

## Research article

# Morality shifting in the context of intergroup violence

BERNHARD LEIDNER<sup>1\*</sup> AND EMANUELE CASTANO<sup>2</sup>

<sup>1</sup>University of Massachusetts Amherst, Amherst, USA; <sup>2</sup>New School for Social Research, New York, USA

### Abstract

We propose morality shifting as a mechanism through which individuals can maintain a moral image of the ingroup. We argue that a shift from the moral principles of harm and fairness to those of loyalty and authority occurs when assessing a potentially threatening event, particularly among high ingroup glorifiers. Three studies confirmed this hypothesis using three different methodologies. Study 1 compared the use of language related to four moral foundations formulated in moral psychology in response to ingroup- and outgroup-committed wrongdoings. Results showed that loyalty- and authority-related words were used more, whereas harm- and fairness-related words were used less in response to ingroup- compared with outgroup-committed wrongdoings. Study 2 replicated this effect with regards to the cognitive accessibility of these moral principles. Study 3 confirmed that morality shifting is a motivated response to social identity threat, rather than a response to mere activation of social identity. Finally, as predicted, Study 3 demonstrated the effect of morality shifting to be moderated by ingroup glorification but not ingroup attachment. Implications and consequences for intergroup and individual wrongdoings, as well as for intergroup relations, are discussed. Copyright © 2011 John Wiley & Sons, Ltd.

In one of the key passages in Graham Greene's book, *The Quiet American*, the intellectual, serious, and idealistic CIA operative Alden Pyle explains to British journalist Thomas Fowler that the bombing of a crowd in the center of a Vietnamese city he has just organized should not be understood as an atrocity but rather as a necessary and morally appropriate act that will change the course of history. Generally, the reader tends to sympathize with Fowler, rather than Pyle, and condemns the bombing for the suffering that it causes. To discount Pyle's view as abhorrent, however, would be naïve, for morality is less of an absolute than we would like to think.

When evaluating the morality of an event, people use various moral foundations (e.g., harm, fairness, loyalty, authority; Haidt & Graham, 2007). Thus, the same event can be judged as utterly immoral or its very opposite, depending upon which moral foundation is guiding the evaluation of the event. Because a judgment of the morality of a specific event/behavior can be consequential for the self, which moral foundation guides the evaluation can be influenced not only by culture (e.g., liberals versus conservatives; Americans versus Hindu Indians) but also by the immediate social context (e.g., the motives of the perceiver). We propose the concept of *Morality Shifting* to account for this self-serving process. Although this process is thought to operate at the individual level as well, we focus here on the collective level.

## SOCIAL IDENTITY AND MORALITY

Social identity research and theorizing have shown that membership in social groups provides individuals with social identities that are constitutive of their sense of the self (Tajfel, 1982). In their quest to maintain a positive social identity, individuals display ingroup favoritism (for a review see Brewer, 1979) and other biases (e.g., Branscombe & Wann, 1994; Ellemers, Spears, & Doosje, 1999), and they use strategies to maintain the integrity of the ingroup and its positive image (e.g. Castano, Paladino, Coull, & Yzerbyt, 2002; Castano, Yzerbyt, Bourguignon & Seron, 2002; Tedeschi & Bond, 2001).

Although research in the last decade has shown that individuals tend to organize social judgments largely along the dimensions of competence and sociability (e.g., Judd, James-Hawkins, Yzerbyt, & Kashima, 2005), recent evidence indicates that a fundamental dimension on which groups are evaluated is morality (Leach, Ellemers & Barreto, 2007). Immoral behavior of the ingroup can thus pose a threat to the individual. How do ingroup members react to such threats?

## REACTIONS TO INGROUP ATROCITIES

Social psychological theories describe how people morally disengage from threats to their psychological equanimity

\*Correspondence to: Bernhard Leidner, Department of Psychology, University of Massachusetts Amherst, 135 Hicks Way, Amherst, MA 01003. E-mail: bleidner@psych.umass.edu

(Bandura, 1999; Opatow, 1990; Staub & Bar-Tal, 2003), and recent research shows that when confronted with atrocities committed by the ingroup, individuals employ moral disengagement strategies (Bandura, 1999; Branscombe & Miron, 2004). People inhumanize outgroup members more strongly when the ingroup is seen as responsible rather than not responsible for their deaths (Castano & Giner-Sorolla, 2006); they display more exonerating cognitions for ingroup-committed atrocities in times of intense (versus less intense) conflict with an outgroup (Roccas, Klar, & Liviatan, 2006); and, in response to ingroup wrongdoings, those who glorify their country minimize outgroup victims' emotions and dehumanize them, which in turn leads to weaker demands to redress the wrongdoings (Leidner, Castano, Zaiser, & Giner-Sorolla, 2010).

Most of the moral disengagement strategies investigated in the context of intergroup violence change the way in which the events or the victims are perceived. The actions might be immoral, but if the magnitude of the events is played down or the victims are seen as sub-human, such actions are less distressing. Changing the moral frame within which the events are understood is a qualitatively different mechanism. Rather than acknowledging the actions as immoral but downplaying them in various ways, through morality shifting the actions come to be perceived as *not* immoral (or even as moral) in the first place.

## MORALITIES

As research on morality suggests, morality is best understood in the plural (Kohlberg, 1969; Shweder, 1982; Turiel, 1983). Bandura (1986), like Opatow (1990) and Staub (1990), and in contrast to proponents of stage theories of morality (e.g., Kohlberg, 1976), argued that morality is flexible; that "the standards for moral reasoning are much more amenable to social influence than stage theories would lead one to expect" (Bandura, 1986, p. 493). Generally, people condemn the killing of other human beings. Yet, they may frame some situations as ones in which the moral principle that applies is not that of not harming other human beings, but rather that of fostering the strength and well-being of the ingroup. Shifting the moral foundations through which one looks at an event might influence perceptions and judgments of the event.

The most well-developed model of morality in psychological theory is Haidt's social intuitionist model (Haidt & Graham, 2007), in which four moral foundations are identified: harm, fairness, loyalty, and authority.<sup>1</sup> *Harm* morals demand that people do not harm others, and *fairness* commands people to treat others fairly and justly. *Loyalty* morals reflect a tendency to see something as moral to the extent that it benefits one's ingroup. Finally, the moral foundation of *authority* consists of values related to subordination, such as duty, obedience and conformity to ingroup norms.

<sup>1</sup>The model also includes a fifth moral foundation, purity, but we do not consider it in this work for two reasons. First, in our preliminary attempts to measure it using the Moral Foundations Questionnaire (Haidt & Graham, 2007), it showed very low internal consistency. Second, and more importantly, purity is not of much relevance to the context in which we investigated the morality shifting process.

The two moral foundations of harm and fairness can be seen as "intuitive" (Haidt & Graham, 2007) or "default" morality and, at least in Western societies, are the most important (Kohlberg, 1969; Miller, 2006; Shweder, 1982; Turiel, 1983) and most frequently used (Smetana, Schlagman, & Adams, 1993). This notion of "default" moral foundations that are usually dominating other moral foundations is also in accordance with the sizeable literature on protected values (e.g., Tetlock, Kristel, Elson, Green, & Lerner, 2000) and moral convictions and moral mandates (Skitka, Bauman, & Mullen, 2004; Skitka, Bauman, & Sargis, 2005), which suggests a certain degree of intransigence in moral reasoning—at least among some people, some of the time. Under specific circumstances, however, we argue that other moral foundations (e.g., loyalty) can come to the foreground and guide the interpretation of behavior, rendering the avoidance and prevention of harm less of an absolute (see also Hare, 1981).

The existence of several moral foundations means that very different conclusions concerning the morality of a certain event or behavior can be drawn. Thus, although external constraints may well exist to the use of moral foundations in the interpretation of an event, their use may also depend on the motivation of the individual. Because an important motivation is that of seeing oneself as a moral being, we can expect that individuals will shift from one moral foundation to another in order to protect themselves from threats to their group's moral standing.

## MORALITY SHIFTING

The shift from one moral foundation to another need not be conceived of as a deliberate and thoughtful strategy. Haidt (2001) proposed that moral foundations are inaccessible to conscious awareness and lead to what he called, in his social intuitionist model of morality, moral intuitions: evaluative feelings of right or wrong resulting in automatic moral judgments in the face of transgressions that often precede deliberative, rational processes of moral reasoning (Haidt & Bjorklund, 2008). The morality shifting process that we propose can be conceived of as part and parcel of the automatic framing that occurs when an individual first approaches a certain piece of information. As noted earlier, we focus here on morality shifting in the context of threats to the morality of one's social identity. Intergroup behavior that violates principles of harm and fairness may threaten one's view of oneself as moral, because of the connection between the self and the ingroup. Individuals are thus likely to be motivated to reduce or eliminate such a threat.

It is our contention that when people detect violations of internalized moral norms committed by fellow ingroup members, the accessibility and importance of moral foundations shift: loyalty and authority come to the foreground, whereas harm and fairness recede to the background. As a consequence, the ingroup's behavior is less likely to be perceived as immoral (or it may even be perceived as moral) and the objective of protecting one's social identity is met (cf. Mummendey, Linneweber, & Loeschper, 1984; Tedeschi, Smith, & Brown, 1974).

Our rationale is also in line with research on motivated reasoning and standard shifting (e.g. Biernat & Manis, 1994; Kunda, 1987; Norton, Vandello, & Darley, 2004; Uhlmann & Cohen, 2005). It has been shown, for instance, that American political conservatives shift from deontological to consequentialist standards (i.e., ends justify the means) when Americans harm Iraqis but not when Iraqis harm Americans (Uhlmann, Pizarro, Tannenbaum, & Ditto, 2009) and that high identifiers set higher injustice standards for ingroup-committed harm and therefore need more evidence to conclude that their ingroup engaged in acts of injustice (Miron, Branscombe, & Biernat, 2010). Although this research fits well into our theoretical framework, it is different in that it is about shifts in evaluative standards for already salient moral foundations (e.g. harm and fairness) rather than about shifts in the moral foundations themselves (e.g. away from harm and fairness and toward loyalty and authority), which we propose and investigate here.

Three studies were conducted to test hypotheses derived from our rationale.

## STUDY 1

If ingroup-committed violations of internalized moral norms that trigger a shift in the accessibility and importance of moral foundations, individuals should use different words when describing the atrocities committed by the two groups. Study 1 tested this hypothesis.

## METHOD

### Participants

The sample consisted of 135 participants, recruited through the Study Response Project (Stanton, 2006; Wallace, 2004). After eliminating 18 people who were not born in the USA and 10 people who did not correctly remember the identity of the perpetrating group, 107 participants (73 women, 32 men, 2 missing; age  $M = 33.87$ ,  $SD = 12.52$ , range = 19–64) were retained for subsequent analyses. There were no univariate or multivariate outliers.

### Procedure

The experiment was conducted online. After giving consent, participants read a news report on four alleged cases of coalition military personnel torturing and killing Iraqi civilians in a prison near Baghdad. Although names were changed, the cases were based on confirmed cases of abuse committed by coalition military personnel in Iraq. The reported mistreatments of prisoners included water torture, beatings, stress positions, and humiliating acts. In all four cases, mistreatment and torture eventually led to the deaths of the prisoners. In the ingroup-atrocity condition, the perpetrators were described as US soldiers, whereas in the outgroup-atrocity condition, the perpetrators were described as Australian soldiers; otherwise, the text remained identical across conditions.

After reading the news report, participants were asked to summarize the article they had read for a third person. We asked for summaries to be as unobtrusive and subtle as possible, but in reality the aim was to obtain participants' subjective accounts of the atrocities they had read about. This was successful: rather than giving objective and neutral summaries, participants gave their own opinions and interpretations of the events. After writing the summary, participants filled out demographic questions.

## RESULTS

The summaries participants provided were spellchecked and proofread by two research assistants who were blind to the hypotheses and to the conditions the summaries came from. Then each summary was analyzed via the Linguistic Inquiry and Word Count (LIWC) program (Pennebaker, 1993), using Graham, Haidt, and Nosek's (2009) moral foundations dictionary. This dictionary consists of word lists for each of the moral foundations we were interested in: harm, fairness, loyalty, and authority. The word count indices provided by LIWC were then used to compute a harm/fairness score and a loyalty/authority score. The words for the four moral foundations were collapsed into two each, for two reasons. Consistent with our theoretical viewpoint, which makes the same predictions for loyalty and authority on the one hand, and harm and fairness on the other, collapsing seems an appropriate strategy. Besides, the word frequencies in each four categories were low; collapsing yielded sufficient frequencies for each of the two resulting variables.

A repeated measures ANOVA with condition as between- and moral foundation (harm/fairness versus loyalty/authority) as within-participant factor on the LIWC scores was conducted. Two effects were significant. The main effect of moral foundation,  $F(1, 105) = 26.93$ ,  $p < .001$ ,  $d = 1.01$ , revealed higher frequencies of use of harm/fairness ( $M = 0.55$ ,  $SD = 0.45$ ) than loyalty/authority ( $M = 0.29$ ,  $SD = 0.29$ ), supporting the claim of cross-cultural literature that harm and fairness are more prevalent than loyalty and authority in Western societies. The interaction between condition and moral foundation was also significant,  $F(1, 105) = 6.52$ ,  $p = .012$ ,  $d = 0.50$ . Fewer words reflecting the moral foundations of harm/fairness were used in the ingroup-perpetrator ( $M = 0.46$ ,  $SD = 0.40$ ) than in the outgroup-perpetrator condition ( $M = 0.63$ ,  $SD = 0.50$ ),  $F(1, 105) = 3.59$ ,  $p = .061$ ,  $d = 0.37$ . The opposite trend was observed for loyalty/authority,  $F(1, 105) = 2.78$ ,  $p = .099$ ,  $d = 0.33$ , for which somewhat greater frequencies emerged in the ingroup-perpetrator ( $M = 0.33$ ,  $SD = 0.32$ ) as compared with the outgroup-perpetrator condition ( $M = 0.24$ ,  $SD = 0.26$ ).

## DISCUSSION

The findings revealed that words related to harm/fairness appeared less frequently in a narrative about immoral ingroup behavior than in immoral outgroup behavior. The opposite pattern emerged for loyalty/authority words. We interpret

these findings as a shift that occurred when participants read about the ingroup-committed atrocity, as opposed to the outgroup-committed atrocity. The main effect of moral foundations we observed is also consistent with the cross-cultural notion of harm and fairness being more prevalent, at least in Western societies, than loyalty and authority.

## STUDY 2

Study 1 provided initial evidence for enhanced accessibility of loyalty and authority over harm and fairness foundations after consideration of ingroup-committed atrocities. Study 2 aimed to yield further substantiation of the morality shifting hypothesis by using a different methodology, providing a more direct test of our claim that moral violations of the ingroup lead to a shift in the relative cognitive accessibility of moral foundations.

Rationalist approaches see moral judgments as reached by conscious reasoning (e.g., Kohlberg, 1969; Piaget, 1932/1965; Turiel, 1983), but more recent thinking, such as the social intuitionist model (Haidt, 2001), conceptualize moral judgments as quick and automatic evaluations (for integrative approaches see Pizarro & Bloom, 2003). Thus, in Study 2, we relied on a measure of accessibility of the moral foundations after participants were confronted with atrocities committed by either the ingroup or an outgroup. The use of a non-intrusive and non-transparent measure has the advantage of avoiding potential reactance, self-presentation concerns, or self-censorship on the part of the participants, which might constitute a problem when using explicit measures such as the Moral Foundations Questionnaire of Haidt and Graham (2007).

## METHOD

### Participants

Forty-one participants were recruited via the internet (www.craigslist.org) to participate in a laboratory study, allegedly on reading comprehension. One participant who was not born in the USA was excluded, leaving 40 participants (14 men, 26 women; 6 left-handed, 3 ambidextrous, 31 right-handed; age:  $M = 33.20$ ,  $SD = 13.45$ ).

### Procedure

The study was conducted in the laboratory on a computer with MediaLab and DirectRT (Jarvis, 2004a, 2004b). First, participants read two news reports, the first on a morality-unrelated topic (plants) and the second one on prisoner abuse similar to the news report used in Study 1. Again, in the ingroup-atrocity condition, the perpetrators were described as US soldiers, whereas in the outgroup-atrocity condition, the perpetrators were described as Australian soldiers; otherwise, the text remained identical across conditions.

Next, participants completed a Lexical Decision Task, presented as a reading comprehension task, but aimed at

measuring accessibility of moral foundations. They then filled out the demographics, including handedness, were debriefed and received \$10 in reward for their participation.

*Lexical Decision Task (LDT)*. Participants were presented with a string of letters on the computer screen and had to decide, as quickly and accurately as possible, whether it was a word or a non-word. Each string was displayed until the participant pressed the right-shift key (word) or the left-shift key (non-word). The inter-trial interval was 1 second. After 12 practice trials with six non-words and six words that were unrelated to morality and not overlapping with the other morality-unrelated words comprising the actual stimuli, participants were presented, in randomized order, with 20 pretested morality-related words (five harm-related words, e.g., *harm*, *abuse*; five fairness-related words, e.g., *fairness*, *justice*; five loyalty-related words, e.g., *loyalty*, *solidarity*; and five authority-related words, e.g., *authority*, *leader*), 20 morality-unrelated words, and 40 non-words. The morality-unrelated words were matched to the morality-related words in length and frequency (Francis & Kucera, 1982). For each word, a non-word was constructed by changing one letter per syllable in the real word; this ensured words and non-words to be equal in length, orthography, and pronounceability; pseudohomophones were avoided. Reaction times and accuracy of participants' responses were recorded.

## RESULTS

Two per cent of the LDT responses were incorrect and therefore excluded from analysis. There was no difference in the number of correct responses between conditions,  $F_s(1, 38) < 2.00$ ,  $p > .05$ , ruling out the possibility of ambiguity in the meaning of the reaction times (RTs) in the remaining correct trials (cf. Salthouse & Hedden, 2002). RT means and *SDs* were computed for each participant in each category (harm, fairness, loyalty, authority, unrelated, non-words, practice). Observations from nine (out of 280) RTs that were greater or smaller than three *SDs* above or below the mean were eliminated (cf. Tabachnick & Fidell, 2007). As it is customary when working with RTs (cf. Schmiedek, Oberauer, Wilhelm, Süß, & Wittmann, 2007) transformation was necessary and resulted in considerably more normal distributions (e.g. lower skewness, from 1.73 to 1.31, and kurtosis, from 2.96 to 0.57).<sup>2</sup> Then, a ratio between each category and the unrelated words was computed (cf. Norris, 2009; Ratcliff, Gomez, & McKoon, 2004; C. Barry, personal communication, June 9, 2009). A mixed repeated-measures analysis was con-

<sup>2</sup>We transformed each of the reaction time variables so that the variable would be as normally distributed as possible. For harm and fairness, this required square root transformation, for loyalty, authority, and unrelated words log transformation. After transforming the variables, we then standardized them to bring them on the same scale, allowing for the subsequent computation of ratio scores and the statistical comparisons. Although this strategy might be uncommon because it is oftentimes not necessary, it provides the best possible reduction of null hypothesis testing decision errors. Our analytical strategy is also in line with literature on data transformation in general (e.g., Tabachnick & Fidell, 2007) and handling of reaction time data in particular (e.g., Norris, 2009; Ratcliff et al., 2004).

ducted on these ratio scores, with perpetrator identity (ingroup versus outgroup) as between- and moral foundation (harm versus fairness versus loyalty versus authority) as within-participants factor, controlling for gender and handedness.

The predicted interaction of perpetrator identity by moral foundation was significant,  $F(3, 106) = 3.96, p = .01, d = 0.22$  (see Figure 1). Given our hypothesis (and the pattern of means), we computed a planned contrast comparing loyalty and authority against harm and fairness. As predicted, after reading about ingroup-committed atrocities, loyalty/authority were more accessible ( $M = -0.25, SE = 0.24$ ) than harm/fairness ( $M = 0.27, SE = 0.25$ ), whereas after reading about outgroup-committed atrocities loyalty/authority were less accessible ( $M = 0.33, SE = 0.26$ ) than harm/fairness ( $M = -0.24, SE = 0.25$ ),  $t(106) < -3.44, p < .001, d = 0.39$ . The main effects of perpetrator identity,  $F(1, 34) = 0.04, p = .84, d = 0.02$ , and moral foundation,  $F(3, 106) = 0.01, p = 1.00, d = 0.01$ , were not significant; neither were gender,  $F(1, 34) = 0.70, p = .41, d = 0.09$ , or handedness,  $F(2, 34) = 0.09, p = .92, d = 0.03$ .<sup>3</sup>

## DISCUSSION

Results of Study 2 yielded further support for our hypothesis. Loyalty- and authority-related words were recognized more quickly among participants confronted with ingroup-committed atrocities as compared with participants confronted with outgroup-committed atrocities. For harm- and fairness-related words the opposite was the case. Although what matters from our perspective is the difference between the accessibility of these moral foundations after reading of ingroup- versus outgroup-committed atrocities, we believe that outgroup-committed atrocities are to be considered as a sort of “baseline,” and that, as the name of our proposed process suggests, a shift occurs when individuals are presented with ingroup-committed atrocities. Two observations support this argument. First, as noted above, research on morality suggests that harm and fairness are the moral principles of choice in Western societies, and thus one would expect greater accessibility of these moral principles. The relative higher accessibility of harm and fairness in the outgroup-atrocity condition is also consistent with this view. Second, this interpretation is in line with previous work claiming that it is ingroup-committed atrocities that elicit defense mechanisms (Castano & Giner-Sorolla, 2006). We will return to this point later.

We conceive of morality shifting as an automatic process that is driven by the need of individuals to protect their social identities from threat to their moral status. It is possible, however, that such an automatic process need not be triggered by a potential threat, but instead is simply the consequence of social identity activation even in the absence of a threat to the moral status of the ingroup. We thus conducted a follow-up

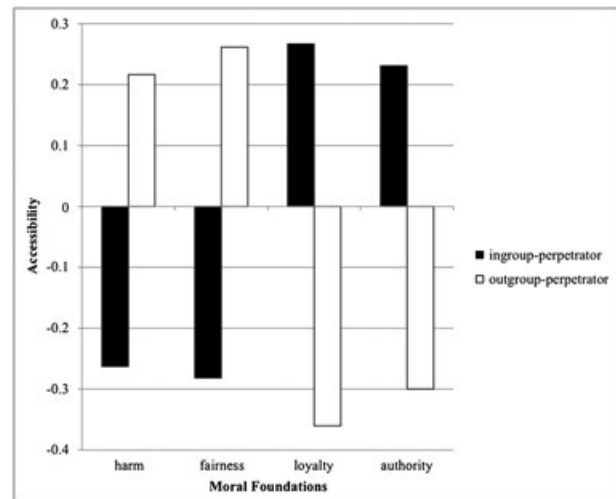


Figure 1. Accessibility of the moral foundations of harm, fairness, loyalty, and authority in ingroup- and outgroup-perpetrator condition of Study 2. For ease of interpretation, the reaction time scores were reversed, so that higher scores in this figure indicate higher accessibility

study in which the ingroup (or outgroup) identity was made salient but no immoral behavior was depicted.

## STUDY 3

Findings from Study 1 and 2 are consistent with our morality shifting hypothesis. Their convergence is particularly encouraging in light of the fact that two very different dependent variables were used. In Study 3, we aimed at further investigating this effect using yet another way of assessing the accessibility of moral foundations. We adapted the Affect Misattribution Procedure to this end (Payne & Cameron, 2010; Payne, Cheng, Govorun, & Stewart, 2005). A second goal of Study 3 was to clarify whether morality shifting occurs specifically in response to identity threat or whether it occurs in response to mere social identity activation. A third goal was to investigate the moderating role of type and strength of social identification.

Previous research (e.g., Leidner et al., 2010; Roccas et al., 2006) demonstrated ingroup glorification (but not ingroup attachment) to be positively related to moral disengagement and outgroup derogation in response to ingroup-committed violence against an outgroup. Presumably because those who strongly glorify their ingroup believe in the ingroup's superiority over other groups and submit to group norms and authorities, leaving no room for criticism of the ingroup. Thus, high glorifiers should be more threatened than low glorifiers by moral violations committed by their ingroup. Consequently, we expect morality shifting to occur particularly, and possibly exclusively, among high glorifiers—especially should it turn out that morality shifting is indeed serving as a psychological defense against identity threat, as Study 1 and 2 tentatively suggested. Low glorifiers, on the other hand, might not be prompted by ingroup transgressions to shift towards loyalty/authority; because low glorifiers allow for criticism of the ingroup, they are psychologically better equipped to deal with negative or norm-violating ingroup behavior. In fact, they may even be more critical to

<sup>3</sup>The same analyses conducted on untransformed data revealed no significant effects. To rule out the possibility that the above-reported pattern was a mere artifact of a condition effect on the unrelated words, we also conducted one-way analyses of variance on the untransformed and transformed RTs of unrelated words, with condition as the independent variable. These effects were not significant.

ingroup- than to outgroup-committed transgressions, and therefore possibly shift even more toward harm and fairness when evaluating ingroup-committed transgressions.

Different predictions were made for ingroup attachment, which refers to commitment and belongingness to the ingroup. Past research found ingroup attachment to be unrelated or even negatively related to moral disengagement and outgroup derogation (e.g., Leidner et al., 2010; Roccas et al., 2006). Accordingly, once its shared variance with glorification is taken into account, it should not be associated to a shift in moral foundations.

Study 3 also aimed to provide a stronger test of whether morality shifting occurs generally in any ingroup context, or whether it occurs in response to a threat to the ingroup identity. Thus, Study 3 used both ingroup- and outgroup-committed atrocities (as in Study 2) and ingroup and outgroup non-atrocities, crossing group (ingroup versus outgroup) with event (atrocities versus neutral).

## METHOD

### Participants

The sample consisted of 142 participants. Two participants who did not follow instructions and eight who reported to be able to speak and/or read Chinese were excluded (see below). Five multivariate outliers were excluded (cf. Tabachnick & Fidell, 2007), leaving 127 participants (88 women, 39 men; 28 18–25 year olds, 25 26–35 year olds, 19 36–45 year olds, 29 46–55 year olds, and 26 older than 55 years) for subsequent analyses.

### Procedure

The experiment was conducted online. After giving consent, participants read a newspaper article either reporting on ingroup- or outgroup-committed atrocities (identical to Study 1 and 2), or reporting on the national park system of the ingroup or an outgroup. The report on the national park system described how national parks were created and maintained through the history of the USA (ingroup) or Australia (outgroup). To ensure that the articles would make national identity (USA or Australia) salient, the articles discussed national parks as a “uniquely American (Australian) idea”, related it to the Declaration of Independence (Federal Constitution) and mentioned Abraham Lincoln (Edmund Barton). After reading the article, the moral foundations accessibility measure was administered. Before the debriefing, participant also answered ingroup attachment and glorification scales and reported demographic questions.

### Moral foundations

We measured the accessibility of four moral foundations of interest (i.e., *harm/fairness*, *loyalty/authority*) by adapting the affect misattribution procedure (Payne et al., 2005). We presented participants with five Chinese characters in random order. Participants guessed the meaning of each character out of four response alternatives. Response alternatives consisted of one harm-related, one fairness-related, one loyalty-related, and one authority-related word, taken from the moral

foundation dictionary of Graham et al. (in press). The words of the response alternatives of different Chinese characters were used only once for each participant. Nine participants did not fill out this measure, leaving 118 responses for this measure.

### Ingroup attachment and glorification

National attachment (e.g., “It is important to me to contribute to my nation”) and glorification (e.g., “The U.S. is better than other nations in all respects”) were measured using the scales by Roccas et al. (2006), adapted to refer to the American identity. The responses were given on a continuum ranging from 1 (*strongly disagree*) to 9 (*strongly agree*).

## RESULTS

Difference scores between the number of times a participant chose a loyalty- or authority-related word and the number of times a participant chose a harm- or fairness-related word were calculated. Thus, the resulting accessibility of moral foundations index ( $M = -0.45$ ,  $SD = 1.87$ ) was positive for participants who chose more loyalty/authority-related than harm/fairness-related words, zero for participants who chose as many harm/fairness-related words as loyalty/authority-related words, and negative for participants who chose more harm/fairness-related than loyalty/authority-related words. Attachment ( $M = 6.90$ ,  $SD = 1.43$ ,  $\alpha = .93$ ) and glorification ( $M = 4.96$ ,  $SD = 1.27$ ,  $\alpha = .81$ ) scores were computed, and entered as predictors (centered) alongside group (ingroup versus outgroup) and event (atrocities versus neutral) in a general linear model, using the accessibility index as dependent variable. The only significant effects were the two-way interactions between group and glorification,  $F(1, 102) = 6.32$ ,  $p = .014$ ,  $d = 0.45$ , and between event and glorification,  $F(1, 102) = 4.31$ ,  $p = .041$ ,  $d = 0.37$ , as well as the expected three-way interaction between group, event, and glorification,  $F(1, 102) = 12.96$ ,  $p = .001$ ,  $d = 0.64$ . The two-way interaction between glorification and attachment reached marginal significance,  $F(1, 102) = 2.94$ ,  $p = .090$ , but none of the follow-up tests did. All other effects were not significant.

The follow-up tests on the group by glorification interaction revealed that, among high glorifiers (i.e., participants one standard deviation or more above the mean of glorification), regardless of the type of event (atrocities or neutral), the accessibility index was higher in the ingroup ( $M = 0.85$ ,  $SE = 0.64$ ) than in the outgroup conditions ( $M = -2.27$ ,  $SE = 1.34$ ),  $t = 2.61$ ,  $p = .010$ ,  $d = 0.47$ . That is, high glorification was associated with higher relative accessibility of loyalty/authority principles when the ingroup rather than an outgroup was concerned. Among low glorifiers (i.e., participants one standard deviation or more below the mean of glorification), on the other hand, if anything, the opposite was the case ( $M_{ingroup} = -1.16$ ,  $SE_{outgroup} = 0.52$ , and  $M_{outgroup} = 0.67$ ,  $SE_{outgroup} = 1.27$ ),  $t = -1.76$ ,  $p = .082$ ,  $d = .31$ .

The follow-up tests on the event by glorification interaction found that, among low glorifiers the accessibility index was higher in the atrocities conditions ( $M = 0.82$ ,  $SE = 1.16$ ) than in

the neutral conditions ( $M = -1.32$ ,  $SE = 0.62$ ),  $t = 2.05$ ,  $p = .043$ ,  $d = .37$ . Among high glorifiers, on the other hand, if anything, the opposite was the case ( $M_{\text{atroc}} = -1.69$ ,  $SE_{\text{atroc}} = 1.42$ , and  $M_{\text{neutral}} = 0.27$ ,  $SE_{\text{neutral}} = 0.56$ ),  $t = -1.64$ ,  $p = .104$ ,  $d = 0.29$ .

Most importantly, disentangling the three-way interaction between group, event, and glorification revealed that, among high glorifiers, the accessibility index was higher in the ingroup-perpetrator condition ( $M = 1.78$ ,  $SE = 0.68$ ) than in both the ingroup-neutral condition ( $M = -0.08$ ,  $SE = 0.59$ ),  $t = -2.08$ ,  $p = .040$ ,  $d = 0.37$ , and the outgroup-perpetrator condition ( $M = -5.17$ ,  $SE = 2.15$ ),  $t = -3.08$ ,  $p = .003$ ,  $d = 0.55$ . That is, loyalty/authority was more accessible in the ingroup-perpetrator condition than in the two other comparable conditions. The contrast between the outgroup-perpetrator condition ( $M = -5.17$ ,  $SE = 2.15$ ) and the outgroup-neutral condition ( $M = 0.62$ ,  $SE = 0.52$ ) was also significant,  $t = 2.61$ ,  $p = .010$ ,  $d = .47$ , indicating a shift in the opposite direction (i.e., toward harm/fairness rather than loyalty/authority). Among low glorifiers, the accessibility index was lower in the ingroup-perpetrator condition ( $M = -1.73$ ,  $SE = 0.50$ ) than in the outgroup-perpetrator condition ( $M = 3.38$ ,  $SE = 1.82$ ),  $t = 2.71$ ,  $p = .008$ ,  $d = 0.48$ , and, as a trend, the ingroup-neutral condition ( $M = -0.60$ ,  $SE = 0.53$ ),  $t = 1.56$ ,  $p = .123$ ,  $d = 0.28$ .

## DISCUSSION

The main goals of Study 3 were to further clarify the morality shifting hypothesis and to investigate the moderating role of attachment and glorification. Consistent with findings emerging in related areas of research (Roccas et al., 2006; Leidner et al., 2010), the enhanced accessibility of loyalty/authority when individuals were confronted with an ingroup-committed atrocity (as compared with an outgroup-committed atrocity or an ingroup-neutral event) was only observed among high glorifiers. Low glorifiers showed, if anything, the opposite pattern. This pattern of findings also confirms our hypothesis that morality shifting occurs specifically in response to identity threat triggered by moral ingroup violations, rather than generally in response to any activation of ingroup identity.<sup>4</sup>

## GENERAL DISCUSSION

Thomas Fowler is not convinced by Pyle's argument that the killing of a handful of civilians can be construed as moral.

<sup>4</sup>The results in the ingroup and outgroup-neutral conditions of Study 3 were also replicated in a study not reported in this paper, using the same experimental material for these conditions as in Study 3, and the same methodology as in Study 2 (Lexical Decision Task). In this study, containing only the two neutral conditions, we also found that all moral foundations, across the board, were more accessible when ingroup rather than outgroup identity was salient. Although not of specific interest to the phenomenon under investigation here, this finding is in line with evidence that the domain of morality is particularly important for people's evaluation of ingroups (Leach et al., 2007). At the same time, however, this raises the question of why such an effect was not observed in Study 2. One possible answer follows from the central claim of our theory: unlike in morally neutral contexts, in the context of moral transgressions such as intergroup violence, morality will not be more important for evaluations of ingroups (as compared with outgroups) in general, but it will depend on the type of morality—loyalty and authority will be relatively more important, whereas harm and fairness will be relatively less important.

The reason for this might be that Fowler is British, and sees Pyle's actions for the atrocities they are, whereas Pyle is American (and the perpetrator, and likely more of a glorifier than Fowler), and sees his actions as service to his country. Both Fowler and Pyle are interpreting the same event utilizing different moral foundations: Fowler is using harm and fairness, whereas Pyle is using loyalty and authority.

In this contribution, we put this idea to test. Building on theories of morality and social identity, we suggested that when people are confronted with ingroup-committed violence, the principles of loyalty and authority come to the foreground, and thus become more accessible, relative to the principles of harm and fairness, which generally, or at least in case of outgroup-committed violence, tend to be more accessible and more utilized. In other words, we hypothesized a *morality shift* from harm/fairness toward loyalty/authority.

The findings of the studies presented here confirmed our hypothesis: compared with being confronted with outgroup-committed atrocities, participants who read about ingroup-committed atrocities subsequently used more loyalty/authority words than harm/fairness words (Study 1), and recognized loyalty/authority words more quickly than harm/fairness words (Study 2). Furthermore, in line with literature on moral disengagement, high glorifiers (but not low glorifiers, or low or high attached people) guessed an unknown Chinese ideogram as having a loyalty/authority meaning more frequently after reading about ingroup-committed atrocities as compared with outgroup-committed atrocities or ingroup-neutral scenarios (Study 3). Taken together, these findings support our conjecture that confrontation with information that threatens the moral view of the ingroup, and thus one's social identity, is associated with an increased relative accessibility of the moral foundations of loyalty/authority.

## IS MORALITY SHIFTING A REACTION TO IDENTITY OR IDENTITY THREAT?

We proposed that people do not shift their moral foundations upon mere activation of their social identity. Rather, the shift is a motivated response to a threat to the morality of their group/collective self. In other words, only ingroup contexts that involve a moral violation committed by the ingroup should trigger morality shifting. The findings of Study 3 support this conjecture. Using a modified version of the affect misattribution procedure, we found evidence that it was information about an immoral action carried out by the ingroup that triggered the shift, as opposed to simple activation of one's social identity.

Yet, it is still possible that situations in which one's social identity (particularly national or ethnic identity) is made salient may trigger shifts similar to the one observed in Study 1–3, even in the absence of information that blatantly threatens the morality of the ingroup. Research stemming from social identity theory has shown that individuals tend to conceive of intergroup situations as inherently comparative and oftentimes confrontational. For instance, Wildschut, Insko, and Gaertner (2002) demonstrated in a prisoner's dilemma game that individuals adhering to (perceived) norms and standards

of their ingroup behave more competitively towards an outgroup (see also Wildschut & Insko, 2006). Investigating behaviors reflecting moral principles that condone or value hostility towards outgroups, Cohen, Montoya, and Insko (2006, Study 1) found that, in preindustrial societies, loyalty to the ingroup covaried with tendencies to value violence between ingroup and outgroups (relative to violence within the ingroup), to engage in more external than internal conflict, and to enjoy conflict. It is thus possible that intergroup situations perceived as comparative and confrontational “by nature” would lead to morality shifting as a response to mere activation of ingroup identity, as opposed to the threat-specific form of morality shifting that we have demonstrated here. Further research is needed to better understand the boundary conditions of this phenomenon.

## CONSEQUENCES OF MORALITY SHIFTING

### Consequences for the individual

Be it threat-specific or not, enhanced accessibility of a framework that privileges loyalty to the ingroup and submission to its authority at the expense of the principles of harm and fairness may have important consequences. If the shift in moral foundations occurs in response to moral transgressions, then it has positive consequences for the individual. One, which can also be seen as a cause of the shift, is the reduction of the discomfort stemming from a perceived inconsistency between personal beliefs about how the ingroup should behave and the actual ingroup behavior (intragroup dissonance; Glasford, Dovidio, & Pratto, 2009). Similarly, morality shifting will likely avoid or at least reduce psychologically costly self-sanctions. Thus, morality shifting protects one’s identity and thus helps maintain psychological equanimity. However, people who only rarely correct for their “moral bias” may also suffer negative consequences in the long run. They might be socially excluded by others, have a smaller social support network, and the like. These aspects of social interconnectedness might provide a powerful corrective, forcing people to keep their automatic biases in check, and possibly correct for them *a posteriori*.

### Consequences for intergroup relations

Although we conceive of morality shifting as motivated by the need to protect one’s identity, and thus as a beneficial mechanism to the individual, we expect it to have much more negative consequences for intergroup relations and for society at large. It can give more leeway in the mistreatment of outgroup members, or lead to their exclusion from the scope of justice (Opatow, 1990), reducing the chance of seeing such mistreatment as violating principles of harm and fairness. Morality shifting can thus be seen as a mechanism that allows people to make a virtue of evil (see Reicher, Haslam, & Rath, 2008). Once the shift occurs, further actions are even more likely to be interpreted from a loyalty/authority perspective rather than from a harm/fairness perspective. That is, next to desensitization and habituation, morality shifting may be an

important mechanism manipulated via propaganda as well as in the direct training of those who are to commit violence against outgroup members.

Although morality shifting in the context investigated here has largely positive consequences at the individual level but negative consequences at the intergroup level, loyalty/authority morals are neither unhealthy nor undesirable per se. Indeed, they are to some extent important and positive for the group. For instance, they can suppress extreme selfishness (Haidt, 2008). It is *shifts* toward loyalty/authority in the way we investigated them here, which have mostly negative consequences for *intergroup* (rather than *intragroup*) relations.

## MODERATING EFFECTS AND BOUNDARY CONDITIONS

Morality shifting is likely to be moderated by a series of factors. One of these is ingroup identification, as it is associated with social identity threat (e.g., Ellemers, Spears, & Doosje, 2002) and defensive reactions (e.g. Branscombe, Ellemers, Spears, & Doosje, 1999), so that high identifiers might be more prone to a morality shift as demonstrated in the present context. Two specific aspects of ingroup identification were investigated in Study 3: glorification and attachment. As predicted, attachment did not moderate morality shifting, whereas glorification did. A shift toward loyalty/authority was observed for high glorifiers, but the opposite trend was observed for low glorifiers. Both results were expected, and consistent with the literature. High glorifiers are, by definition, less capable and willing to integrate negative information about the ingroup, and thus more likely to react defensively to identity-threatening information (Castano, 2008; Leidner et al., 2010). Low glorifiers, on the other hand, if anything, shift toward harm/fairness in response to ingroup-committed wrongdoings, reflecting a more critical stance toward norm-violating ingroup behavior that is likely due to their lower threat perception and lower defensiveness. Research has shown that the moderating effects of glorification are independent of two well-established ideological individual differences, social dominance order (SDO) and right-wing authoritarianism (RWA) (Leidner et al., 2010), but further investigation of the concept of glorification and its correlates is certainly needed.

## ALTERNATIVE PROCESSES

Instead of, or in addition to, morality shifting, individuals can respond to threats to their sense of moral collective self in a variety of ways. As mentioned earlier, they can engage in moral disengagement strategies, for instance by dehumanizing the victims (Castano & Giner-Sorolla, 2006; Giner-Sorolla, Leidner, & Castano, in press; Leidner et al., 2010). They can resort to de-moralizing (see Rozin, 1999) so as to render the questionable event not a question of morality, engage in moral equilibration (Staub, 1989, 1990), or shift justice standards (Miron & Branscombe, 2008; Miron et al., 2010). In situations where group membership to a given group or the violated norm is not central to a person’s identity or morality, or where



the norm violators can easily be seen as deviants, processes such as disidentification (Aronson, Blanton, & Cooper, 1995; and national de-identification in particular, Verkuyten & Yildiz, 2007), or distancing from the ingroup by rendering ingroup deviants as black sheep in an effort to protect oneself (Eidelman & Biernat, 2003) or the group (Castano et al., 2002; Marques, Yzerbyt, & Leyens, 1988), can render morality shifting unnecessary.

### MORALITY SHIFTING AS A DEFENSE MECHANISM TO ANY SELF-RELATED VIOLATIONS

In the present contribution, we proposed *morality shifting* as a mechanism through which individuals can maintain a moral image of their ingroup, and, in turn, of themselves, when this image is threatened by the behavior of fellow ingroup members. The same process might occur on individual level, where personal norm violations are concerned. A shift from harm and fairness morals to loyalty and authority morals related to the personal self might also help maintain a moral view of oneself as an individual. The phenomenon of moral hypocrisy, for instance, describes moral judgments as stemming from egocentric and automatic evaluations as right or wrong and therefore being self-serving in nature (Batson, Kobrynowicz, Dinnerstein, Kampf, & Wilson, 1997; Batson, Thompson, Seufferling, Whitney, & Strongman, 1999; Batson, Thompson, & Chen, 2002; Epley & Caruso, 2004). This phenomenon might result, among other things, from morality shifting. As such, morality shifting is likely not exclusive to ingroup wrongdoings as investigated here, but rather a general defense mechanism occurring in any transgression related to the self. The preliminary evidence presented here supports our conjecture, but future research is needed to investigate the individual level, moderating factors, boundary conditions, and consequences of this phenomenon.

### ACKNOWLEDGEMENTS

Research reported here as well as the writing of this paper was supported by a Fellowship for Young Scientists of the Gottlieb Daimler- und Karl Benz-Foundation (02/15/06), to the first author, and by National Science Foundation Research Grant BCS-0545801, to the second author. We wish to thank Chris Barry for his assistance with the Lexical Decision Task, Galen Bodenhausen and three anonymous reviewers of another paper for bringing up points discussed here and examples used to illustrate morality shifting, and Caitlin Thompson and Rebecca Mielczarek for comments on earlier drafts of this manuscript.

### REFERENCES

- Aronson, J., Blanton, H., & Cooper, J. (1995). From dissonance to disidentification: Selectivity in the self-affirmation process. *Journal of Personality and Social Psychology*, 68(6), 986–996.
- Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory*. Englewood Cliffs, NJ: Prentice-Hall, Inc.
- Bandura, A. (1999). Moral disengagement in the perpetration of inhumanities. *Personality and Social Psychology Review*, 3(3), 193–209.
- Batson, C. D., Kobrynowicz, D., Dinnerstein, J. L., Kampf, H. C., & Wilson, A. D. (1997). In a very different voice: Unmasking moral hypocrisy. *Journal of Personality and Social Psychology*, 72(6), 1335–1348.
- Batson, C. D., Thompson, E. R., & Chen, H. (2002). Moral hypocrisy: Addressing some alternatives. *Journal of Personality and Social Psychology*, 83(2), 330–339.
- Batson, C. D., Thompson, E. R., Seufferling, G., Whitney, H., & Strongman, J. A. (1999). Moral hypocrisy: Appearing moral to oneself without being so. *Journal of Personality and Social Psychology*, 77(3), 525–537.
- Biernat, M., & Manis, M. (1994). Shifting standards and stereotype-based judgments. *Journal of Personality and Social Psychology*, 66(1), 5–20.
- Branscombe, N. R., & Miron, A. M. (2004). Interpreting the ingroup's negative actions toward another group: Emotional reactions to appraised harm. In L. Z. Tiedens & C. W. Leach (Eds.), *The social life of emotions. Studies in emotion and social interaction* (pp. 314–335). New York, NY: Cambridge University Press.
- Branscombe, N. R., & Wann, D. L. (1994). Collective self-esteem consequences of outgroup derogation when a valued social identity is on trial. *European Journal of Social Psychology*, 24(6), 641–657.
- Branscombe, N. R., Ellemers, N., Spears, R., & Doosje, B. (1999). The context and content of social identity threat. In N. Ellemers, R. Spears, & B. Doosje (Eds.), *Social identity: Context, commitment, content* (pp. 35–58). Oxford, UK: Blackwell.
- Brewer, M. B. (1979). In-group bias in the minimal intergroup situation: A cognitive-motivational analysis. *Psychological Bulletin*, 86(2), 307–324.
- Castano, E. (2008). On the perils of glorifying the in-group: Intergroup violence, in-group glorification, and moral disengagement. *Social and Personality Psychology Compass*, 1, 154–170.
- Castano, E., & Giner-Sorolla, R. (2006). Not quite human: Dehumanization in response to responsibility for intergroup killing. *Journal of Personality and Social Psychology*, 90, 804–818.
- Castano, E., Paladino, M.-P., Coull, A., & Yzerbyt, V. (2002a). Protecting the ingroup stereotype: Ingroup identification and the management of deviant ingroup members. *British Journal of Social Psychology*, 41(3), 365–385.
- Castano, E., Yzerbyt, V. Y., Bourguignon, D., & Seron, E. (2002b). Who may enter? The impact of ingroup identification on ingroup-outgroup categorization. *Journal of Experimental Social Psychology*, 38, 315–322.
- Castano, E., Yzerbyt, V. Y., Paladino, M. P., & Sacchi, S. (2002c). I belong therefore I exist: Ingroup identification, ingroup entitativity, and ingroup bias. *Personality and Social Psychology Bulletin*, 28, 135–143.
- Cohen, T. R., Montoya, R. M., & Insko, C. A. (2006). Group morality and intergroup relations: Cross-cultural and experimental evidence. *Personality and Social Psychology Bulletin*, 32(11), 1559–1572.
- Eidelman, S., & Biernat, M. (2003). Derogating black sheep: Individual or group protection? *Journal of Experimental Social Psychology*, 39(6), 602–609.
- Ellemers, N., Spears, R., & Doosje, B. (1999). *Social identity: Context, commitment, content*. Oxford, UK: Blackwell.
- Ellemers, N., Spears, R., & Doosje, B. (2002). Self and social identity. *Annual Review of Psychology*, 53(1), 161–186.
- Epley, N., & Caruso, E. M. (2004). Egocentric ethics. *Social Justice Research*, 17(2), 171–187.
- Francis, W. N., & Kucera, H. (1982). *Frequency analysis of English usage*. Boston, MA: Houghton Mifflin.
- Giner-Sorolla, R., Leidner, B., & Castano, E. (in press). Dehumanization, demoralization, and morality shifting: Paths to moral certainty in extremist violence. In M. A. Hogg & D. L. Blaylock (Eds.), *Extremism and the psychology of uncertainty*. Boston: Wiley-Blackwell.
- Glasford, D. E., Dovidio, J. F., & Pratto, F. (2009). I continue to feel so good about us: In-group identification and the use of social identity-enhancing strategies to reduce intragroup dissonance. *Personality and Social Psychology Bulletin*, 35(4), 415–427.
- Graham, J., Haidt, J., & Nosek, B. A. (2009). Liberals and conservatives rely on different sets of moral foundations. *Journal of Personality and Social Psychology*, 96(5), 1029–1046.
- Haidt, J. (2001). The emotional dog and its rational tail: A social intuitionist approach to moral judgment. *Psychological Review*, 108(4), 814–834.
- Haidt, J. (2008). Morality. *Perspectives on Psychological Science*, 3(1), 65–72.
- Haidt, J., & Bjorklund, F. (2008). Social intuitionists answer six questions about moral psychology. In W. Sinnott-Armstrong (Ed.), *Moral psychology, Vol. 2: The cognitive science of morality: Intuition and diversity* (pp. 181–217). Cambridge, MA: MIT Press.
- Haidt, J., & Graham, J. (2007). When morality opposes justice: Conservatives have moral intuitions that liberals may not recognize. *Social Justice Research*, 20, 98–116.
- Hardy, S. A., & Carlo, G. (2005). Identity as a source of moral motivation. *Human Development*, 48(4), 232–256.

- Hare, J. (1981). *Moral thinking*. Oxford, UK: Clarendon Press.
- Jarvis, B. G. (2004a). *DirectRT [Computer software]*. New York, NY: Empirisoft.
- Jarvis, B. G. (2004b). *MediaLab [Computer software]*. Columbus, New York, NY: Empirisoft.
- Judd, C. M., James-Hawkins, L., Yzerbyt, V., & Kashima, Y. (2005). Fundamental dimensions of social judgment: Understanding the relations between judgments of competence and warmth. *Journal of Personality and Social Psychology, 89*(6), 899–913.
- Kohlberg, L. (1969). Stage and sequence: The cognitive-developmental approach to socialization. In D. A. Goslin (Ed.), *Handbook of socialization theory and research*. Chicago, IL: Rand McNally.
- Kohlberg, L. (1976). Moral stages and moralization. In T. Lickona (Ed.), *Morality, moral behavior and moral development: Basic issues in theory and research* (pp. 52–73.). New York, NY: Holt, Rinehart & Winston.
- Kunda, Z. (1987). Motivated inference: Self-serving generation and evaluation of causal theories. *Journal of Personality and Social Psychology, 53*(4), 636–647.
- Leach, C. W., Ellemers, N., & Barreto, M. (2007). Group virtue: The importance of morality (versus competence and sociability) in the positive evaluation of in-groups. *Journal of Personality and Social Psychology, 93*(2), 234–249.
- Leidner, B., Castano, E., Zaiser, P., & Giner-Sorolla, R. (2010). *Ingroup glorification, moral disengagement, and justice in the context of collective violence*. *Personality and Social Psychology Bulletin*.
- Marques, J. M., Yzerbyt, V., & Leyens, J.-P. (1988). The “Black Sheep Effect”: Extremity of judgments towards ingroup members as a function of ingroup identification. *European Journal of Social Psychology, 18*(1), 1–16.
- Miller, J. G. (2006). Insights into moral development from cultural psychology. In M. Killen & J. G. Smetana (Eds.), *Handbook of moral development* (pp. 375–398.). Mahwah, NJ: Lawrence Erlbaum.
- Miron, A. M., Branscombe, N. R. (2008). Social categorization, standards of justice, and collective guilt. In A. Nadler, T. E. Malloy, & J. D. Fisher (Eds.), *The social psychology of intergroup reconciliation* (pp. 77–96.). New York, NY: Oxford University Press.
- Miron, A. M., Branscombe, N. R., & Biernat, M. (2010). Motivated shifting of justice standards. *Personality and Social Psychology Bulletin, 36*(6), 768–779.
- Mummendey, A., Linneweber, V., & Loeschper, G. (1984). Actor or victim of aggression: Divergent perspectives—divergent evaluations. *European Journal of Social Psychology, 14*(3), 297–311.
- Norris, D. (2009). Putting it all together: A unified account of word recognition and reaction-time distributions. *Psychological Review, 116*(1), 207–219.
- Norton, M. I., Vandello, J. A., & Darley, J. M. (2004). Casuistry and social category bias. *Journal of Personality and Social Psychology, 87*(6), 817–831.
- Opatow, S. (1990). Moral exclusion and injustice: An introduction. *Journal of Social Issues, 46*, 1–20.
- Payne, B. K., & Cameron, C. D. (2010). Divided minds, divided morals: How implicit social cognition underpins and undermines our sense of social justice. In B. Gawronski & B. K. Payne (Eds.), *Handbook of implicit social cognition: Measurement, theory, and applications* (pp. 445–460.). New York, NY: Guilford Press.
- Payne, B. K., Cheng, C. M., Govorun, O., & Stewart, B. D. (2005). An inkblot for attitudes: Affect misattribution as implicit measurement. *Journal of Personality and Social Psychology, 89*(3), 277–293.
- Pennebaker, J. W. (1993). Putting stress into words: Health, linguistic, and therapeutic implications. *Behaviour Research and Therapy, 31*(6), 539–548.
- Piaget, J. (1932/1965). *The moral judgment of the child*. New York, NY, US: Free Press.
- Pizarro, D., & Bloom, P. (2003). The intelligence of the moral intuitions: A comment on Haidt (2001). *Psychological Review, 110*(1), 193–196.
- Ratcliff, R., Gomez, P., & McKoon, G. (2004). A diffusion model account of the lexical decision task. *Psychological Review, 111*(1), 159–182.
- Reicher, S., Haslam, A., Rath, R. (2008). Making a virtue of evil: A five-step social identity model of the development of collective hate. *Social and Personality Psychology Compass, 2*(3), 1313–1344.
- Roccas, S., Klar, Y., & Liviatan, I. (2006). The paradox of group-based guilt: Modes of national identification, conflict vehemence, and reactions to the ingroup’s moral violations. *Journal of Personality and Social Psychology, 91*(4), 698–711.
- Rozin, P. (1999). The process of moralization. *Psychological Science, 10*(3), 218–221.
- Salthouse, T. A., & Hedden, T. (2002). Interpreting reaction time measures in between-group comparisons. *Journal of Clinical and Experimental Neuropsychology, 24*(7), 858–872.
- Schmiedek, F., Oberauer, K., Wilhelm, O., Süß, H.-M., & Wittmann, W. W. (2007). Individual differences in components of reaction time distributions and their relations to working memory and intelligence. *Journal of Experimental Psychology, 136*(3), 414–429.
- Shweder, R. A. (1982). Beyond self-constructed knowledge: The study of culture and morality. *Merrill-Palmer Quarterly, 28*(1), 41–69.
- Skitka, L. J., Bauman, C. W., & Mullen, E. (2004). Political tolerance and coming to psychological closure following the September 11, 2001, terrorist attacks: An integrative approach. *Personality and Social Psychology Bulletin, 30*(6), 743–756.
- Skitka, L. J., Bauman, C. W., & Sargis, E. G. (2005). Moral conviction: Another contributor to attitude strength or something more? *Journal of Personality and Social Psychology, 88*(6), 895–917.
- Smetana, J. G., Schlagman, N., & Adams, P. W. (1993). Preschool children’s judgments about hypothetical and actual transgressions. *Child Development, 64*, 202–214.
- Stanton, J. M. (2006). The Study Response Open Recruitment Method. (Tech. Rep. No. 13007). Syracuse, NY: Syracuse University, School of Information Studies.
- Staub, E. (1989). Individual and societal (group) values in a motivational perspective and their role in benevolence and harming. In N. Eisenberg, J. Reykowski, & E. Staub (Eds.), *Social and moral values: Individual and societal perspectives* (pp. 45–61.). Hillsdale, UK: Lawrence Erlbaum.
- Staub, E. (1990). Moral exclusion, personal goal theory and extreme destructiveness. *Journal of Social Issues, 46*, 47–64.
- Staub, E., & Bar-Tal, D. (2003). Genocide, mass killing and intractable conflict: Roots, evolution, prevention and reconciliation. In D. O. Sears, L. Huddy, & R. Jervis (Eds.), *Oxford handbook of political psychology* (pp. 710–751.). New York, NY: Oxford University Press.
- Tabachnick, B. G., & Fidell, L. S. (2007). *Using multivariate statistics* (5th ed.). Boston, MA, US: Allyn & Bacon.
- Tajfel, H. (1982). Social psychology of intergroup relations. *Annual Review of Psychology, 33*, 1–39.
- Tedeschi, J. T., & Bond, M. H. (2001). Aversive behavior and aggression in cultural perspective. In R. M. Kowalski (Ed.), *Behaving badly: Aversive behaviors in interpersonal relationships* (pp. 257–293.). Washington, DC, US: American Psychological Association.
- Tedeschi, J. T., Smith, R. B., & Brown, R. C. (1974). A reinterpretation of research on aggression. *Psychological Bulletin, 81*(9), 540–562.
- Tetlock, P. E., Kristel, O. V., Elson, S. B., Green, M. C., & Lerner, J. S. (2000). The psychology of the unthinkable: Taboo trade-offs, forbidden base rates, and heretical counterfactuals. *Journal of Personality and Social Psychology, 78*(5), 853–870.
- Turiel, E. (1983). *The development of social knowledge: Morality and convention*. Cambridge, UK: Cambridge University Press.
- Uhlmann, E. L., & Cohen, G. L. (2005). Constructed criteria: Redefining merit to justify discrimination. *Psychological Science, 16*(6), 474–480.
- Uhlmann, E. L., Pizarro, D. A., Tannenbaum, D., & Ditto, P. H. (2009). The motivated use of moral principles. *Judgment and Decision Making, 4*(6), 476–491.
- Verkuyten, M., & Yildiz, A. A. (2007). National (dis)identification and ethnic and religious identity: A study among Turkish-Dutch Muslims. *Personality and Social Psychology Bulletin, 33*(10), 1448–1462.
- Wallace, C. (2004). The Study Response project: A Description and Evaluation of Using Standing Panels of Participants for Psychological Research. (Tech. Rep. No. 13006). Syracuse, NY: Syracuse University, School of Information Studies.
- Wildschut, T., & Insko, C. A. (2006). A paradox of individual and group morality: Social psychology as empirical philosophy. In P. A. M. Van Lange (Ed.), *Bridging social psychology* (pp. 377–384). Mahwah, NJ: Lawrence Erlbaum.
- Wildschut, T., Insko, C. A., & Gaertner, L. (2002). Intragroup social influence and intergroup competition. *Journal of Personality and Social Psychology, 82*, 975–992.