

The Preference for Indirect Harm

Edward B. Royzman¹ and Jonathan Baron^{1,2}

We presented subjects pairs of hypothetical scenarios. The action in each scenario harmed some people in order to aid others. In one member of the pair, the harm was a direct result of the action. In the other member, it was an indirect byproduct. Subjects preferred the indirect harm to the direct harm. This result could not be fully explained in terms of differences in judgments about which option was more active, more intentional, more likely to cause harm, or more subject to the disapproval of others. Taken together, these findings provide evidence for a new bias in judgment, a tendency to favor indirectly harmful options over directly harmful alternatives, irrespective of the associated outcomes, intentions, or self-presentational concerns. We speculate that this bias could originate from the use of a typical but somewhat unreliable property of harmful acts, their directness, as a cue to moral evaluation. We discuss the implications of the bias for a range of social issues, including the distinction between passive and active euthanasia, legal deterrence, and the rhetoric of affirmative action.

KEY WORDS: tragic moral choices; omission bias; direct vs. indirect harm; Kant.

Judgments about decisions often favor options that lead to worse consequences than the alternatives. In the endowment effect, people are unwilling to give up their endowment for what they would otherwise prefer to it (Kahneman *et al.*, 1990). In omission bias, people are unwilling to bring about a preferred outcome through harmful action, when the alternative involves omission that is even more harmful (Ritov and Baron, 1990; Spranca *et al.*, 1991). Many preference reversals show that some means of eliciting choices must lead to *nonoptimal outcomes*, that is, outcomes that the decision maker herself would find less desirable than the available alternatives (e.g., Hsee *et al.*, 1999).

¹Department of Psychology, University of Pennsylvania, Philadelphia, Pennsylvania.

²All correspondence should be addressed to Jonathan Baron, Department of Psychology, University of Pennsylvania, 3815 Walnut Street, Philadelphia, Pennsylvania 19104-6196; e-mail: baron@psych.upenn.edu.

Such judgments seem to result from the application of intuitive principles of judgment, such as “avoid losses” or “do no harm (through action).” When decision makers follow such principles, nonoptimal outcomes result, so long as no other factors compensate for the biases in question (Baron, 1994, 1998). Such findings of bias away from the optimal outcome are of interest for two reasons. First, they inform us about our basic methods for making judgments and decision. Judgments that always favor the best outcome are not as informative, because they can be explained in so many ways (histories of genetic, cultural, or individual adaptation, or abstract understanding of the optimal relation between decisions and outcomes). Second, findings of bias suggest ways of improving decision making so as to produce better outcomes. (Of course, such suggestions need to be examined: explicit attempts to produce the best consequences are sometimes selfdefeating, and reliance on intuitive judgments may produce better results in some situations.)

These arguments do not distinguish between moral and nonmoral judgments. Biases against producing the best consequences are found in moral judgments as well as in nonmoral ones (Baron, 1994, 1995; Ritov and Baron, 1990), and the distinction between moral and nonmoral judgments may be slippery. In the present studies, we ask about preference without specifying whether the basis of preference should be moral or nonmoral. The issues of interest are, however, historically derived from moral philosophy. If such preferences lead to more harm than would otherwise occur, philosophical questions about the justification for this preference become more pointed.

Omission bias, for example, seems to lead to worse consequences when it plays out in individual and policy decisions (Baron, 1998; Meszaros *et al.*, 1996; Ritov and Baron, 1990). Harm is an act, but failing to prevent harm is an omission. People are more tolerant of harmful omissions than of harmful acts. Many people think that it is wrong to vaccinate a child against a fatal disease if the vaccine also causes fatalities, even if the fatalities from the disease greatly exceed those from the vaccine (Meszaros *et al.*, 1996; Ritov and Baron, 1990). The magnitude of the bias may be measured by asking how many deaths from the vaccine are acceptable, at most, to prevent 100 deaths from the disease. A utilitarian would answer 99. Many people answer considerably less, and some say “0.”

Spranka *et al.* (1991) also found a bias towards omission in a distinctly moral context. One of their hypothetical scenarios told a story of a villainous tennis player (John) who wound up playing his final against a much stronger opponent (Ivan). The scenario went on to say that, as John and Ivan sat down for dinner the night before the final, John (knowing that Ivan was allergic to cayenne pepper) decided to recommend Ivan the salad dressing containing the cayenne pepper in the hopes that Ivan will get sick and will lose the match. In one (harm through commission) version of the story, John makes the recommendation as planned. In the alternative (harm through omission) version, Ivan orders the dressing himself just before John is about to recommend it (and John, of course, says nothing). About a third of the

subjects who considered this scenario said that John behavior's was worse when he actually made the recommendation than when he "allowed" Ivan to choose the dressing himself.

Our concern here is with the preference for *indirect harm* over *direct harm*, when the consequences are the same (or even when they are worse for indirect harm). By "direct," we mean that the harm is perceived as part of the action, difficult to separate from the action without redefining what the action is. For example, if Jill pushes Jack off the top of the hill, that is a *direct* cause of Jack's fall. If Jack is leaning on a wobbly fence and Jill pushes the fence over, thus causing Jack to fall because he has nothing to lean on, that is a case of harm brought about in an *indirect* way. In the second case, the direct object of Jill's action is the fence, not Jack, though her entire operation is calculated to engage other factors in such a way as to bring about Jack's fall. Yet, in the second case, we can describe the action as "pushing the fence down," and this description would still be valid if Jack were not leaning on the fence (so that no harm would come to him from this act). It is more difficult (but not impossible) to imagine circumstances under which Jack would be unharmed in the case of the first action "pushing Jack off the hill." Notice that the distinction between direct and indirect is a matter of perception. In general, when harm is indirect, other causes of the harm, aside from a decisionmaker's choice, are salient, such as the fact that Jack was leaning on an unreliable fence. Furthermore, both direct harm and indirect harm are clearly distinct from "harm through omission," which, returning to the Jack-and-Jill scenario, would take place if Jill, though in the position to observe the fence slowly recede under Jack's weight, did or say nothing to prevent the fall.

A preference for indirect harm over direct harm can make consequences worse. For example, when a mother's life is threatened by a pregnancy, some Catholic hospitals will permit a hysterectomy to save the mother but not an abortion. The fetus dies in either case, but, in the case of the hysterectomy (which, unlike the abortion, leaves the mother unable to bear another child), the killing is seen as an indirect byproduct (Bennett, 1966; Kuhse, 1987). In other cases, the idea of harming someone directly to help others is invoked as a reason for moral restraint, as in the treatment of human subjects in research. The principle of informed consent is often justified as a way of making sure that subjects are not treated as means and harmed *directly* in order to help others.

Historically, opposition to direct harm (as opposed to indirect harm) may be traced to the doctrine of the double effect in Catholic moral theory, attributed to Thomas Aquinas (1947, 2 III, Q. 64, art. 7) as part of his explanation of why killing in self-defense can be morally acceptable:

Nothing hinders one act from having two effects, only one of which is intended, while the other is beside the intention. Now moral acts take their species according to what is intended, and not according to what is beside the intention, since this is accidental. . . . Accordingly the act of self defense may have two effects, one is the saving of one's life, and the other is

the slaying of the aggressor. Therefore this act, since one's intention is to save one's own life, is not unlawful, seeing that it is natural to everything to keep itself in "being," as far as possible. And yet, though proceeding from a good intention, an act may be rendered unlawful, if it be out of proportion to the end.

As later writers attempted to make sense out of this injunction, the role of intention was downplayed and another distinction came to the fore, that between indirect or accidental harm, on the one hand, and direct harm that is a means to an end is another. Quinn (1989), for example, says,

... a new and better formulation of the doctrine [of the double effect] ... distinguishes between agency in which harm comes to some victims, at least in part, from the agent's deliberately involving them in something in order to further his purpose precisely by way of their being so involved ... and harmful agency in which either nothing is in that way intended for the victims or what is so intended does not contribute to their harm.

Quinn uses the term "direct" and "indirect" for these two kinds of harm, and we shall use these terms.

The injunction against direct harm may also be traced to Kant's second formulation of the Categorical Imperative: "Act so that you treat humanity, whether in your own person or in that of another, [1] always as an end and [2] never as a means only" (Kant, 1785/1983). The negative injunction, "never as a means only," has been the source of much speculation. Because it contains the word "only," it is fully consistent with the view that those used as means must also be considered as part of the overall accounting. This view, the utilitarian view, would make no distinction between direct and indirect harm. If Kant had accepted this utilitarian view, however, he would simply have said that everyone should count as an end, and the negative injunction would not have been needed. Perhaps Kant was drawing on a more common intuition against the idea of using people as means at all: use of people as means suggests harming them directly for the benefit of others.

We show here that people favor indirect over direct harm, just as they favor harmful omissions over harmful acts. We use hypothetical scenarios, similar to those used to study omission bias, to attain maximum control over the stimuli. In the third experiment, we find both indirectness and omission biases in the same subjects. In the discussion section, we speculate about the psychological basis and the link between direct and indirect harm and omission, and draw the implications of the indirect bias for a range of social issues, including the distinction between passive and active euthanasia, legal deterrence, and the rhetoric of affirmative action.

STUDY 1

We report here some results from an initial study of two scenarios, which is inconclusive but which illustrates the basic effect. This study presented eight scenarios to different subjects. Each scenario contrasted direct and indirect harm and asked the subjects for a preference and a justification. Consequences were

equated between the two types of harm. Small samples were sufficient to show the effects, but they were not sufficient to test for effects of other variables.

Method

All the subjects (176 in total) were undergraduate students at the University of Pennsylvania, solicited through advertisements at various campus locations. Every subject was paid \$5 for his/her participation.

Each subject read a basic scenario followed by two different endings, each involving an action that harmed some people in order to aid others. In one ending, the harm was a direct result of the action. In the other ending, it was an indirect byproduct. After reading the scenarios, the subjects indicated a preference for one of the three options direct harm, indirect harm, or “no preference,” and offered a justification for their choice. The order of the two endings was counterbalanced across subjects. The experimenter (E.R.) was available to answer questions.

The numbers of subjects varied from 20 to 40 for different scenarios. The Airplane scenario (below) had 24, and the Mall had 20.

The Airplane scenario (much longer in its written form) asked the subject to take the role of a “senior officer at an U.S. military base facing a major emergency situation. A missile has just been mistakenly fired from your base at what is known to be a civilian aircraft, carrying on board a total of 200 passengers, including women and children.” The direct action was to “alter the course of the private jet flying nearby so that it is placed in the path of the missile. In this case the missile will collide with the jet before it can do any harm to the aircraft. (The jet, however, will be utterly destroyed).” The indirect action was to fool the missile’s tracking device “by altering the course of the civilian aircraft so that it is positioned directly behind the private jet,” in which case “the missile will collide with the jet and, thus, will explode before it can do any harm to the aircraft. (The jet, however, will be utterly destroyed).”

In the Mall scenario, the subject takes the role of walking through a crowded mall, when he notices that someone is about to shoot at him. In the direct action, the subject can position himself behind someone else who will take the bullet. In the indirect action, the subject can leap aside, in which case someone else, standing behind the subject, will take the bullet. (The story makes it clear that, in any case, exactly one person will die.)

Results and Discussion

Most subjects favored indirect harm over direct harm. Across all eight scenarios, 88% of the responses favored indirect harm, 12% were neutral, and less than 1% (one response out of 226) favored direct harm. Each scenario showed a significant preference for indirect harm by a binomial test ($p < 0.001$).

Justifications indicated that subjects understood the scenarios as we intended and did not assume that the consequences were any different for the direct and indirect option. For example, in the Airplane scenario, nine out of 19 subjects who favored the indirect option justified their response with a simple moral assertion about the direct option, for example, that it was “not right,” “morally reprehensible,” “more unethical,” “wrong,” “morally wrong,” or “something [one] could not live with.” Other common justifications (favored by a total of six subjects) were in terms of disrespect for the victim, for example, treating the victim as “less than human.” The remaining four subjects who were in favor of the indirect option either did not provide any explanation or gave an explanation that was circular in nature (e.g., “this seems like a better option”).

Likewise, in the Mall Scenario, eight out of 17 subjects who favored the indirect option justified their response with a simple moral assertion (e.g., saying that the direct option was “unethical” or “not right”). Four subjects justified their preference in terms of disrespect for the victim entailed by the direct option. The remaining five (out of 17) subjects could not explain why they favored the indirect option.

Justifications of indifference were either that the issue was obvious or that the consequences were the same. Justifications were similar in all the other scenarios.

The scenarios had several problems. Some were unrealistic. This may have made subjects try to think of whether the consequences were equally probable for the two options being compared. Subjects’ justifications for the judgments did not mention such differential probability, but we ask about it explicitly in the next study. The scenarios also asked the subjects to imagine themselves making the decision, and the scenarios described the motivation they were supposed to have. Subjects did not object to this demand, but the next study does not mention motives and asks the subjects to evaluation decisions made by someone else.

STUDY 2

The purpose of Study 2 was to test three alternative explanations of the preference for the indirect option. Each explanation assumes that subjects perceived the two versions differently in some way that they see as relevant but not the same as the distinction between direct and indirect harm. We tested the alternatives by asking the subject whether descriptions corresponding to each explanation apply differentially to the direct and indirect options. We also asked the subjects which descriptions were reasons for their judgments. We also used the third person and removed all descriptions of motivation. The alternative explanations were:

1. *Intention.* Part of the Doctrine of the Double Effect is that direct harm is more intended. We tried to control intention in Study 1 by making the actor’s motives constant, as well as the actor’s beliefs about expected consequences. However, subjects may still see the harm to the victim as more intended when the action is direct.

2. *Impression management (appearance)*. Direct action may look worse in the eyes of others. For example, in the indirect option of the Mall Scenario, others may think that the actor did not know someone was behind him when he leapt aside. We know it, in judging him, but we may judge him partly through appearances.
3. *Probability of harm*. Despite the instructions, subjects may think that the direct option is more likely to cause harm.

We also examined the role of action vs. omission. This is not an alternative explanation. Rather, direct actions may be seen as more actionlike, and this may explain why they are seen as morally worse. Spranca *et al.* (1991), for example, found that harmful actions were often judged to be morally worse than otherwise-equivalent harmful omissions. We explore this issue further in Study 3.

Method

Fifty-four subjects (ages 17–69, median age 34, 22% students, 69% females) completed a questionnaire on the World Wide Web for \$3. Subjects found the site mostly through links on other web pages and through search engines. Subjects who wished to be paid had to provide their name, e-mail address, address, and (for U.S. residents) their Social Security Number. Several studies (Birnbaum, 1999, 2000; McGraw *et al.*, 2000) find little or no difference between web subjects and other subjects.

After each of eight scenarios, the subject indicated the preferred option and then answered five more questions, concerning intention, appearance, probability, action–omission, and a check on the manipulation of directness.

Two representative scenarios are presented below, with the questions and response options—identical for all scenarios—given only after the first. (In the actual questionnaire, each response option was represented as a separate button that a subject could click to indicate his or her choice). The option we counted as indirect is marked with an * (not in the questionnaire itself, nor was “Morality question”). The dilemmas were presented in a random order, randomized separately for each subject. The responses were made with buttons, and the subject could not continue without answering all questions. The subject had to choose one button for each question.

Scenario 1: Organ Transplant

X is in charge of a computer database controlling the distribution of available organ transplants. The first person in line for a difficult-to-get liver transplant is Mr. Y. Mr. Y was an alcoholic, and his drinking ruined his liver. Y no longer drinks. The rules say that past alcohol use should not be considered, but X still thinks that

Y should not get priority, so he decides to break the rules and prevent Y from getting the next liver. He can do this in two ways:

- *A. X can raise everyone else's priority score by 20 points.
- B. X can lower Y's priority score by 20 points.

In either case, X is sure that the database program will rearrange the list, moving Y to the end of it, and that nobody will find out what he has done.

[Morality question:] Which option is more wrong, morally?

- A is much more wrong
- A is a little more wrong
- B is much more wrong
- B is a little more wrong
- Equal

Directness. Something is a *direct* cause of harm if the harm is part of the action itself. Which of the following is more true about X's behavior leading to harm?

- More direct in A, and that makes A more wrong than B.
- More direct in A, but that does not make A more wrong than B.
- More direct in B, and that makes B more wrong than A.
- More direct in B, but that does not make B more wrong than A.
- Equally direct in A and B.

Intention. Which of the following is more true about X's behavior leading to harm?

- More intentional in A, and that is why A is more wrong than B.
- More intentional in A, but that does not make A more wrong than B.
- More intentional in B, and that is why B is more wrong than A.
- More intentional in B, but that does not make B more wrong than A.
- Equally intentional in A and B.

Appearance. Which of the following is more true about how X's behavior appears to others?

- Appears more wrong in A, and that makes A more wrong than B.
- Appears more wrong in A, but that does not make A more wrong than B.
- Appears more wrong in B, and that makes B more wrong than A.
- Appears more wrong in B, but that does not make B more wrong than A.
- Appears equally wrong (or not wrong) in A and B.

Action Vs. Omission. Which of the following is more true about X's behavior leading to harm?

- More from an action (vs. omission) in A, and that makes A more wrong than B.
- More from an action in A, but that does not make A more wrong than B.

- More from an action in B, and that makes B more wrong than A.
- More from an action in B, but that does not make B more wrong than A.
- Equally from an action in both A and B.

Probability. Which of the following is more true of X’s behavior?

- More likely to cause harm in A than in B.
- More likely to cause harm in B than in A.
- Equally likely to cause harm in A and B.

(At this point, the instructions read “You must answer all questions. You can change your mind before you click to go on.”)

Scenario 2: Zoo

A zoo has been created to conserve 200 species of wild animal that have become extinct elsewhere. The zoo is now threatened with a parasitic disease that infects the animals. X, the zookeeper, has two options:

- A. Painlessly poison the animals in which the parasite reproduces, thus saving the other animals. Five species will become extinct.
- *B. Poison the parasites. The same poison will cause five animal species to become extinct. In both cases, X is sure that he will save most of the species and lose five. The five lost are equally valuable in the two cases.

Results and Discussion

Subjects considered the direct option to be wrong more often. They judged the direct option more immoral 26.2% of the time, and the indirect option 9.5% ($t_{53} = 5.12, p = 0.0000$, one-tailed; across the eight cases, $t_7 = 3.47, p = 0.0052$, one-tailed).

Table I shows the proportions of responses to each question, and the correlation of each question with the morality question, over all cases and subjects (hence 8×54 observations). The morality question was coded as 1 when the direct action was judged to be worse, -1 when the indirect action was judged worse, and 0 for “equal.” The other questions were coded 1 for the predicted response [e.g., that the direct act was more of an action or led to the more probability outcome], 0 for “equal,” and -1 for a response opposite from that predicted. Questions about intent, etc. were divided into two parts, according to whether the subject indicated that the property in question was a reason for the moral judgment or not. The answer could count for only one part. For example, a difference in intent could not be both a reason and “not a reason.” If subject chose “More intentional in A, and that is why A is more wrong than B,” and if A was the direct action, then the response was coded as favoring the hypothesis for “Intent, reason,” and neutral for “Intent, not a reason.”

Table I. Proportions of Responses to Each Question, According to Whether or Not It Was Cited as a Reason for a Moral Judgment (Except for Probability), and Its Correlation With the Morality Question Across All Observations

Question	Predicted	Opposite	Correlation
Probability	0.120	0.049	0.467
Intent			
Reason	0.150	0.035	0.649
Not a reason	0.174	0.051	0.099
Appearance			
Reason	0.157	0.032	0.609
Not a reason	0.271	0.063	0.157
Omission			
Reason	0.160	0.028	0.553
Not a reason	0.222	0.065	0.092
Directness			
Reason	0.169	0.039	0.698
Not a reason	0.324	0.056	0.012

It is apparent that all questions correlated in the expected direction. (They also correlated with each other. All correlations in Table I above 0.092 are “significant” at $p < 0.05$, although the correlations reported are presented for descriptive purposes only.)

Subjects did not always agree with our judgments about which option was more direct. They agreed with us in 49.3% of all cases (including cases in which directness was a reason for moral judgment and cases in which it was not) and disagreed in 9.5%. In subsequent analyses, we consider only those cases in which subjects agreed with us about which option was more direct (the 49.3%).

The design of this study allowed us to ask whether the effect of directness on moral judgment could be explained in terms of intention, appearance, and probability of harm. To test this, we discounted all responses in which the subject indicated that the difference between the options in intention or appearance were seen as reasons for a judgment that one option was less moral, or in which the subject indicated that the bad outcome had a different probability. (That is, we assigned these cases, 71% of the total a response of “equally immoral,” whatever the subject said about their relative immorality.) For example, we discounted responses of “More intentional in A, and that is why A is more wrong than B,” or the reverse (A and B switched), but we did not discount responses that ended with “but that does not make A more wrong than B” (or the reverse).

Subjects now judged in direct option more immoral 4.9% of the time, and the indirect option 0.9%. This is a stringent test, and the results were much reduced in magnitude, but the direct option was still judged less moral, and the result was significant across subjects ($t_{50} = 3.24$, $p = .0011$, one tailed) and scenarios $t_7 = 6.26$, $p = .0002$).

In sum, the effect of directness on moral judgment cannot be fully explained in terms of the effect of directness on judgments of appearance, probability, or

intention that were thought to be relevant to moral judgment. In this tightly controlled experiment, the effect is small, possibly because directness did affect the perception of appearance, probability, and intention, but also possibly because subjects feel the need to justify their moral judgment in several ways, through a kind of “belief overkill” (Jervis, 1976).

STUDY 3

Study 2 left open the question of whether the indirectness bias is a manifestation of omission bias. Study 3 attempts to demonstrate both biases in the same study. Each scenario involves a decision with three options: an action leading to direct harm, an action leading to indirect harm, and an omission that leads to harm. Subjects compared the options two at a time. We hypothesize that, when the harm is equal, people will prefer the options in this order: omission, indirect, direct. We also expect that they will sometimes order the options this way even when the magnitude of harm leads to the opposite ordering. Such a result would imply that omission bias exists for actions that lead to indirect harm, but also that the bias is stronger when harm is direct.

Method

Sixty-nine subjects completed a questionnaire on the World Wide Web for \$2 each. The subjects were 70% female, 20% students, and their ages ranged from 17 to 69 (median 34).

The questionnaire, titled “Tragic choices,” began:

The following five stories concern the sorts of events that occur rarely in human history, happily. But they illustrate moral questions of more general interest. The stories concern bad events such as deaths. You get seven questions about each story. The first three concern one option. The second three concern another option. The last concerns both options. Within each group of three, you must be consistent. For example, you cannot say that the options are equally attractive no matter how many deaths they cause. But you need not try to be consistent between the questions about the first option and the second, or either of these and the last question. The options are of different types. Your answers may be different too.

There were five items in total. One item read as follows, and the questions were analogous for all the other items:

Scenario 1

Scientists planted a forest in order to preserve endangered species of trees and other plants. Most of the species in the forest are extinct everywhere else. The forest is now threatened by an infestation of insects. You are one of the scientists. If nothing is done, 10 species will become extinct.

Suppose that all you can do is spray the forest with a chemical that will destroy the plant species in which the insects make their nests, thus killing the insects and saving the other plants.

- Question 1: Would you do this if this action would cut the number of species extinctions in half? Yes No I would be completely indifferent
- Question 2: Would you do this if this action would lead to the same number of species extinctions? Yes No I would be completely indifferent
- Question 3: Would you do this if this action would double the number of species extinctions? Yes No I would be completely indifferent

Now suppose that all you can do is spray the forest with a chemical that will destroy the insects. The same chemical will kill some of the plant species.

- Question 4: Would you do this if this action would cut the number of species extinctions in half? Yes No I would be completely indifferent
- Question 5: Would you do this if this action would lead to the same number of species extinctions? Yes No I would be completely indifferent
- Question 6: Would you do this if this action would double the number of species extinctions? Yes No I would be completely indifferent

Now suppose you have two options:

1. You can spray the forest with a chemical that will destroy the plant species in which the insects make their nests, thus killing the insects and saving the other plants.
2. You can spray the forest with a chemical that will destroy the insects. The same chemical will kill some of the plant species.

Question 7: Either option would cut the number of species extinctions in half. Which would you choose?

- Option 1 Option 2 I would be completely indifferent

The subjects had to choose one of three response buttons after each choice: “Yes,” “No” (or “Option 1,” “Option 2”) or “. . . completely indifferent.” Questions 1–3 were designed to measure omission bias for the first option, which in this item is direct, 4–6, for the second (indirect) option. Questions 3 and 6 were included, along with 1 and 4, so that the group of questions did not bias the subjects for or against action. Within the three questions concerning omission bias, they subjects were not allowed to respond more favorably when the outcome was worse, and the subjects could not respond “indifferent” more than once.

Question 7 pitted direct against indirect. Option 1 involves direct harm, because the species are killed as a means to the end of saving other species. Option 2 is indirect because the species are killed as a byproduct.

We scored the responses as 0 for indifferent, 1 for the option that agreed with our hypothesis, and -1 for the other option. That is, we assigned 1 when the response to Questions 1–6 favored omission, or when the response to Question 7 favored the indirect option. Omission bias was thus the sum of the “no” responses minus the sum of the “yes” responses in each group of three: 1–3 and 4–6. Recall that subjects were not allowed to express indifference more than once, nor to say “yes” to an action with worse consequences than one to which they said “no.” Thus, the absence of bias toward action or omission would be “yes” to Question 1, “indifferent” to Question 2, and “no” to Question 3.

The five items were presented in random order. Two additional representative items the following (“Direct” and “Indirect” labels not presented to the subjects):

1. You are a U.S. military base commander. A missile has just been mistakenly fired from your base at a commercial airliner. If nothing is done, 100 passengers will die.

Indirect: You can alter the course of the commercial airliner. The airliner will be safe, but the missile will destroy another commercial airliner flying right behind the airliner.

Direct: you can alter the course of another commercial airliner flying nearby so that it is placed in the path of the missile. The first airliner will be safe, but the missile will destroy the airliner whose course you change.

2. There is an outbreak of a strange infectious disease on an island in the South Pacific. The disease causes cerebral hemorrhage and death in children under the age of 12. A child can carry the disease and pass it on to others for months, but it cannot be diagnosed until just before it becomes fatal, so the present carriers cannot be identified and quarantined. Many children on the island are already infected and beyond help. If nothing is done, many more children will become infected and die. You are a public health official.

Indirect: Release special “immunizing fumes” into the air. This will cause immunity in all those who are not yet infected. However, the fumes will kill all of the children who are already infected.

Direct: Release special “lethal fumes” into the air. This will kill all of the children who are already infected, checking any further spread of the disease and saving the still uninfected children.

Results and Discussion

Both hypotheses were supported. We found omission bias when omission was pitted against the direct option. The mean omission bias over all three questions was 0.22 (where 1 would indicate favoring omission on every question and 0 would indicate neutrality toward acts and omissions). This was significant across subjects

($t_{68} = 5.77$, $p = 0.0000$ —in all cases t tests were “paired”) and across the five scenarios ($t_4 = 3.06$, $p = 0.0377$). The mean omission bias for the indirect option was 0.07 ($t_{68} = 2.09$, $p = 0.0407$, across subjects, $t_4 = 1.50$, ns, across scenarios).

Omission bias was greater for direct action than for indirect action ($t_{68} = 5.17$, $p = 0.0000$, across subjects; $t_4 = 3.51$, $p = 0.0247$, across scenarios). Finally, subjects also showed indirectness bias by favoring indirect action in the question that pitted the direct option against the indirect option (mean of 0.15, where 0 represents indifference and 1 represents favoring indirect action every time; $t_{68} = 5.10$, $p = 0.0000$, across subjects; $t_4 = 2.3227$, $p = 0.0809$, across scenarios).

The measure of indirectness bias based on Question 7 was not significantly correlated with the mean of omission bias across subjects ($r = 0.11$). However, this measure of indirectness bias was correlated with the difference between omission bias for direct actions minus that for indirect actions ($r = 0.44$, $t_{67} = 3.98$, $p = 0.0002$). Looking at this another way, the correlation between indirectness bias and direct omission bias was 0.27 ($p = 0.0272$), while that between indirectness bias and indirect omission bias was -0.10 (ns). (The correlation between indirect omission bias and direct omission bias was 0.64.)

We could summarize the results by saying that omission bias for indirect actions is very small. Omission bias exists largely for direct actions. *It is apparently the dislike of causing harm through direct action that seems to account for the indirectness bias and for most of the omission bias.* Individual differences in this dislike account for individual differences in both biases.

GENERAL DISCUSSION

We found evidence for a new bias in judgment, a tendency to favor indirectly harmful options over directly harmful options. Studies 1 and 2 showed this bias when consequences were the same. Study 3 showed that omission bias exists largely for direct actions.

Perhaps, the indirect bias is entirely for direct actions, since even the slight evidence for indirect omission bias may be accounted for by the fact that some of our putative cases of indirect action incorporate elements of direct harm. For example, some unintended side effects (as the anticipated dying of the infected children, the inevitable byproduct of releasing the “immunizing fumes” in the “infectious disease” scenario) may be seen as directly caused harm. Possibly, both “omission bias” and “indirect harm bias” are reducible to a single preference toward *not* causing harm through direct action, with individual differences in this preference accounting for individual differences in both biases.

How can we explain this preference? One possible explanation of most of our results depends on the distinction between harm as a mediator and harm as a side effect. In the direct option of the “infectious disease” scenario, for example, the act (releasing the “lethal fumes”) causes harm, and the harm, in turn causes the

desired outcome (the fumes kill the sick children, thereby checking further spread of the disease). In the indirect option of the same scenario, the act (releasing the “immunizing fumes”) causes both the desired outcome (checks further spread of the disease) and the harm (killing the sick children as a side effect), but the harm is not instrumental to the desired outcome.

In the first (direct) case, it is difficult to imagine getting rid of the harm and keeping the desired outcome. In the case of side effects, we can easily imagine getting rid of the harm and preserving the causal link from the act to the desired outcome. Although we can imagine it, the situation does not allow it. Because it is imaginable, though, the harm is less a part of the act, and therefore less connected with the decision to act. Notice that this is a psychological distinction, a matter of how we perceive things rather than a matter of how they are. For example, in Bennett’s case of abortion vs. hysterectomy to save the mother, we must think of the death of the fetus as a bad event in the abortion, for it is necessary and part of the cure. But in hysterectomy, we must imagine that the uterus could be removed but the fetus somehow put back and allowed to grow. Still, this may be how we think of it, even though this is impossible.

Of course, there is an alternate account. In light of widespread inhibitions against intraspecific aggression among various animal species (Lorenz, 1967; Midgley, 1995), we may posit (a) that humans have evolved a repertoire of specific aggression-inhibitory mechanisms whose activation produces a sense of “moral wrongness”; (b) that this repertoire is activated by a variety of specific cues, (c) that one such a cue is the perception of acting directly and in an injurious fashion upon the body of a conspecific (such a mode of inhibition would, presumably, serve us quite well in the technologically pristine environment of the Pleistocene, that is, in an environment devoid of mechanical weapons, where harming another person is virtually synonymous with causing direct damage to another’s flesh), and (d) that our capacity for symbolic thought allows us to process a situation in which another’s body is *not* the direct object of our action (in the database scenario, a person’s name serves as a standin for the person herself) as one in which the inhibitory mechanism acts as if it were such an object. On this view, the harm through omission and causally indirect harm both work to palliate the sense of moral wrongness by allowing us to ensure the production of harm without requiring us to operate directly on the harmed person or her symbolic representation (e.g., the person’s name). And this may be the common, even if implicit, psychological significance of the two types of indirect harm to the experimental participant. Further work is needed to test this possibility.

The indirectness bias is similar to the status quo bias or endowment effect (Kahneman *et al.*, 1990; Samuelson and Zeckhauser, 1988), and to omission bias. These effects can be seen as resulting from lay deontological principles, that is, principles that favor one option over another, on the basis of general properties of actions, even when the consequences are the same. More generally, all of these biases may result from over-application of heuristic principles that are reasonable

in many situations. People may overapply these principles because they forget, or never understood, the relation between the principles and their underlying purpose of doing good and preventing harm (Baron, 1994).

These principles express themselves in the historical development of public morality, so each person's adherence to them is, we suspect, partly a result of cultural transmission. It seems unlikely, though, that cultural transmission is their only source. More likely they originate in individual attempts to generalize very basic moral injunctions to a broader class of cases. When parents scold a child for a harmful act such as hitting a sibling, the child may not see the wrongness of the act as just a matter of choosing an option that causes harm. Rather, the child may focus on other properties of the choice made, such as its activeness, the fact that it is a change from the status quo (since the sibling was not in pain before), and the fact that it is direct.

We should note that our results are limited to one kind of task, in which the subjects make explicit comparisons of hypothetical cases presented side by side. We did not look at choices with real consequences. However, the fact that indirectness seems relevant to written laws and rules (such as the case of Catholic hospitals described in the introduction) suggests it would influence real consequences.

We also did not compare direct and indirect options presented at different times or to different subjects. Our results therefore concern people's explicit views of how they ought to make judgments rather than how they would in fact make judgments about cases presented one at a time. Choices in which indirect and direct options are part of the same choice set are quite rare (the hospital example being one of the few we know). Our results are thus most relevant to questions about explicit policy formulation. Usually, though, when people show explicit effects such as those we have found, they also show similar effects when presented with one option at a time (although such effects may be more difficult to detect because of increased variability).

A number of further questions can and should be asked in regard to the reported findings. These may include the following: Is the indirectness bias found when options are presented one at a time? Would it hold when the consequence is not harmful? Would it reverse for benefit? Would it be found if the harm is fully justified (rather than unfortunate and unavoidable), as in the case of punishment? Would the findings hold for children and for people raised in very different cultures?

Also, insofar as the indirectness bias seems similar to the omission bias, we may ask if it is mediated by similar processes. There is evidence that regret, at least short-term regret (e.g., Zeelenberg *et al.*, 1996), is greater for actions than inactions (e.g., Baron and Ritov, 1994), and, insofar as people's decisions seem to be informed by the avoidance of anticipated regret (Zeelenberg *et al.*, 1996), people's preference for omissions over equally harmful commissions may be motivated by the greater anticipated potency of action regret (Baron and

Ritov, 1994; Ritov and Baron, 1990). Although we know of no empirical evidence or argument suggesting that indirectly harmful actions (including those with clearly malevolent intentions) are less regrettable than (equally malevolent) directly harmful actions, we believe this to be a fruitful topic for future research.

Our results have implications for other issues. One of these is the debate over the distinction between killing and letting die, as expressed in the controversy over active euthanasia (Steinbock, 1995). Philosophical and legal discussions of such issues often make use of hypothetical scenarios of the sort we have used (e.g., Bennett, 1966; Davis, 1995; Trammel, 1995). Some of these scenarios involve killing or harming one person in order to save or prevent harm to others. For example, consider the following two cases:

1. A: We have only five grains of a scarce medicine. We can use it to save five people, each of whom needs only a one grain dose, thereby allowing the death of one person who needs a five grain dose.
 B: We can save five people if [and only if] we kill a sixth person and make a medicine out of his body (adopted from Davis, 1995).
2. A: There is a runaway train headed in the direction of one person presently walking on the tracks. There are also five other people on a bridge above the tracks. By throwing a switch you can cause the train “to jump onto tracks on the bridge,” thus, saving one person below but killing the five on the bridge tracks.
 B: Everything is the same, except that now you find yourself standing on a bridge “watching a trolley hurtling” toward five innocent people, and there is an incredibly fat man standing right next to you. By pushing the man onto the tracks below you can impede the further progress of the trolley, thus, saving five innocent people (adopted from Petrinovich and O’Neill, 1996).

We can see that in both of these classic killing/letting die scenarios “the killing” option (part B) requires us to treat a victim as a mere means as a walking supply of lifesaving chemicals in B of 1 or, as a roadblocking device in B of 2. It is this requirement that, one can argue, contributes greatly to our possibly seeing the commission in part B of all of the above scenarios as more morally problematic than the omissions in part A.

If we are correct that indirectness bias contributes to these intuitive judgments, then moral and legal discussion must address the moral and legal relevance of indirectness. Utilitarian philosophers, of course, would find it irrelevant in itself, for they are concerned only with expected consequences.

A second issue is the harnessing of our preference for indirect harm in the service of mass evil. Milgram (1974) was plainly aware of the possibility that most

people find it morally easier to inflict harm indirectly than directly and held it to be one of the major reasons for the scrupulous division of labor found in many a modern bureaucracy:

It is typical of modern bureaucracy, even when it is used for destructive purposes, that most people involved in its organization do not carry out any destructive actions. They shuffle papers or load ammunition or perform some other act which, though it contributes to the final destructive effect, is remote from it in the eyes and mind of the functionary. (p. 121)

These considerations led Milgram (1974) to predict that

Any force that is placed between the subject and the consequences of shocking the victim, any factor that will create distance between the subject and the victim, will lead to a reduction of strain on the participant and thus lessen disobedience. (p. 121)

To test this prediction, Milgram carried out a version of his basic obedience procedure in which “the act of shocking was removed from the naive subject and placed in the hands of another participant (a confederate),” with the naïve subject passing the information from another location. This manipulation resulted in a substantial increase in obedience compared to the basic condition where the subject himself was depressing the lever (Milgram, 1974, exp. 18).

Although both Milgram’s procedure and his explanation for its real-world analog confound indirect bias proper with proximity, we concede his fundamental point that our greater moral tolerance toward the infliction of indirect harm can be utilized in the service of mass evil.

A third issue is deterrence in general, and capital punishment in particular. Deterrence can be seen as harming one person in order to prevent harm to others. The criminal is punished in order to deter other potential wrongdoers. Such is the utilitarian account of punishment. But people may resist this understanding because it is the result of direct harm in order to bring about a good end. The aversion to the idea of direct harm may cause people to reframe punishment as retribution, the restoration of balance rather than a harm committed for a beneficial end.

In using a person as a deterrence-object, in capital punishment, one is essentially reducing him or her to the level of a special device that the Law and the State may wield in their campaign of homicide prevention. Consequently, the concept of capital punishment merely as a form of “retributive justice” may be more appealing (Rachels, 1986) not so much because of people’s inherent lust for revenge, but because, unlike the utilitarian justification of executions, it doesn’t demand the violation of the intuitive injunction against instrumental harm in its most basal form.

Finally, our admissions scenario suggests an interpretation of the debate between those for and against affirmative action, or provision of free care to the poor by hospitals, or anything that involves providing a benefit to someone when the total amount of benefits is limited. In these cases, provision of the benefit to one person means denying it to someone else. Those favoring the extra benefit may see the harm to others as indirect, hence not so bad. Some of those who oppose

the benefit may think in terms of consequences alone, pointing out that the limitation on total benefits due to special treatment would have the same effect as a direct denial to whoever would get the benefit in place of the person given special treatment.

ACKNOWLEDGMENTS

Thanks to John Sabini, Paul Rozin, Lila Gleitman, and the graduate students attending the 1997 beer and cheese seminar for useful comments. This work was supported by National Science Foundation Grant No. SES 9876469.

REFERENCES

- Aquinas, T. (1947). *Summa Theologica* (translated by the Fathers of the English Dominican Province), Benziger Brothers. Retrieved from <http://www.newadvent.org/summa/>
- Baron, J. (1994). Nonconsequentialist decisions (with commentary and reply). *Behav. Brain Sci.* 17: 142.
- Baron, J. (1995). Blind justice: Fairness to groups and the donoharm principle. *J. Behav. Decis. Making* 8: 7183.
- Baron, J. (1998). *Judgment Misguided: Intuition and Error in Public Decision Making*, Oxford University Press, New York.
- Baron, J., and Ritov, I. (1994). Reference points and omission bias. *Organ. Behav. Hum. Decis. Process.* 59: 475–498.
- Bennett, J. (1966). Whatever the consequences. *Analysis* 26: 83–102.
- Birnbaum, M. H. (1999). Testing critical properties of decision making on the Internet. *Psychol. Sci.* 10: 399–407.
- Birnbaum, M. H. (ed.) (2000). *Psychological Experiments on the Internet*, Academic Press, New York.
- Davis, N. A. (1995). The priority of avoiding harm. In Steinbock, B., and Norcross, A. (eds.), *Killing and Letting Die*, 2nd edn., Fordham University Press, New York, pp. 301–302.
- Hsee, C. K., Loewenstein, G. F., Blount, S., and Bazerman, M. H. (1999). Preference reversals between joint and separate evaluations of options: A review and theoretical analysis. *Psychol. Bull.* 125: 576–590.
- Jervis, R. (1976). *Perception and Misperception in International Politics*, Princeton University Press, Princeton, NJ.
- Kahneman, D., Knetsch, J. L., and Thaler, R. H. (1990). Experimental tests of the endowment effect and the Coase theorem. *J. Pol. Econ.* 98: 1325–1348.
- Kant, I. (1983). Grounding for the metaphysics of morals. In Kant, I. (ed.), *Ethical Philosophy* (Ellington, trans.), J. W., Hackett, Indianapolis, IN. (Originally published 1785)
- Kuhse, H. (1987). *The Sanctity of Life Doctrine in Medicine: A Critique*, Oxford University Press, Oxford, UK.
- Lorenz, K. (1963). *On aggression* (M. K. Wilson, trans.). Harcourt, Brace & World, New York.
- McGraw, K. O., Tew, M. D., and Williams, J. E. (2000). The integrity of Web-delivered experiments: Can you trust the data? *Psychol. Sci.* 11: 502–506.
- Meszaros, J. R., Asch, D. A., Baron, J., Hershey, J. C., Kunreuther, H., & Schwartz-Buzaglo, J. (1996). Cognitive processes and the decisions of some parents to forego pertussis vaccination for their children. *J. Clin. Epidem.* 49: 697–703.
- Midgley, M. (1995). *Beast and man: The roots of human nature*. Routledge, New York.
- Milgram, S. (1974). *Obedience to Authority*, Harper and Row, New York.
- Petrinovich, L., and O'Neill, P. (1996). Influence of wording and framing effects on moral intuitions. *Ethol. Sociobiol.* 17: 145–171.

- Quinn, W. (1989). Actions, intentions, and consequences: The doctrine of double effect. *Philos. Public Aff.* 18: 334–351.
- Rachels, J. (1986). *The Elements of Moral Philosophy*, McGraw Hill, New York.
- Ritov, I., and Baron, J. (1990). Reluctance to vaccinate: Omission bias and ambiguity. *J. Beh. Decis. Making* 3: 263–277.
- Samuelson, W., and Zeckhauser, R. (1988). Statusquo bias in decision making. *J. Risk Uncert.* 1: 759.
- Spranca, M., Minsk, E., and Baron, J. (1991). Omission and commission in judgment and choice. *J. Exp. Soc. Psychol.* 27: 76–105.
- Steinbock, B. (1995). Introduction. In Steinbock, B., and Norcross, A. (eds.), *Killing and Letting Die*, 2nd edn. Fordham University Press, New York.
- Trammell, R. (1995). Saving life and taking life. In Steinbock, B., and Norcross, A. (eds.), *Killing and Letting Die*, 2nd edn., Fordham University Press, New York, pp. 290–298.
- Zeelenberg, M., Beattie, J., van der Plicht, J., and de Vries, N. K. (1996). Consequences of regret aversion: Effects of expected feedback on risky decision making. *Organ. Behav. Hum. Decis. Process.* 65: 148–158.

Copyright of Social Justice Research is the property of Kluwer Academic Publishing and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.