# The Effects of Moral and Pragmatic Arguments Against Torture on Demands for Judicial Reform

<table>
<thead>
<tr>
<th>Journal:</th>
<th>Political Psychology</th>
</tr>
</thead>
<tbody>
<tr>
<td>Manuscript ID</td>
<td>POPS-OA-09-2015-2491.R1</td>
</tr>
<tr>
<td>Wiley - Manuscript type:</td>
<td>Original Article</td>
</tr>
<tr>
<td>Broad Area of Study:</td>
<td>Group Processes and Intergroup Relations, Identity, Public Opinion</td>
</tr>
<tr>
<td>Epistemological and Methodological Preference:</td>
<td>Experimental &lt; Quantitative</td>
</tr>
<tr>
<td>Specific Research Interests:</td>
<td>Ethics/morality/moral emotions, Group processes</td>
</tr>
<tr>
<td>Geographical Specialization:</td>
<td>USA and Canada</td>
</tr>
</tbody>
</table>

**Abstract:** Torture can be opposed on the basis of pragmatic (e.g., torture does not work) or moral arguments (e.g., torture violates human rights). Three studies investigated how these arguments affect US citizens’ attitudes toward US-committed torture. In Experiment 1, participants expressed stronger demands for redressing the injustice of torture when presented with moral rather than pragmatic or no arguments against torture. Experiment 2 replicated this finding with an extended justice measure and also showed the moderating role of ingroup glorification and attachment. Moral arguments increased justice demands among those who typically react most defensively to ingroup-committed wrongdoings: the highly attached and glorifying. Experiment 3 showed that the effect of moral arguments against torture on justice demands and support for torture among high glorifiers is mediated by moral outrage and empathy but not guilt.
Abstract

Torture can be opposed on the basis of pragmatic (e.g., torture does not work) or moral arguments (e.g., torture violates human rights). Three studies investigated how these arguments affect US citizens’ attitudes toward US-committed torture. In Experiment 1, participants expressed stronger demands for redressing the injustice of torture when presented with moral rather than pragmatic or no arguments against torture. Experiment 2 replicated this finding with an extended justice measure and also showed the moderating role of ingroup glorification and attachment. Moral arguments increased justice demands among those who typically react most defensively to ingroup-committed wrongdoings: the highly attached and glorifying. Experiment 3 showed that the effect of moral arguments against torture on justice demands and support for torture among high glorifiers is mediated by moral outrage and empathy but not guilt.

Word count: 132.

Keywords: torture; morality; judgment; outrage; empathy; justice; ingroup-committed violence.
ARGUMENTS AGAINST TORTURE

The Effects of Moral and Pragmatic Arguments Against Torture on Demands for Judicial Reform

Following the release of the U.S. Senate Select Committee on Intelligence’s study of the CIA’s detention and interrogation program in 2014 (Senate Select Committee on Intelligence, 2014), in 2015 the U.S. Senate and the American Psychological Association have re-stated their opposition to torture. Despite this institutional opposition to torture as well as its prohibition by international declarations and conventions, public opinion still supports torture to a large extent. A poll found that in December 2014, Americans supported torture by a margin of almost 2 to 1 (59% support vs. 31% opposition, Washington Post/ABC News, 2015). Similar numbers have emerged in other countries, such as France, the UK, or South Korea (e.g., Lester, 2005). Even more striking is that since the Abu Ghraib scandal broke in 2004 (Hersh, 2004), and despite official condemnations of the policies under the Bush and Obama administrations, Americans’ support for torture has increased and opposition to it has decreased (Pew Research Center, 2014).

The consistency of public support for torture appears to be multiply determined by interlinked factors. One prominent factor is that when torture is perpetrated by their own (rather than another) country, people are motivated to protect their image of their country, which often leads them to use moral disengagement strategies such as dehumanization or blaming the victims (Leidner, Castano, Zaiser, & Giner-Sorolla, 2010; Viki, Osgood, & Phillips, 2013; see also Kelman, 2005, and Zimbardo, 2007). As a result, people can experience reduced empathy for torture victims, and come to see torture not only as not immoral but even as moral (Tarrant, Branscombe, Warner, &
This perception of torture as not immoral directly affects the perceived costs and benefits of its use. As Liu and Ditto have shown in a sample of more than 1,800 Americans, the more people believed that “forceful interrogation of terrorist suspects” is not wrong, the more effective they believed it to be (Liu & Ditto, 2013, Study 2).

Together, these psychological processes explain why the belief that torture is effective is rather widespread and popular, if erroneous (Janoff-Bulman, 2007; see also Costanzo & Gerrity, 2009). Beliefs in the effectiveness and morality of torture can further explain the longevity of the public’s support for torture. Importantly and problematically, in a representative sample of Americans both beliefs were stronger when people were under the impression that torture had been practiced for a longer (rather than shorter) amount of time (Crandall, Eidelman, Skitka, & Morgan, 2009). Similarly, and similarly problematically, people’s support for torture leads them to see information obtained through torture as more valuable, which in turn strengthens their support for torture (Ames & Lee, 2015).

While the recent research reviewed above furthered our understanding of why people support torture and why public support for torture remains high despite vigorous public debate, rigorous empirical research on the type of arguments that may be successful in eroding this support is lacking. This state of affairs is particularly troubling considering the increasing number of arguments against torture, and their prevalence in political and societal discourse over the past decade. Thus, in three experiments we investigated the effects of different arguments against torture on support for, or opposition to, torture among people with varying degrees of attachment to and glorification of their country. In doing so, the research reported here contributes to an
ARGUMENTS AGAINST TORTURE

emerging knowledge base on attitudes toward torture in particular (Ames & Lee, 2015; Liu & Ditto, 2013; Tarrant et al., 2012; Viki et al., 2013), and to the knowledge base on how to develop a culture of peace and respect for human rights in general (Di Lellio & Castano, 2016; de Rivera, Kurrien, & Olsen, 2007; Kelman, 2010, 2012; Leidner & Li, 2015; Leidner, Tropp, & Lickel, 2013). While the studies were conducted in the context of attitudes toward torture, they also speak to our understanding of attitudes of public opinion more generally, and how these attitudes can shift as a result of communication (i.e., moral and pragmatic arguments about a societal issue) and identity (i.e., attachment and glorification). In this sense, the present research also takes up the challenge to bridge social psychological research on peace and conflict with mainstream social psychology (see Kelman, 2012; Leidner et al., 2013).

Moral versus Pragmatic Arguments Against Torture

Arguments against torture featuring in political and societal discourse fall into two main categories. One is pragmatic, stating that torture can lead to mistreatment of or retaliation against ‘our’ soldiers when they are captured by ‘the enemy’ (Costanzo, Gerrity, & Lykes, 2006). The other is moral, stating that torture violates the constitution, human rights, humanitarian law, ethical values and ideals (Malinowski, 2008; Skoll, 2008). The pragmatic argument does not oppose torture because torture violates universal moral standards or the victims’ rights. Rather, it does so because any possible benefits of torture are not seen as outweighing its costs. These costs include the risk of retaliation against members of the group that perpetrates torture, the risk of torture producing unreliable information, loss of the group’s reputation and credibility in the eyes of its
own members and/or outsiders, and the mental and emotional toll on torturers or members of their group (Costanzo & Gerrity, 2009). Regardless of which cost is emphasized by different pragmatic arguments, what unites them is the point that torture is detrimental to the group that tortures. According to Vice President Joe Biden (2007), torture needs to be opposed because, ‘at the end of the day,’ it “does not work”—meaning, it is not useful to ‘us’ (i.e., the ingroup). As such, pragmatic arguments may reduce support for torture because they draw attention to the costs or repercussions that torture can have for the perpetrator group.

The moral argument, on the other hand, opposes torture on the basis of deontological morality and ethics, rather than utilitarian-consequentialist morality and cost-benefit analyses. Although it could be argued that the moral argument entails utilitarian-consequentialist elements and/or self-serving motives, too, the moral argument arguably does so to a lesser extent than the pragmatic argument. For instance, a person could use the moral argument out of a motivation to protect or reassert ingroup norms and values, and in this sense endorse the moral argument on utilitarian grounds. Yet, the pragmatic argument can only be viewed through a utilitarian-consequentialist lens, whereas the moral argument may partially be viewed through a utilitarian-consequentialist lens, but always also through a deontological lens.

**Which Type of Argument Will Be More Effective?**

Many factors – from cognitive accessibility to affect to values and individual differences – influence the effectiveness of arguments in changing attitudes or opinion (for a review see Petty, Wegener, & Fabrigar, 1997). Thus, it is generally as likely for
ARGUMENTS AGAINST TORTURE

moral arguments against torture to effectuate a change in attitudes toward torture as it is for pragmatic ones. Pragmatic arguments, by definition, do not draw attention to an internalized moral standard that was violated. They do, however, draw attention to benefits (or cost-avoidance) for the ingroup. A large body of research on group processes suggests that people want to provide benefits and reduce costs for the ingroup, especially in intergroup situations (e.g. Brewer, 1979). Thus, pragmatic arguments opposing the idea that torture is effective and therefore would benefit the ingroup (see Costanzo & Gerrity, 2009) might lead to stronger opposition to torture. Yet, recent research shows that people’s belief in the effectiveness of torture decreases, the more they believe torture is inherently immoral (Liu & Ditto, 2013). This suggests that people’s belief in the effectiveness of torture does not only inform attitudes towards torture; attitudes towards torture inform people’s belief in its effectiveness (see also Ames & Lee, 2015). In other words, people do not adjust their attitudes in line with torture’s purported effectiveness as much as they adjust their beliefs in its effectiveness in line with their (pre-existing) attitudes. In this way, then, pre-existing support for torture and beliefs in its effectiveness might actually inoculate people against new, pragmatic arguments about the ineffectiveness of torture. If this is true, pragmatic arguments against torture – calling into question torture’s effectiveness – should have limited effect.

Moral arguments against torture, on the other hand, may be more effective in reducing people’s support for torture. Making salient the notion that an internalized moral standard has been violated, moral arguments against torture should motivate people to demonstrate the ingroup’s virtue and morality through opposing torture more strongly (e.g. Batson, Chao, & Givens, 2009; Mullen & Skitka, 2006), demanding to live up to
and reaffirm moral standards by redressing past and preventing future injustice (e.g. Darley & Pittman, 2003; de Rivera, Gerstmann, & Maisels, 2002; Mikula, Scherer, & Athenstaedt, 1998). This hypothesis is in line with research showing that attitudes such as people’s support for democracy are related to perceptions of the attitude object (e.g. democracy) as fulfilling a value-expressive rather than ego-defensive function (Gastil, 1992). Further, a wide variety of literature in social psychology – on ingroup-committed violence, moral convictions, ingroup dissent and collective action – shows that people at times oppose their group’s group-level behavior, when this behavior clashes with internally held values or convictions (for reviews in the different literatures see Packer, 2008; Skitka & Mullen, 2002; van Zomeren, Postmes, & Spears, 2008). For example, people can oppose (rather than defend) past or anticipated ingroup-committed transgressions such as the 2003 U.S. war against Iraq when they see their group’s image at stake (Iyer, Schmader, & Lickel, 2007).

Such ‘attitudinal (and sometimes actual/behavioral) rebellion’ (Hornsey, 2005) against the ingroup is driven by emotional responses to ingroup-committed transgressions (for reviews see Thomas, McGarty, & Mavor, 2009a, 2009b). For instance, anger or outrage directed at the ingroup, as well as group-based guilt, motivate people to right ingroup-committed wrongs (Doosje et al., 1998; Iyer et al., 2007; Leach et al., 2006; Lickel et al., 2005; van Zomeren et al., 2004). Similarly, empathy motivates people to end others’ suffering, including that of outgroup members (for reviews see Castano, 2012; Dovidio, Johnson, Gaertner, Pearson, Saguy, & Ashburn-Nardo, 2010; Stephan & Finlay, 1999). It is thus possible, even in situations of ingroup-committed wrongdoing in which people are often motivated to defend the ingroup’s violation of generally held
ARGUMENTS AGAINST TORTURE

moral standards, that moral arguments and their focus on the violations of these standards lead to less support for torture, and more desire to redress its injustice, by enhancing group-level emotional responses such as outgroup-directed empathy, ingroup-directed anger/outrage, and group-based guilt. In the context of political action intentions, however, group-based guilt might be less predictive than anger/outrage or empathy, because guilt is more self-reflective (as opposed to other-directed) than empathy, and more passive (as opposed to action-oriented) than anger/outrage (see Iyer et al., 2007; Lodewijkz, Kersten, & van Zomeren, 2008; Wakslack, Jost, Tyler, & Chen, 2007). Consistent with this view of the effects of emotional responses to ingroup-committed transgressions on political action intentions, group-based guilt did not predict political action intentions after controlling for anger or other variables (Iyer et al., 2006, 2007; Leach et al., 2006). Nevertheless, when investigating the emotional processes underlying the effects of arguments against torture on people’s opinion on their group’s use of torture (Study 3), we included group-based guilt alongside ingroup-directed anger/outrage and outgroup-directed empathy in order to provide a comprehensive test of these processes.

Which Type of People Will Be More Convinced?

Moral arguments critical of the group’s actions may be even more powerful to the extent that people are psychologically invested in the moral image of their group. This hypothesis is in line with research demonstrating that information-processing and attitude change are likely to be biased when the information is personally relevant and people have a vested interest in it (Chen et al., 1992; Hutton & Baumeister, 1992; Liberman &
Chaiken, 1992; Petty & Cacioppo, 1979). More specific to group processes, our hypothesis is also supported by findings that morality is a crucial dimension on which the ingroup is evaluated (Leach, Ellemers, & Barreto, 2007), and that people who strongly identify with their group can at times be more (rather than less) likely to challenge their own group, in an effort to save the group from “morally faltering” (Packer, 2009; for a reviews see Packer, 2008; van Zomeren et al., 2008; see also Kelman, 2006). From this perspective, yielding to moral arguments against torture can be part of people’s attempts to protect the ingroup’s morality, as well as their self to the extent that it includes the group. Similarly, different views of the role that dissonance plays in attitude change (for a review see Petty et al., 1997) suggest that people adjust their attitudes when they observe a gap between behavior and their attitude/belief system (Festinger, 1957; Cooper & Fazio, 1984) or self-concept (Aronson, 1968; Steele, 1988). Thus, moral arguments against torture might effectuate stronger opposition to torture, perhaps even more so the more invested people are in their group’s moral image.

Again, we reasoned that pragmatic arguments would be relatively less likely than moral arguments to change people’s opinion on ingroup-committed torture, even among those strongly identifying with their group. This reasoning was based on Ames and Lee’s (2015) finding that pre-existing support for torture strengthens beliefs in its effectiveness. If high identifiers are generally more supportive of their countries’ policies (including ‘enhanced interrogation’), and if this pre-existing attitude strengthens the belief that ‘torture works,’ then exposure to arguments that ‘torture does not work’ may not be very effective even for high identifiers who otherwise – if exposed to arguments that ‘torture is wrong,’ for example – may be swayed.
ARGUMENTS AGAINST TORTURE

However, group identification is not a unidimensional construct; it is better assessed by looking at both attachment with and glorification of the group (Roccas, Klar, & Liviatan, 2006). While attachment refers to commitment to and perceived importance of the group, glorification refers to the belief that the group is superior to other groups (especially in terms of morality) and prescribes deference to group norms and authorities. Thus, glorification has a comparative component; it focuses on ingroup superiority, and emphasizes loyalty and obedience to the ingroup. Attachment is not comparative in nature; it is ingroup-focused and allows for a more critical evaluation of the ingroup and its actions. Glorification is related to greater use of exonerating cognitions for and moral disengagement from the wrongdoings of the ingroup, whereas attachment typically shows either no effect or the opposite (Leidner & Castano, 2012; Leidner et al., 2010; Roccas et al., 2006). We thus expected high ingroup glorification, but not attachment, generally to be associated with lower demands to redress the injustice of ingroup-committed torture.

Importantly, however, we also expected glorification to moderate the effects of arguments against torture on people’s support for or opposition to torture. If high identifiers are more sensitive to moral than pragmatic arguments – for the reasons just explained – then this sensitivity should be particularly driven by glorification, as it is the glorified image of the group that takes pride in the group’s purported morality. This expected moderation of effects of arguments by glorification should be such that low glorifiers should display rather high demands for justice regardless of whether the argument is moral or pragmatic. High glorifiers, on the other hand, should increase their demands in response to a moral argument in particular (as opposed to one that is pragmatic).
Overview of the Studies

We present three experiments conducted in 2009 (Study 1), June/July and November 2010 (Study 2 and 3, respectively), that investigated the effectiveness of moral (as compared to pragmatic) arguments on people’s attitudes towards specific events involving torture perpetrated by their own country, as well as torture itself, while also testing possible moderators and mediators. First, we report the results of a pilot study that tested the effectiveness of the manipulation used in the subsequent experiments. Study 1 then provides a first experimental test of the effects of moral vs. pragmatic arguments against torture. Study 2 tested the moderation of these effects by people’s glorification of their group, and Study 3 tested the mediation of the effects by outgroup-directed empathy, ingroup-directed anger/outrage, and group-based guilt (moderated by ingroup glorification).

Pilot Study

From a psychological as well as methodological perspective, it is important that the arguments we theorized earlier to be pragmatic are actually seen by people as “pragmatic,” and that the arguments we theorized earlier to be moral are actually seen by people as “moral.” Thus, we ran a pilot study testing whether moral arguments against torture elicit perceptions of a moral critique of torture, and whether pragmatic arguments elicit perceptions of a pragmatic critique of torture. To this end, we presented forty-two participants recruited through Amazon Mechanical Turk (MTurk) for monetary compensation with a brief statement reminding them that U.S. soldiers had allegedly
ARGUMENTS AGAINST TORTURE

tortured Iraqi detainees, followed by a statement from a U.S. General commenting on the events (see Online Appendix). The arguments used by the general to condemn the soldiers’ behavior were either pragmatic (e.g., “Torturing prisoners is against the interests of our own military”) or moral in nature (e.g., “Torturing prisoners is against the ideals of the United States of America and its military”). Next, participants conveyed, on visual analog scales ranging from 1 to 9, the extent to which they (a) perceived the arguments as being rather pragmatic (1) or rather moral (9) in nature; agreed or disagreed with the arguments; liked or disliked the arguments; and found the arguments convincing. The latter three dimensions were assessed to ensure that pragmatic and moral arguments did not vary on other important dimensions, but only on the dimension of perceived pragmatism vs. morality.

Of the 42 participants, three were not born in the United States and one participant answered “9” to all questions in the survey. After eliminating these four participants, we ran analyses on the remaining 38 participants (age: $M = 30.70$, range = 19-56; gender: 16 male, 21 female, one did not indicate gender). Participants saw the moral arguments as more moral ($M = 6.88$) than the pragmatic ones ($M = 3.51$), $F(1, 36) = 16.73, p < .001$. T-tests further confirmed that the perceived morality of the moral arguments was significantly higher than the midpoint of the scale, $t(20) = 3.74, p = .001$, whereas the perceived morality of the pragmatic arguments was significantly lower than the midpoint of the scale, $t(16) = -2.21, p = .042$. Importantly, participants only differed in how moral they perceived the arguments, but not in how much they agreed with them, liked them, or were convinced by them, $F$s < 0.64, $ps > .430$. Furthermore, the difference in perceived morality remained significant even when controlling for agreement, liking,
or level of being convinced, $F(1, 33) = 15.10, p < .001$. Thus, we found strong support for the validity of our conceptualization and operationalization of moral and pragmatic arguments.

**Experiment 1**

Experiment 1 focused on the torture of prisoners at the hands of American soldiers in Iraq, and compared the effects of moral and pragmatic arguments to a baseline without arguments. Rather than focusing on support for or opposition to torture in the abstract, we focused on it in the concrete, assessing people’s demands to redress the injustice of several instances of torture reminiscent of Abu Ghraib.

**Method**

**Participants**

The sample consisted of 215 people recruited via the internet website Craigslist. After excluding fourteen people for whom the U.S. was not the ingroup, and ten people who failed to remember the identity of the soldiers who perpetrated torture in the manipulation check, the remaining sample comprised of 191 people (age: $M = 35.94$, $SD = 13.41$, range = 18-78; gender: 54 male, 132 female, 5 did not indicate gender; conservatism: $M = 3.35$, $SD = 2.00$, range = 1-9). Participants in the different conditions did not significantly differ in age ($M$s = 35.95, 35.03 and 36.70 for moral, pragmatic, and no-argument condition, respectively), $F(2, 186) = 0.24, p = .787$, conservatism ($M$s = 3.34, 3.43 and 3.16, respectively), $F(2, 185) = 0.26, p = .771$, or gender, $\chi^2(2) = 0.96, p = .617$, indicating that participants were demographically comparable across conditions.

**Procedure**
ARGUMENTS AGAINST TORTURE

The experiment was conducted online. Participants were randomly assigned to read one of three fictitious news articles reporting on four cases of U.S. soldiers torturing Iraqi prisoners in a Baghdad jail. The article was comparable to real events and presented in the layout of online articles of The New York Times. Depending on condition, the article concluded with criticism of the ingroup’s actions on pragmatic grounds or on moral grounds, with the exact same excerpts used in the pilot study reported above. In the third, no-argument condition, the article did not end in any criticism of the ingroup’s actions.

Measures

Demands for justice. After reading the article, participants responded to eight items measuring demands for retributive (e.g., “Independently from any other kind of punishment, these U.S. soldiers should be fired by the army”) and restorative aspects of justice (e.g., “The families of the victims should receive an apology by the U.S. government”, “The families of the victims should receive financial compensation”), using a 1-9 visual analog scale (Leidner et al., 2010). The first retributive justice item asked whether the soldiers should be sent to prison and participants answered yes or no. If participants answered “yes,” they were then asked to indicate the recommended sentence on a second item (1=minimum by law; 9=maximum by law); if they answered “no” on the first item, they scored zero on this second item. All subsequent items were answered by everyone, regardless of the responses to the first or second item, on a scale ranging from 1 to 9.

Manipulation checks. Finally, in open-ended questions participants were asked “Where did the prisoners come from? (which country)”, “Where did the soldiers come
from? (which country)”, and “In which country or city was the prison located the newspaper article was about?”.

Results and Discussion

We first factor-analyzed all demands for justice items but the first, dichotomous item. Factor analysis indicated that the items loaded on one factor (97% of the variance explained; $\alpha = .85$). We thus averaged them into a composite score; before doing so, however, we had to standardize the items due to the different scale for the second item.\(^1\) The resulting score was used as a dependent variable ($M = -.01, SD = .75$) in an analysis of variance with the experimental manipulation as the independent variable (IV: no-argument vs. pragmatic vs. moral). The effect of argument was significant, $F(2, 188) = 5.15, p = .007, \eta^2 = .05$. Participants in the moral argument condition ($M = .24, SD = .47$) scored significantly higher than participants in the no-argument ($M = -.12, SD = .89$), $t(188) = 2.80, p = .006, d = .04$, and pragmatic argument conditions ($M = -.13, SD = .77$), $t(188) = 2.76, p = .006, d = .04$, while the latter two did not differ, $t(188) = 0.00, p = .947$ (see Figure 1). For additional analyses, see Online Supplemental Materials. Altogether, the results suggest that moral but not pragmatic arguments against torture change people’s opinions on torture.

Experiment 2

Experiment 2 aimed to replicate and extend Experiment 1. In Experiment 1, it remained unclear whether or not the observed pattern of findings holds among people who strongly identify with their group. This moderating factor was thus tested in

\(^1\) The results reported below did not change when only using the six items that shared the same scale (and thus did not necessitate standardization when being used without item No. 2).
ARGUMENTS AGAINST TORTURE

Experiment 2. As explained earlier, we predicted that high glorifiers would be particularly swayed by moral arguments, because morality is the most important domain people focus on when evaluating their group (Leach et al., 2007), and high glorifiers are particularly invested in the moral image of their group (Roccas et al, 2006). Thus, the expected moderation of effects of arguments by glorification should be such that low glorifiers should display rather high demands for justice regardless of (the type of) argument, whereas high glorifiers should increase their demands in response to a moral argument. In other words, the effects of moral arguments found in Experiment 1 should only emerge among high glorifiers.

In addition to the moderating factors, Experiment 2 also included a more comprehensive scale measuring justice demands: willingness to provide immediate help and relief, as well as sustained financial support for victims; the necessity to express remorse and ask for forgiveness – two important dimensions of symbolic compensation that affect conflict resolution positively (Cehajic, Brown, & Castano, 2008; Noor, Brown, Gonzalez, Manzi, & Lewis, 2008); and, most importantly, demands for structural and policy reforms to prevent torture from recurring in the future.

Experiment 2 also differed in the way the sample was recruited. Instead of Craigslist, we used Amazon’s Mechanical Turk (MTurk). MTurk participants have been found to be more representative of the U.S. population than standard internet samples and significantly more diverse than college samples (Buhrmester, Kwang, & Gosling, 2011). Furthermore, data tends to be of the same quality and reliability as data obtained via traditional samples (Buhrmester et al., 2011; Casler, Bickel, & Hackett, 2013; Paolacci & Chandler, 2014), and replicates effects found in traditional samples, be it in survey
designs (Goodman, Cryder, & Cheema, 2013; Horton, Rand, & Zeckhauser, 2011; Lutz, 2016; Paolacci, Chandler, & Ipeirotis, 2010), diary studies (Boynton & Richman, 2014), or bogus partner interactions (Summerville & Chartier, 2013). Importantly, online samples have been used in both correlational and experimental studies in clinical, personality, social and political psychology (Andover, 2014; Campbell & Kay, 2014; Costa, Ntoumanis, & Bartholomew, 2015; Craemer, 2010; Jonason, Wee, Li, & Jackson, 2014; Leidner, 2015; Leidner & Castano, 2012; Leidner, Castano, & Gingles, 2013; Leidner et al., 2010; Lickel, Kushlev, Savalei, Matta, & Schmader, 2014), including research on attitudes (e.g., Adelman, Leidner, Ünal, Nahhas, & Shnabel, in press; Bizer, Hart, & Jekogian, 2012; Rovenpor, Leidner, Kardos, & O’Brien, 2016).

Compared to traditional research, however, participants in online research tend to be less attentive to the research material (Goodman et al., 2013). This is likely to constitute a problem in research studies as those reported here, where it is imperative that participants attend to the information presented in the manipulation (e.g., the alleged newspaper article) for the theorized attitude change and its underlying psychological processes to emerge. We thus recorded the time participants spent processing the information. Since even fast readers cannot read over 600 words per minute without considerable loss of comprehension (Carver, 1985), participants who read at a rate faster than 600 words per minute (i.e. low reading time) were to be excluded from the sample before conducting data analysis. In our research context of attitudes toward torture, the advantages of online samples (heterogeneity, diversity, age range, etc.) outweigh this disadvantage – i.e. that some participants may pay less-than-ideal attention to stimulus/manipulation materials – especially given that laboratory studies with college
ARGUMENTS AGAINST TORTURE

student samples would have more serious disadvantages in this context (e.g., threat to ecological validity, generalizability, etc.).

Method

Participants

324 participants were recruited via MTurk. Of these, 71 were eliminated because the time spent reading the article was below the minimum for comprehension, as indicated above. Seven participants who were univariate outliers on the time they took to read the article (i.e., they spent a disproportionate amount of time, indicating that they were likely interrupted during the study) were also eliminated (Tabachnick & Fidell, 2007), together with ten participants who reported not being born in the U.S. and three participants who failed manipulation checks (see below; two remembered the victims’ nationality incorrectly, one remembered the perpetrators’ nationality incorrectly). The final sample comprised 233 participants (age: $M = 31.13$, $SD = 11.58$, range = 18-81; gender: 78 male, 152 female, 3 did not indicate gender; conservatism: $M = 4.29$, $SD = 2.25$, range = 1-9). Participants in the different conditions did not significantly differ in age ($Ms = 31.02, 31.86$ and $30.68$ for moral, pragmatic, and no-argument condition, respectively), $F(2, 227) = 0.19, p = .826$, conservatism ($Ms = 4.18, 4.17$ and $4.19$, respectively), $F(2, 227) = 0.52, p = .600$, or gender, $\chi^2(2) = 0.40, p = .819$, indicating that participants were demographically comparable across conditions.

The initial deletion of participants with below-threshold reading time for the manipulation material resulted in the elimination of a hefty 22% of the sample. Inclusion of these participants in our analyses did not produce the same pattern of results reported below. However, this discrepancy is consistent with the rationale behind the selection;
namely that these participants could not be affected by the manipulation because they did not attend to it and did not adequately process the information for attitude change to occur. Further, exclusion of 22% of the sample is within normal range for online studies and deemed necessary to ensure high data quality (Chandler, Mueller, & Paolacci, 2013).

**Procedure**

Participants were randomly assigned to read one of the three news articles used in Experiment 1, and filled out the measures described below, followed by demographics.

**Measures**

All answers were given on an analogue visual scale ranging from $1 = \text{no, absolutely not}$ to $9 = \text{yes, absolutely}$. Factor analyses suggested that all the scales were unidimensional, as indicated by scree plots and the eigenvalue-greater-one criterion, with factor loadings $> .40$, explaining 75% or more of the variance for each scale.

*Manipulation checks* consisted of the same questions as in Experiment 1.

*Ingroup glorification and attachment* were measured with the scales by Roccas et al. (2006). Attachment ($\alpha = .94$, $M = 6.71$, $SD = 1.79$) included eight items such as “It is important for me to serve my country” and “Being an American is an important part of my identity”. Glorification ($\alpha = .83$, $M = 4.72$, $SD = 1.44$) included eight items such as “Other nations can learn a lot from us” and “It is disloyal for Americans to criticize the U.S.”.

*Demands for justice* were measured with: six statements regarding the individual ingroup perpetrators (e.g., “These U.S. soldiers should be punished for the things they have done); four items measuring participants’ willingness to provide immediate help to the victims (e.g., “The U.S. should make sure the basic needs of the victims’ family...
ARGUMENTS AGAINST TORTURE

members are met, so they can mourn in peace”); three items asking participants about financial reparations (e.g., “The families of the victims should receive financial compensation from the U.S.”); five items measuring participants’ demands for apologies, remorse, and forgiveness (e.g., “The families of the victims should receive an apology by the U.S. government”); and five items measuring participants’ demands for reform (e.g., “To prevent prisoner abuse in the future, the U.S. governmental and military structures should be reorganized”). All but one item with a low factor loading (about capital punishment for the U.S. soldiers) were averaged into a composite score ($\alpha = .94, M = 6.66, SD = 1.42$).

Results

We first ran the same analysis as in Experiment 1. A marginally significant effect of the type of argument on participants’ demands for justice emerged, $F(2, 230) = 2.74, p = .067$, $\eta^2 = .023$. Whereas no difference emerged between participants in the no-argument and pragmatic argument conditions, $t(230) = 0.00, p = .960, d = .00$, participants in the moral argument condition demanded significantly more justice ($M = 6.94$) than participants in the no-argument condition ($M = 6.48$), $t(230) = 2.08, p = .039, d = .32$, and marginally significantly more than participants in the pragmatic argument condition ($M = 6.49$), $t(230) = 1.90, p = .059, d = .31$ (see Figure 2). While the omnibus test and the preplanned contrast between moral and no-argument reached only marginal significance (unlike the contrast between moral and pragmatic argument, which reached full significance), this pattern by and large reproduced the results from Experiment 1.

---

2 Using a statistically conservative threshold for the notion of “marginal significance,” we refer to p values between .05 and .07 as marginally significant, and p values greater than .07 as non-significant. Yet, these verbal descriptors should not detract from the fact that p values are, of course, on a continuum from 0 to 1, and that any descriptors are based on arbitrary conventions of language and less nuanced than the values they describe.
Further, when adding age or gender as a covariate to the model, the omnibus test as well as the two critical contrasts became fully significant.

To test for the focal hypothesis of Study 2 (i.e. the hypothesis of a joint effect of argument and glorification on demands for justice), we computed a moderated regression with condition, glorification, attachment, and their interactions as additional predictors. Before doing so, we ensured that neither glorification nor attachment was itself affected by the type of argument, $F$s(2, 229) = 0.06 and 0.91, $ps = .939$ and .402, respectively, and then centered both to allow for their use as moderators in our regression model. In this model, the main effect of glorification was significant, $F(1, 220) = 11.73, p < .001$, $\eta^2 = .05$, showing that it had a strong negative effect on demands for justice, $\beta = -.41$. The only other significant effect was the three-way interaction, $F(2, 220) = 4.87, p = .003$, $\eta^2 = .06$. The effect of condition at low and high levels of attachment (+/- 1 SD) and glorification (+/- 1 SD) was thus computed. Only for individuals high in glorification and attachment, demands for justice differed between conditions. Among this set of individuals, justice demands differed significantly between the moral condition ($M = 6.95$) and the no-argument condition ($M = 6.17$), $t(220) = 2.30, p = .023, d = .36$, but not between the pragmatic condition ($M = 6.42$) and the no-argument condition, $t(220) = .67, p = .504$, nor between the pragmatic and the moral condition, $t(220) = -1.42, p = .159$.

Among any other set of individuals (i.e., individuals high in either glorification or attachment, or low in both), none of the conditions differed from each other. In other words, only for individuals who were strongly attached to and strongly glorifying of their country, moral arguments against torture increased justice demands (compared to no
ARGUMENTS AGAINST TORTURE

argument; see Figure 3). These results remained virtually the same when adding age or
gender as covariates to the model.

Discussion

Experiment 2 largely reproduced (or even fully reproduced, when adding age or
gender as covariates) the findings of Experiment 1 using a more comprehensive measure
of demands for justice. Compared to participants who read about accounts of torture
followed by no argument or pragmatic arguments, participants exposed to moral
arguments voiced (somewhat) stronger demands for justice. While the omnibus test and
one of the pre-planned contrasts reached only marginal statistical significance, the effect
sizes of both the overall effect of argument and the comparisons between moral versus
neutral/pragmatic argument were small but meaningful—suggesting the effects observed
in Study 1 and 2 are consistent and reproducible (see Cohen, 1994; Fisher, 1925; Kirk,
1996, 2003). Importantly, the pattern of results in the three conditions and their three
pair-wise comparisons matched the pattern in Study 1.

Extending Study 1, the effect of argument on demands for justice was moderated
by glorification and attachment. While we expected glorification alone to be a significant
moderator, the expected moderating effect of glorification occurred only among strongly
(but not weakly) attached individuals. As this result was partially unexpected given
previous findings showing that attachment often has the opposite effect of glorification
(but see Leidner, 2015, Study 2), we explored it further in Experiment 3.

Experiment 3
Experiment 3 expanded on our earlier operationalization of opinions on torture as demands to redress its injustice, but also addressed more directly the main issue at hand with a direct measure of support for / opposition to torture. Second, it investigated mediating factors of the effects observed in Experiment 1 and 2: moral outrage, empathy, and guilt. Because in Experiment 1 and Experiment 2 the pragmatic condition did not differ from the no-argument condition, Experiment 3 focused on the moral argument condition (compared to no argument) only.

Method

Participants

234 participants were recruited through MTurk. Of these, thirty-three were eliminated because the time spent reading the article was below the minimum for comprehension, as in Experiment 2. Five participants were identified as univariate outliers (Tabachnick & Fidell, 2007) for taking significantly longer than average to read the article – possibly being interrupted during the study. Five other participants not born in the U.S., and one participant who specified the identity of the perpetrators incorrectly, were also eliminated. 190 participants were thus retained for subsequent analyses (age: $M = 33.34$, $SD = 12.32$, range = 18-72; gender: 50 male, 133 female, 7 did not indicate gender; conservatism: $M = 4.37$, $SD = 2.12$, range = 1-9; education: one reported less than high school, 21 reported high school / GED, 67 reported some college, 68 reported to have a college degree, 23 reported to have a Master’s degree, 3 reported to have a postdoctoral degree, 7 did not indicate; religion: 24 self-identified as agnostic, 26 as atheist, 3 as Buddhist, 36 as Catholic, 5 as Jewish, 1 as Muslim, 37 as Protestant, 50 as “other,” and 8 did not indicate; 94 reported ties through family or friends to the U.S.
ARGUMENTS AGAINST TORTURE

Army, 89 reported not to have any such ties, 7 did not indicate). Participants in the moral condition and participants in the no-argument condition did not significantly differ in age ($M_s = 33.68$ and 33.02 for moral and no-argument condition, respectively), $F(1, 180) = 0.13, p = .720$, conservatism ($M_s = 4.46$ and 4.29, respectively), $F(1, 181) = 0.29, p = .591$, gender, $\chi^2(1) = 1.57, p = .210$, education, $\chi^2(5) = 6.83, p = .234$, religion, $\chi^2(7) = 7.16, p = .412$, or ties to the U.S. Army, $\chi^2(1) = 1.19, p = .275$. Again, this indicated that participants were demographically comparable across conditions.

Procedure

The experiment was conducted online. Participants were randomly assigned to read one of the two articles (moral vs. no-argument).

Measures

Participants completed items measuring empathy (six items, e.g. “I am moved by the plight of the prisoners”; $\alpha = .93, M = 5.93, SD = 1.96$), collective guilt (five items, e.g. “As an American, thinking about what these prisoners endured makes me feel guilty”; $\alpha = .98, M = 5.35, SD = 2.52$), moral outrage (three items, e.g. “I am morally outraged by the events I read about in the news report”; $\alpha = .97, M = 7.03, SD = 2.11$), and support for torture (five items, “To what extent the torture of the prisoners was… ‘professional,’ ‘appropriate,’ ‘understandable,’ ‘fair,’ and ‘justifiable’”; $\alpha = .93, M = 2.19, SD = 1.54$). Demands for justice ($\alpha = .95, M = 6.62, SD = 1.55$), attachment ($\alpha = .92, M = 6.79, SD = 1.71$) and glorification ($\alpha = .85, M = 4.86, SD = 1.49$) were measured using the same items as in Experiment 2. The composite score for demands for justice was computed with all but one item; the same item that was excluded in Experiment 2, about capital punishment for the U.S. soldiers, also had a poor factor
loading, below .40, in this experiment. All answers were given on a visual analogue scale ranging from 1 to 9.

**Manipulation check.** In open-ended questions, participants were asked to think about the news report and recall which country the prisoners (“Which country did the prisoners come from?”) and the soldiers (“Which country did the soldiers come from?”) came from. All participants passed the manipulation check.

**Results**

Neither glorification nor attachment was affected by the type of argument, $F_s < 1$. We centered them and included them as independent (continuous) variables in subsequent analyses, together with argument (moral vs. no-argument) and their interaction term. Age, education, religion, and ties to the U.S. Army had no effects on any dependent variables (DVs), nor did they change the results reported below when entered as covariates in the analyses. Gender and conservatism, however, did have effects on all DVs, and did change the results reported below when entered as covariates in the analyses. Importantly, the use of gender and conservatism as covariates did not decrease but increase the key interaction effect (of argument type by glorification) on all DVs (in the predicted direction). Below we first and foremost report the analyses without covariates, as those were what we had planned in our a priori data-analytical strategy. Analyses where the addition of gender and conservatism made a difference in conventional significance levels compared to the analyses without covariates are reported in the Online Supplemental Materials.

**Demand for justice.** Glorification had a significant main effect, $F(1, 178) = 29.84$, $p < .001$, $\eta^2 = .14$. More glorification led to less justice motives ($\beta = -.73$). The
interaction between glorification and argument was also significant, $F(1, 178) = 5.02, p = .026, \eta^2 = .027$. Low glorifiers’ justice demands did not differ between the moral ($M = 7.13, SE = .25$) and the no-argument ($M = 7.47, SE = .24$) condition, $t(178) = -1.00, p = .321$. High glorifiers, however, demanded significantly more justice in the moral ($M = 6.26, SE = .28$) than in the no-argument condition ($M = 5.40, SE = .26$), $t(178) = 2.26, p = .025, d = .34$ (see Figure 4). All other effects were non-significant, $Fs < 2.4, ps > .13$.

**Support for torture.** Glorification and attachment predicted support for torture in opposite ways, $F(1, 178) = 32.52, p < .001, \eta^2 = .15, \beta = .74$, and $F(1, 178) = 3.57, p = .060, \eta^2 = .02, \beta = -.26$, respectively. The interaction between glorification and condition was significant, $F(1, 178) = 9.69, p = .002, \eta^2 = .05$. Low glorifiers did not differ in their support for torture in the moral ($M = 1.74, SE = .24$) versus no-argument condition ($M = 1.15, SE = .23$), $t(178) = 1.77, p = .079$. High glorifiers, on the other hand, supported torture significantly less in the moral ($M = 2.41, SE = .26$) than in the no-argument condition ($M = 3.42, SE = .25$), $t(178) = -2.79, p = .006, d = .41$ (see Figure 5). No other effect reached significance.

**Empathy.** Glorification had a significant main effect, $F(1, 178) = 24.04, p < .001, \eta^2 = .12, \beta = -.84$, and so did attachment, $F(1, 178) = 5.91, p = .016, \eta^2 = .03, \beta = .44$. The interaction effect of glorification and condition was significant, $F(1, 178) = 5.44, p = .021, \eta^2 = .03$. Low glorifiers showed no difference in empathy between the moral ($M = 6.35, SE = .32$) and the no-argument condition ($M = 6.80, SE = .31$), $t(178) = -1.00, p = .319$. High glorifiers, on the other hand, displayed significantly more empathy in the moral ($M = 5.46, SE = .35$) than in the no-argument condition ($M = 4.31, SE = .33$), $t(178) = 2.39, p = .018, d = .35$ (see Figure 6). No other effect reached significance.
Collective guilt. Glorification and attachment had opposite effects, $F(1, 178) = 16.81$, $p < .001$, $\eta^2 = .09$, $\beta = -.94$, and $F(1, 178) = 10.22$, $p = .002$, $\eta^2 = .05$, $\beta = .77$, respectively. The interaction of glorification and argument on collective guilt did not reach significance, $F(1, 178) = 2.02$, $p = .157$, $\eta^2 = .01$. No other effect reached significance.

Moral outrage. Glorification and attachment had opposite, significant effects, $F(1, 178) = 31.53$, $p < .001$, $\eta^2 = .15$, $\beta = -1.03$, and $F(1, 178) = 21.97$, $p < .001$, $\eta^2 = .11$, $\beta = .91$, respectively. The interaction of glorification and argument was marginally significant, $F(1, 178) = 3.63$, $p = .058$, $\eta^2 = .02$. Among low glorifiers, moral outrage did not differ between the moral ($M = 7.80$, $SE = .34$) and the no-argument condition ($M = 8.16$, $SE = .33$), $t(178) = -.76$, $p = .447$. High glorifiers, on the other hand, reported significantly more moral outrage in the moral ($M = 6.44$, $SE = .37$) than in the no-argument condition ($M = 5.41$, $SE = .35$), $t(178) = 2.00$, $p = .047$, $d = .30$ (see Figure 7). No other effect reached significance.

Mediational analyses. We tested whether the effect of moral arguments among high glorifiers, on both demands for justice and support for torture, was mediated by empathy and outrage (as guilt was not affected by argument in the first place), running a moderated multiple mediation analysis with 5,000 bootstrap samples and 95% confident intervals (Hayes, 2012). To be consistent with our prediction and our above tests of a two-way interaction of condition by glorification, and because we needed now to test for the indirect effects at low and high levels of one moderator (glorification) across the levels of a second moderator (attachment) while still accounting for this second moderator, we used Hayes’s “model 8” with added control terms of attachment and its
ARGUMENTS AGAINST TORTURE

interactions with condition and glorification. Hayes’s “model 10,” while allowing for a more straightforward way to include a second moderator, in this case would not have allowed us to test indirect effects at low and high levels of glorification across levels of attachment (i.e., among high glorifiers low and high in attachment) while still accounting for attachment, nor does it provide an index of moderated mediation. Thus, using model 8, we introduced argument (coded as ‘-1’ for participants in the no-argument and ‘+1’ for participants in the moral condition) as independent variable, empathy and outrage as mediators, justice as DV, glorification as moderator, and attachment and its interaction terms as covariates. Collective guilt was not entered as mediator because it was not jointly affected by condition and glorification to begin with, in the analysis without covariates (and even with covariates it was only marginally affected). As hypothesized, empathy and outrage only mediated the effect of argument among high glorifiers but not among low glorifiers, controlling for attachment and its interactions with condition and glorification. For high glorifiers, the indirect effects through empathy (boot coefficient = .190, lower CI = .059, upper CI = .393) and outrage (boot coefficient = .172, lower CI = .091, upper CI = .367) were both significant. For low glorifiers, neither indirect effect was significant (empathy: boot coefficient = -.085, lower CI = -.234, upper CI = .039; outrage: boot coefficient = -.066, lower CI = -.214, upper CI = .067; see Figure 8). Further, the difference between these (significant) indirect effects at high levels of glorification and these (non-significant) indirect effects at low levels of glorification was also significant (empathy: boot coefficient = 137, lower CI = .040, upper CI = .273; outrage: boot coefficient = .119, lower CI = .065, upper CI = .265).
The same mediational model with support for torture as the DV (instead of demands for justice) revealed a similar pattern: empathy and outrage mediated the effect of argument among high but not low glorifiers, controlling for attachment and its interactions with condition and glorification. For high glorifiers, the indirect effects through empathy (boot coefficient = -.119, lower CI = -.277, upper CI = -.035) and outrage (boot coefficient = -.188, lower CI = -.408, upper CI = -.013) were both significant. For low glorifiers, neither indirect effect was significant (empathy: boot coefficient = .053, lower CI = -.018, upper CI = .166; outrage: boot coefficient = .070, lower CI = -.065, upper CI = .255; see Figure 9). Again, the difference between these (significant) indirect effects at high levels of glorification and these (non-significant) indirect effects at low levels of glorification was also significant (empathy: boot coefficient = -.086, lower CI = -.192, upper CI = -.026; outrage: boot coefficient = -.130, lower CI = -.300, upper CI = -.003).³

Discussion

Experiment 3 replicated the effect of moral arguments against torture on demands for justice and extended previous results on a measure of support for torture. While Experiment 2 found moral arguments to affect only high glorifiers who were also high in attachment, Experiment 3 found this effect among high glorifiers across levels of attachment (i.e., among high glorifiers high or low in attachment). Most importantly, Experiment 3 investigated the mediating role of three variables: moral outrage, empathy, and guilt. The first two showed very similar patterns. Namely, while high glorifiers in general conveyed less moral outrage and empathy towards the victims of ingroup-

³ For both demands for justice as well as support for torture, the moderated mediation through empathy and outrage remained significant when entering gender and conservatism as additional covariates.
ARGUMENTS AGAINST TORTURE

committed torture, they were affected by moral arguments: greater moral outrage and empathy was observed among high glorifiers who read a moral argument against torture as opposed to no argument against torture. Also, both of these variables emerged as mediators of the effect of moral arguments on demands for justice and support for torture among high glorifiers.

General Discussion

In 1987, the United Nations Convention against Torture and Other Cruel, Inhuman or Degrading Treatment or Punishment (UNCAT) came into force. Article 2 of the convention prohibits use and justification of torture without any exceptions, and commands all participating countries to prevent the use of torture in their territories. Yet, public opinion in the U.S. and worldwide supports the use of torture in large numbers, even majorities. We thus set out to investigate which arguments against torture are most likely to resonate with the public and lead to stronger demands to redress the injustice of torture and to weaker support for torture it. Specifically, we compared the effectiveness of moral and pragmatic arguments against torture, the moderating role of glorification of (and attachment to) the nation, and, in an effort to unveil the factors underlying this process, the mediating role of moral outrage, empathy, and guilt.

Findings from Experiment 1 suggest that only moral arguments are effective in prompting people to oppose torture by redressing its injustice. This finding is consistent with Ames and Lee (2015), who found that rather than the lack of benefit of torture reducing support for torture, pre-existing support for torture increases the perceived benefit of torture. As such, it may then not be surprising that pre-existing support for
torture and beliefs in its effectiveness inoculate people against new, pragmatic arguments about the ineffectiveness of torture. Experiment 2 further demonstrated that the effect of moral arguments is only evident among people who strongly glorify the ingroup (and are strongly attached to it). Again, pragmatic arguments were ineffective, regardless of whether people were high or low in glorification or attachment. Experiment 3 replicated this effect for all high glorifiers (across both low and high levels attachment), and found that they support torture less strongly after moral arguments because these arguments elicit more moral outrage over torture as a violation of moral standards, and more empathy towards victims (but not more guilt). The mediation results were consistent with the rationale that moral arguments draw attention to the victims and the wrong done to them. As a consequence, the perceiver empathizes more with the victims, and becomes more outraged when pondering the violation of moral standards. This, in turn, leads the perceiver to reduce his/her support for torture, and to increase his/her support of the pursuit of justice against the perpetrators and in aid of the victims.

What makes moral arguments effective?

H. G. Wells wrote in *The Rights of Man - What Are We Fighting For?* that the allied soldiers in WWII “had been stirred profoundly by those outrages upon human dignity perpetrated by the Nazis.” Moral outrage is one powerful motivation that pushes people to pursue justice and oppose perceived injustice. Oftentimes, however, when the same cruelty is committed by members of the ingroup, this outrage and its corresponding motivation are missing – especially among people who hold an immaculate and unrealistically positive image of the ingroup (Leidner et al., 2010). Our findings show that this can be redressed via the presentation of a moral critique of the ingroup’s actions,
ARGUMENTS AGAINST TORTURE

which then increases outrage over violated moral standards. This finding is also in line with research on moral mandates; if people’s moral convictions are violated, they react with anger (Mullen & Skitka, 2006). It appears this moral mandate effect can even extend to situations in which people would otherwise disengage from the moral violation (i.e. ingroup-committed wrongdoings), as long as it is combined with moral arguments against the standard violation.

Moral arguments also exert their effect by enhancing empathy toward the victims. Our finding on empathy, a ubiquitous and automatic response that humans have towards other humans (Rifkin, 2010; Decety & Ickes, 2009), is often curtailed as a consequence of social categorization processes that separate ‘us’ from ‘them’ (Castano, 2012), and by motivational processes that are functional to maintaining psychological equanimity in the face of atrocities involving the collective self (Castano, 2011; Castano & Giner-Sorolla, 2006; Cehajic et al., 2010; Leidner et al., 2010; Leidner & Castano, 2012). A moral critique of such atrocities likely functions as a reminder of the victims’ humanity and their place in the moral circle (rather than outside of it; Singer, 1981), and thus of the connection that ingroup and outgroup members have as humans despite any group boundaries on other dimensions such as nationality, ethnicity, or religion.

Limitations and future research

Our findings are of both theoretical and practical importance. Practical, because they identified which type of arguments against torture are most effective in curbing public support for torture, and thus should perhaps be used more in the public debate over torture. Theoretical, because they illuminate communication and persuasion processes that are usually influenced by bias (i.e., self-relevant attitudes), and how and why
attitudes can be changed even among those people who are usually most biased (i.e., high glorifiers). Both advances highlight the great potential that our understanding of communication, identity and attitudes of the public’s opinion has to build a culture of peace and respect for human rights in society. While we focused on torture, our rationale could be extended to any norm violations committed by the ingroup at the expense of outgroup members. Economic exploitation of third world countries or support for violent authoritarian regimes are but two examples which lend themselves to the same analysis we presented here for the case of torture. Future research is needed to assess whether our findings for the effectiveness of moral but not pragmatic arguments against torture generalizes to other norm violations.

While the effects were remarkably robust across studies and substantial, it is unclear whether people’s attitudes really changed. From a design perspective, a within-participants design with attitudes measures pre- and post-argument might add further evidence for attitude change. But ultimately the between-participants experimental designs we employed here yield the same evidentiary value. More importantly, a within-participants design would further compound the real issue at hand: Did participants’ self-reports reflect actual attitude change, or pressure or social desirability concerns? While our data is not able to speak to this question, it does not critically restrict the interpretability of our data with respect to public opinion on torture. First, when people report their attitudes in public opinion polls, social pressure or social desirability concerns are likely to operate similar as in our self-report measures—perhaps even to a greater degree. Second, in the case of public opinion and its consequences for policy and decision making at the group level, it matters first and foremost what people report their
ARGUMENTS AGAINST TORTURE

attitudes to be; whether or not such self-reports match the “real attitude” is secondary in this context.

A factor our studies did not account for is the perceived likelihood of possible (positive or negative) consequences of torture. The effectiveness of pragmatic arguments in particular might depend on people’s subjective assessments of questions such as “How likely will our use of torture put our own troops or our country at risk?” Thus, it is possible for pragmatic arguments to be effective in reducing support for torture under more specific circumstances than the ones explored in our studies (e.g. when the costs of torture are particularly high and likely to be incurred). Yet, given the more general circumstances of our studies, it seems safe to say that without special circumstance, moral arguments against torture yield more promise to reduce public support for torture than pragmatic ones.

Another open question concerns the potency of moral arguments. Is it specific to condemnation of torture, or does it generalize to the endorsement of torture as well? More specifically, can moral arguments only reduce high glorifiers’ support for torture when condemning its immorality, or could they also increase high glorifiers’ support for torture when embracing its morality? While this question needs to be answered by future research, there is reason to believe that moral arguments for torture will be less potent than moral arguments against torture. This is because the effects we observed were unique to high glorifiers—that is, those group members who usually defend morally questionable ingroup behavior ‘by default’ (Bilali, 2013; Leidner & Castano, 2012; Leidner et al., 2010; Roccas et al. 2006). This default naturally restricts the space and likelihood for moral arguments for torture to increase high glorifiers’ support for torture.
To conclude, in three experiments we found moral arguments against ingroup-perpetrated torture to be successful in increasing people’s demands for justice on behalf of the victims, as well as in decreasing support for torture. These findings are of particular interest because they move precisely those people who are most problematic from the perspective of social justice and peaceful intergroup relations—that is, high glorifiers. Rather than naïve or hopelessly idealistic, appealing to the moral and good in human nature may be worthwhile and effective after all, and help strengthen the normative power of international humanitarian law (see also Di Lellio & Castano, 2016).
ARGUMENTS AGAINST TORTURE

References


ARGUMENTS AGAINST TORTURE


ARGUMENTS AGAINST TORTURE


ARGUMENTS AGAINST TORTURE


ARGUMENTS AGAINST TORTURE


ARGUMENTS AGAINST TORTURE


ARGUMENTS AGAINST TORTURE

Figure 1: Demands for justice as a function of argument type (none vs. pragmatic vs. moral) in Study 1.
Figure 2: Demands for justice as a function of argument type (none vs. pragmatic vs. moral) in Study 2.
ARGUMENTS AGAINST TORTURE

Figure 3: Demands for justice as a function of argument type (none vs. pragmatic vs. moral), ingroup attachment and glorification in Study 2.
Figure 4: Demands for justice as a function of argument type (none vs. pragmatic vs. moral) and ingroup glorification in Study 3.
ARGUMENTS AGAINST TORTURE

Figure 5: Support for torture as a function of argument type (none vs. pragmatic vs. moral) and ingroup glorification in Study 3.
Figure 6: Empathy as a function of argument type (none vs. pragmatic vs. moral) and ingroup glorification in Study 3.
ARGUMENTS AGAINST TORTURE

Figure 7: Moral outrage as a function of argument type (none vs. pragmatic vs. moral) and ingroup glorification in Study 3.
Figure 8: Indirect effects of argument type (none vs. moral) and ingroup glorification through empathy and moral outrage on demands for justice (Study 3).

**High glorifiers**
Indirect effect through empathy: boot coefficient = -.085, lower CI = -.234, upper CI = .039
Indirect effect through outrage: boot coefficient = -.066, lower CI = -.214, upper CI = .067

**Low glorifiers**
Indirect effect through empathy: boot coefficient = 1.137, lower CI = .040, upper CI = .273
Indirect effect through outrage: boot coefficient = .119, lower CI = .065, upper CI = .265
ARGUMENTS AGAINST TORTURE

High glorifiers
Indirect effect through empathy: boot coefficient = -.119, lower CI = -.277, upper CI = -.035
Indirect effect through outrage: boot coefficient = -.188, lower CI = -.408, upper CI = -.013

Low glorifiers
Indirect effect through empathy: boot coefficient = .053, lower CI = -.018, upper CI = .166
Indirect effect through outrage: boot coefficient = .070, lower CI = -.065, upper CI = .255

Figure 9: Indirect effects of argument type (none vs. moral) and ingroup glorification through empathy and moral outrage on support for torture (Study 3).
Online Supplemental Materials

Experiment 1

We also analyzed the effect of condition for the first justice item, regarding the appropriateness of a prison term for the perpetrators, separately. Chi-square analysis and ANOVA (coding -1 for “no prison” and +1 for “prison”) gave virtually identically results, so we report the ANOVA results for simplicity. Condition (neutral vs. pragmatic argument vs. moral argument) had a significant effect, $F(2, 187) = 4.92, p = .008, \eta^2 = .05$. Participants in the moral argument condition ($M = .97, SD = .18$) reported significantly stronger justice demands than participants in the no-argument ($M = .79, SD = .41$), $t(187) = 2.62, p = .010$, or pragmatic argument condition ($M = .82, SD = .39$), $t(187) = 2.81, p = .006$. The pragmatic and the no-argument condition did not differ significantly, $t(187) = 0.30, p = .763$.

Age and conservatism, but not gender, had an effect on demands for justice when entered as additional covariates in the above analyses. Age, but not gender or conservatism, had an effect on the first justice item. In no case, however, the inclusion of these covariates substantially changed the effects of condition reported in the analyses above.

Experiment 3

*Collective guilt.* When adding gender and conservatism as covariates, the interaction between argument and glorification became marginally significant, $F(1, 173) = 3.45, p = .065, \eta^2 = .02$. Low glorifiers showed no difference in collective guilt between the moral ($M = 5.60, SE = .43$) and the no-argument condition ($M = 6.00, SE = .40$), $t(173) = -0.71, p = .478$. High

---

1 The within-degrees of freedom in the analyses with covariates and those without covariates differ by more than the number of covariates due to a few additional missing values on the covariates.
glorifiers, on the other hand, displayed significantly more collective guilt in the moral ($M = 4.77, SE = .48$) than in the no-argument condition ($M = 3.54, SE = .44$), $t(173) = 1.96, p = .052, d = .29$.

**Moral outrage.** When adding gender and conservatism as covariates, the interaction between argument and glorification reached full significance, $F(1, 173) = 4.68, p = .032$; tests of the simple effects produced the same patterns as those of the analysis without covariates, with the same conventional significance levels.
Online Appendix

Main text of the scenario used in Study 1-3 (the pilot study only reminded participants of U.S. torture of detainees, without presenting this whole text):

In recent years there has been growing evidence that prisoners in U.S.-run jails have been treated in manners forbidden by the Geneva Convention on Human Rights and, more broadly, by international humanitarian law. Episodes of violence close to, or even beyond, torture have been reported.

Recently, the results of an investigation conducted by the human rights organization Amnesty International were made public and reported internationally by several newspapers. This information provides detailed evidence of four cases of severe maltreatment of Iraqi prisoners in Camp Cropper, a high-value detention site (HVD) near Baghdad International Airport, operated by the United States Army. According to the AI report, one of the prisoners in Camp Cropper, Arkan Ali Sulaiman, was severely beaten by U.S. military police while being interrogated. His head was submerged in cold water several times until he lost consciousness. Revived by punches and kicks from the U.S. soldiers, Sulaiman suffered from multiple internal injuries, which, as subsequent medical analyses demonstrated, eventually led to his death.

Another prisoner, Karim Jassim Mohammad, died after being stripped naked and beaten about the head and body with belt buckles by three U.S. soldiers from the 115th Military Police Battalion. Mohammad’s head was so badly injured that identification was possible only by means of DNA samples retrieved from his home. The third case is that of Abdel Rahaman Ahmed. When this prisoner repeatedly refused to bend over and walk like a dog, as the U.S. military personnel had ordered him to do, he was tied to a chair and his toes and feet were hit with a hammer. His feet having been reduced to a pulp, Ahmed not only lost a great deal of blood but later contracted an infection. The International Committee of the Red Cross was prevented from monitoring Ahmed’s treatment for several months, giving rise to suspicions that military personnel were hiding something. When medical personnel finally diagnosed the infection, Ahmed was treated, but they could not save his life. He died on May 3, 2004, and his family was only notified one week later, when his body had already been buried. The report mentions that such a delay might have been intentional to minimize the chances of an autopsy.

The last case is Khudir Abbas al-Obeidi, who tried to commit suicide before being killed by U.S. military police. When prison attendants noticed that al-Obeidi was trying to hang himself with a rolled-up sheet, they repeatedly hit his head against a wall. With blood all over his face, he was suspended by his ankles, a position which caused major loss of blood. When al-Obeidi passed out, the U.S. soldiers beat and kicked him to wake him up until they realized that he was already dead. It is not clear whether he died of blood loss, sheer exhaustion, or the beating and kicking; most likely it was a combination of the factors.

A few weeks ago, perhaps anticipating that Amnesty International was going to publicize its findings, the U.S. Army initiated an investigation. An officer close to the investigation, who asked not to be publicly identified, revealed that new evidence was still emerging and that two soldiers who worked at the U.S. prison at the time events allegedly took place were being investigated. These allegations come at a difficult moment for the U.S. Army in Iraq, which is facing an increase in the number of deadly attacks by militia groups. Yesterday a rudimentary
device exploded at a café in central Baghdad, killing eight people and critically wounding several others.

Additional text for the moral argument conditions of Study 1-3 as well as the pilot study (text passages that differed between conditions are bolded):

Several high-ranking military officials have expressed great concern about the situation, promised an extensive and thorough investigation, and clearly stated that no leniency whatsoever will be used towards the personnel involved in these events. “These events are a disgrace for the U.S. military,” commented a military source that wished to remain anonymous. “Whether or not torture if effective in gathering information, torturing prisoners is against the ideals of the United States of America and its military. On a basic human level, regardless of where they were born or what they believe, all men are equal and are to be treated with respect – this is the idea our country is build upon.” Gen. Philip A. Starwood, USMC, expressed similar views when interviewed earlier this week. From his experience in the Vietnam conflict, Gen. Starwood said, “the torture of Viet Cong prisoners in prisons run by the U.S. military in Vietnam was an unacceptable departure from American principles of fair treatment, leading to questions of whether we were morally any better than the Viet Cong.” By straying from American ideals and ignoring humanitarian rights and laws, General Starwood said, policies allowing torture in the war in Iraq have also cost the United States heavily in moral authority. “In the military we are honorable people. Our policy should show respect for the dignity and well-being of enemy combatants who surrender,” he continued. “We should never torture our prisoners.”

Additional text for the pragmatic argument conditions of Study 1-3 as well as the pilot study (text passages that differed between conditions are bolded):

Several high-ranking military officials have expressed great concern about the situation, promised an extensive and thorough investigation, and clearly stated that no leniency whatsoever will be used towards the personnel involved in these events. “These events are a disgrace for the U.S. military,” commented a military source that wished to remain anonymous. “Whether or not torture if effective in gathering information, torturing prisoners is against the interests of our own military. When our soldiers are captured by the enemy, it is much more likely that they will be subjected to the same treatment that we subject our prisoners to.” Gen. Philip A. Starwood, USMC, expressed similar views when interviewed earlier this week. From his experience in the Vietnam conflict, Gen. Starwood said, “news of the torture of Viet Cong
prisoners in prisons run by the U.S. military in Vietnam led to an increase of torture of U.S. soldiers at the hand of the Vietnamese and Viet Cong.” By straying from previous American rules of engagement, General Starwood said, policies allowing torture in the war in Iraq have also cost the United States numerous casualties, by making enemy combatants less willing to surrender in battle. “In the military we are professionals. Our policy should achieve our objectives with the lowest cost to our troops,” he continued. “We should never torture our prisoners.”