past 2 years*; 8 = *within the past 5 years*; 9 = *within the past 10 years*. Similar methods have been implemented to characterize the actual time that events occurred in the past (e.g., Escobedo & Adolphs, 2010; Stanley et al., 2017). The other objective time variable (hereafter referred to as $time_B$) indicates the number of months that have passed since the remembered event occurred, starting with remembered events that occurred in the same month as the experimental session coded as 0. The alpha level for all statistical tests was set at .05.

Results

Figure 1 depicts means for each variable of interest as a function of the memory cue (i.e., morally right vs. morally wrong). An initial linear mixed-effects model was computed to ensure that the remembered actions generated from the morally wrong cue were, in fact, judged to be more morally wrong on the 7-pt scale than the remembered actions generated from the morally right cue. This expectation was confirmed (b = 3.80, SE = .08, t = 48.80, $p < .001$, 95% CI = [3.64, 3.95]).

We then investigated the relationship between the cued morality of the action (binary, fixed factor: morally wrong vs. morally right) and the extent to which participants judged themselves to be *different* people now than they were when the event occurred. A linear mixed-effects model revealed a significant effect of the cued morality of the action on the
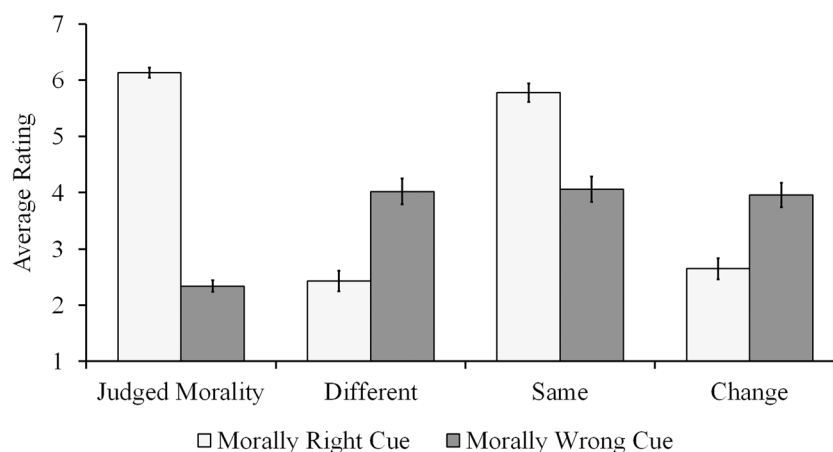
Fig. 1 Means and standard error bars are depicted for the judged morality of the remembered actions, different judgments, same judgments, and change judgments, all as a function of the memory cue (morally right or morally wrong)

extent to which participants believe they are different people now than they were when the event occurred ($b = 1.59$, $SE = .14$, $t = 11.06$, $p < .001$, 95% CI = [1.28, 1.88]). Participants judged that, when the remembered action occurred, they were more dissimilar to their current selves for the morally wrong condition relative to the morally right condition. This effect held at a similar magnitude even after statistically controlling for time$_A$ ($b = 1.34$, $SE = .13$, $t = 10.34$, $p < .001$, 95% CI = [1.07, 1.60]) and time$_B$ ($b = 1.25$, $SE = .13$, $t = 9.55$, $p < .001$, 95% CI = [1.01, 1.50]) in separate models (see Supplemental Materials for full models).

Next, we investigated the relationship between the cued morality of the action (binary, fixed factor: morally wrong vs. morally right) and the extent to which participants judged themselves to be the *same* people now that they were when the event occurred. A linear mixed-effects model revealed a significant effect of the cued morality of the action on the extent to which participants believe they are the same people now as they were when the event occurred ($b = -1.72$, $SE = .14$, $t = -12.28$, $p < .001$, 95% CI = [-2.00, -1.45]). Participants judged that, when the remembered action occurred, they were more similar to their current selves in the morally right condition relative to morally wrong condition. This effect held at a similar magnitude even after statistically controlling for time$_A$ ($b = -1.48$, $SE = .13$, $t = -11.52$, $p < .001$, 95% CI = [-1.74, -1.25]) and time$_B$ ($b = -1.37$, $SE = .13$, $t = -10.66$, $p < .001$, 95% CI = [-1.62, -1.12]) in separate models (see Supplemental Materials for full models).

Finally, we investigated the relationship between the cued morality of the action (binary, fixed factor: morally wrong vs. morally right) and the extent to which participants judged themselves to have *changed* since the event

occurred. An initial linear mixed-effects model revealed a significant effect of the cued morality of the action on the extent of perceived change since the event occurred ($b = 1.30$, $SE = .14$, $t = 9.38$, $p < .001$, 95% CI = [1.04, 1.60]). Participants judged that they had changed more since the remembered morally wrong actions occurred than they had since the morally right actions occurred. This effect held at a similar magnitude even after statistically controlling for time$_A$ ($b = 1.03$, $SE = .12$, $t = 8.50$, $p < .001$, 95% CI = [.78, 1.27]) and time$_B$ ($b = .95$, $SE = .12$, $t = 7.64$, $p < .001$, 95% CI = [.72, 1.21]) in separate models (see Supplemental Materials for full models).[1]

Discussion

Taken together, the results of Study 1 suggest that after remembering their own morally wrong actions from the personal past, participants perceive their current selves to be more dissimilar to those past selves than they do after recollecting morally right actions; they also perceive themselves to have undergone greater change since those past transgressions occurred. These effects remained at a similar magnitude even after statistically controlling for objective, calendar time.

Study 2

The purpose of Study 2 was to provide further support for the roles of recalling morally right and wrong actions in service of fostering a morally good view of the self across diverse domains of morality. To this end, we investigated individuals' perceptions of change in the self over time by directly comparing memories involving honesty versus dishonesty, helping versus harming, fairness versus unfairness, and loyalty versus disloyalty. We chose these matched pairs because people report frequently engaging

---

[1] As suggested by an anonymous reviewer, we also conducted additional analyses investigating possible interaction effects between the cued morality of the remembered behavior and objective time variables. These results are reported in the Supplemental Materials.

in morally right and wrong actions within these domains in everyday life (Hofmann et al., 2014).

Materials and method

Participants One-hundred and ten individuals voluntarily participated in this study via AMT for monetary compensation. Participant recruitment was restricted to individuals in the USA with a prior approval rating above 85%. Eleven participants were excluded for failing to provide at least one memory, for providing a memory of an event that occurred more than 10 years ago, or for providing the incorrect type of memory based on the cue. As such, data were analyzed with the remaining 99 participants (M_{age} = 33.78 years, SD = 9.78, age range = 20–69 years, 47 females, 51 males). All participants reported being fluent English speakers. Those who participated in Study 1 were automatically prevented from participating in Study 2. Informed consent was obtained from each participant in accordance with the protocol approved by the Duke University Campus Institutional Review Board.

Procedure The study was self-paced. Participants were asked to recall a total of eight distinct actions, one at a time, from their personal pasts that occurred within the past 10 years. Participants were provided with a unique cue for each of the eight memories: (1) recall a specific past experience in which you were honest with another person, and you believe your action was morally right; (2) recall a specific past experience in which you were dishonest with another person, and you believe your action was morally wrong; (3) recall a specific past experience in which you helped another person, and you believe your action was morally right; (4) recall a specific past experience in which you harmed another person, and you believe your action was morally wrong; (5) recall a specific past experience in which you treated another person fairly, and you believe your action was morally right; (6) recall a specific past experience in which you treated another person unfairly, and you believe your action was morally wrong; (7) recall a specific past experience in which you were loyal to another person, and you believe your action was morally right; and (8) recall a specific past experience in which you were disloyal to another person, and you believe your action was morally wrong. We randomized the order in which these cues were presented across participants. This cueing procedure ultimately produced four distinct matched pairs of remembered actions – one morally right and the other morally wrong – for each particular kind of behavior: (1) honesty-dishonesty; (2) helping-harming; (3) fairness-unfairness; and (4) loyalty-disloyalty.

For each remembered action, participants described the event in two to five sentences. Participants then reported when the event occurred (the same two measures used in Study 1 were also used in Study 2). As a manipulation check, participants answered the following: how morally wrong or morally

right was your behavior in this instance? (1 = *very morally wrong*, 7 = *very morally right*). Participants then answered the following three questions in random order: as you think about this memory now, do you feel like you are a *different* person now than you were then? (1 = *definitely no*, 7 = *definitely yes*); as you think about this memory now, do you feel like you are the *same* person now that you were then? (1 = *definitely no*, 7 = *definitely yes*); how much have you *changed* as a person since this event occurred? (1 = *not at all*, 7 = *a lot*). After completing the study, participants were monetarily compensated for their time.

Data analyses Data were analyzed in two ways. First, separate paired-samples t-tests were computed to investigate mean differences in the four outcome variables (i.e., judged morality, different, same, and change) as a function of memory cue (morally right vs. morally wrong). Second, using R (R Development Core Team, 2009) and the lme4 software package (Bates, Maechler, Bolker, & Walker, 2015), data were fit to linear mixed-effects models that included the memory cue as a predictor of the different, same, and change outcome variables in separate models. $Time_A$ and $time_B$ were included as controls in separate models (due to multi-collinearity, they could not both be included in the same model), and subject was included as a random effect (random intercepts only) in all models. As in Study 1, significance for fixed effects was assessed using Satterthwaite approximations to degrees of freedom, and 95% CIs around beta-values were computed using parametric bootstrapping. The alpha level for all statistical tests was set at .05.

Results

Four separate paired-samples t-tests were initially computed to verify that the remembered actions generated from the morally wrong cue were, in fact, judged to be more morally wrong than the remembered actions generated from the morally right cue. Confirming this expectation, remembered actions generated from the morally wrong cue were rated as more morally wrong on the 7-pt scale than those generated from the morally right cue for each matched pair: honesty-dishonesty (M_{diff} = 3.75, $t(98)$ = 24.41, $p < .001$, 95% CI = [3.44, 4.05]), helping-harming (M_{diff} = 4.08, $t(98)$ = 29.06, $p < .001$, 95% CI = [3.80, 4.36]), fairness-unfairness (M_{diff} = 3.82, $t(98)$ = 30.04, $p < .001$, 95% CI = [3.57, 4.07]), and loyalty-disloyalty (M_{diff} = 3.99, $t(98)$ = 23.70, $p < .001$, 95% CI = [3.66, 4.32]). All reported 95% CIs are for mean differences. See Table 1 for descriptive statistics.

For each matched pair, we investigated the extent to which participants believed they were *different* people when they recalled the event than they were when the event actually occurred as a function of recalling a morally right or wrong past action. Paired-samples t-tests revealed that participants

Table 1 Summary of means (SDs) for each measured variable as a function of memory cue

Memory cue	Judged Morality	Different	Same	Change
Honesty	6.05 (0.81)	2.97 (1.98)	5.16 (1.92)	3.08 (2.02)
Dishonesty	2.30 (1.14)	4.18 (2.18)	3.93 (2.15)	4.07 (2.12)
Helping	6.27 (0.75)	2.39 (1.78)	5.91 (1.65)	2.51 (1.81)
Harming	2.19 (1.08)	4.49 (2.10)	3.55 (2.05)	4.48 (2.12)
Fairness	6.11 (0.87)	2.53 (1.79)	5.56 (1.69)	2.59 (1.74)
Unfairness	2.29 (0.80)	4.52 (2.07)	3.60 (2.04)	4.13 (2.01)
Loyalty	6.20 (0.94)	3.08 (1.92)	5.07 (1.84)	3.28 (1.86)
Disloyalty	2.21 (1.04)	4.36 (2.08)	3.68 (2.08)	4.38 (2.11)

Note. 7-pt scales were used for all four measured variables

felt more dissimilar to their past selves after recalling morally wrong actions relative to morally right actions for honesty-dishonesty (M_{diff} = 1.21, $t(98)$ = 4.32, p < .001, 95% CI = [.66, 1.77]), helping-harming (M_{diff} = 2.11, $t(98)$ = 7.86, p < .001, 95% CI = [1.58, 2.64]), fairness-unfairness (M_{diff} = 1.99, $t(98)$ = 7.71, p < .001, 95% CI = [1.48, 2.50]), and loyalty-disloyalty (M_{diff} = 1.28, $t(98)$ = 4.48, p < .001, 95% CI = [.71, 1.85]). See Table 1. Subsequent LMERs were then computed, each with the cued morality of the past action (morally wrong vs. morally right) predicting the extent to which participants believed they were different people now than they were when the event actually occurred, all while statistically controlling for when the past experiences actually occurred. For each remembered pair of events, participants felt more dissimilar to their past selves in the morally wrong condition relative to the morally right condition, even after statistically controlling for time$_A$ and time$_B$ in separate models (all ps < .001; see Supplemental Materials for full models).

Next, for each matched pair, we investigated the extent to which participants believed they were the *same* people upon recalling the event as they were when the event actually occurred as a function of recalling a morally right or wrong past behavior. Paired-samples *t*-tests revealed that participants were more likely to feel the same as their past selves after recalling morally right actions relative to morally wrong actions for honesty-dishonesty (M_{diff} = 1.23, $t(98)$ = 4.42, p < .001, 95% CI = [.68, 1.79]), helping-harming (M_{diff} = 2.36, $t(98)$ = 9.25, p < .001, 95% CI = [1.86, 2.87]), fairness-unfairness (M_{diff} = 1.96, $t(98)$ = 7.63, p < .001, 95% CI = [1.45, 2.47]), and loyalty-disloyalty (M_{diff} = 1.39, $t(98)$ = 4.99, p < .001, 95% CI = [.84, 1.95]). See Table 1. Subsequent LMERs were then computed, each with the cued morality of the past action (morally wrong vs. morally right) predicting the extent to which participants believed they were the same people now that they were when the event occurred, all while statistically controlling for when these past experiences actually occurred. For each remembered pair of events, participants were more likely to feel the same as their past

selves in the morally right condition relative to the morally wrong condition, even after statistically controlling for time$_A$ and time$_B$ in separate models (all ps < .001; see Supplemental Materials for full models).

Finally, for each matched pair, we investigated how much participants believed they had *changed* since the remembered event occurred as a function of whether a morally right or wrong past action was recalled. Paired-samples *t*-tests revealed that participants believed that they had changed more since the remembered event occurred after recalling morally wrong actions relative to morally right actions for honesty-dishonesty (M_{diff} = .99, $t(98)$ = 3.59, p = .001, 95% CI = [.44, 1.54]), helping-harming (M_{diff} = 1.98, $t(98)$ = 7.61, p < .001, 95% CI = [1.46, 2.50]), fairness-unfairness (M_{diff} = 1.55, $t(98)$ = 6.26, p < .001, 95% CI = [1.06, 2.04]), and loyalty-disloyalty (M_{diff} = 1.10, $t(98)$ = 3.90, p < .001, 95% CI = [.54, 1.66]). See Table 1. Subsequent LMERs were then computed, each with the cued morality of the past action (morally wrong vs. morally right) predicting the extent to which participants believed they had changed since the remembered event occurred, all while statistically controlling for when these past experiences actually occurred. For each remembered pair of events, participants reported that they had changed more since the remembered events occurred in the morally wrong condition relative to the morally right condition, even after statistically controlling for time$_A$ and time$_B$ in separate models (all ps < .004; see Supplemental Materials for full models).[2]

Discussion

The results from Study 2 corroborate and extend our findings from Study 1. As in Study 1, we found that after recalling their own morally wrong actions from the personal past, individuals perceived greater change and dissimilarity in the self since the events occurred than they did after recalling their own morally right actions. This pattern of results was obtained for each matched pair: honesty versus dishonesty, helping versus harming, fairness versus unfairness, and loyalty versus disloyalty. In all cases, these effects remained even after statistically controlling for objective, calendar time.

Study 3

The results from Studies 1 and 2 provide some evidence that recollecting past immoral actions is associated with stronger feelings of dissimilarity and change in the sense of self over time than remembering past moral actions. Building upon our

---

[2] As suggested by an anonymous reviewer, we also conducted additional analyses investigating possible interaction effects between the cued morality of the remembered behavior and objective time variables. These results are reported in Supplemental Materials.

findings in Studies 1 and 2, the purpose of Study 3 is twofold. First, the within-subjects design in the previous studies could have created a demand characteristic: participants might have assumed that the experimenter wanted them to answer the outcome measures differently for remembered morally wrong versus morally right behaviors. To circumvent this issue, Study 3 implements a between-subjects design. Second, the effects in Study 1 could have been produced by some other process related to how people are perceived more generally, rather than a motivational process about the self. In order to provide more direct support for the explanation that these effects resulted from a motivational process about the self, an additional condition was included: some participants were asked to provide memories of witnessing *other* people commit morally right or wrong actions. If the difference between morally right and wrong conditions in different, same, and change measures is exclusive to memories about the self committing morally right and wrong actions, then the explanation for our pattern of results is likely motivational.

Materials and method

Participants Two-hundred and sixty individuals voluntarily participated in this study via AMT for monetary compensation. Participant recruitment was restricted to individuals in the USA who had completed at least 50 HITs on AMT and had obtained an approval rating above 85%. Twenty-two participants were excluded for failing to provide at least one memory, for recalling an event that occurred more than 10 years ago, or for providing the incorrect type of memory based on the cue. As such, data were analyzed with the remaining 238 participants (M_{age} = 36.26 years, SD = 10.51, age range = 18–70 years, 118 females, 119 males). All participants reported being fluent English speakers. Those who participated in the previous studies were prevented from participating in this study. Informed consent was obtained from each participant in accordance with the protocol approved by the Duke University Campus Institutional Review Board.

Procedure The study was self-paced. We manipulated both the nature of the remembered behavior (morally wrong vs. morally right) and the person who committed the act (self vs. other). Participants were randomly assigned to one of these four conditions in a between-subjects fashion: (1) participants recalled three morally right behaviors that they themselves committed, (2) three morally wrong behaviors that themselves committed, (3) three morally right behaviors committed by another person, or (4) three morally wrong behaviors committed by another person. Participants were told that these remembered actions must have occurred within the past 10 years on a particular day in a particular place.

For each remembered behavior, participants described the event in two to five sentences. Participants then reported when the event occurred (the same two measures used in Studies 1 and 2 were also used in Study 3). As a manipulation check, participants answered the following: how morally wrong or morally right was the remembered behavior in this instance? (1 = *very morally wrong*, 7 = *very morally right*). Participants then answered the *different*, *same*, and *change* questions in a random order. *Different*, *same*, and *change* judgments were always made about the individual who behaved morally or immorally. Upon completion of the study, participants were monetarily compensated for their time.

Data analyses Using R (R Development Core Team, 2009) and the lme4 software package (Bates, Maechler, Bolker, & Walker, 2015), data were fit to linear mixed-effects models. Fixed effects and outcome variables differed depending upon the model, but subject was included as a random effect (random intercepts only) in all models. $Time_A$ and $time_B$ were included as controls in separate models (due to multi-collinearity, they could not both be included in the same model). As in previous studies, significance for fixed effects was assessed using Satterthwaite approximations to degrees of freedom, and 95% CIs around beta-values were computed using parametric bootstrapping. The alpha level for all statistical tests was set at .05.

Results

An initial linear mixed-effects model was computed to ensure that the remembered actions generated from the morally wrong cue were, in fact, judged to be more morally wrong on the 7-pt scale than the remembered actions generated from the morally right cue. This expectation was confirmed (b = 3.36, SE = .16, t = 21.45, p < .001, 95% CI = [3.04, 3.68]).

We investigated the effects of the cued morality of the action (binary, fixed factor: morally wrong vs. morally right) and the person who committed the action (binary, fixed factor: self vs. other) on *different*, *same*, and *change* judgments in separate linear mixed-effects models. The interaction between the cued morality of the action (morally wrong coded as 1, morally right coded as 0) and the person committing the action (self coded as 1, other coded as 0) was significant for all three different outcome variables (i.e., different, same, and change; all ps < .005). Table 2 depicts full results from these three models.

To further interrogate these interaction effects, we computed follow-up linear mixed-effects models for self and other conditions taken separately. In the *self* condition, there was a significant effect of the cued morality of the action on the extent to which people believe they are different now relative to who they were when the event occurred (b = 1.43, SE = .30, t = 4.71, p < .001, 95% CI = [.81, 2.00]), the extent to which people believe they are the same now relative to who they were when the event occurred (b = -1.27, SE = .27, t = -

Table 2 The results of three different linear mixed-effects models with the cued morality and the person committing the action as predictors of different, same, and change judgments, respectively

	b	SE	t-value	p-value	95% CI
Outcome: Different					
Cued Morality	-.18	.29	-.60	.55	[.77, .41]
Person Committing the Action	-.53	.31	-1.73	.08	[-1.16, .06]
Cued Morality × Person Committing the Action	1.63	.43	3.76	< .001	[.82, 2.45]
Outcome: Same					
Cued Morality	-.12	.26	-.46	.65	[-.66, .40]
Person Committing the Action	.43	.27	1.55	.12	[-.14, .94]
Cued Morality × Person Committing the Action	-1.13	.39	-2.93	.004	[-1.87, -.35]
Outcome: Change					
Cued Morality	-.26	.31	-.85	.39	[-.86, .40]
Person Committing the Action	-.20	.32	-.63	.53	[-.98, .45]
Cued Morality × Person Committing the Action	1.31	.45	2.89	.004	[.34, 2.33]

Note. All 95% CIs are for the beta-estimates

4.76, $p < .001$, 95% CI = [-1.83, -.76]), and the extent to which they believe they have changed since the event occurred (b = 1.05, SE = .32, t = 3.26, p = .001, 95% CI = [.50, 1.67]; see Fig. 2). When participants recalled their own morally wrong actions relative to their own morally right actions, they judged themselves to be more dissimilar to their past selves who committed the actions and they judged themselves to have changed more since the event occurred. These effects persisted at a similar magnitude even after statistically controlling for time$_A$ and time$_B$ in separate models (all ps < .004; see Supplemental Materials for full models).

In the *other* condition, linear mixed-effects models revealed no significant effects of the cued morality of the action on different (b = -17, SE = .31, t = -.55, p = .58, 95% CI = [-.80, .51]), same (b = -.19, SE = .28, t = -.67, p = .51, 95% CI = [-.77, .36]), or change (b = -.21, SE = .32, t = -.66, p = .51, 95% CI = [-.83, .42]) judgments. So, the effect of recalling morally wrong versus morally right actions on different, same, and change judgments are evident in the *self* but not in the *other* condition.

Discussion

The pattern of results from Studies 1 and 2 was replicated in Study 3 using a between-subjects design instead of a within-subjects design. That is, participants' recollections of their own past immoral actions were accompanied by stronger feelings of dissimilarity and change in the self over time relative to recollections of their own past moral actions. This effect persisted even after statistically controlling for objective, calendar time. Study 3 also provides more direct support for a motivational explanation for our pattern of results that is about the self. Participants perceived greater change and dissimilarity in their selves after recalling their own past moral

transgressions relative to their own past morally praiseworthy deeds, but there were no significant differences between remembered morally right and wrong actions committed by other people.

Study 4

In the previous three studies, we found that people report perceiving greater change and dissimilarity in their selves after recalling their own past immoral behaviors relative to moral behaviors. In Study 4, we investigate whether people specifically report greater *positive* change in particular moral traits (e.g., honesty, loyalty) over time after recollecting their past immoral behavior relative to moral behaviors. While recalling an immoral behavior from the personal past in which they were dishonest, participants might be motivated to believe they are more honest now than they were when the event occurred. By believing that they have become more honest since they acted dishonestly, they can still foster a positive view of the current self as morally good by perceiving their moral improvement over time. In contrast, after recalling a morally good behavior from the personal past in which they were honest, there would be no motivation to perceive positive change since the event occurred. We further expect that the perceived changes in the possession of particular moral traits after recalling morally right versus wrong past behaviors will remain after accounting for variance attributable to objective, calendar time.

Materials and method

Participants Eight-hundred and sixty individuals voluntarily participated in this study via AMT for monetary
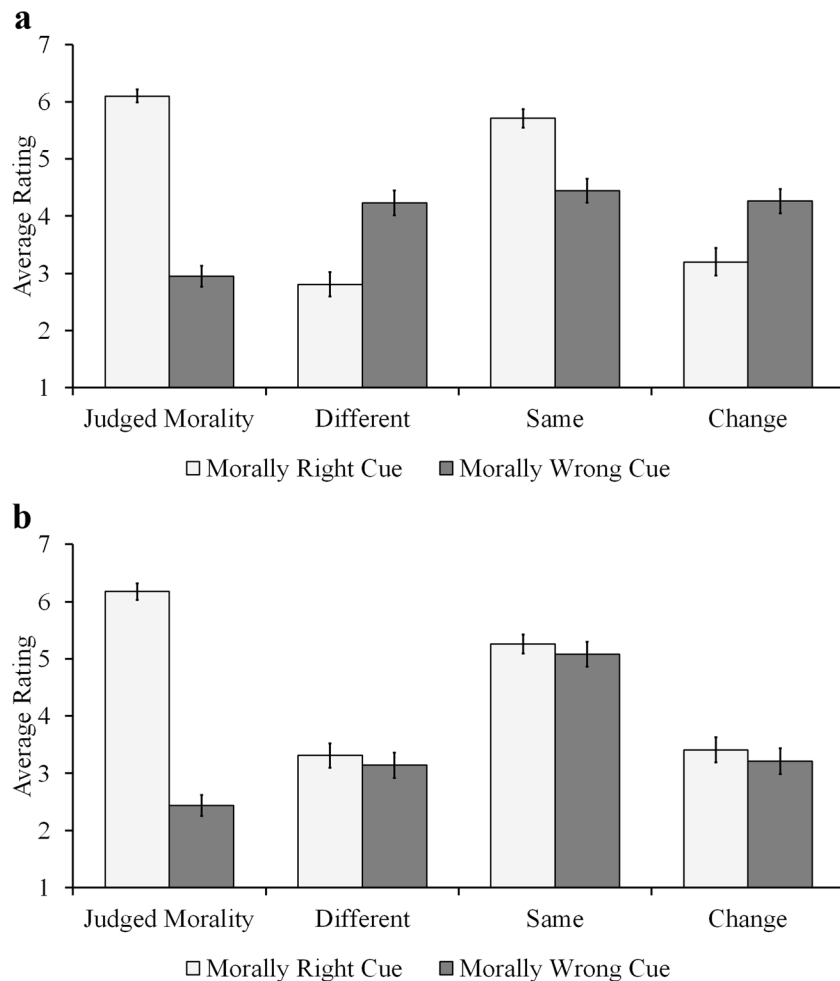
a

b

Fig. 2 For the *self* (**a**) and *other* (**b**) conditions, means and standard error bars are depicted for the judged morality of the remembered actions, different judgments, same judgments, and change judgments, all as a function of the memory cue (morally right or morally wrong)

compensation. Participant recruitment was restricted to individuals in the USA who had completed at least 50 HITs on AMT and had obtained an approval rating above 85%. Fifty-six participants were excluded for failing to provide at least one memory, for providing a memory of an event that occurred more than 10 years ago, or for providing the incorrect type of memory based on the cue. As such, data were analyzed with the remaining 804 participants (M_{age} = 34.68 years, SD = 10.78, age range = 19–86 years, 322 females, 480 males). All participants reported being fluent English speakers. Those who participated in any of the previous studies were automatically prevented from participating in Study 4. Informed consent was obtained from each participant in accordance with the protocol approved by the Duke University Campus Institutional Review Board.

Procedure The study was self-paced. Participants were asked to recall one event from the personal past that occurred within the past 10 years. There were eight different conditions to

which participants were randomly assigned in a between-subjects fashion: (1) recall a specific past experience in which you were honest with another person, and you believe your action was morally right; (2) recall a specific past experience in which you were dishonest with another person, and you believe your action was morally wrong; (3) recall a specific past experience in which you helped another person, and you believe your action was morally right; (4) recall a specific past experience in which you harmed another person, and you believe your action was morally wrong; (5) recall a specific past experience in which you treated another person fairly, and you believe your action was morally right; (6) recall a specific past experience in which you treated another person unfairly, and you believe your action was morally wrong; (7) recall a specific past experience in which you were loyal to another person, and you believe your action was morally right; and (8) recall a specific past experience in which you were disloyal to another person, and you believe your action was morally wrong. As in Study 2, this cueing procedure ultimately produced four distinct matched pairs of remembered actions – one morally right and the other morally wrong – for each

particular kind of behavior: (1) honesty-dishonesty; (2) helping-harming; (3) fairness-unfairness; and (4) loyalty-disloyalty. However, unlike Study 2, all comparisons between matched pairs of remembered actions were between-subjects.

After recalling an event from the personal past, participants described the event in two to five sentences. Participants then reported when the event occurred (the same two measures used in the three previous studies were also used in Study 4). As a manipulation check, participants answered the following: how morally wrong or morally right was your behavior in this instance? (1 = *very morally wrong*, 7 = *very morally right*). For our outcome variables of interest, participants made one of four different ratings depending upon the particular kind of behavior remembered. Participants who recalled an event involving honesty or dishonesty answered the following question: do you believe you are a more honest person now than the person you were when this event occurred? (1 = *definitely no*, 7 = *definitely yes*). Participants who recalled an event involving helping or harming answered the following question: do you believe you are a more helpful person now than the person you were when this event occurred? (1 = *definitely no*, 7 = *definitely yes*). Participants who recalled an event involving fairness or unfairness answered the following question: do you believe you are a more fair person now than the person you were when this event occurred? (1 = *definitely no*, 7 = *definitely yes*). And participants who recalled an event involving loyalty or disloyalty answered the following question: do you believe you are a more loyal person now than the person you were when this event occurred? (1 = *definitely no*, 7 = *definitely yes*). All participants also answered the following questions about the phenomenology of their memories: how detailed is your memory of this event? (1 = *not at all detailed*, 7 = *very detailed*); how vivid is your memory of this event? (1 = *very vague*, 7 = *very vivid*); how clear is your memory of this event (1 = *very unclear*, 7 = *very clear*). These phenomenology measures were included to help conceal the purpose of the study. After completing the study, participants were monetarily compensated for their time.

Data analyses Data were analyzed in two ways. First, separate independent-samples t-tests were computed to investigate mean differences in the outcome variables (i.e., perceived changes in honesty, helpfulness, fairness, and loyalty since the remembered event occurred) as a function of memory cue (morally right vs. morally wrong). Second, data were fit to regression models that included the memory cue (morally wrong coded as 1, morally right coded as 0) as a predictor of perceived changes in honesty, helpfulness, fairness, and loyalty in separate models. $Time_A$ and $time_B$ were then included as controls in separate regression models (due to multi-collinearity, they could not both be included in the same model). The alpha level for all statistical tests was set at .05.

Results

Four separate independent-samples t-tests were initially computed to verify that the remembered actions generated from the morally wrong cue were, in fact, judged to be more morally wrong than the remembered actions generated from the morally right cue. Confirming this expectation, remembered actions generated from the morally wrong cue were rated as more morally wrong on the 7-pt scale than those generated from the morally right cue for each matched pair: honesty-dishonesty ($n = 199$, $M_{diff} = 3.62$, $t(197) = 22.34$, $p < .001$, 95% CI = [3.30, 3.94]), helping-harming ($n = 197$, $M_{diff} = 3.95$, $t(195) = 25.58$, $p < .001$, 95% CI = [3.65, 4.26]), fairness-unfairness ($n = 203$, $M_{diff} = 3.51$, $t(201) = 22.51$, $p < .001$, 95% CI = [3.21, 3.82]), and loyalty-disloyalty ($n = 205$, $M_{diff} = 3.73$, $t(203) = 20.49$, $p < .001$, 95% CI = [3.37, 3.38]). All reported 95% CIs are for mean differences.

Honesty-dishonesty For memories involving honesty and dishonesty, we investigated the effect of the cued morality of the action (morally wrong vs. morally right) on the extent to which participants judged themselves to be more honest at the time of recalling the event than they were when the event actually occurred. Relative to participants cued to provide a memory of a morally right, honest behavior ($M = 4.27$, $SD = 1.72$), those participants cued to provide a memory of a morally wrong, dishonest behavior ($M = 4.80$, $SD = 1.79$) judged themselves to be more honest at the time of recalling the event than they were when the event actually occurred ($n = 199$, $M_{diff} = .53$, $t(197) = 2.13$, $p = .028$, 95% CI = [.04, 1.02]). This effect remained at a similar magnitude even after statistically controlling for $time_A$ and $time_B$ in separate regression models (both $ps < .05$; see Supplemental Materials for full models).

Helping-harming For memories involving helping and harming, we investigated the effect of the cued morality of the action (morally wrong *versus* morally right) on the extent to which participants judged themselves to be more helpful at the time of recalling the event than they were when the event actually occurred. Relative to participants cued to provide a memory of a morally right, helpful behavior ($M = 4.68$, $SD = 1.57$), those participants cued to provide a memory of a morally wrong, harmful behavior ($M = 5.21$, $SD = 1.65$) judged themselves to be more helpful at the time of recalling the event than they were when the event actually occurred ($n = 197$, $M_{diff} = .53$, $t(195) = 2.30$, $p = .022$, 95% CI = [.08, .98]). However, this effect was rendered non-significant after statistically controlling for $time_A$ ($p = .052$) and $time_B$ ($p > .10$) in separate regression models (see Supplemental Materials for full models).

Fairness-unfairness For memories involving fairness and un-fairness, we investigated the effect of the cued morality of the action (morally wrong vs. morally right) on the extent to which participants judged themselves to be more fair at the time of recalling the event than they were when the event actually occurred. Relative to participants cued to provide a memory of a morally right, fair behavior ($M = 4.33$, $SD = 1.66$), those participants cued to provide a memory of a morally wrong, unfair behavior ($M = 5.11$, $SD = 1.74$) judged themselves to be more fair at the time of recalling the event than they were when the event actually occurred ($n = 203$, $M_{diff} = .78$, $t(201) = 3.25$, $p = .001$, 95% CI = [.31, 1.25]). This effect remained at a similar magnitude even after statistically controlling for time$_A$ and time$_B$ in separate regression models (both ps < .006; see Supplemental Materials for full models).

Loyalty-disloyalty For memories involving loyalty and disloy-alty, we investigated the effect of the cued morality of the action (morally wrong vs. morally right) on the extent to which participants judged themselves to be more loyal at the time of recalling the event than they were when the event actually occurred. Relative to participants cued to provide a memory of a morally right, loyal behavior ($M = 4.56$, $SD = 1.66$), those participants cued to provide a memory of a mor-ally wrong, disloyal behavior ($M = 5.16$, $SD = 1.72$) judged themselves to be more loyal at the time of recalling the event than they were when the event actually occurred ($n = 205$, $M_{diff} = .61$, $t(203) = 2.56$, $p = .011$, 95% CI = [.14, 1.07]). This effect remained at a similar magnitude even after statis-tically controlling for time$_A$ and time$_B$ in separate regression models (both ps < .03; see Supplemental Materials for full models).

Discussion

Overall, the results from Study 4 indicate that people report greater positive change in particular moral traits over time after recalling their past immoral relative to moral behaviors. This effect was obtained for all four moral traits investigated: honesty, helpfulness, fairness, and loyalty. After recollecting memories of morally wrong behaviors, there seems to be a motivation to perceive positive change in the self since the events occurred. That is, to counteract a threat to a morally good view of self, participants tend to perceive positive change over time in the extent to which they possess moral traits. In contrast, after recollecting their morally right behav-iors, there is no need to perceive moral improvement over time in service of repairing a negative view of the self. Furthermore, for memories involving honesty-dishonesty, fairness-unfairness, and loyalty-disloyalty, the perceived changes in the possession of these moral traits (after recalling morally right versus wrong past events) remained at a similar magnitude even after statistically controlling for objective,

calendar time. So, regardless of how long ago the event actu-ally occurred, people still perceived meaningful, positive psy-chological change in the self.

General discussion

Autobiographical memories are not literal records of what has occurred in the personal past (Conway, 2005; D'Argembeau & Van der Linden, 2008; Schacter, 1999; Schacter, Guerin, & St. Jacques, 2011). Current motivations affect which personal past experiences are recalled, how those events are recon-structed, and how those events are interpreted within a life narrative. In four studies, we found consistent evidence for a particularly strong bias in how past experiences are exploited to foster a positive moral self-image. After recollecting their own moral transgressions from the personal past, participants reported feeling dissimilar to their past selves and perceiving significant change in the self since those events occurred. In contrast, after recalling their own morally praiseworthy past actions, participants reported feeling more similar to those past selves and perceiving less change in the self since those events occurred. Not only did these effects hold for diverse domains of morality (i.e., honesty/dishonesty, helping/ harming, fairness/unfairness, and loyalty/disloyalty), but they also persisted after statistically controlling for when the events actually occurred in the past (objective, calendar time). Supporting a motivational explanation, we consistently found effects of recalling morally right versus wrong actions on judgments of change over time when those past actions were committed by the participants themselves; however, there were no significant effects when participants recalled morally right and wrong actions committed by other people.

Overall, the current findings support our hypothesis that people strategically construct a positive, morally good sense of self in the present (1) by perceiving similarity in the self after reflecting on their own morally praiseworthy actions from the past and (2) by perceiving meaningful change or transformation in the self after reflecting on their own immoral actions from the past. Moreover, these perceptions of similar-ity and change in the self after recollecting past moral and immoral behaviors manifested in both general and specific ways. More generally, participants reported perceiving greater global change and dissimilarity in their selves after recalling their past moral transgressions relative to their past morally praiseworthy deeds (Studies 1–3). More specifically, partici-pants reported perceiving greater positive change in the extent to which they exemplified particular moral traits (e.g., hones-ty, loyalty) over time after recalling their past immoral relative to moral behaviors (Study 4). These results add to a substantial literature suggesting that the ways in which we remember and interpret our past experiences serve self-enhancement and self-protective functions (Leary, 2007; Wilson & Ross,

2003). For example, people tend to believe they have undergone personal improvement over time to a greater extent than other people (Wilson & Ross 2001), people tend to take greater personal responsibility for their own past successes than their own past failures (Blaine & Crocker 1993), and people are more likely to engage in self-enhancing deceptions involving their current selves than their past selves (Robinson & Ryff, 1999). Going beyond this existing literature, our findings suggest that how we remember and interpret our past experiences helps to construct a current sense of self that is not just positive but also morally good. This result is particularly important because moral traits and characteristics are considered to be the most important, fundamental, or central part of personal identity, or the self (Chen, Urminsky, & Bartels, 2016; De Freitas et al., 2017; Molouki & Bartels, 2017; Strohminger & Nichols, 2014; Strohminger et al., 2017).

Philosophers and psychologists have used the terms "self" and "identity" in several distinct ways. For our purposes, there are at least two ways of thinking about the self, or personal identity, that are worth differentiating: (1) numerical identity is the sense in which a single thing persists over time and (2) qualitative identity refers to the sharing of particular properties over time (Starmans & Bloom, 2018). In assessing numerical identity, we might say that infant Neil Armstrong is, for example, identical to adult Neil Armstrong. But in assessing qualitative identity over time, we might say that Neil Armstrong prior to walking on the moon is dissimilar to, or not the same as, Neil Armstrong after walking on the moon, because walking on the moon must be a particularly transformative experience. As in our previous work (Stanley et al., 2017), we believe that the current findings specifically inform individuals' judgments about qualitative identity over time. That is, in one sense, a given person is still the same person before and after committing a serious moral transgression, so numerical identity remains fixed over time. But, in another sense, after reflecting on a particular moral transgression several months later, a person might report feeling very different from his past self or feeling as though he has undergone significant change since that remembered transgression occurred; so, qualitative identity changes in meaningful, predictable, and systematic ways over time.

Almost all existing research on morality and the self has utilized stylized vignettes, not actual memories of personal past experiences. Although this existing research has produced valuable insights into the relationship between morality and the self, we suggest that investigating autobiographical memories of moral and immoral behaviors that actually occurred in the past is necessary for obtaining a complete picture of how people construct and understand their moral selves. By remembering and reflecting on events from the personal past, we come to construct and understand who we believe ourselves to be in the present (Conway, 2005; McAdams, 2013;

Wilson, Gunn, & Ross, 2009). Biases, distortions, and motivations in how we remember our personal pasts play critical roles in how we view our selves in the present, and these biases, distortions, and motivation are unlikely to play a role in vignette-based investigations of morality and the self.

There are positive consequences of viewing the self accurately and truthfully, even if that that view of the self isn't particularly flattering. Maintaining an accurate, truthful view of the self can facilitate sensible future planning, enable goal success, and cultivate personal and social well-being (Strube, 1990; Trope, 1986). Nevertheless, converging lines of evidence now indicate that enhancing or protecting a positive view of the self is frequently pursued at the expense of obtaining an accurate, truthful view of the self. Of course, views of the self do not always have to be veridical to positively affect intentions and behavior: an unrealistically favorable self-view induces feelings of efficacy and a willingness to engage in decisive action (Taylor, Lerner, Sherman, Sage, & McDowell, 2003). But sometimes there are significant costs that accompany an overriding desire view the self in a positive way. In the moral domain, people seem particularly susceptible to forgetting, dismissing, or *post hoc* justifying their own moral transgressions for the sake of maintaining a positive, but inaccurate, view of the self as morally upstanding. Many people who consider themselves to be morally upstanding still cheat, steal, lie, deceive, and harm others. In fact, utilizing ecological momentary assessment, Hofmann et al. (2014) recently showed that people behave immorally with surprising frequency. Moral transgressions, even those that are relatively minor, often have deleterious social and financial consequences (Kouchaki & Gino, 2016; Shalvi, Gino, Barkan, & Ayal, 2015). An important goal of our future research will be to develop strategies for encouraging people to develop accurate, truthful views of the moral self, specifically in those contexts where accurate, truthful views of the moral self are better suited to impact intentions and behavior in a positive way.

Finally, it is worth mentioning a possible alternative interpretation of our results. Because the tasks we employed instructed participants to recall moral and immoral autobiographical memories, it is possible that reported impressions of change (or lack thereof) in the self across time simply reflect the extent to which they embrace or disavow the remembered action. As such, our results may reflect participants' feelings toward the remembered action rather than a definitive perception of change or sameness in the self across time.[3] The problem with this alternative account is that it is not clear whether embracing or disavowing a remembered action is independent of, or even explanatorily prior to, a person's perception of change in the self over time. Do we disavow immoral actions

_____

[3] We thank an anonymous reviewer for suggesting this alternative interpretation.

because we feel that we are different than we were before, or do we feel that we are different than we were before because we disavow such immoral actions? It is reasonable to think that people are as motivated to disavow a past immoral action because, upon reflection, they do not think of themselves as the same person who committed it, as they are to think of themselves as a different person than they were before precisely because they are motivated to disavow the remembered action. If these two explanations are truly distinct, then it is an empirical question to uncover right order of explanation. This, we think, is another fruitful avenue for future research on the self and moral memories.

Compliance with ethical standards

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.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effect models using lme4. *Journal of Statistical Software, 67*, 1-48.

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

Blaine, B., & Crocker, J. (1993). Self-esteem and self-serving biases in reactions to positive and negative events: An integrative review. Self-esteem (pp. 55-85). Springer, Boston, MA.

Bluck, S. (2003). Autobiographical memory: Exploring its functions in everyday life. *Memory, 11*(2), 113-123.

Chen, S. Y., Urminsky, O., & Bartels, D. M. (2016). Beliefs about the causal structure of the self-concept determine which changes disrupt personal identity. *Psychological Science, 27*, 1398-1406.

Conway, M. A. (2005). Memory and the self. *Journal of Memory and Language, 53*(4), 594-628.

Conway, M. A., Singer, J. A., & Tagini, A. (2004). The self and autobiographical memory: Correspondence and coherence. *Social Cognition, 22*, 491-529.

D'Argembeau, A., & Van der Linden, M. (2008). Remembering pride and shame: Self-enhancement and the phenomenology of autobiographical memory. *Memory, 16*, 538-547.

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.

Demiray, B., & Janssen, S. M. (2015). The self-enhancement function of autobiographical memory. *Applied Cognitive Psychology, 29*, 49-60.

Escobedo, J. R., & Adolphs, R. (2010). Becoming a better person: Temporal remoteness biases autobiographical memories for moral events. *Emotion, 10*, 511-518.

Fivush, R., Habermas, T., Waters, T. E., & Zaman, W. (2011). The making of autobiographical memory: Intersections of culture, narratives and identity. *International Journal of Psychology, 46*(5), 321-345.

Gebauer, J. E., Broemer, P., Haddock, G., & Von Hecker, U. (2008). Inclusion-exclusion of positive and negative past selves: Mood congruence as information. *Journal of Personality and Social Psychology, 95*(2), 470.

Greenwald, A. G. (1980). The totalitarian ego: Fabrication and revision of personal history. *American Psychologist, 35*(7), 603.

Heiphetz, L., Strohminger, N., & Young, L. L. (2017). The role of moral beliefs, memories, and preferences in representations of identity. *Cognitive Science, 41*(3), 744-767.

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

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

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

Kunda, Z., Fong, G. T., Sanitioso, R., & Reber, E. (1993). Directional questions direct self-conceptions. *Journal of Experimental Social Psychology, 29*(1), 63-86.

Kunda, Z., & Sanitioso, R. (1989). Motivated changes in the self-concept. *Journal of Experimental Social Psychology, 25*(3), 272-285.

Leary, M. R. (2007). Motivational and emotional aspects of the self. *Annu. Rev. Psychol., 58*, 317-344.

Libby, L. K., & Eibach, R. P. (2002). Looking back in time: Self-concept change affects visual perspective in autobiographical memory. *Journal of Personality and Social Psychology, 82*, 167-179.

Markus, H., & Wurf, E. (1987). The dynamic self-concept: A social psychological perspective. *Annual Review of Psychology, 38*(1), 299-337.

McAdams, D. P. (2013). The psychological self as actor, agent, and author. *Perspectives on Psychological Science, 8*(3), 272-295.

Molouki, S., & Bartels, D. M. (2017). Personal change and the continuity of the self. *Cognitive Psychology, 93*, 1-17.

Pasupathi, M., Mansour, E., & Brubaker, J. R. (2007). Developing a life story: Constructing relations between self and experience in autobiographical narratives. *Human Development, 50*(2-3), 85-110.

Robinson, M. D., & Ryff, C. D. (1999). The role of self-deception in perceptions of past, present, and future happiness. *Personality and Social Psychology Bulletin, 25*(5), 596-608.

Ross, M. (1989). The relation of implicit theories to the construction of personal histories. *Psychological Review, 96*, 341-357.

Ross, M., & Wilson, A. E. (2002). It feels like yesterday: Self-esteem, valence of personal past experiences, and judgments of subjective distance. *Journal of Personality and Social Psychology, 82*, 792-803.

Ryff, C. D. (1991). Possible selves in adulthood and old age: A tale of shifting horizons. *Psychology and Aging, 6*, 286-295.

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

Sanitioso, R. B., & Wlodarski, R. (2004). In search of information that confirms a desired self-perception: Motivated processing of social feedback and choice of social interactions. *Personality and Social Psychology Bulletin, 30*(4), 412-422.

Schacter, D. L. (1999). The seven sins of memory: Insights from psychology and cognitive neuroscience. *American Psychologist, 54*(3), 182.

Schacter, D. L., Guerin, S. A., & St. Jacques, P. L. (2011). Memory distortion: An adaptive perspective. *Trends in Cognitive Sciences*, *15*, 467-474.

Sedikides, C. (1993). Assessment, enhancement, and verification determinants of the self-evaluation process. *Journal of Personality and Social Psychology, 65*, 317-338.

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

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.

Stanley, M. L., Henne, P., Iyengar, V., Sinnott-Armstrong, W., & De Brigard, F. (2017). I'm not the person I used to be: The self and autobiographical memories of immoral actions. *Journal of Experimental Psychology: General*, *146*(6), 884.

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(5): 787-795.

Starmans, C., & Bloom, P. (2018). Nothing personal: What psychologists get wrong about identity. *Trends in Cognitive Sciences, 22*, 566-568.

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*, *1*, 159–171.

Strohminger, N., & Nichols, S. (2015). Neurodegeneration and identity. *Psychological Science*, *26*, 1469-1479.

Strube, M. J. (1990). In search of self: Balancing the good and the true. *Personality and Social Psychology Bulletin*, *16*(4), 699-704.

Taylor, S. E., Lerner, J. S., Sherman, D. K., Sage, R. M., & McDowell, N. K. (2003). Are self-enhancing cognitions associated with healthy or unhealthy biological profiles?. *Journal of Personality and Social Psychology*, *85*(4), 605.

Trope, Y. (1986). Self-enhancement and self-assessment in achievement behavior. In R. M. Sorrentino & E. T. Higgins (Eds.), Handbook of motivation and cognition: Foundations of social behavior (pp. 350-378). New York, NY, US: Guilford Press.

Wilson, A. E., Gunn, G. R., & Ross, M. (2009). The role of subjective time in identity regulation. *Applied Cognitive Psychology*, *23*(8), 1164-1178.

Wilson, A. E., & Ross, M. (2000). The frequency of temporal-self and social comparisons in people's personal appraisals. *Journal of Personality and Social Psychology*, *78*, 928-942.

Wilson, A. E., & Ross, M. (2001). From chump to champ: People's appraisals of their earlier and present selves. *Journal of Personality and Social Psychology*, *80*, 572-584.

Wilson, A. E., & Ross, M. (2003). The identity function of autobiographical memory: Time is on our side. *Memory*, *11*, 137-149.