The Motivated Use of Moral Principles

David A. Pizarro
Cornell University

Eric Luis Uhlmann
Yale University

David Tannenbaum and Peter H. Ditto
University of California, Irvine

DRAFT: PLEASE DO NOT CITE WITHOUT PERMISSION OF AUTHORS

WORD COUNT [MAIN TEXT]: 4000

WORD COUNT [ABSTRACT]: 121

REFERENCES: 30

CONTACT:
David Pizarro
Department of Psychology
Cornell University
Phone: (607) 255-3835
Email: dap54@cornell.edu
Abstract

Three studies demonstrate that people shift their moral principles to rationalize desired judgments. In Study 1, college students confronted with a footbridge dilemma were more willing to endorse sacrificing an innocent life to save the lives of many when the innocent person was White than when he was Black. In Study 2, political conservatives were more permissive regarding the killing of innocent civilians when the victims were Iraqis, while liberals were more willing to condone the killing of innocents when the victims were American. In Study 3, nonconsciously priming patriotism vs. multiculturalism led participants to shift their moral standards in favor of American forces and Iraqi insurgents, respectively. Implications for the roles of reason and intuition in moral judgment are discussed.
The Motivated Use of Moral Principles

Most people believe that harming innocent children is wrong, as is cheating on an exam or breaking a promise. More controversially, some people believe that abortion is wrong, that the death penalty is unjust, or that animals should not be killed and eaten. These moral judgments are unlike other social judgments in an important way. Not only do we believe that our moral judgments are correct, but we believe that (unlike our attitudes toward, say, chocolate ice cream) everyone else should agree with us (Haidt, Rosenberg, & Hom, 2003; Skitka, Bauman, & Sargis, 2005). However, a problem arises when defending moral judgments. Defending a moral judgment by appealing to our subjective preferences (e.g., “abortion is wrong because I don’t like it”) is unpersuasive, inasmuch it fails to provide a compelling reason why others should agree. And unlike factual beliefs (e.g., that the world is round), there is usually no objective set of facts that can be used to evaluate a moral claim. These features make disagreement in the moral domain a tricky problem (Pizarro & Uhlmann, 2006; Sturgeon, 1994; Sunstein, 2006).

What individuals often do, however, is defend a specific moral judgment by appealing to a general moral principle. Principles have the advantage of being foundational rules that can guide judgment across a wide variety of situations, making these judgments appear to be less like ad hoc preferences and more like rational facts. A principle serves as a first step—once there is agreement about a principle, whether or not the specific moral claim is an instantiation of the principle can be deduced. Reasoning one’s way to specific moral judgments by using general principles is not only the way moral philosophers do things, it is also a sign of mature moral reasoning according to developmental psychologists (e.g., Kohlberg, 1984). Individuals at the highest stages of moral reasoning, according to Kohlberg, reason their way from a set of
universal principles to making judgments about specific dilemmas they encounter in everyday life.

Of course, there has been significant debate within moral philosophy as to which principles should be endorsed. In particular, moral philosophers are quite divided as to whether a consequentialist or a deontological normative ethic is most defensible (Smart & Willams, 1973). Consequentialism holds that acts are morally right or wrong to the degree that they maximize good outcomes, and that the means to such maximization are irrelevant. Deontologists, on the other hand, believe that there are constraints against action independent of consequences—some acts are wrong in-and-of themselves. Such constraints often include injunctions not to break promises, not to lie, and in general not to harm innocent others.

Motivated Moral Reasoning

Like logical principles, moral principles are presumed to be invariant (e.g., the transitivity of identity holds when speaking of integers, apples, or widgets; likewise consequentialism should hold across an equally wide variety of moral scenarios). It would violate objectivity, for instance, to argue that breaking one’s promises is morally impermissible, while making an exception for yourself or a family member. Indeed, much research on judgment utilizes this normative principle of invariance (e.g., Plous, 1993). When trivial changes are made in describing events—such as the way the information is framed—it is presumed that the rational respondent should ignore the extraneous information when making a judgment (although this is often not the case; Kahneman & Tversky, 1973).

While the successful defense of moral judgments through the use of principles depends on the objective application of moral principles, it is unclear whether individuals actually go about making their moral judgments this way. There is a large body of evidence in the social
cognition literature which demonstrates that we often mis-use reasoning in order to defend judgments we hold dear (e.g. Ditto & Lopez, 1992; Kunda, 1987; Pyszczynski & Greenberg, 1987). One of the most interesting types of motivated reasoning occurs when individuals construct or change their general standards to rationalize a specific judgment. For instance, people define merit in a self-serving fashion, asserting criteria of excellence that “happen” to put their own idiosyncratic credentials in a positive light (Dunning, Leuenberger, & Sherman, 1995; Dunning & Cohen, 1992). And when making hiring decisions, male evaluators inflate the importance of credentials a male applicant happens to have, and downplay the importance of those credentials he lacks (Norton, Vandello, & Darley, 2004; Uhlmann & Cohen, 2005).

Recently, Haidt and his colleagues have argued that our use of moral reasoning is similarly motivated (Haidt, 2001; Haidt & Hersh, 2001; Haidt, Koller, & Dias, 1993). Rather than reason our way to a moral judgment, we make an intuitive judgment and use post-hoc reasoning to justify it. Consistent with this, when people are given descriptions of disgusting but harmless moral acts they condemn such acts as morally wrong despite their inability to justify their judgment (Haidt et al., 1993).

Could objective moral principles be used in such a motivated manner? Consider the moral defense of the Hiroshima bombing. Individuals who argue that such actions were defensible on the grounds that the killing of innocent civilians for the sake of a greater good is permissible should not argue that the same actions are impermissible in a similar situation (e.g., Palestinians’ targeting of Israeli civilians as a means of achieving their nationalistic goals). Such motivated rationalizations are evident in current political discourse. For instance, some have argued that sending American troops overseas in order to liberate oppressed people is unjustified—America has no duty to sacrifice its military force to intervene in the affairs of
another country. Political conservatives made this very point when President Clinton sent American troops into Bosnia—that it would be hard to “Explain to the mothers and fathers of American servicemen that may come home in body bags why their son or daughter had to give up their life” (Sean Hannity, Fox News, 4/6/99). Liberal political thinkers, on the other hand use the very same non-interventionist reasoning to condemn the sending of troops to Iraq. Belief in the principle of non-intervention seems influenced by the political affiliation of the current President.

Current Studies

In the current set of studies, we sought to investigate whether individuals consistently apply a set of general moral principles across similar scenarios, or whether they recruit whichever principles support their initial moral judgments. The latter finding would support the hypothesis that reasoning about moral principles often rationalizes rather than mediates moral judgment. At the same time, it would demonstrate that not only descriptive, but also prescriptive moral beliefs are subject to shifting standards. In testing these hypotheses, we constructed scenarios in which consequentialist rules (specifically, consequentialism about the sacrificing of innocent lives for the sake of a greater good) were pitted against deontological rules. While it is not always the case that these two moral philosophies conflict, we (as well as many philosophers and psychologists before us) utilized examples in which the two broad ethical principles offered conflicting answers regarding which judgment was appropriate. By manipulating the content of the scenarios across groups, we sought to discover whether people are “invariantists” in their use of principles, as many would hold moral reasoners to be, or whether they shift their moral standards, using moral principles as post hoc justifications. Switching the use of a moral
principle that is defined as “universal” simply to support a desired conclusion would be evidence of motivated moral reasoning.

Study 1: Trolley with a Twist

In Study 1 we sought to test the notion that individuals endorse a moral principle when it is consistent with a moral judgment that they find agreeable, but reject the same moral principle when it is used to defend a moral judgment they find disagreeable. We used a modified version of the oft-cited trolley/footbridge dilemma, in which an individual must decide whether or not to sacrifice one innocent individual to save a group of people who will be killed by a trolley headed in their direction. This example is often used in thought experiments by philosophers concerned with determining whether consequentialism is an appropriate normative ethical theory, and has been used in a number of psychological experiments on the nature of moral judgment (e.g., Cushman, Young, & Hauser, in press; Greene, Somerville, Nystrom, Darley, & Cohen, 2001).

It is known that there is a strong disdain among American college students for harboring feelings that may be considered prejudiced (Czopp & Monteith, 2003; Monin & Miller, 2001; Tetlock, Kristel, Elson, Green, & Lerner, 2000). Therefore, we surreptitiously varied the race of the characters in the scenarios to see whether this would influence participants’ judgments concerning the appropriate moral action—a variable that most would deem morally irrelevant when deciding to save lives.

Method

Participants. Participants were 96 undergraduate students at the University of California, Irvine, who participated for course credit.
Materials and Procedure. Participants received one of two scenarios involving an individual who has to decide whether or not to throw a large man in the path of a trolley (he is described as large enough that he would stop the progress of the trolley) in order to prevent the trolley from killing 100 innocent individuals trapped in a bus. Half of the participants received a version of the scenario in which the agent could choose to sacrifice an individual named “Tyrone Peyton” to save 100 members of the New York Philharmonic, and the other half received a version in which the agent could choose to sacrifice “Chip Ellsworth III” to save 100 members of the Harlem Jazz Orchestra. In both scenarios the individual decides at the end to throw the person onto the trolley tracks. While we did not provide specific information about the race of the individuals in the scenario, we reasoned that Chip and Tyrone were stereotypically associated with Caucasian and African American individuals respectively, and that the New York Philharmonic would be assumed to be majority Caucasian, and the Harlem Jazz Orchestra would be assumed to be majority African American.

All participants were then provided with the following items intended to assess the endorsement of general principle of consequentialism, on 7-point scales: (1) Is sacrificing Chip/Tyrone to save the 100 members of the Harlem Jazz Orchestra/New York Philharmonic justified or unjustified? (2) Is sacrificing Chip/Tyrone to save the 100 members of the Harlem Jazz Orchestra/New York Philharmonic moral or immoral? (3) It is sometimes necessary to allow the death of an innocent person in order to save a larger number of innocent people, (4) We should never violate certain core principles, such as the principle of not killing innocent others, even if in the end the net result is better, and (5) It is sometimes necessary to allow the death of a small number of innocents in order to promote a greater good.

Results and Discussion
Because all of the dependent variables were constructed as a general probe into the endorsement of consequentialism, and because they were highly correlated (Cronbach’s $\alpha = .78$), we combined the items to form an index of consequentialism. Participants who were presented with the version of the story in which Tyrone Peyton could be sacrificed to save 100 members of the New York Philharmonic were significantly less likely to endorse consequentialism in general ($M_s = 2.98$ and $2.55$, $SD_s = 0.79$ and $0.78$, respectively), $t(94) = 2.66$, $p < .01$, $d = 0.55$. Table 1 presents the means and standard deviations for the individual items. As can be seen, when participants were led to infer that the “innocent” in the trolley scenario was White, they rated the sacrificing of this individual to save 100 others (who were most likely Black) as more morally justified than when they were led to infer that the “innocent” was Black and the individuals to be saved were White. The motivation to be egalitarian (or at least, avoid appearing inequalitarian) appears to cause a differential endorsement of the principle of consequentialism across these scenarios.

**Study 2: Military Action, Collateral Damage, and Political Orientation**

Study 1 provided initial evidence that individuals shift their moral principles to rationalize their moral judgments, but some questions still remain. For instance, would individuals with differing initial moral attitudes differentially endorse moral principles consistent with these attitudes? That is, would people flip-flop in their use of moral principles to justify these attitudes (as seems apparent in the sorts of everyday political judgments described in the introduction)? In Study 2 we relied on the fact that conservative versus liberal participants would have pre-existing differences in the extent to which they favored military intervention in Iraq. We hypothesized that these biases would influence the extent to which participants found the
killing of innocents in the course of warfare morally permissible, depending on the nationality of the perpetrator.

**Method**

**Participants.** Participants were 144 undergraduates—98 at the University of California, Irvine, and 46 at California State University, Fresno—who participated in the study for course credit.

**Materials and procedure.** Participants were given one of two scenarios in which military leaders in Iraq initiated an action that foreseeably but unintentionally killed innocent civilians. Half of the participants received a version in which American troops attacked Iraqi insurgents, and the other half received a version in which Iraqi insurgents attacked American troops. The scenario described American/Iraqi leaders deciding to carry out an attack to stop key leaders of the Iraqi insurgency/American military in order to prevent future deaths of Iraqi insurgents/American troops. It was stated in each scenario that while the decision-makers were aware of the possibility of innocent deaths, they reasoned that sometimes it is necessary to sacrifice innocent people for the sake of a greater good (in this case the saving of many future lives). It also specifically stated that the decision makers did not intend the death of any innocent civilians, they merely foresaw it as an unwanted consequence of their military actions.

Participants were then asked to what extent they agreed with the military leaders’ position on the deaths of innocents, and to what extent they endorsed the consequentialist principle that was used by the military leaders to justify such action. We asked participants whether they generally agreed with the position of the American/Iraqi leaders as well as their agreement with two of the general consequentialism items used in Study 1 (*It is sometimes necessary to allow the death of a small number of innocents in order to promote a greater good,*
We should never violate certain core principles, such as the principle of not killing innocent others, even if in the end the net result is better), and added a question asking participants to evaluate the reasoning behind the American/Iraqi leaders’ decision (How well-reasoned was the position of the American [Iraqi] leaders?). All responses were indicated on 7-point scales:
Finally, all participants were asked about their political orientation: Politically, I consider myself: (1 = Conservative, 7 = Liberal).

Results and Discussion

Because the scenarios contained a section in which the military leaders explicitly endorse consequentialism as a general principle, the items about participants’ agreement with the action and the quality of the reasoning behind it were taken as an endorsement of the consequentialist ethic, along with the two general consequentialism items. We thus combined all items to form an index of consequentialism (Cronbach’s α = 0.75). All subsequent analyses utilized this index.

There was no significant main effect of experimental condition, $F(1, 136) = 2.67, p = .10$, nor of data collection location, $F < 1$. In order to test our predictions that conservatives and liberals would respond differently to our scenarios, we divided our sample by constructing a median split on our political orientation question (using a regression analysis with the political orientation item as a continuous variable yielded identical results). A main effect of political orientation emerged, with conservatives ($M = 3.93, SD = 1.10$) reporting a greater endorsement of consequentialism than liberals ($M = 3.39, SD = 1.32$), $F(1, 136) = 7.55, p < .01$, suggesting that political conservatives have a generally more permissible view of collateral casualties in the course of warfare than do political liberals. As predicted, however, an interaction between experimental condition and political orientation emerged, $F(1, 136) = 7.62, p < .01 \eta^2 = 0.05$ (represented in Figure 1). Specifically, conservatives showed a nonsignificant tendency to
endorse consequentialism more after reading a scenario in which Americans killed innocent Iraqis than one in which Iraqis killed innocent Americans. Liberals, on the other hand, showed a strong reversal of this tendency, being significantly more likely to endorse consequentialism in the Iraqis-kill-Americans scenario than in the Americans-kill-Iraqis scenario, $t(62) = 2.67, p = .01, d = 0.68$. A key aspect of the observed effects was conservatives and liberals differential evaluation of the American transgression. Conservatives were much more likely to endorse consequentialism than liberals when Americans killed Iraqi civilians, $t(67) = 4.07, p < .001, d = 1.00$.

Study 3: Political Priming and Military Action

Across two studies, we demonstrated that individuals shift their moral standards to justify their moral position. While appearing to embrace a universal principle, participants are responding to (normatively) irrelevant details about the scenarios. However, up to this point, we have been forced to rely on the fact that individuals bring their moral biases with them. While this is certainly an appropriate way in which to examine the motivated use of moral principles, an ideal test would be to manipulate an individual’s moral judgments and demonstrate that this manipulation can cause a change in endorsement of moral principles. In Experiment 3 we sought to provide such a test by nonconsciously priming participants (Bargh & Chartrand, 1999; Dijksterhuis & Bargh, 2001) with either patriotism or multiculturalism—values thought to underlie support for American troops and Iraqi insurgents, respectively.

Use of a nonconscious priming procedure provides a particularly strong test of the hypothesis that automatic, intuitive processes drive moral judgments, which are then justified on a post-hoc basis by moral reasoning (Haidt, 2001; Haidt et al., 1993). To our knowledge, no study has provided an experimental demonstration in which 1) a moral attitude/judgment is
caused by a nonconscious process and which 2) the judgment is then rationalized using conscious moral principles.

**Method**

**Participants.** Ninety-two UC Irvine undergraduates participated in the study for course credit.

**Materials and Procedure.** Participants were randomly assigned to one of two priming conditions, *patriotic* or *multicultural*. Participants were primed utilizing a sentence unscrambling procedure (Srull & Wyer, 1979). All participants were asked to unscramble 11 sentences and remove one word that did not belong in each of the sentences. One group of participants received 6 neutral sentences, and 5 sentences that contained words pertaining to patriotism (i.e., *patriots, American, U.S.A., flag, and loyal*). The other group of participants received the same 6 neutral sentences, and 5 additional sentences that contained words pertaining to multiculturalism (i.e., *multicultural, include, diversity, equal, and minority*).

Immediately following the sentence-unscrambling task, participants were asked to read the scenarios about military action in the Middle East utilized in Study 2. The study was thus a 2 (Primes: Patriotic vs. Multicultural) X 2 (Military scenario: Americans attack Iraqis vs. Iraqis attack Americans) between-subjects design. Following the scenarios, participants were asked to judge the actions of the military leaders by responding to the same 5 dependent variables utilized in Study 2.

**Results and Discussion**

As in Study 2, we combined all of the items to form an index of consequentialism (Cronbach’s α = .67). There was no main effect of scenario, or priming condition, $F$s < 1. However, the predicted scenario x priming condition interaction emerged, $F(1, 88) = 3.91$, $p =$
.05, $\eta^2 = 0.04$ (see Figure 2). Mirroring the pattern found in Study 2 with measured political ideology, individuals who received the patriotic (i.e., conservative) primes were more likely to endorse consequentialism after reading the scenario in which Americans killed innocent Iraqi civilians, while those who received the multicultural (i.e., liberal) primes were more likely to endorse consequentialism after reading the scenario in which Iraqis killed American civilians. Neither of the simple effects associated with these comparisons, however, were statistically significant. As in Study 2 a key aspect of the effect was the differential endorsement of consequentialism after reading the scenario in which Americans killed Iraqis between participants who received the patriotic prime ($M = 4.17$, $SD = 0.77$) and participants who received the multicultural prime ($M = 3.59$, $SD = 0.86$), $t(41) = 2.26, p < .05, d = 0.71$.

**General Discussion**

People shift their moral principles to rationalize their judgments. In Study 1, college students confronted with a trolley car dilemma were more willing to endorse sacrificing an innocent White person in order to save the lives of many African Americans than vice versa. It seems that participants’ judgments were driven by egalitarianism (or, at least, a concern with appearing inegalitarian). When consequentialism provided support for this motivation, it was endorsed. If not, consequentialism was rejected as a principle for guiding judgment. In Study 2, political conservatives were more permissive regarding the killing of innocent civilians when the victims were Iraqis, while liberals were more willing to condone the killing of innocents when the victims were American. Feelings of patriotism appear to have influenced whether participants endorsed a consequentialist position on “collateral damage.” Consequentialism *per se*, of course, was not the causal factor behind the judgment—political ideology led to the recruitment of the principle most consistent with the judgment the participant wanted to make.
These findings indicate that not only descriptive beliefs about how the world is (Dunning & Cohen, 1992; Kunda, 1987; Norton et al., 2004; Uhlmann & Cohen, 2005) but also prescriptive beliefs about how the world ought to be can shift to rationalize social judgments.

To our knowledge, the present research provides the first direct, experimental support for the idea that nonconscious processes can cause moral judgments and that the explicit reasons given for those judgments take the form of post hoc rationalizations (Haidt, 2001). A number of studies have shown experimentally that intuitions can drive moral judgments (e.g., Greene, 2001; Wheatley & Haidt, 2001), but demonstrated no evidence of rationalization using principles. Others have shown both intuition as the proximal cause of a judgment and rationalization during a post hoc defense, but using an interview rather than an experimental paradigm (Haidt, 2001; Haidt et al., 1993). In the present Study 3, unobtrusively priming patriotism versus multiculturalism influenced sympathy for U.S. forces over Iraqi insurgents, and participants rationalized their biases accordingly.

These results have implications for philosophers and psychologists who seek to extrapolate general moral principles based on the intuitions elicited by a specific moral scenario. Rather than being true moral generalists, people may be moral “particularists” who make judgments based on case-specific information, yet rationalize their judgments based on general moral principles. An overreliance on individual scenarios to tap moral intuitions may make people look more principled and consistent than they actually are. Instead, it appears as if participants have a “moral toolbox” with which they can draw the arguments necessary to support their biases.

Does this mean that moral judgments are always based on intuitive processes, desires, or motivations, and that conscious reasons are inevitably rationalizations for the “true” causes of the
judgment? Such may well be the modal case of moral judgment. But reason may still exert a strong *a priori* influence on moral judgment in cases in which people are highly motivated to make an accurate judgment or good decision (e.g., while serving on a jury). In addition, reason-based moral judgments may be statistically rare but nonetheless critically important (Pizarro & Bloom, 2003). Moreover, if a judgment is based on motivations, desires, or intuitions it does not necessarily follow that reason is inert. Intuitions can serve as a starting point for the reasoning process, for example by providing an input into whether a reasoned principle is reliable. At the same time, well-reasoned decisions can exert an influence over motivations and intuitions through a process of automatization over time. The disturbing implication of recent research, however, is that we have no simple way of verifying whether the justifications we passionately invoke for our moral judgments are our real reasons, or merely empty rationalizations.
References


Table 1. Study 1: Means and Standard Deviations

<table>
<thead>
<tr>
<th></th>
<th>Chip Ellsworth/Harlem Jazz</th>
<th></th>
<th>Tyrone Peyton/NY Philharmonic</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( M )</td>
<td>( SD )</td>
<td>( M )</td>
<td>( SD )</td>
</tr>
<tr>
<td>Is sacrificing Chip/Tyrone justif( \text{ied})?</td>
<td>2.89</td>
<td>1.24</td>
<td>2.40</td>
<td>1.17</td>
</tr>
<tr>
<td>Is sacrificing Chip/Tyrone immoral/moral?</td>
<td>2.30</td>
<td>1.09</td>
<td>2.02</td>
<td>0.97</td>
</tr>
<tr>
<td>It is sometimes necessary to allow the death of an innocent person in order to save a larger number of innocent people.</td>
<td>3.40</td>
<td>1.01</td>
<td>2.94</td>
<td>1.13</td>
</tr>
<tr>
<td>We should never violate certain core principles, such as the principle of not killing innocent others, even if in the end the net result is better.</td>
<td>2.89</td>
<td>0.98</td>
<td>3.23</td>
<td>1.15</td>
</tr>
<tr>
<td>It is sometimes necessary to allow the death of a small number of innocents in order to promote a greater good.</td>
<td>3.18</td>
<td>1.13</td>
<td>2.64</td>
<td>1.11</td>
</tr>
</tbody>
</table>
Figure 1. Study 2: Mean Endorsement of Consequentialism by Political Orientation
Figure 2. Study 3: Mean Endorsement of Consequentialism by Priming Condition