How the source, inevitability, and means of bringing about harm interact in folk-moral judgments

Bryce Huebner, Marc D. Hauser, Phillip Pettit

Abstract: Means-based harms are frequently seen as forbidden, even when they lead to a greater good. But, are there mitigating factors? Results from five experiments show that judgments about means-based harms are modulated by: 1) Pareto considerations (was the harmed person made worse off?), 2) the directness of physical contact, and 3) the source of the threat (e.g., mechanical, human, or natural). Pareto harms are more permissible than non-Pareto harms, Pareto harms requiring direct physical contact are less permissible than those that do not, and harming someone who faces a mechanical threat is less permissible than harming someone who faces a non-mechanical threat. These results provide insight into the rich representational structure underlying folk-moral computations, including both the independent and interacting roles of the inevitability, directness and source of harm.

When a doctor gave his patient an excessive dose of a drug, killing him in order to harvest his organs (McKinley, 2008), his actions evoked a great deal of moral outrage. The doctor had treated this patient as a mere means to some further end, summarily dismissing his desires about how to live, or whether to go on living. Such cases suggest that it is (typically) objectionable to use a person merely as a means to furthering your own ends (Kant, 2002; Rawls, 1971; Smart & Williams, 1973). However, our moral psychology has a rich internal structure that suggests caution in moving too quickly over cases like this (Kamm, 2005). When an obese woman was trapped in a narrow passage in the Cango caves of South Africa (BBC, 2007), it seems on both rational and intuitive grounds that it would have been morally permissible to take her life if this was the only way to save the 22 people who were trapped behind her (including a diabetic man who would have died without his insulin). Fortunately, this real life dilemma turned out okay for all: she was dislodged with paraffin and a pulley after ten hours, and everyone else, including the diabetic man, walked away without a scratch.

It is perhaps unsurprising that the consequences of an action often play an important role in our moral intuitions. But consider the case where a person is used as a means to some further end while not making her worse off. Although it is a tragic choice, many people think that a person who is hiding in a basement during wartime with several others can permissibly smother her baby if this is the only way for everyone else in the basement to avoid being killed by oncoming soldiers (e.g., Greene et al., 2004). In this case, the participants who treat this action as acceptable are likely to do so on the (perhaps unconscious) basis that the baby is not made worse off by smothering; if she continued to cry, she would be killed anyway, along with the others. Scenarios like the cry-baby case have led some philosophers to suggest that Paretian considerations might play an important role in determining when it is permissible to use someone as a means to some greater good (Pareto, 1906; see Kamm, 2005, for a review). Many people seem to think that it is more permissible to harm someone if it does not make them worse off than if it does (Adler & Posner, 2006; Arrow, 1968; Pareto, 1906).

In this paper, we focus on the role of Paretian considerations in moral judgments of unfamiliar moral dilemmas. To explore the generality of Paretian considerations, as well as the possible mediating role of other factors, we test subjects' intuitive judgments in a wide variety of contexts, manipulating both the source of threat and whether it occurs by means of direct or indirect means. These additional factors have, to some extent, been independently explored in other studies (Cushman et al., 2006; Greene et al., 2009). Our contribution is to explore both the generality of Paretian considerations across contexts, as well as the potential interaction between the evitability of the harms, the source of threat, and the extent of physical contact.

1. Paretian Preliminaries:

In a recent experiment, Moore and colleagues (2008) explored a facet of the Paretian consideration by examining whether, in cases involving personal harm, the individual involved faced an evitable or inevitable death. They found that participants tended to see means-based harms as more permissible when the victim's fate was inevitable. This result, combined with other studies (e.g., Cushman et al., 2006, Mikhail 2007) suggests the interesting possibility that our folk intuitions operate over a series of distinctions that include, minimally, whether the harm results from using a person as a means as opposed to seeing her as a side effect, and whether the harm makes the individual worse off. At this stage, however, it is unclear how these distinctions, and others, interact across different contexts to guide our folk moral intuitions. Therefore, to set the stage for our investigation of Paretian considerations, we take advantage of a data set collected by Greene et. al. (2004), which focused primarily on the difference between utilitarian calculations and deontological considerations. However, it also included cases of personal harm that involved Paretian considerations and others that did not.

We re-analyzed participants judgments about whether various actions were acceptable (yes; no) and found that although means-based harms tended to elicit predominantly negative responses (Cushman et al., 2006; Greene et al., 2001; Hauser et al., 2007; Mikhail et al., 2007), participants were significantly more likely to judge as acceptable actions that yielded Pareto improvement than actions that did not, c^2 (2, N=714)=59.85, p<.001. However, the extent to which it is permissible to so use a person is also sensitive to a range of other features that were woven into the texture of these highly varied scenarios. For example, participants tended to consider it acceptable to harm one person as a means to a greater good in emergency cases (e.g., in the bomb threat and a vaccine cases, 93% and 79% of participants, respectively, endorsed harming one person). Similarly, making a person slightly worse off to save the lives of others, but not killing him, tended to be seen by a significant minority as morally acceptable (Vitamins, 35%). More surprisingly, a prohibition against cannibalism attenuated the salience of Paretian considerations when a young boy who would die anyway was killed *and* eaten (Plane crash, 30%); and considerations of kinship also seemed to trump the salience of Pareto considerations in one case (Sacrifice, 43%). Finally, the use of direct physical contact to bring about the greater good seems to interface in important ways with considerations of Pareto improvement, again attenuating the significance of the consequences (Crying Baby, 54%; Lifeboat, 58%). Planed comparisons for each scenario are reported in Table 1, and the text of the scenarios available at http://moral.wjh.harvard.edu/ParetoSupp).

	Mean	c^2 Value		Mean	c^2 Value
Transplant	.05**	(1, N=41)=33.39, p<.001	Crying baby	.54	(1, N=41)=.22, p=.639
Footbridge	.13**	(1, N=40)=22.50, p<.001	Plane Crash	.30**	(1, N=40)=6.40, p=.011
Country Road	.02**	(1, N=41)=37.10, p<.001	Lifeboat	.58	(1, N=38)=.95, p=.330
Architect	.02**	(1, N=42)=38.10, p<.001	Ecologists	.81**	(1, N=32)=12.50, p<.001
Hard times	.03**	(1, N=36)=32.11, p<.001	Submarine	.83**	(1, N=41)=17.78, p<.001
Hired Rapist	.02**	(1, N=41)=37.10, p<.001	Sophie	.68*	(1, N=41)=5.50, p=.019
Infanticide	.02**	(1, N=42)=39.10, p<.001	Sacrifice	.43	(1, N=41)=.90, p=.343
Bomb	.93**	(1, N=42)=30.86, p<.001	Euthanasia	.75*	(1, N=40)=10.00, p=.002
Vitamins	.35	(1, N=40)=3.60, p=.058			
Vaccine Test	.79**	(1, N=38)=12.74, p<.001			

Table 1: A reanalysis of Greene et al's (2004) judgment data for Pareto (grayed cells) and Non-Pareto scenarios. (One Sample Analysis, Hypothesized Mean = .50)

Summarizing from this initial set of analyses, we conclude that Paretian considerations are always likely to yield some increase in permissibility ratings, but as our discussion of these scenarios reveals, other factors can either enhance or attenuate the significance of such considerations in folk moral judgments. These data suggest that there are a number of ways in which Paretian considerations might be attenuated or overridden; and, all of these factors should ultimately be submitted to empirical scrutiny. However, rather than examining all of the ways in which Paretian considerations are embedded in our moral psychology, this paper focuses on whether Paretian considerations apply across different contexts, the extent to which they are modulated by the source of an ongoing threat, and whether the harm is caused by direct physical contact or more indirect means.

To explore these questions, we developed a series of scenarios in which a person is harmfully used as a means to some greater good, varying 1) the extent to which this person is made worse off by an action, 2) the extent to which direct physical contact is used to achieve this end, and 3) the source or nature of the ongoing threat. We hypothesized, based in part on previous work (Greene et al., 2001, 2004, 2009; Cushman et al., 2006) that the degree to which direct physical contact is required to secure a desired outcome will often play a role in folk moral judgments; however, we contend that the extent to which it does is also likely to be modulated by the source of the ongoing threat. For example, we predicted that it would matter whether the source of the ongoing threat is a person (animate agent), a trolley (inanimate object), or a fire (natural threat). This prediction is based on the fact that these different forms of threat differ in terms of the likelihood that they will bring about a particular consequence; moreover, they also differ in the extent that the cause of such differences are internally (e.g., an agent's mental state) or externally (e.g., physical constraints on where an inanimate object can move) generated.

2. General Experimental Methodology:

For each of the five studies reported below, participants voluntarily logged on to the Moral Sense Test (MST) website (http://www.moral.wjh.edu). Previous research, presenting moral dilemmas of this kind, has demonstrated that there are no substantive differences between judgments obtained from participants who complete Web-based questionnaires and participants who complete more traditional pen-and-paper questionnaires (Hauser et al., 2007). All procedures were conducted in accordance with the Institutional Review Board of Harvard University, and followed the testing procedures of other web-based research projects. Each participant was asked to complete the test without interruption, to read each scenario and associated question carefully, and to answer the questions solely on the basis of the information provided. Each experiment included six distinct dilemmas targeting Paretian considerations as well as the directness of physical contact (the complete text of all test materials is available at http://moral.wjh.harvard.edu/ParetoSupp).

After reading the text of a dilemma, each participant was asked to judge whether the protagonist's action was morally permissible; and, participants responded with either a 'Yes' or a 'No'. Between 17 December 2007 and 6 July 2008, 3371 responded to two dilemmas drawn from the five experiments reported below. Each dilemma was presented as part of a longer questionnaire, separated either by a number of intervening vignettes that targeted the acceptability of different sleeping arrangements among members of a family (thus ensuring that the intervening material did not include any other moral dilemmas); or, as part of a longer questionnaire that included a number of additional scenarios that were unrelated to the target cases presented here.

The dilemmas presented in these five experiments varied in the extent to which using a person as a means to some greater good made her worse off (including both cases where harm was inevitable and cases where a new harm was introduced); as well as the extent to which direct physical contact was required to bring about a greater good (including both cases where an actor could throw a rock at a person and cases where a person could be physically pushed). Each experiment included 1) a dilemma in which a Pareto improvement could be achieved by *throwing a rock at a person*, causing him to scream, and thereby allowing five others to escape an ongoing threat; 2) a dilemma in which a Pareto improvement could be achieved by *pushing a person* who was already doomed, causing him to scream, and thereby allowing five others to escape an ongoing threat; 3) a dilemma in which a *dead body* could be used to prevent harm that would otherwise occur as the result of an ongoing threat; 4) a dilemma in which a *person who's death was immanent* could be used to prevent harm that would otherwise occur as the result of an ongoing threat; 5) a variant of the well-known footbridge dilemma, where a person could be forced into harms way, killing him but saving the lives of five others; and 6) a dilemma in which a person could be coerced to step into harms way, killing him but saving the lives of five others.[1]

3. Trolleys and Wrecking Balls

We first examined the pattern of responses to these types of dilemmas in the context of a pair of mechanical threats (each based on the now familiar trolley problem). In Experiment 1, we examined a set of dilemmas in which 'trolleys' were replaced with 'empty boxcars' to prevent participants from assuming that the trolleys could be carrying or driven by additional people, thus generating additional harms. In Experiment 2, we examined a set of dilemmas that occurred in the context of a novel mechanical threat: a wrecking ball that is swinging out of control on a construction site. We begin with Boxcar cases because of the substantial amount of theoretical and empirical data that has already been collected for Trolley cases. However, given that our key interest concerns the extent to which the pattern of responses obtained for these cases generalized across different types of threats, we also used of the Wrecking Ball cases to explore the extent to which the popularity of trolley problems makes them unusual, and the extent to which or whether other cases of mechanical harm behave similarly. We hypothesized that in the context of a

[1] NB: the term 'coercion' here is not intended as compulsion by a threat, but merely inducement with a positive reward to do something that a person would not otherwise do.

novel mechanical threat, the pattern of responses are likely to remain similar to the pattern of judgments provided for the Boxcar dilemmas.

3.1 Results:

In line with our reanalysis of the data from Greene et al (2004), and the results reported by Moore et al (2008), a significantly higher proportion of participants who were presented with Boxcar dilemmas judged that it was permissible to harm a person when this brought about a Pareto improvement (81%) as opposed to making that person worse off (37%), c^2(1, n=1236) = 257.075, p<.001. Moreover, a significantly greater proportion of people judged that it was permissible to push a person in front of an oncoming boxcar in the context of a Pareto improvement as opposed to making this person worse off, c^2(1, n=448)=80.526, p<.001. In line with previous work showing the importance of physical contact in folk-moral judgments, we also found that a significantly greater (89%) proportion of participants judged that it was permissible to throw a rock at a person who was about to be hit by a boxcar (bringing about a Pareto improvement) than the proportion of participants who judged that it was permissible to push a person who was about to be hit by a boxcar (70%; c^2(1, n=393) =22.423, p<.001). Further, the proportion of participants who judged that it was permissible to use a dead person to save the lives of others (85%) was significantly greater than the proportion of participants who judged that it was permissible to use a dying person (65%) to save the lives of others, c^2(1, n=405)=22.215, p<.001. Lastly, the proportion of participants who judged that it was permissible to push a person in front of a boxcar in a non-Pareto context (27%) trended toward, but did not differ significantly from the proportion who judged that it was permissible to coerce someone onto the tracks (20%; c^2(1, n=438) = 3.039, p=.081).

	Boxcar	Wrecking Balls	Boxcar cases vs Wrecking Ball cases
Pareto, Mediated contact	89	90	c^2 (1, n=160) = 0.163, p=.686
Dead person, Direct contact	85	90	*c^2 (1, n=198) = 4.535, p=.033
Pareto, Direct contact	70	78	*c^2 (1, n=185) = 6.184, p=.013
Dying person, Direct contact	65	65	c^2 (1, n=192) = 0.015, p=.904
Worse off, Direct contact	27	23	c^2 (1, n=165) = 1.319, p=.251
Worse off, Coercion	20	20	c^2 (1, n=189) = 0.021, p=.884

Table 2: Proportion of affirmative responses for each of the Boxcar and Wrecking Ball Cases, Grayed cells represent Pareto cases

Similarly, in the case of an out of control wrecking ball, a significantly larger proportion of participants judged that it was permissible to harm a person where this brought about a Pareto improvement (86%) as opposed to making a person worse off (36%; c^2(1, n=1089) = 283.799, p<.001). Similarly, using direct physical contact to harm a person in the context of a Pareto improvement (78%) was once again judged to be more permissible than pushing someone in a non-Pareto context (23%; c^2(1, n=350)=107.093, p<.001). Moreover, the proportion of participants who judged that it was permissible to throw a rock at a person who was about to be killed, bringing about a Pareto improvement (90%) was significantly greater than the proportion who judged that it was permissible to push a person who was about to be killed (78%; c^2(1, n=345) = 8,522, p=.004). And, the proportion of participants who judged that it was permissible to use a dead person (90%) to save the lives of others was significantly greater than the proportion of participants who judged that it was permissible to use a dying person (65%) to save the lives of others,

c^2(1, n=390)=37.498, p=.003. Finally the proportion of participants who judged that it was permissible to push a person who was not about to be killed (23%) was not significantly different from the proportion who judged that it was permissible to coerce someone into harms way (20%; c^2(1, n=354)=.629, p=.428). Planned comparisons examining the relationships between the individual dilemmas in this experiment (reported in Table 2) revealed that two of the six dilemmas differed.

3.2 Discussion

The pattern of judgments that were offered for the dilemmas presented in these experiments confirm the hypothesis that means-based harms that result in a Pareto improvement will be judged more permissible than means-based harms where there is no Pareto improvement, thereby replicating and extending the results reported by Moore et al (2007). Moreover, because these experiments allowed us to compare dilemmas that differed only in the source of the ongoing threat (a boxcar vs. a wrecking ball), this provides initial evidence that Paretian considerations are likely to play similar roles in familiar and unfamiliar sorts of cases. However, although the overall pattern of responses that were offered in response to these two sets of dilemmas was nearly identical, there were two dilemmas for which differences in judgments emerged. A significantly greater proportion of participants judged it permissible to push a dead person in front of a wrecking ball as compared to a boxcar; and, a significantly greater proportion of participants judged that it was permissible to push a healthy person in front of a wrecking ball as compared to a boxcar where there was a Pareto improvement. Given that these two sorts of cases are nearly identical, the difference between these cases is unexpected.

As we suggested above, the relative familiarity with the Boxcar dilemmas might lead people to offer judgments that differ slightly from the judgments that they make in less familiar cases. In part, such an effect might be driven by the increased accessibility of deliberative principles in making moral judgments for these cases (cf., Mussweiler 2002). Building on a suggestion by Epley and Gilovich (2006), we suggest that when conflict is generated between an intuitive moral judgment, and a prior deliberative response to a moral question, this could yield moral judgments that are 'close enough' to approximate plausible responses but that are generated by a deliberate and effortful search of possible options. That is, in cases where such a conflict arises, people may feel the need to adjust their initial, reflexive response, but they will stop adjusting once they have reached something that feels like a plausible moral judgment (Epley and Gilovich 2006). On the basis of these initial data, we suggest that differences in judgments might arise as a result of changing the context in which a harm occurs. Alternatively, participants may have been more likely to suppose that a human body could stop a wrecking ball than to suppose that a human body could stop an oncoming trolley. Perhaps the supposition that a body could stop a wrecking ball could lead participants to judge these two actions more permissible. However, because these two sorts of cases are almost indistinguishable, we cannot be sure whether such differences will always be evoked by differences in context.

In studies where this many dilemmas are examined, it is often difficult to tell just how important such effects are to the overall all pattern of judgments that people are likely to make to various dilemmas. It is, thus, important to acknowledge that although there are small differences between the pattern of judgments that were offered in the context of the boxcar and wrecking ball scenarios, a clear overall pattern was displayed by these two sorts of cases. Yet, we cannot be sure whether more robust differences in this pattern of judgments will arise where the differences in the source of a threat are more pronounced. Thus, in the remaining experiments, we turn to three parallel sets of dilemmas that manipulate the nature of the impending threat while maintaining the overall structure of the dilemmas that are presented.

4. Travelers and Rough Neighborhoods

We next turned to the context of an ongoing threat that was caused by an intentional agent. To examine the generality of our results from Experiments 1 and 2, we developed two additional sets of scenarios, each of which was based on Williams' (Smart & Williams, 1973) discussion of the 'Jim and the Indians' scenario.[2] The dilemmas in Experiment 3 involved an intentional agent who was threatening the lives of others in the context of a foreign country; the dilemmas in Experiment 4 involved an intentional agent who was threatening the lives of others in the context of a more familiar rough neighborhood. We chose to use this contrast to test for the possibility that the permissibility of an agent's actions depends, in part, on whether moral judgments about the permissibility of action are stable across cultural contexts or whether they are constrained by cultural setting. If familiarity plays a role in folk-moral judgments, these two types of cases will both differ from one another, as well as differing from the pattern of judgments that were elicited by the Boxcar dilemmas. Alternatively, if only the predictability of the behavior of a morally degenerate agent is integral to moral judgment, these dilemmas should elicit responses that are similar to one another, but different from the Boxcar dilemmas.

4.1 Results:

In Experiment 3, we presented a case in which an unfortunate traveler in a foreign country stumbles upon a military leader who is about to execute a number of villagers. Here, a significantly larger proportion of participants judged that it was permissible to harm a person where this brought about a Pareto improvement (66%) as opposed to making a person worse off (40%; c^2(1, n=1236) = 85.248, p<.001). Similarly, using direct physical contact to harm a person in the context of a Pareto improvement was once again judged to be more permissible than pushing someone in a non-Pareto context, c^2(1, n=412)=33.511, p<.001. However, in contrast to the results that we obtained for the Boxcar scenarios in Experiment 1, the proportion of participants who judged that it was permissible to throw a rock at a person who was about to be killed, bringing about a Pareto improvement (64%), did not differ significantly from the proportion that judged it permissible to push a person who was about to be killed (63%), c^2(1, n=415) = .091, p=.838. The proportion of participants who judged that it was permissible to use a dead person (71%) to save the lives of others was once again significantly greater than the proportion of participants who judged that it was permissible to use a dying person (52%) to save the lives of others, c^2(1, n=433)=16.954, p<.001. Finally the proportion of participants who judged that it was permissible to push a person who was not about to be killed in front of the executioner (34%) did not differ significantly from the proportion who judged that it was permissible to coax someone out in front of the executioner (32%), c^2(1, n=388) = .161, p=.688. Planned comparisons examining the relationships between

[2] The case runs as follows. Jim happens into the central square of a small South American town where 20 indigenous tribesmen are about to be executed. The executioner, seeing Jim as an honored foreigner, makes the following offer: If Jim kills one person himself, the others will go free; if he refuses, all 20 will be executed as planned. Williams asks whether it is morally permissible to shoot the person and castigates the utilitarian who would treat the issue as a matter of simple arithmetic (supplemented by calculations over the probability of success and comparative judgments about alternative possibilities). However, Williams acknowledges that, tragically, it would be morally right in this hard case to shoot the person since this person is already doomed and the outcome will be better than it would have been otherwise.

the individual dilemmas in this experiment with their counterparts from the Boxcar case (reported in Table 3) revealed that each of the six dilemmas differed significantly as a result of the context in which the dilemma occurred.

	Boxcar	Traveler	Traveler VS. % from Boxcar cases
Pareto, Mediated contact	89	64	c^2 (1, n=217)=137.992, p<.001
Dead person, Direct contact	85	71	c^2 (1, n=220)=34.260, p<.001
Pareto, Direct contact	70	63	c^2 (1, n=198)=5.127, p=.024
Dying person, Direct contact	65	52	c^2 (1, n=213)=16.703, p<.001
Worse off, Direct contact	27	34	c^2 (1, n=214)=5.492, p=.019
Worse off, Coercion	20	32	c^2 (1, n=174)=16.144, p<.001

Table 3: Proportion of affirmative responses for each of the Boxcar and Traveler Cases, Grayed cells represent Pareto cases

Presented with the agent based dilemmas that occurred in the context of a rough, but presumably more familiar environment, a significantly larger proportion of participants judged that it was permissible to harm a person where this brought about a Pareto improvement (72%) as opposed to making a person worse off (48%), c^2(1, n=1089) = 65.072, p<.001. However, in this case, although using direct physical contact to harm a person in the context of a Pareto improvement (53%) was typically judged to be more permissible than pushing someone in a non-Pareto context (44%), this difference was not statistically significant, c^2(1, n=358) = 3.050, p=.081. Once again, the proportion of participants who judged that it was permissible to throw a rock at a person who was about to be killed, bringing about a Pareto improvement (69%), was significantly greater than the proportion who judged that it was permissible to push a person who was about to be killed (53%), c^2(1, n=555)=10.104, p=.001. And, the proportion of participants who judged that it was permissible to use a dead person (98%) to save the lives of others was significantly greater than the proportion of participants who judged that it was permissible to use a dying person (67%) to save the lives of others, c^2(1, n=387) = 8.081, p=.004. Finally the proportion of participants who judged that it was permissible to push an otherwise safe person in front of the executioner (44%) was significantly greater than the proportion who judged that it was permissible to coerce someone out in front of the executioner (31%), c^2(1, n=354)=53.603, p<.001. Planned comparisons (reported in Table 4) revealed that five of the six dilemmas differed significantly from their Boxcar counterparts, and four of the six differed significantly from their Traveler counterparts.

	Rough Neighborhood	VS. % from Traveler cases	VS. % form Boxcar cases
Pareto, Mediated contact	69	c^2 (1, n=192)= 1.880, p=.170	c^2 (1, n=192) =80.421, p<.001
Dead person, Direct contact	98	c^2 (1, n=165)= 56,598, p<.001	c^2 (1, n=165)= 20.466, p<.001
Pareto, Direct contact	53	c^2 (1, n=198)= 8.443, p=.004	c^2 (1, n=198)= 27.152, p<.001
Dying person, Direct contact	67	c^2 (1, n=189)= 17.485, p<.001	c^2 (1, n=189)= .401, p=.527
Worse off, Direct contact	44	c^2 (1, n=160)= 6.778, p=.009	c^2 (1, n=160)= 22.775, p<.001
Worse off, Coercion	31	c^2 (1, n=185)= .036, p=.850	c^2 (1, n=185)= 14.899, p<.001

Table 4: Proportion of affirmative responses for each of the Rough Neighborhood cases, Grayed cells represent Pareto cases

4.2 *Discussion:*

The pattern of judgments for the dilemmas presented in Experiments 3 and 4 confirm the hypothesis that means-based harms that result in a Pareto improvement will be judged more permissible than means-based harms where there is no Pareto improvement. Moreover, given that these dilemmas differ with respect to the source of the ongoing threat (a mechanical device vs. a person), it is clear that Paretian considerations are likely to be operative across domains of harm. However, these data also reveal that although the directness of physical contact (i.e., whether a person was pushed or hit with a rock) had a significant effect on judgments in the context of a Pareto-improvement where the source of the harm was mechanical dilemmas, we found no similar effect for the Traveler dilemmas. Yet, in neither case did pushing as opposed to coercing a person into harms' way evoke a significant difference between judgments in either case. Thus, although the overall pattern of responses that we found in response to each of these sets of dilemmas was broadly similar, these data also reveal that there are important differences between the responses that people tend to provide in response to Boxcar as opposed to Traveler-based dilemmas. Given these results, it seems that the source of an ongoing threat is likely to have a significant effect on folk-moral judgments about the moral permissibility of an action, even though it does not trump the role of Paretian considerations. Although the Traveler cases replicated the role of Pareto improvement on folk-moral judgments, the proportion of affirmative responses that was offered for each of the Pareto dilemmas was significantly reduced in the Traveler cases as compared to the Boxcar cases. Moreover, actions that made a person worse off were more frequently judged to be permissible in Traveler dilemmas than in Boxcar dilemmas.

With the Rough neighborhood cases, we recovered the same overall pattern of responses, including the interaction between considerations of Pareto improvement and considerations of the directness of physical contact. However, in this case, a notably smaller proportion of participants judged that the actions that resulted in a Pareto improvement, but required direct physical contact, were permissible. This raises an important theoretical question: Why would the source of an ongoing threat have a significant effect on the pattern of responses that is provided to a series of dilemmas? More specifically: Why would the source of an ongoing threat have any significant effect on the way in which considerations of Pareto-improvement and the directness of physical contact interact?

To our minds, there are a number of reasons why this might be the case, all of which should be further pursued to flesh out the richness of the representations underlying moral judgments. First, it seems to engage in an action that conforms to the will of the executioner. This being the case, even though some actions in the Traveler dilemmas are likely to bring about Pareto-improvements, the good-making feature of the action (i.e., bringing about the Pareto Improvement) is likely to be mitigated by the bad-making feature of these cases (i.e., acting in accordance with the desires of a bad agent). Second, there are likely to be differences in the probabilities of accurately predicting the outcome of an intervention in the Traveler dilemmas compared to the Boxcar dilemmas. While it is reasonable to suppose that one's actions will have the predicted effect in a Boxcar dilemma where there are no other agents involved, it is more difficult to predict what is likely to occur when another agent, specifically a morally degenerate agent, is the source of an ongoing threat. In the Traveler cases, the actor must both decide to act and be willing to accept that the executioner will hold up his end of the bargain. Third, because an agent is the source of the ongoing threat in the Traveler case, bringing about a Pareto improvement (e.g., by throwing a rock at a person) may also put the protagonist at risk.

Engaging in the relevant action might, for example, anger the executioner and cause him to kill the protagonist. Fourth, in the Traveler cases, it is possible that even if an action brings about a Pareto improvement in the short-term, the people whose lives are saved might be captured again and killed at some future time. Put briefly, Traveler cases seem to entail a higher level of uncertainty about actions and outcomes than do Trolley cases, and this uncertainty may generate the observed variation. Finally, the Traveler cases take place in an unfamiliar country, and people may have been unsure whether an executioner in this context could be believed or trusted—that is, the unfamiliarity of the case may have exacerbated the difficulty that participants face in predicting the behavior of a morally degenerate agent. The increased difficulty in predicting the behavior of unknown agents (or individuals from distinctively different 'out-groups') may play an integral role in structuring folk-moral judgments about moral permissibility.

When a human agent is the source of an ongoing threat, it is necessary to examine both the overall structure of the case as well the intentions of the person who is about to bring about the relevant harm. However, in the case of a more familiar gang-leader, it might be easier to predict his behavior. For example, throwing a rock at one person may introduce a threat to self, risking death by the gang-leader. Moreover, in acting in accordance with the gang-leaders' request, it is unclear whether you can trust him to follow through on his claim that he will release the others—but it might be your safest bet if you don't want to upset the gang leader. Although we ask participants not to incorporate any additional considerations into their judgments, considerations such as these may be immediately recruited by our moral psychology in cases where an intentional agent is the source of an ongoing threat. If so, we should expect that dilemmas that rely on an intentional agent as the cause of an ongoing threat will show patterns of responses that deviate from those in which a threat is mechanical in nature; and this may come about both because of the uncertainty concerning the agent's mental states and the probability that certain consequences will emerge. We cannot, of course, know which of these factors is operative in driving the differences between these two sorts of dilemmas. However, the pattern of results in these two experiments invites a closer examination of the effects of varying the source of an ongoing threat on the computations that are carried out in processing means-based harms that include a Pareto improvement. We thus turn to a final set of cases that target a plausible hypothesis about the effects of the source of an ongoing threat on folk-moral judgments.

5. Burning Houses

The dilemmas in Experiments 1-4 differ most clearly in the source of the ongoing threat (a boxcar or a wrecking ball vs. a person) and the familiarity of the context in which the threat occurs (relatively familiar: railroad tracks and rough neighborhoods vs. unfamiliar: remote South American village). However, they also differ in the extent to which engaging in an action requires conforming to the will of a morally degenerate agent. To examine the role of this variable in folk-moral judgments, we designed a final set of dilemmas in the context of a more familiar and a more plausible threat: a burning house where a single life can be sacrificed to save the lives of others. Our Burning House dilemmas include a threat that is likely to be familiar; however, it is also important to note that these cases retain much of the predictive uncertainty that we find in cases where an agent is the source of an ongoing threat. The trajectory of a fire is not as easily predictable as the trajectory of a runaway boxcar, but a fire has no intentional states of its own. Thus, although this case maintains the unpredictability of the threat and the outcome, as well as the inanimacy of the source of threat, it is not necessary to make assumptions about the extent to which the decision to act brings one into line with the aversive desires of an evil agent.

5.1 Results:

Presented with the Burning House dilemmas, a significantly higher proportion of participants judged that it was permissible to harm a person when this brought about a Pareto improvement (87%) as opposed to making that person worse off (52%; c^2(1, n=1046) = 157.339, p<.001). Moreover, a significantly higher proportion of people judged that it was permissible to push a person in the context of a Pareto improvement as opposed to pushing a person where this made him worse off, c^2(1, n=378)=106.224, p<.001. Surprisingly, given the results of Experiments 1 and 2, the proportion of participants who judged that it was permissible to throw a rock at a person to bring about a Pareto Improvement (74%) was significantly smaller than the proportion who judged that it was permissible to push a person to bring about a Pareto improvement (93%; c^2(1, n=358) = 23.735, p<.001). Once again, however, the proportion of participants who judged that it was permissible to use a dead person (95%) to save the lives of others was significantly greater than the proportion of participants who judged that it was permissible to use a dying person (78%) to save the lives of others, c^2(1, n=323)=20.220, p<.001. Moreover, in line with our previous results, the proportion of participants who judged that it was permissible to push a person in a non-Pareto context (43%) did not differ significantly from the proportion who judged that it was permissible to coerce someone into harms way (37%; c^2(1, n=365) = 1.475, p=.225). Planned comparisons examining the relationships between the individual dilemmas in this experiment with their counterparts from Experiments 1 and 2 (reported in Table 4) revealed that each of the six dilemmas differed significantly from their Boxcar counterparts, and five of the six differed significantly from their Traveler counterparts.

	Burning House	VS. % form Boxcar cases	VS. % form Traveler cases
Pareto, Mediated contact	74	c^2 (1, n=167)=40.179, p<.001	c^2 (1, n=167)= 6.754, p=.009
Dead person, Direct contact	95	c^2 (1, n=166)=13.494, p<.001	c^2 (1, n=166)= 47.140, p<.001
Pareto, Direct contact	93	c^2 (1, n=191)=46.744, p<.001	c^2 (1, n=191)= 72.132, p<.001
Dying person, Direct contact	78	c^2 (1, n=157)=12.288, p<.001	c^2 (1, n=157)= 43.653, p<.001
Worse off, Direct contact	43	c^2 (1, n=187)=25.256, p<.001	c^2 (1, n=187)= 7.232, p=.007
Worse off, Coercion	37	c^2 (1, n=178)=32.449, p<.001	c^2 (1, n=178)= 2.110, p=.146

Table 3: Proportion of affirmative responses for each of the Burning House cases, Grayed cells represent Pareto cases

4.2 Discussion:

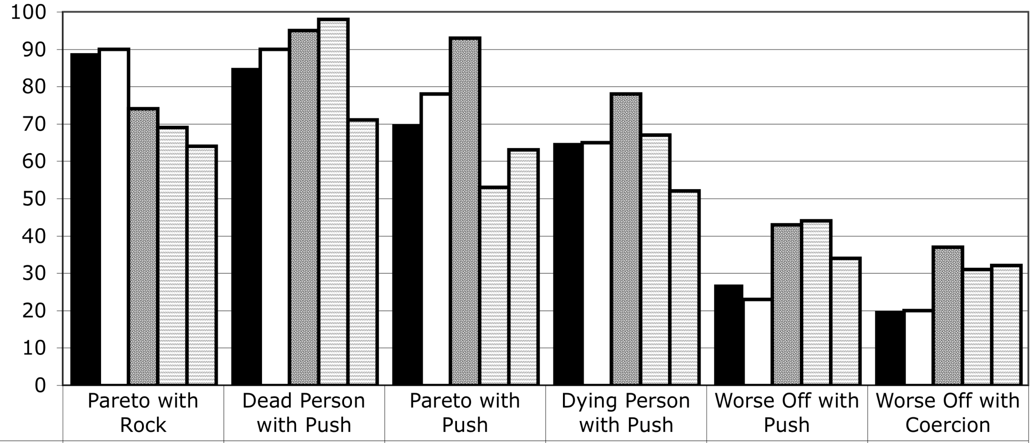
Although the Burning House dilemmas may have been more intuitively realistic than the boxcar, wrecking ball, or traveler cases, they nonetheless evoked a similar overall pattern to the judgments evoked by these scenarios. However, there is one point that is worthy of discussion: participants in this experiment were less likely to judge it permissible to throw a rock at a person to bring about a Pareto improvement than to push a person to bring about a Pareto improvement. One factor that might evoke this sort of effect is the temporal ordering of harms (Kamm, 2005; Sinnott-Armstrong, Mallon, Hull, & McCoy, forthcoming). In the parallel boxcar case, a lone person will be killed before the five others who are threatened further down the track. So, throwing a rock at the person leaves the order of harms unaltered. But, in the Burning House case, the fire threatens the lone person at exactly the same time as it threatens the people in an adjacent room. So, throwing a rock at the person introduces a new temporal ordering of harms; the harm caused by hitting this person with the rock causes an earlier harm. As Hart and Honoré (1985) argue, where an actor voluntarily intervenes in the causal chain stretching from an initial threat to the outcome of an action, our perception of the causal structure of the case tends to terminate at the point where the intervention occurs. The voluntary act then functions as 'a barrier and a goal in tracing back causes' (Hart & Honore, 1985, p. 44). This fact about intervention in the order of harms, is likely to make temporal ordering more salient—first this person is hit by a rock, then he is killed by the fire. And salience may enhance or attenuate certain aspects of the dilemma, at least with respect to our initial intuitions. We suggest that voluntary intervention into the order of harms could manipulate the structural description of the action, thereby introducing a new *intentional harm* that the person did not antecedently face.[3]

Indeed, this hypothesis is further confirmed by the results that we obtained for the Wrecking Ball cases, which displayed an almost indistinguishable pattern from the one observed in the Boxcar cases (represented graphically in Figure 1). We suggest, therefore, that Paretian considerations interact in substantive ways with the source of an ongoing threat, attenuating or enhancing the perceived permissibility of the morally relevant action.

5. General Discussion:

Although there has been a considerable amount of philosophical scrutiny (Foot, 1967; Kamm, 1989, 2005; Smart & Williams, 1973; Thomson, 1985) applied to questions of permissible harm, the empirically based psychology of means-based harms is only just beginning. Recent work in moral psychology has shown that, *ceteris paribus*, participants consistently judge means-based harms as worse than side effects, providing support for the hypothesis that some form of the doctrine of double effect plays an important role in folk-moral intuitions (Cushman et al., 2006; Greene et al 2009; Hauser, 2006; Hauser et al., 2007; Mikhail, 2007; Abarbanell & Hauser, 2010). Moore et al. (2008) further showed that when death to one person is inevitable, using this person as a means to bring about a greater good for others is perceived as more acceptable than if the death is evitable (Moore et al., 2008). This distinction between inevitable and evitable harms closely maps to our implementation of Paretian considerations. Furthermore, in each of our experiments, participants consistently judged actions that entailed a Pareto improvement to be more permissible than parallel cases in which a person was made worse off (i.e., non-Pareto). That being said, we found that there were important differences that also emerged across our five types of scenarios. This suggests that there are important ways in which the different factors that are operative in a morally

[3] Additionally, there is a possible issue of believability for the Pareto improvement cases that included direct contact. In the Pareto improvement with direct contact case, it was claimed that Dave, the protagonist, sees that there is a person immediately behind a door that he could knock down to get the other people out of the room safely. However, because this person was described as standing behind the door, it is not obvious how the protagonist could 'see' the person behind the door. While we intended the term 'sees' to be understood metaphorically, if people intuitively interpreted the term literally, they may have treated the action that brought about the harm in the case of the Pareto-improvement as irrelevant and focused only on the issue of saving the five people. Of course, we cannot be sure if this is the case; however, it is clear that people saw Dave's action as far more permissible than the other comparable harms that we have examined.



	Pareto with Rock	Dead Person with Push	Pareto with Push	Dying Person with Push	Worse Off with Push	Worse Off with Coercion
■ Boxcars	89	85	70	65	27	20
□ Wrecking balls	90	90	78	65	23	20
▩ Burning Houses	74	95	93	78	43	37
▤ Rough Neighborhood	69	98	53	67	44	31
▤ Unfortunate Travelers	64	71	63	52	34	32

Figure 1: Proportion of affirmative responses in Experiments 1-5

significant scenario can interact to constrain and regulate folk-moral judgments. We conclude, therefore, by raising some of theoretical and methodological issues that emerge from these analyses.

The results presented here confirm the importance of Paretian considerations for folk-moral judgments. We suggest that Paretian considerations should be treated as an *abstract principle* that is operative in folk-moral judgment across a wide variety of contexts, involving different sources of threat and different degrees of contact. From the perspective of folk-morality, it matters whether someone is made worse off, *independently* of the context in which this occurs. This finding helps to inform current work in moral psychology that is aimed at testing the hypothesis that humans, as a species, are endowed with a universal moral grammar (Mikhail, 2007; Hauser, Young, & Cushman, 2007; Dwyer, Huebner, & Hauser, in press) or a set of abstract principles (Greene et al., 2009) that guide moral judgment; but, it is also of considerable importance to the extent that it demonstrates that folk-morality is not as insensitive to consequentialist considerations as non-consequentialists might have hoped (Anscombe 1967; Taurek 1977; though see Kamm, 1993). Pareto cases are likely to elicit judgments of permissibility because there is no doubt, in such cases, that the relevant action will bring about the best overall consequences. This principle does not merely turn on weighing the imposition of a harm that worsens the fate of one person against mitigating a harm to a number of others. Rather it turns on weighing the imposition of harm to one person against another situation in which that person, as well as others whose welfare or fortune is already at stake, will suffer a similar fate.

Returning to the case of Jim and the Indians, Kamm (1999, 2005) argues that this is a Pareto case: it pits the life of a single person who is about to be killed against the lives of many others who are also about to be killed. Although this case is typically seen as an argument against utilitarian moral theories, even Williams concedes that *in this case*, the right thing to do is to kill the individual to save the others. Such a judgment requires a comparative analysis of the relative value of the loss of one life in saving many others; and this computation provides a reason for action *even for the non-consequentialist*. No matter how strongly he is willing to argue against the utilitarian, even Williams sees that Paretian considerations are sufficiently integral to our commonsense judgments about right and wrong to mitigate the impermissibility of taking the life of one person in order to save many others. Such considerations make it right, albeit tragically so, to shoot the one person. Analogously, Kamm's (1999, 2005) anti-consequentialist arguments leave room for the comparative analyses required for considerations of Pareto improvement. This fact about the elaboration of non-consequentialist moral theories, coupled with the data that we have presented in this paper, provide strong support for the claim that Paretian considerations play an integral role in folk-morality.

This convergence between philosophical moral theories and folk-morality provides an important insight into a possible constraint on the space of viable moral theories. Whether or not Kamm and other moral theorists succeed in establishing that Paretian considerations should be relevant for the non-consequentialist, it is significant that scholars on both sides of the debate are compelled to recognize that Paretian considerations are *significant in moral deliberation*. This provides strong evidence for the claim that some aspects of folk-morality are so deeply engrained in our cognitive architecture that they will continue to have significant implications for structuring the range of plausible moral theories regardless of what sorts of reflective moral theories are also adopted. The key claim here is that our intuitive folk-morality is grounded in a set of abstract principles that are likely to be species-typical, and that plausibly account for the robust patterns of similarities and differences that we are likely to find in both explicitly articulated philosophical theories, as well as folk-morality (Dwyer, 1999, In press; Hauser, 2006; Mikhail, 2007, 2008a).

The brute complexity of the decision procedure that must be carried out in making moral judgments must involve an intricate set of computations operating over at least some principles that are sensitive to the structure of a causal chain and the intentionality of an action. Theoretically speaking, this point has been aptly defended by Kamm (2005) in her philosophical work, and by Mikhail (2007) in his attempt to provide a set of transformational rules that provide the structure for the folk-moral domain. Of course, the intuitions that we find in philosophical theories are likely to be more thorough, more precise, and more coherent than their folk-moral counterparts. However, the emerging experimental data, including the data that we have reported in this paper, strongly suggest that some of the moral principles that philosophers have examined are indeed integral to folk morality in a way that cannot be easily over-ridden. In short, folk-morality provides us with a rough-and-ready tool that allows us to competently navigate our social world, as well as to make rapid moral judgments about unfamiliar cases. In order for this to be possible, however, it seems that these intuitive judgments need only operate over a relatively small set of abstract principles, which (when interfaced in the right way) provide us with the resources that are required for making practical judgments about most of the morally salient situations that we are likely to face (Greene, 2007; Greene & Haidt, 2002; Greene et al., 2004; Greene et al., 2001, 2009; Hauser, 2006; Mikhail, 2007, 2008a). Based on the work that has been conducted to date (Cushman et al, 2006, Cushman, 2008; Baron, Moore et al., 2008, Greene et al., 2009), and the evidence that we have presented here, it is clear that folk-morality must minimally contain principles that allow moral judgments to be sensitive to the following six distinctions:

1. Actions vs omissions:
2. Means vs side-effects;
3. Physical contact vs no physical contact;
4. Linear vs the non-linear relationships between an intention and an outcome;
5. The use of personal force vs a mediated causal relationship; and
6. Considerations of Pareto-improvement.

In making such a claim about these distinctions, and in adding Paretian considerations in particular, we do not intend to argue that Paretian considerations must be a part of the fundamental grammar of our moral psychology. Even on the assumption that there is a Universal Moral Grammar (Rawls, 1971, Harman, 1978, Dwyer, 1999, Hauser, 2006, Mikhail 2007), it could be that Pareto considerations are insufficiently abstract to qualify as a fundamental principle of folk-morality. In fact, it remains an open possibility that Pareto considerations play an important role in modulating judgments about what counts as harm (this option is discussed briefly below), computing over the structure of the outcome prior to the operation of distinctly moral cognition *per se*. Some support for the claim that Paretian considerations are likely to serve a more modulatory role in our psychology has been adduced by Moore et. al. (2008b), who suggest that although the inevitability of harm attenuates the aversion to harming another person, it only does so when there is a clear benefit to others. As we have argued, this principle may only play a role in cases where Pareto considerations are genuinely relevant. Regardless of the precise role that Paretian considerations play in folk-morality, there are complexities in the data that still call for analysis.

To begin with, although we found few significant differences between dilemmas targeting a single variable in our boxcar cases and our wrecking ball cases, there were significant differences between the judgments that were offered in response to these highly mechanical dilemmas and the responses that were offered for dilemmas in which a harm was brought about in some non-mechanical way. Specifically, although participants who were presented with the Boxcar and Wrecking ball cases tended to see it as more permissible to throw a rock at someone than to push him where this brought about a Pareto improvement, no similar effect was seen where the source of the ongoing threat was a fire in a burning house or a paramilitary leader in an unfamiliar country. Thus, although a robust general pattern of responses emerged for Paretian considerations across our manipulations, considerations of the directness of physical contact seem to have been made more salient where the impending harm originated in a mechanical as opposed to a non-mechanical source. Thus, our results seem to demonstrate that folk-moral judgments are significantly affected by the source of an ongoing threat, especially as this is related to the plausibility of the success of a particular intervention in the causal chain. To our knowledge, such issues have not, hitherto, been addressed empirically, and further analyses are required along these different dimensions. However, such considerations do seem to suggest an alternative account of the psychology of means-based harms.

Second, because the inevitability of a harm is only brought on-line where harming another person is a means to helping others (Moore et al., 2008), one could argue that the modulation of our moral judgments by such considerations results from a recoding of the action in a way that removes it from the domain of harm. In each of the Pareto cases that we developed, a direct and immediate physical harm is done to a person either by hitting him with a rock or pushing him to the ground. In each of these cases, however, it is made clear from the beginning that this person's interests are already doomed—he is sure to be killed by the oncoming threat. This being the case, the fact that a direct, immediate, and physical harm is done to a person may be of a different order of magnitude than the unavoidable harm that the person already faces in having all of her or his interests set back by death. Therefore, it might be that the initial aversion to treating a person harmfully as a means to a further end persists, but in some cases, an initial harm is *recoded* in a way that allows it to be treated as nothing more than a physical 'hurt.' On this view, a genuine harm requires that a significant number of interests are set-back (Feinberg, 1984).

To make this final theoretical point clear, consider the fact that in order to save 5 people down-track, it is more permissible to throw a rock at one person (making him scream, and thereby treating him as an alarm to alert the others of the impending threat) than to push a dying man onto the tracks, even though both are used as a means to save the 5, and even though both will inevitably die. It may initially seem as though a greater harm is done to the person who is used as an alarm call–after all, he presumably would have had a much longer life ahead of him were it not for his current circumstances. However, as our participants read this case, it presumably became clear that all of this person's interests were already *immediately* doomed by the impending threat of the runaway boxcar. Thus, merely adding the added hurt of being hit in the head with a rock fails to be a significant harm to this person beyond the harm that would soon be caused by the runaway boxcar. Things are, however, quite different with the dying man. As described, life support has been removed and this person does not wish to be resuscitated —he wants to die. However, as the case is described, this man may have an interest in watching the sun set one last time, in watching a final train roll past, or perhaps even in awaiting a final fade to black. Thus, it is possible that pushing him onto the track may set back some of this man's interests. However, far fewer of this man's interests are set back than are set back by pushing a healthy backpacker onto the tracks. The hypothesis that some apparent *harms* are recoded as mere *hurts* is, thus, consistent with the pattern of data that we see in our Pareto cases. However, as this hypothesis was not targeted by the current research, we cannot be sure whether this is actually how the computations required for making such moral judgments are implemented.

With these theoretical points in mind, we close with a methodological point about the study of our moral psychology. Trolley dilemmas have become the norm for research into the folk-psychological processes that yield moral judgments (Greene et al, 2001, 2004, 2009; Hauser et al., 2007; Sinnott-Armstrong et al., forthcoming; Valdesolo & DeSteno, 2006). We have also used a set of trolley car cases because they readily lend themselves to targeted inquiry into the abstract principles that are operative in our moral psychology; further, thanks to the work of philosophers such as Thomson, and especially Kamm, we have numerous trolley cases already on offer. However, relying on the responses of participants to these sorts of cases cannot *by itself* demonstrate the presence of an abstract principle or factor in our moral psychology. Of course, other researchers have used other sorts of dilemmas (Cushman et al., 2006; Greene et al., 2004; Waldmann & Dietrich, 2007). Our point is, more specifically, that in order to demonstrate the significance of a psychological distinction or abstract principle within the moral domain, multiple sets of cases, all of which target one central parameter or moral consideration, must be developed.

Our results suggest that there are some kinds of folk-moral considerations that play a role in our moral psychology across a variety of contexts, ranging from artificial cases (trolleys) to more natural and plausible cases (burning houses); further, they apply when the means of harming are mechanical or non-mechanical. However, in some cases the results from the trolley set do not generalize to other situations. Consequently, we suggest that appeals to trolley cases may not, in every case, yield a result that generalizes as widely as might be hoped. We contend that the only way to show that an abstract principle or distinction is operative in our moral psychology is to demonstrate that it mediates judgments over a wide range of contexts.

Georgetown University, Dept. of Philosophy
Harvard University, Depts. of Psychology and Human Evolutionary Biology
Princeton University, Dept. of Philosophy

6. References:

Adler, M., & Posner, E. 2006: New foundations of cost-benefit analysis. Cambridge: Harvard University Press.

Arrow, K. 1968: Economic Equilibrium In D. Sills (Ed.), International Encyclopedia of the Social Sciences (Vol. 4, pp. 376–388). London and New York: Macmillan and the Free Press.

BBC. 2007, 2 January 2007: Stuck woman traps SA cave group. BBC News Retrieved 04 November 2008, from http://news.bbc.co.uk/2/hi/africa/6225301.stm

Cushman, F., Young, L., & Hauser, M. 2006: The Role of Reasoning and Intuition in Moral Judgments: Testing three principles of harm Psychological science, 17.

Dwyer, S. 1999: Moral Competence. In K. Murasugi & R. Stainton (Eds.), Philosophy and Linguistics (pp. 169-190). Boulder, CO: Westvew Press.

Dwyer, S. In press: Moral dumbfounding and moral psychology. Mind and language, XX, yy-zz.

Epley, N., & Gilovich, T. 2006: The anchoring-and-adjustment heuristic. *Psychological science*, 17 (4), 311-318.

Feinberg, J. 1984: Harm to others. Oxford: Oxford University Press.

Foot, P. 1967: The Problem of Abortion and the Doctrine of Double Effect. Oxford Review, 5, 5-15.

Greene, J.D., Cushman, F.A., Stewart. L.E., Lowenberg, K., Nystrom, L.E., and Cohen, J.D. 2009: Pushing moral buttons: The interaction between personal force and intention in moral judgment. Cognition, Vol. 111 (3), 364-371

Greene, J. 2007: Why are VMPFC patients more utilitarian?: A dual-process theory of moral judgment explains. Trends in cognitive science, 11, 322-323.

Greene, J., & Haidt, J. 2002: How (and where) does moral judgment work? Trends in cognitive science, 6, 517-523.

Greene, J., Nystrom, L., Engell, A., Darley, J., & Cohen, J. 2004: The neural bases of cognitive conflict and control in moral judgment. Neuron, 44(389-400).

Greene, J., Sommerville, R., Nystrom, L., Darley, J., & Cohen, J. 2001: An fMRI investigation of emotional engagement in moral judgment. Science, 293, 2105-2107.

Hart, H. L. A., & Honore, A. 1985: Causation in the law. Oxford: Oxford University Press.

Hauser, M. 2006: Moral Minds. New York: Harper Collins.

Hauser, M., Young, L., & Cushman, F. 2007: Reviving Rawls' linguistic analogy. In W. Sinnott-Armstrong (Ed.), Moral Psychology (Vol. 1: The evolution of morality). Cambridge, MA: MIT Press.

Kamm, F. 1989: Harming Some to Save Others. Philosophical studies, 57, 227-260.

Kamm, F. 1999: Responsibility and collaboration. Philosophy and public affairs, 28, 169-20

Kamm, F. 2005: Intricate Ethics. Oxford: Oxford University Press.

Kant, I. 2002: Groundwork for the metaphysics of morals (A. Wood, Trans.). New Haven:

Yale University Press.

McKinley, J. 2008, 02/08/2008: Surgeon Accused of Speeding a Death to Get Organs. New York Times,

Mikhail, J. 2007: Universal Moral Grammar. Trends in cognitive science, 11, 143-152.

Mikhail, J. 2008a: Moral cognition and Computational theory. In W. Sinnott-Armstrong (Ed.), Moral Psychology: The Neuroscience of Morality: Emotion, Disease, and Development. Cambridge, MA: MIT Press.

Mikhail, J. 2008b: The poverty of the moral stimulus. In W. Sinnott-Armstrong (Ed.), Moral Psychology: The evolution of morality. Cambridge, MA: MIT Press.

Mikhail, J., Hauser, M., Cushman, F., Young, L., & Kang-Xing Jin, R. 2007: A Dissociation Between Moral Judgments and Justifications. Mind and language, 22(1), 1-21.

Moore, A., Clark, B., & Kane, M. 2008: Who shall not kill? Individual Differences in Working Memory Capacity, Executive Control, and Moral Judgment. Psychological science, 19(6), 549-557.

Mussweiler, T. 2002: The malleability of anchoring effects. *Experimental Psychology*, 49, 67-72.

Pareto, V. 1906: Manual of Political Economy. New York: Augustus M. Kelley.

Petrinovich, L., O'Neil, P., & Jorgensen, M. 1993: An empirical study of moral intuitions: Toward an evolutionary ethics. Journal of Personality and Social Psychology, 64, 467-478.

Rawls, J. 1971: A theory of justice. Cambridge: Cambridge University Press.

Schaich Borg, J., Hynes, C., van Horn, J., Grafton, S., & Sinnott-Armstrong, W. 2006: Consequences, Action, and Intention as Factors in Moral Judgments: An fMRI Investigation. Journal of Cognitive Neuroscience, 18, 803-817.

Sinnott-Armstrong, W., Mallon, R., Hull, J., & McCoy, T. forthcoming: Intention, Temporal Order, and Moral Judgments. Mind and language.

Smart, J., & Williams, B. 1973: Utilitarianism: For and Against. Cambridge: Cambridge University Press.

Thomson, J. 1985: The Trolley Problem. Yale Law Journal, 94, 1395-1415.

Valdesolo, P., & DeSteno, D. 2006: Manipulations of Emotional Context Shape Moral Judgment. Psychological science, 17, 476-477.

Waldmann, M., & Dietrich, J. 2007: Throwing a bomb on a person versus throwing a person on a bomb: Intervention myopia in moral intuitions. Psychological science, 18(3), 247-253.