

Protected values and omission bias as deontological judgments

Jonathan Baron*

Ilana Ritov

University of Pennsylvania

Hebrew University of Jerusalem

July 1, 2008

DRAFT of invited article for D. Medin, L. Skitka, D. Bartels, & C. Bauman (Eds.), *Psychology of Learning and Motivation*, Vol. 50.

*Email: baron@psych.upenn.edu. We thank Dan Bartels and Doug Medin for discussion about the design of one experiment and for comments on two drafts. The research was supported by a grant from the U.S.-Israel Bi-national Science Foundation to the authors.

Contents

Introduction	5
Biases as deviations from utilitarianianism	5
Overview	7
Protected values (PVs)	8
Relation of PVs to other types of judgment	8
Moralistic values	8
Moral realism	9
Authority independence (vs. social norms or conventions)	9
Omission bias	10
Relation of omission bias to other biases	12
Default bias and status-quo bias	12
Indirectness and the double effect	12
Agent relativity	13
Naturalism	14
Physical proximity and contact	14
Study 1: Relation of PVs to omission bias	14
Method	16
Results	18
Discussion	20
Study 2: Relation to emotion	21
Method	21
Results	25
Follow-up	26
Study 3: The nature of omission bias	27
Method	28

<i>Protected values and omission bias</i>	3
Results	32
Conclusion	34

Abstract

We review the major findings concerning omission bias and protected values (PVs). PVs are values that are absolute, hence protected from trade-offs with other values. We argue that PVs against omissions are relatively rare, since a prohibited omission would be an injunction to act, which could create infinite obligations, and we provide support for this argument. We also replicate and extend earlier findings of a correlation between PVs and omission bias, the bias to favor harms of omission over equal or greater harms from action. We discuss the nature of omission bias and its relation to other biases. We also find that, although emotional responses are correlated with biases, we can manipulate apparent emotional responses without affecting PVs or omission bias. We thus argue that emotions are not the only cause of biases.

Introduction

The study of heuristics and biases in judgments and decisions, begun by Daniel Kahneman and Amos Tversky, has been extended to judgments and decisions that count as moral, if only because they concern choices that affect others aside from the decision maker. The main focus of this chapter is the bias that favors harms of omission over equal or lesser harms from action, and its relation to values that are taken to be absolute, that is, protected from trade-offs with other values. The bias favoring harmful omissions — omission bias — by itself is found in both moral and non-moral judgments and decisions. Indeed, an early study (Ritov & Baron, 1990) found essentially no difference between cases that involved decisions for others and those that involved decisions for the self. But protected values seem more uniquely moral, although they need not be so.

Biases as deviations from utilitarianism

In what sense are these biases?

Our concern is decisions about what to do, and judgments of such decisions, such as “which option is a better choice.” The field of “moral judgment” includes much more than such decisions. We regard judgments about permissibility, praise, blame, character, virtues, and other judgments as derived, not fundamental.

Given this focus, it is natural to think about moral judgment as a type of decision making. A common approach to the study of decision making is to compare people’s judgments and choices to normative models. A normative model provides a standard for evaluation of judgments. For example, the psychological study of errors in logical syllogisms has used formal logic as a normative model for over 100 years. Judgments of logical syllogisms depart from this normative model, and much research has tried to explain these departures.

In decision making, a natural normative model is utility theory, which holds that the best options are those that maximize the utility of their consequences, or the expectation of that utility taking probabilities into account. “Utility” is simply the amount of goodness or “good” (Broome, 1991). The concept assumes that goodness can be compared across different types of outcomes, or even outcomes that affect different people. (See Baron, 1993, and Hare, 1981, for a defense of some of these assumptions.) Note that utility is an interval scale, like time, not a ratio scale like most physical measures; it has no natural zero point. All we need to make decisions are differences, because we always compare one option to some other option.

When this model is applied to decisions that affect many people, it amounts to utilitarianism. Utilitarianism is taken to be a theory of moral decision making because it applies to decisions that affect other

people aside from the decision maker. It thus conflicts with other moral theories, which are loosely grouped under the heading of “deontology.”¹ Deontological theories concern properties of options other than their consequences, such as whether they involve action or omission and whether they bring about consequences directly or indirectly.

One other property that is relevant in some deontological theories is intention or motivation. We regard intention as irrelevant to the question of what to do, because that question is asked by a decision maker who presumably wants to form an intention to choose a particular option and is asking for advice from the theory. Yet, in experiments about judgments of options, we must be careful not to let subjects’ judgments be contaminated by their judgments of intention, so this is something that require careful experimental control. (Such control is very difficult in between-subject designs. The easiest way to control intention is to give a background story that explains the intention before asking about different ways of implementing it, as done by Spranca et al., 1991.)

Utilitarianism is controversial because of these conflicts. Deontological theories have a long history and are sometimes more consistent with moral intuitions than utilitarian theories. We want to suggest here that at least some of these intuitions are the result of cognitive biases that have been found elsewhere in the study of judgment and decision making. Indeed, one of the biases of interest — the bias toward harms of omission over smaller harms of taking action — is found in judgments and decision that affect the self only, as well as those that affect others. It is even possible that reliance on intuitions in the development of philosophical normative theories is subject to these biases, which then become systematized into the theories (as suggested by Greene, 2007, and Greene and Baron, 2001).

The study of biases away from utilitarian judgment might be useful even if utilitarianism is an incorrect moral theory. Let us assume the “simple-effects hypothesis”: decisions made by trying to follow principle P usually yield outcomes more consistent with P than are the outcomes of decisions made in other ways. In particular, decisions that try to produce the best consequences usually do just that. Conversely, and most importantly, decisions made on the basis of deontological principles usually lead to results that are not as good as the best that could be achieved. If “usually” means “almost always,” then the study of biases away from utilitarian decisions can help us understand how consequences — things that matter to us — might be

¹We ignore, for simplicity, theories that are non-utilitarian but still consequentialist in the sense they evaluate decisions in terms of goodness of consequences but not in terms of total good or expected good. And we ignore the fact that some types of moral theory are neither consequentialist nor deontological, e.g., because they are concerned with virtues rather than with particular decisions. Deontological theories are the ones that are directly opposed to utilitarianism.

improved by making decisions differently. If such improvement results in violating some true (deontological) principle of morality, and if we follow that principle, then at least we will know the cost of following it.

Note that utilitarianism could be incorrect in a more interesting way than simply being inconsistent with some moral truth. It could turn out that trying to make utilitarian decisions is self-defeating (Parfit, 1984). We might maximize utility by deciding never to commit adultery or to kill innocent civilians in a terrorist attack, rather than by reflecting on the costs and benefits of such options. But it could also be that we would never do these things even if we *did* weigh the costs and benefits, so long as our reflection took into account the probability of error in our own judgments (Baron, 2008). The utilitarian model would have to include the consequences of such errors. This argument strengthens the appeal of the simple-effects hypothesis by showing that possible exceptions to it may not be exceptions at all.

Utilitarianism is often criticized because of the “excessive calculation” required to consider all the probabilistic outcomes and their effects on different people. Yet, in practice, in the real world, “trying to maximize utility” can amount simply to asking oneself which option would maximize utility. If two options are so close in the practical decision maker’s mind that calculation seems to be required, then choosing the wrong one would probably result in little utility loss. A much bigger loss could result from following a deontological principle that favors an option that is obviously inferior to the option that seems to maximize utility.

Overview

In this chapter, we shall focus on a few biases (away from utilitarianism) that are inconsistent with utilitarianism because they are driven by deontological intuitions or principles. We have no definitive list of such biases, in part because they are defined negatively, as simply principles that refer to properties of options other than their expected utility. The ones we describe are those we have studied. We shall try to give a psychological account of the nature of these biases. This account will be incomplete, but we think we have made some progress in the last 20 years (since the research described by Ritov & Baron, 1990).

We shall first discuss what we called “protected values” (PVs) because they are protected from trade-offs with other values (Baron & Spranca, 1997). We define PVs according to this property of absoluteness, but we find that they tend to have other properties, and we speculate about why this is so. One property is that they tend to involve prohibitions of actions rather than “prohibitions of omissions” (i.e., injunctions to act). This property implies that PVs ought to be associated with a stronger bias toward harmful omissions as opposed to less harmful acts, the bias we call omission bias. We then move to the discussion of omission bias

itself, which is, in turn related to other biases such as the status-quo bias and the bias toward indirect harm rather than direct harm. We conclude with some speculations about the origin and nature of deontological principles in general.

Protected values (PVs)

People claim to value some goods so strongly that they are unwilling to destroy any of these goods, no matter how great the benefits. Baron and Spranca (1997) called such attitudes “protected values” (PVs) because they were protected from trade-offs with other values. Many of these values involve morally questioned behaviors, such as cloning people for reproduction or destroying the natural environment. Other researchers have used the term “sacred values” for a similar phenomenon (Fiske & Tetlock, 1997; Tetlock, Lerner, & Peterson, 1996).

It appears that such rules are often overgeneralizations. When people are asked to try to think of counterexamples, cases in which the benefits would be great enough to justify taking the prohibited action, they can usually do so, and they change their mind about whether the rule is absolute (Baron & Leshner, 2000). Rules of this sort, if they were taken seriously, would cause great difficulty for evaluation of public policies through measurement of utility, because they would amount to infinite utilities and would thus allow one person to determine a decision for everyone — unless someone else had a conflicting rule, in which case the choice could not be determined. People who had more than one such rule — and many people say they have several — could find themselves in a similar dilemma. We thus consider PVs to be a form of bias, even though the argument against lexical preferences is not theoretically strong.

Relation of PVs to other types of judgment

PVs are closely related to other types of values, which express themselves in other ways.

Moralistic values

Moralistic goals or values are those that people want others to follow, regardless of whether the others endorse the same values and regardless of whether the consequences are, on the whole, worse as a result.

Baron (2003) found that people endorsed moralistic goals for banning actions like the following:

- testing a fetus for IQ genes and aborting it if its expected IQ is below average

- cloning someone with desired traits so that these may be passed on, such as an athletic champion or brilliant scientist
- modifying the genes of an embryo so that, when it is born, it will have a higher IQ
- giving a drug (with no side effects) to enhance school performance of normal children

In many cases (22% of examples like these), subjects (recruited on the World Wide Web) would ban these actions even if the consequences of allowing the actions were better on the whole than the consequences of banning them, if the subjects could imagine that the consequences might be better, and if “almost everyone in a nation thought that the behavior should be allowed.” In sum, they were willing to impose their moral(istic) principles on others, whatever the consequences, and whatever the others wanted. Moralistic values, taken seriously, are protected from trade-offs with values involving consequences, so they are likely to be thought of as PVs.

Moral realism

People differ in whether they think that moral judgments can be objectively true, in the way that “ $2 + 2 = 4$ ” is true, or in the way in which scientific knowledge can be true. The alternative is that moral judgments are subjective, more like tastes, and that they legitimately differ from person to person. Goodwin and Darley (2008) found that many adults believe that moral principles are objective or real in this sense. Piaget (1965) previously argued that moral realism was found in an early stage of moral development. When moral judgments are seen as subjective, they are less likely to be seen as absolute, other things being equal.

Authority independence (vs. social norms or conventions)

Turiel (1983) and others following in his tradition (e.g., Nucci, 2001) found that even children can distinguish moral rules and conventional rules. Turiel and his collaborators asked the children whether it would be okay to call one’s teacher by her first name if everyone thought it was okay, or to bow instead of shaking hands, or whether it would be okay for a bully to push another kid off the top of a slide if everyone thought it was okay. Even most second graders said it would be okay to change the conventions but not the moral rules: It would still be wrong to push someone off of a slide, even if everyone thought it was not wrong.

Bicchieri (2006) argues, in essence, for a distinction between two types of judgment that for Turiel and Nucci would both fall under the heading of convention. She would distinguish conventions and social norms.

True conventions are rules that can be followed without cost, and typically with some benefit, provided that enough others are following them. Examples of conventions are using the alphabet for ordering, driving on the right (or left) side of the road. It is to your advantage to drive on the right if everyone else is driving on the right.

Social norms require cost. They are like conventions in that they are contingent on the behavior of others. A person who embraces a social norm is willing to pay some personal cost to abide by it, provided that enough others are doing the same. Examples are norms of fashion or dress (some of which require considerable effort), norms of politeness, and codes of ethics (written or unwritten), such as those that forbid sexual relationships between students and teachers. A person who embraces a norm against such relationships will not regard the issue as fully moral (that is, applicable regardless of what people think and do) but will abide by the rule if it is expected and if others live up to the expectation. Moral rules also involve some cost, but people who endorse these rules think they should be followed regardless of what others think.

Omission bias

Some vaccines cause (in essence) the disease they prevent, but the chance of getting the disease is lower with the vaccine than without it. Note that, in this case and others like it, if we vaccinate, then we are hurting some people — those who will get sick from the vaccine — in order to help others, those who would get sick without it. We just don't know who will be hurt and who will be helped. If our moral intuition says that it is wrong to hurt some and help others, we might consider this from the perspective of the individual. For each affected person, the decision is between a risk of illness from the vaccine and a higher risk of illness without it. Expected utility would dictate that each person would want the vaccine. A refusal to provide the vaccine — on the grounds of not wanting to hurt some in order to help others — would reduce everyone's expected utility.

We can thus argue that intuition favoring harmful omission over less harmful acts is nonnormative even without defending all of utilitarianism (Baron, 1993, ch. 7). The simple argument is the Golden Rule. If you were a child, which would you prefer: a 10/10,000 chance of death from a disease or a 5/10,000 chance of death from a vaccine? Would it matter to you whether the chance resulted from an act or omission? More generally, each of us has reason to endorse a principle that others should make decisions so as to minimize our chance of death (other things being equal).

Spranca, Minsk, and Baron (1991) found a bias toward omissions in situations that were more obviously moral than those discussed so far. In one scenario, John, the best tennis player at a club, wound up playing the final of the club's tournament against Ivan Lendl (then ranked first in the world). John knew that Ivan was allergic to cayenne pepper and that the salad dressing in the club restaurant contained it. When John went to dinner with Ivan the night before the final, he planned to recommend the house dressing to Ivan, hoping that Ivan would get a bit sick and lose the match. In one ending to the story, John recommended the dressing. In the other, Ivan ordered the dressing himself just before John was about to recommend it, and John, of course, said nothing. When asked whether John's behavior is worse in one ending or the other, about a third of the subjects said that it was worse when he acted. These subjects tended to say that John did not cause Ivan's illness by saying nothing. (In reply, it might be said that the relevant sense of "cause" here concerns whether John had control over what Ivan ate, which he did.)

In sum, "omission bias" is the tendency to judge acts that are harmful (relative to the alternative option) as worse than omissions that are equally harmful (relative to the alternative) or even more harmful (as in the vaccination case) (Baron and Ritov, 1994). In any given case, some people display this bias and others do not.

Such a conclusion has strong implications. Most of us have the goal of not hurting people. If we combine this goal with the argument that the distinction between omission and commission is normatively irrelevant, then we must conclude that we ought to desire to help people as well. The reasons for not hurting people and the reasons for helping them are the same: We care about the achievement of their goals.

Omission bias is related to issues of public controversy, such as whether active euthanasia should be allowed. Most countries (and most states of the United States) now allow passive euthanasia, the withholding of even standard medical treatment for those who are judged to be no worse off dead than alive, but active euthanasia is almost everywhere banned even for those who wish to die. Opponents of active euthanasia can, of course, find other arguments against it than the fact that it is "active." But it is possible that these arguments would not be seen as so compelling if the distinction between acts and omissions were not made.

Omission bias could also justify a lack of concern with the problems of others (Singer, 1993). For example, much of the world's population lives in dire poverty today and into the foreseeable future. People — even people who take an interest in social issues — often think that they are not responsible for this poverty and need do nothing about it. It can be argued, however, that with a little effort we can think of all sorts of things we can do that will help the situation immensely at very low cost to ourselves, such as supporting

beneficial policies. Failure to do these things can be seen as a harm, but many people do not see it that way.

Relation of omission bias to other biases

Several other results in the literature overlap with omission bias, in the sense that they are confounded with it, in experiments and in the real world.

Default bias and status-quo bias

The status-quo bias is a bias toward the status-quo or keeping one's endowment (Kaheman, Knetch, & Thaler, 1990; Samuelson & Zeckhauser, 1988). It has been found repeatedly in many situations. In most studies of this bias, keeping the status quo requires no action, and changing the status quo requires action. Ritov and Baron (1992) found that the status-quo bias — the attachment to the status-quo that is common to all of these findings — was largely a consequence of a bias toward the default option. Specifically, when subjects were told that keeping the status-quo required action and that giving it up required inaction, subjects then favored giving up the status-quo. When both options required action, no preference for the status-quo was found, but when both options yielded new outcomes (neither one matching the status-quo), a preference for omissions was still found. Schweitzer (1994) and Baron and Ritov (1994) found both omission bias without a status-quo option. Thus, the status-quo bias as usually measured seems to consist of two different effects, a bias toward the default and (in some cases) a bias toward the status-quo.

The terms “omission bias” and “default bias” have been used interchangeably for various results. However, as originally defined by Ritov and Baron (1990), the omission bias is a bias toward *harms* of omission over equal or greater harms that result from action. The term “default bias” is used more generally for a bias toward the default (which is, of course, often the same as “doing nothing,” hence an omission).

Indirectness and the double effect

The indirectness bias is illustrated in the doctrine of the double effect. For example, when a mother's life is threatened by a pregnancy, some Catholic hospitals will permit a hysterectomy to save the mother, but they will not permit an abortion. The fetus dies in either case, but, in the case of the hysterectomy (which of course leaves the mother unable to bear another child), the killing is seen as an indirect by-product (Bennett, 1966; Kuhse, 1987). In the abortion, however, the death of the fetus is the means to save the mother, so

the fetus is being harmed directly. The indirectness bias is shown in the following scenario (Royzman and Baron, 2002):

A new viral disease is spreading rapidly in a region of Africa. Left alone, it will kill 100,000 people out of 1,000,000 in the region. X, a public health official, has two ways to prevent this.

Both will stop the spread of the virus and prevent all these deaths:

A. Give all 1,000,000 a shot that makes them immune to the first disease. The shot will also cause, as a side effect, a second disease that will kill 100 people.

B. Give all 1,000,000 a shot that gives them another disease, which is incompatible with the first disease. The second disease will kill 100 people.

Most subjects thought that option A was better, because the deaths are a side effect rather than part of the mechanism of the main effect. Later, we suggest that indirectness bias is related to omission bias in that both are affected by perceived causality.

Agent relativity

Agent relativity illustrated in the following scenario used by Baron and Miller (1999). X is one of ten people who could save someone's life by donating bone marrow (a painful but relatively risk-free procedure) to Y. Is X's obligation to donate greater when X is Y's cousin than when X and Y are unrelated? Many people think so. Utilitarians even think so, if they think that family cohesion is a good thing that should be promoted for other reasons. Now consider Z, who is unrelated to X or Y. X, the potential donor, asks Z's advice about whether to donate, and Z knows that X will probably follow the advice offered. Does Z have a greater obligation to advise donation when X and Y are cousins than when X and Y are unrelated? A utilitarian who answered yes to the first question would have to answer yes to this one. After all, it is promoting family loyalty that is at issue, and it doesn't matter whose family it is (without knowing more details, of course). An *agent relative* response, however, would say that only Y needs to worry about family obligations. The obligation is relative to the agent. It differs from person to person. Miller and Baron found no evidence for agent relativity in any of their subjects (who were Indian as well as American). However, many philosophers argue that some obligations are agent relative in this way. (See McNaughton & Rawling, 1991, for a review.)

Omission bias is agent relative when the omission is the result of someone else's action. In the classic case of shooting one prisoner to save ten from the horrible dictator, the choice of not shooting is implicitly agent

relative, because shooting will happen anyway. This is not a pure test of agent relativity, though, because the two options also differ in doing something versus doing nothing.

Naturalism

Naturalism is the bias toward nature. It is also related to omission bias, because “nature” often defines the default situation, the result of inaction, as in the case of the vaccination, where the disease can be assumed to be natural. (Of course, the result of omission is not always natural, as in the case of the dictator just described.) Rozin et al. (1994) found that people prefer natural food even if it is chemically identical to “artificial” food.

Physical proximity and contact

People consider harm to be worse when it involves physical contact with the victim. For example, people regard it as worse to “push a man off of a footbridge and in front of a train in order to cause the man to fall and be hit by the train, thereby slowing it and saving five people ahead on the tracks” than to “pull a lever that redirects a trolley onto a side track in order to save five people ahead on the main track if, as a side-effect, pulling the lever drops a man off a footbridge and in front of the train on the side tracks, where he will be hit” (Cushman, Young, & Hauser, 2006). Actions sometimes involve such contact, more than omissions.

Study 1: Relation of PVs to omission bias

Baron and Spranca (1997) proposed that PVs were not concerned so much with consequences as with the means of bringing about those consequences. In particular, they proposed that people would hold such values for acts but generally not for omissions (except in cases in which the decision maker is “responsible” for prevention of harm). It is possible for someone to hold many PVs for acts — do not clone, do not cut rainforests, and so on — and adhere to all these rules. Yet, a PV for an omission would require taking any action to prevent some bad outcome. It would be difficult to hold more than a few of these omission PVs consistently, as one’s whole life could be consumed with any one of them.

Ritov and Baron (1999; also Baron & Leshner, 2000) found that, indeed, PVs tended to correlate with omission bias when these trade off with harms of action. People who have PVs against the destruction of

animal species tend to be particularly reluctant to take actions that will destroy some species in order to save more species.

Ritov and Baron asked people, for example, how many species they were willing to destroy in order to save 20 species of fish that would become extinct if nothing were done. We found that those who had PVs against the destruction of species had lower thresholds, where the threshold is the maximum number that they were willing to destroy.

Arguably, an even better measure is the unwillingness to destroy *any* species in order to save more. If a PV is absolute, then this is its implication. We found that these *zero threshold* responses were also correlated with PVs. We might think that true PVs are always associated with zero responses. Alternatively, Baron and Leshner (2000) found that PVs were often not so absolute as they seemed to people at first, and people were willing to weaken them if the harm caused by action was very small relative to the harm that action prevented.

The PVs that Ritov and Baron (1999) examined were mostly those that involved acts. Only one or two (e.g., “letting people die from starvation”) were expressed in terms that were not actions. These non-action items also correlated with omission bias.

The present experiment tests the basic claim in another way, by asking PV questions that are explicitly about acts or omissions. The main hypothesis is that PVs for acts are more likely to be found than those for omissions. We suggest that this is for the reason proposed by Baron and Spranca, namely, the difficulty of living with more than one PV for omissions.

In addition, the experiment deals with a finding of Bartels and Medin (2007). They used a different way of measuring omission bias and found that it was negatively correlated with PVs, not positively correlated as found by Ritov and Baron. The measure, used by Connolly and Reb (2003)² is described in the method section below. The present experiment, like that of Bartels and Medin, used both of these measures of omission bias. It differed in two main ways. First, it was done on the World Wide Web, so the subjects were older and more varied than those used by Bartels and Medin. Second, the measure of PVs focused on morality rather than acceptability, following the idea that PVs are primarily moral judgments. Like previous studies, the design is all within-subject; the two measures of omission bias were counterbalanced, with the

²Connolly and Reb failed to find omission bias for vaccination decisions, using this measure. Baron & Ritov (2004) examine possible reasons for this failure and provide data replicating in a new way the omission bias for vaccination. This issue is not relevant here.

PV measure between them.

Method

Subjects were part of a panel who completed questionnaires on the Web for pay, recruited from a variety of sources beginning in 1995. From other studies in which we asked for demographic data, we found that the panel is roughly typical of the U.S. in age, income, and education (although less varied in income, given that the rich do not need the money and the poor do not have computers). Some panel members are not from the U.S. The majority are women. This study was completed by 170 subjects, 21% males, with ages from 18 to 67 (median 43). Each subject was paid \$4. (Six other subjects omitted because their mean response times were outliers on the fast side. Results were similar if they were included.)

The study began with a short introduction: “Each page describes some outcome that is morally controversial. We are interested in what you think of government policies that would affect the frequency this outcome. Most of these question involve choices between two bad outcomes. Please take seriously the possibility that such choices arise. . . .”

The study used two methods. In the RB method (Ritov/Baron), a typical item was:

If the government does nothing, 20 fish species will become extinct due to naturally changing water levels.

Building a dam would save 20 species of fish above the dam but cause 6 species to become extinct downstream.

Should the government build the dam? yes no

What is the largest number of species lost as a result of building the barrier at which the government should build the dam?

0 2 4 6 8 10 12 14 16 18 20

In the CR (Connolly/Reb) method, the last two questions were replaced with nine questions of the form, “Should the government build the dam if the number of species extinctions caused by the dam were 2? yes no.” The numbers went from 2 to 18, in steps of two, in this case. (For 91 subjects, there were ten questions, beginning with 0 instead of 2, to be more parallel to the RB version). The web questionnaire refused to accept responses with a yes answer to a higher number than one with a no answer.

In both conditions the numbers were chosen for the case. The initial level of the harms from action was always 30% of the harms of omission. The cases were as follows, with percent PV-omit in parentheses after each

- 10,000 people will die from a new form of flu. A vaccine that causes a weaker form of flu will prevent these flu deaths by creating immunity, but the vaccine itself will cause 3,000 flu deaths. (17%)
- 100,000 ancient redwood trees will be destroyed by natural fire, Intentionally setting smaller fires will prevent the loss of 100,000 ancient redwoods but cause 30,000 ancient redwoods to be destroyed. (3%)
- 20 fish species will become extinct due to naturally changing water levels. Building a dam would save 20 species of fish above the dam but cause 6 species to become extinct downstream. (7%)
- 1,000 emergency patients in government hospitals will suffer debilitating strokes. Giving a new drug to all emergency patients would prevent 1,000 debilitating strokes but would itself cause 300 debilitating strokes. (20)
- 100 natural miscarriages will occur in women with multiple pregnancies in government hospitals. Selective abortion of some fetuses will prevent 100 miscarriages, but the abortions themselves amount to 30 intentionally caused miscarriages. (29%)
- 500 airplane passengers will die from an accidental guided missile attack on a large jet. Directing a second jet into the path of the missile, to take the hit, would save the 500 passengers but cause 150 passenger deaths. (26%)
- 5,000 workers at a manufacturing plant will lose their jobs. Providing a tax break to this plant would save 5,000 jobs, but would cause the loss of 1,500 other jobs in a competitor's plant give the tax break. (16%)
- 200 people will die in car crashes in one stretch of highway. Building a by-pass to move traffic elsewhere would save 200 deaths in crashes, but cause the death of 60 others in the new area. (13%)
- 10 bird species will become extinct from natural causes, Building a barrier would prevent the extinction of 10 bird species but cause the loss of 3 species lost as a result of building the barrier. (8%)

The order of items was randomized for each subject. The same random order was used for both conditions. The order of the CR and RB sets was also randomized. In between these two conditions, the subjects answered PV questions about acts and omissions. The PV-act (a PV against causing through action) and PV-omit (against failing to prevent) questions were on the same page in alternating order (but the cases were

in the same random order as in the other conditions.) The two items, one for failing to prevent (PV-omit) and one for causing (PV-act) were of the following form:

FAILING TO PREVENT [CAUSING] THE EXTINCTION OF FISH SPECIES FROM NATURALLY CHANGING WATER LEVELS

What do you think of this outcome?

It is not a moral issue.

It is morally acceptable or even a good thing to do.

It is a moral issue, but I cannot say in general whether it is wrong or not.

It is morally wrong but should be allowed in most cases.

It is morally wrong and should not be allowed except in exceptional cases where it leads to a great benefit or prevents a great harm.

It is morally wrong and should not be allowed even if it leads to a great benefit, or prevented a great harm, no matter how great the benefit or harm.

Only the last option was counted as a PV.

Results

As hypothesized, PV-act (last option) was much more frequent than PV-omit: 15% vs. 8%, respectively ($t_{169} = 6.89$, $p = 0.0000$, across subjects; $t_8 = 3.29$, $p = 0.0110$, across the 9 cases). This difference was not significantly affected by sex, age, whether or not the zero item was presented in CR, and likewise for all other effects reported here. Table 1 shows the percent of PV-act for each item used. The difference was, however, affected by the order in which the PV-act and PV-omit questions were asked, in the direction of a carry-over effect from the first response to the second (mean difference of differences 5.5%, $t_{169} = 2.34$, $p = 0.0206$). When we looked only at the first response, the means for these responses were 18% and 7% for PV-act and PV-omit respectively ($t_{169} = 6.19$, $p = 0.0000$).

[Table 1 about here.]

The mean omission bias threshold measures were .33 for CR and .34 for RB (where 0 is the minimum threshold and 1 is the maximum). The proportion of zero thresholds (unwillingness to accept any effect of action) were likewise .33 and .34 for CR and RB, respectively. Although these levels are consistent with

other studies, the present study was not designed to test rigorously for omission bias (as done better by Baron & Ritov, 2004).

In the main analysis, we used mixed models as described by Baayen, Davidson, and Bates (2008). The dependent variables were the PV measures: PV-act and PV-omit (1 for PV, 0 for no PV). To predict these, using a logistic model, we included fixed effects for the subject means of RB threshold, the items means of RB threshold, the sum of the deviation of each response from its subject mean and the deviation of each response from its item mean (labeled RB and CR in Table 1 the order of the PV questions on the page, and the order of the CR and RB conditions for the subject).³ Subject and case were random effects.

Table 1 shows the estimated effects for eight different models. Thresholds are the predictors in the top half, and zero responses are the predictors in the bottom half. The first two rows in each half use RB, and the second half, CR. Dependent variables are PV-act and PV-omit in each pair of rows. It is apparent that both CR and RB threshold measures predicted protected values, with similar coefficients. Most of the “action” was in the variance due to cases (RB.vm and CR.rm). Figure shows the relationship between mean thresholds and PV proportions for PV-act and RB threshold, as an example. Still, the idiosyncratic effects of individual subject responses on individual cases, not accounted for by the subject means or the case means (CR and RB) also played a role. The individual differences among subjects (CR.sm and RB.sm) played a smaller role. The results for zero responses are similar to those for thresholds.

[Figure 1 about here.]

It is interesting, and not expected, that low thresholds are associated with PVs for omission as well as for acts. This result may be partly an artifact of the apparent spillover from the PV-act question to the PV-omit question, as indicated by the order effect for the order of these two items. There was also an order effect for which type of omission bias question (CR or RB) came first in the session.

³We also tested models that included the interaction between these order effects and CR and RB. One of tests was significant (for both threshold and zero responses), that between the effect of the CR condition and order of the CR/RB conditions for PV-omit: the correlation between low threshold and PVs was higher when RB came first. Because we can make no sense of this particular result, we present the results with no interactions. Qualitatively, all the results are the same with the interaction included.

Discussion

The main finding is that PV-act is more frequent than PV-omit. That is, people are more likely to think in terms of absolute prohibitions of action than “prohibitions of omission,” which amount to absolute injunctions to act whenever the opportunity arises. This result can be understood as a result of the difficulty of defining absolutes with respect to consequences alone, because such a rule would cover omissions as well as acts. We must assume that people tend to realize that they cannot be strongly obliged to take very many actions regardless of the costs. Opportunities to act in ways that save endangered species, etc., are almost always present.

The results also replicate those of Ritov and Baron (1999) in finding correlations between PVs and omission bias, and they conflict with Bartels and Medin’s finding that the correlation between PVs and omission bias reversed for the CR measure of omission bias for the threshold measure. (For the zero responses, the results are identical overall for CR and RB — 5 zero responses associated with PVs and 1 not — and, from the data presented in the paper, the overall association between zero responses and PVs, combining all the data, is significant and in the predicted direction.)

We see two possible reasons for the discrepancy, although we cannot explain, and did not replicate, the reversal of the effect for the CR threshold measure. First, our largest effects were the result of variation among cases within subjects. Bartels and Medin relied entirely on differences among subjects. It is possible that the use of several cases calls attention to differences among them. This would be consistent with the results of Kahneman and Ritov (1994; also Ritov and Kahneman, 1997), who found that joint presentation of items causes people to think differently about relative importance.

Second, their sample of items and subjects was quite different from ours. They used college students, and we used a possibly more varied group.⁴ Possibly as a result of this, their overall mean results were different. In general, they found much less omission bias: higher thresholds (roughly twice as high as ours) and far fewer zero responses (6%, compared to 33%). They also found more PVs, at least twice as many as we found.⁵ It may be that the zero responses in particular are doing much of the work in explaining the results

⁴Many of our subjects had never attended college, and they were mostly older. We tested for age effects and found no effects close to statistical significance in any of the studies reported here.

⁵These differences held up for the single item that was similar in both studies, which we called “fish” and they called “river diversion.” For this item, our subjects had 6% PVs (as opposed to 35%), mean thresholds of .40, .12, .39, .09 for PV/RB, no-PV/RB, PV/CR, and no-PV/CR, respectively (vs. .60, .34, .47, and .66), and zero responses of 26% for RB and 27% for CR (vs. 4% and 1%).

for thresholds, especially the subject differences in thresholds, and there were just too few of them in the Bartels/Medin study. Indeed, when we included both zero responses and thresholds in the same regressions of PV-act, we found no significant subject effect of thresholds for either the RB or CR measure (although most other effects remained significant).

Study 2: Relation to emotion

The present experiment again shows a correlation between omission bias and PVs, using mostly different hypothetical decisions.

Importantly, though, it adds questions about emotion, so that we can examine the correlation between emotion, PVs, and omission bias. We also included a manipulation intended to affect emotion without affecting the consequences of the decision. Specifically, we varied the cause of the bad outcomes in each scenario. One cause was relatively benign, usually natural, and the other was more malign, usually nefarious human actions.

Method

The 79 subjects ranged in age from 20 to 79 (median 42); 33% were male. (Three others were omitted because they did the study substantially more quickly than everyone else.) The introduction began:

Moral judgments

Each page describes some outcome that is morally controversial. We are interested in how you would make decisions that would affect the frequency this outcome, if the decision were up to you.

Most of these question involve choices between two bad outcomes. Please take seriously the possibility that such choices arise.

There are 22 pages, each with 5 questions.

An example of a page is:

If nothing is done, 1,000 emergency patients in government hospitals will suffer debilitating strokes FROM NATURAL CAUSES.

Giving a new drug to all emergency patients would prevent 1,000 debilitating strokes but would itself cause some debilitating strokes.

Now suppose that you have the following options:

A: do nothing; 1,000 debilitating strokes will occur.

B: give the new drug to all patients, causing at least 200 debilitating strokes, in order to prevent 1,000 debilitating strokes.

Try to imagine that these are your only options, and you must decide. How upset or disturbed do you feel?

Not at all A little Moderately Very much Extremely upset or disturbed

How angry do you feel?

Not at all A little Moderately Very angry Furious

How sad do you feel?

Not at all A little Moderately Very sad Despondent

Choose the first acceptable response:

I would give the new drug to all patients even if it led to 999 debilitating strokes.

I would give the new drug to all patients even if it led to 800 debilitating strokes.

I would give the new drug to all patients even if it led to 600 debilitating strokes.

I would give the new drug to all patients even if it led to 400 debilitating strokes.

I would give the new drug to all patients even if it led to 200 debilitating strokes.

I would not choose option B if it led to any debilitating strokes at all.

Which of the following best expresses your view about causing debilitating strokes with a drug?

This should be done if the benefits of doing it are greater than the harm.

This should not be done unless the benefits of allowing it are much greater than the harm.

This should never be done, no matter how great the benefits of doing it in some case.

Please write any comments on this page here (up to 255 characters):

The last question concerns PVs, with the third option representing a PV. The next-to-last question concerns omission bias. The bias was the position in the list of the item chosen, with the first counted as 0. The response was re-scaled so that its maximum was 1. The emotion measure was the sum of the responses to the three preceding items, rescaled between 0 and 1.

The full set of items was, with the situation and the two options for each item:

- 10,000 people will die from a new form of flu, WHICH AROSE THROUGH A NATURAL MUTATION [WHICH STARTED WHEN A CARELESS SCIENTIST ALLOWED A VIRUS TO ESCAPE FROM THE LAB].

A vaccine that causes a weaker form of flu will prevent these flu deaths by creating immunity, but the vaccine itself will cause [some] flu deaths.

- 100,000 ancient redwood trees will be destroyed BY NATURAL FIRE CAUSED BY LIGHTNING [BY FIRES STARTED BY A CAMPER WHO VIOLATED THE RULES OF A NATIONAL PARK].

Intentionally setting smaller fires will prevent the loss of 100,000 ancient redwoods but cause [some] ancient redwoods to be destroyed.

- 20 fish species in a river will become extinct DUE TO A NATURAL DECLINE IN THE WATER LEVEL [DUE TO A DECLINE IN THE WATER LEVEL, WHICH STARTED WHEN AN EXPLOSION IN A BOMB FACTORY CHANGED THE RIVER'S COURSE].

Building a dam would save 20 species of fish above the dam but cause [some] species to become extinct downstream.

- 1,000 emergency patients in government hospitals will suffer debilitating strokes FROM NATURAL CAUSES [AS A RESULT OF AN UNTESTED FOOD ADDITIVE].

Giving a new drug to all emergency patients would prevent 1,000 debilitating strokes but would itself cause [some] debilitating strokes.

- 100 miscarriages will occur in women with NATURAL multiple pregnancies in government hospitals [with multiple pregnancies INDUCED BY FERTILITY DRUGS WITH THIS EXCESSIVE RISK WITHHELD FROM DOCTORS BY THE COMPANIES].

Selective abortion of some fetuses will prevent 100 miscarriages, but the abortions themselves amount to [some] intentionally caused miscarriages.

- 500 airplane passengers will die from an ACCIDENTAL missile attack on a large jet, CAUSED BY A COMPUTER MALFUNCTION [BY TERRORISTS].

Directing a second jet into the path of the missile, to take the hit, would save the 500 passengers but cause [some] passenger deaths.

- 5,000 workers at a manufacturing plant will be laid off SO THAT THE COMPANY CAN AVOID BANKRUPTCY FROM PENSIONS FOR RETIRED WORKERS [SO THAT THE COMPANY CAN INCREASE ITS PROFITS].

- Providing a tax break to this plant would save 5,000 jobs but would cause the loss of [some] other jobs in a competitor's plant.

- 1,000 refugees in a camp in Africa will die from a famine CAUSED BY DROUGHT [CAUSED BY ARMED THUGS WHO BURNED CROPS AND FORCED PEOPLE OUT OF THEIR VILLAGES].

An airplane with food, the only one available, is on its way to a second camp with people affected by the same famine. Sending the airplane to the first camp would save 1,000 people but would cause the death of [some] people in the second camp.

- 1,000 people will die of cancer X next year because the INGREDIENTS FOR THE ONLY CURE ARE EXTREMELY EXPENSIVE [A DRUG COMPANY IS TRYING TO MAKE AS MUCH PROFIT AS POSSIBLE from the only cure, and it is too expensive].

A special program provides drugs for cancer Y, which is similar to X, and the funding for this program could be switched to X. Switching the funds from Y to X would save 1,000 people with X but would cause the death of [some] people with Y.

- 100 coal miners will die because of a mine collapse DUE TO AN EARTHQUAKE (IN A REGION THAT HAS NEVER HAD ONE IN RECORDED HISTORY) [DUE TO THE MINING COMPANY'S VIOLATION OF SAFETY RULES].

Cutting a tunnel to rescue the miners would save 100 miners but would cause the death of [some] other miners whose normal exit was blocked by the same collapse; the tunnel construction would block their backup exit.

- 10 bird species will become extinct FROM NATURAL CAUSES [BECAUSE OF ILLEGAL LOGGING WITHOUT REPLACING CUT TREES].

Building a barrier would prevent the extinction of 10 bird species but cause the loss of others.

The items were presented twice, in a random order chosen for each subject. One presentation (randomly chosen) used the first, more benign, cause, and the other used the malign cause.

[Figure 2 about here.]

Results

The measures of interest are the emotion questions (summed and rescaled from 0 to 1, the PV question, with PVs indicated by the last choice (“should never be done, no matter how great the benefits ...”), the omission bias question (rescaled from 0 to 1, with higher numbers indicating more bias), and “zero,” whether or not the subject would tolerate any harm caused by action (1 when the subject chose the last option in the omission bias question, indicating intolerance of any harm). We also examine the effect of the cause manipulation.

The mean scores for these questions were: emotion, .64; PV, .17; omission bias, .69; zero, .22.

As we have found repeatedly in unpublished studies, emotion correlated with PVs and with omission bias. We tested this here using the mixed-model analysis described in the last study. The dependent variables were (in different analyses) PVs, omission bias, and zero responses. The predictors were subject means for emotion, case means for emotion (collapsing across the malevolent and benign causes), and the idiosyncratic emotion measure (the sum of the deviations of the emotion judgment from these two means). Cases (collapsed across the causes) and subjects were random effects. For all three dependent variables, the idiosyncratic effect of emotion was large and highly significant (.14 for PV, .12 for omission bias, .17 for zeros; all t values greater than 4; the 99.9% highest posterior density interval did not include 0), but the subject and case effects were not significant. The variables are all on the same 0–1 scale, so that these are roughly equivalent to slopes of regression lines. Although not significant in the mixed-model analysis that included other effects, the correlation between emotions and PVs across the 11 cases was significant by itself ($r = .53$, $p = .0486$, one tailed) and is shown in Figure 2, where the words are centered on the means for each of the 11 cases.

In sum, emotion is correlated with omission bias and PVs. However, these correlations do not show a causal effect of emotional responses on PVs or omission bias. They could reflect influences of third variables, such as the perceived seriousness of the outcome in each case.

To test this possibility, we examined the effect of the manipulation of cause. To do this, we treated the 22 cases as 11 matched pairs, each pair differing in cause, and analyzed the 11 differences. As a manipulation check, the differences were clear and significant for emotion itself, with a mean difference of .11 ($t_{10} = 8.07$, $p = 0.0000$, across the 11 pair; ($t_{78} = 8.00$, $p = 0.0000$, across subjects). But the cause manipulation had absolutely no effect on PV (mean effect $-.01$ on the 0–1 scale), omission bias ($-.01$), or zero ($-.01$). The arrows in Figure 2 show the effect of the cause manipulation. It is apparent that the emotional judgment

increased for every case, but the PVs did not increase overall.

These results were confirmed with a (theoretically more sensitive) mixed model analysis, using the differences between the malign and benign causes in each variable as the dependent variable, with random effects for subjects and cases, and the intercept as the statistic of interest (i.e., whether the difference is statistically zero). Emotion was strongly affected by the manipulation of cause ($t = 6.1$), but PVs, omission bias, and zero responses were not affected by the manipulation. (The effect of the manipulation was opposite to the hypothesis and not close to significant for all three variables.)

In sum, it appears that omission bias and PVs are not necessarily influenced by emotion — despite the correlations of omission bias and PVs with emotion — but rather by other factors that affect emotion. We can manipulate the apparent emotional response without affecting PVs or omission bias.

Follow-up

We carried out a second study to check the results, using different measures of omission bias and PVs. The PV measure referred specifically to morality.

The 86 subjects ranged in age from 22 to 64 (median 43); 38% were male. The cases were the same as in the last experiment, some slightly modified to make the cause manipulation more consistent for the 10 pairs of cases. An example of the questions was

If nothing is done, 5,000 workers at a manufacturing plant will be laid off SO THAT THE COMPANY CAN INCREASE ITS PROFITS BY INCREASING THE WORK LOAD ON THE REMAINING EMPLOYEES.

Try to imagine reading about this in the news. How upset do you feel?

[Not at all A little Moderately Very upset Extremely upset]

How angry do you feel? [Not at all . . . Extremely angry]

Providing a tax break to this plant would save 5,000 jobs but would cause the loss of 2,000 other jobs in a competitor's plant.

Should the government give the tax break? [yes no]

What is the largest number of lost jobs in the competitor's plant at which the government should give the tax break? (Pick the closest.)

0 500 1,000 1,500 2,000 2,500 3,000 3,500 4,000 4,500 5,000

What do you think of the morality of intentionally causing job losses in manufacturing?

It is not by itself a moral issue.

It is usually immoral.

It is immoral and impermissible, except in exceptional cases where it leads to a great benefit or prevents a great harm.

It is always immoral and impermissible even if it leads to a great benefit or prevents a great harm, no matter how great the benefit or harm.

The main results were replicated. Using mixed model analysis, the cause manipulation affected emotion strongly (mean effect 0.98 on a scale of 8, $t = 4.7$) but had no effect on PVs, omission bias, or zero responses (all t 's less than 1.0). Omission bias and PVs were correlated (e.g., $r = .58$ across the 20 cases, $p = .0076$, confirmed in a mixed-model analysis with subjects and cases as random effects). And emotion strongly predicted PVs, omission bias, and zero responses (all t values > 5 , with emotion as the only fixed effect and with subjects and cases as random effects). Once again we find that we can manipulate the emotional response without affecting other judgments. These judgments must then sometimes be independent of emotional responses.

Study 3: The nature of omission bias

What is it about acts and omissions that leads to their different effect on judgment? Spranca, Minsk, and Baron (1991, Experiments 6) asked subjects whether each of several factors was morally relevant in a direction so as to make an act worse than omission (or the opposite). Factors that subjects saw as relevant were:

Cause: For omissions, the outcome has another cause.

Movement: Acts involve physical movement and omissions do not.

Presence: The outcome would have occurred in the absence of the actor.

Knowledge: The actor would have done the same thing if she did not know about the possibility of making a decision.

Factors that subjects did not see as relevant (but had sometimes been proposed in the literature) were:

Detectability: Acts are more detectable.

Repeatability: Harmful omissions result from special opportunities that are unlikely to be repeated.

Number of ways: There are few ways to perform an act, many ways to perform an omission (since doing most other acts is consistent with not doing a given act).

Some evidence supported an explanation in terms of “responsibility,” but this was confounded with the possibility of an alternative cause. For example, when someone deceives someone else by saying nothing, then the victim of the deception must have acquired a false belief in some other way and may thus be responsible in some sense.

Other evidence supports a role for causality in a different sense. Ritov and Baron (1990) compared a vaccine-failure condition, in which a vaccine failed to prevent a disease, which caused some number of deaths, to an act condition, in which the vaccine caused the same number of deaths from side effects while completely preventing the disease. Clearly these two cases were identical in all of the factors just listed (except cause), but the vaccine-failure case was strongly preferred to the case in which the vaccine caused harm. Baron and Ritov relied completely on this sort of comparison. Royzman and Baron (2002) found substantial omission bias when the effect of the action was direct, but much less omission bias when the effect of the action was indirect. Directness was defined in terms of causality. For example, a drug that killed infected victims of a disease (who would otherwise live) in order to prevent the spread of the disease was worse than a drug that would immunize everyone against the disease but kill the same number as a side effect. (Interestingly, the vaccine scenarios used by Ritov and Baron, 1990, were thus less direct than would be possible.)

We carried out an additional experiment to examine two of the possible factors that might characterize acts and omissions. In particular, we looked at cause and detectability, using a different method from that used by Spranca et al. (1991). We also looked at the relation between omission bias and emotion.

Method

The 86 subjects ranged in age from 24 to 81 (median 44) and were 29% males. The cases are shown below. Each item had two options, the first an action (or more direct action for the last 3 cases) and the second an omission (or a less direct action, for the last 3 cases). The last 3 items were modified from Royzman and Baron (2002). Note that the first 7 of the 14 items involved a conflict between consequences and the act/omission distinction, and the second 7 did not, yielding the same outcome from act and omission (or from direct and indirect action, in the last three).

Flu25: 25% of the U.S. population is expected to get a flu that is sometimes fatal. A government official must decide what to do. A vaccine will prevent the flu, but the vaccine itself causes side

effects that are just as bad as the flu (with the same chance of death), in 5% of those who are vaccinated.

A. The official recommends the vaccine and makes sure there is enough of it. 5% of those vaccinated get the side effects, and 25% of those not vaccinated.

B. The official does not recommend the vaccine. 25% get the flu.

Flu10: 10% of the U.S. population is expected to get a flu that is sometimes fatal. A government official must decide what to do. A vaccine will prevent the flu, but the vaccine itself causes side effects that are just as bad as the flu (with the same chance of death), in 5% of those who are vaccinated.

A. The official recommends the vaccine and makes sure there is enough of it. 5% of those vaccinated get the side effects, and 10% of those not vaccinated.

B. The official does not recommend the vaccine. 10% get the flu.

Fail: 10% of the U.S. population is expected to get a flu that is sometimes fatal. A government official must decide what to do. A vaccine will prevent the flu, but the vaccine itself causes side effects that are just as bad as the flu (with the same chance of death), in 5% of those who are vaccinated. A second vaccine has no side effects but it fails half the time. The result is that 5% of those who are vaccinated will get the flu.

A. The official recommends the first vaccine and makes sure there is enough of it. 5% of those vaccinated get the side effects, and 10% of those not vaccinated get the flu.

B. The official recommends the second vaccine and makes sure there is enough of it. 5% of those vaccinated get the flu, and 10% of those not vaccinated get the flu.

Trolley: A trolley car in the repair yard is rolling down a hill toward 5 workers who do not see it or hear it. It will kill them if nothing is done. The director of the yard can switch the trolley to another track with 1 worker, who would be killed, but the 5 would be saved.

A. The director switches the trolley to the other track, and one worker is killed.

B. The director does nothing, and five workers are killed.

Push: A trolley car in the repair yard is rolling down a hill toward 5 workers who do not see it or hear it. It will kill them if nothing is done. The director of the yard can push another worker into the path of the trolley. This worker would be killed, but the 5 would be saved.

A. The director pushes the worker in front of the trolley. The trolley stops, and this worker is

killed.

B. The director does nothing, and five workers are killed.

Gun: A terrorist is holding 10 Israeli soldiers and will shoot 9 of them, picked by lot, unless an Israeli government official shoots the single other soldier himself at close range.

A. The official shoots the one soldier and 9 soldiers are set free.

B. The official does not shoot. The terrorist shoots the 9 and the one is set free.

Button: A terrorist is holding 10 Israeli soldiers and will shoot 9 of them, picked by lot, unless an Israeli government official executes the single other soldier himself by pushing a button in the next room, causing a gun to fire.

A. The official executes the one soldier and 9 soldiers are set free.

B. The official does not execute the soldier. The terrorist shoots the 9 and the one is set free.

Brick: Joe is angry at a neighbor.

A. When the neighbor's car starts rolling down a hill, Joe sees that a brick in front of it will stop it from rolling. Joe pushes the brick away, and the car suffers expensive damage.

B. When the neighbor's car starts rolling down a hill, Joe sees that it is possible to stop the car by pushing a brick in front of it. Joe does nothing, and the car suffers expensive damage.

Cash: Jane returns a \$50 item in a department store.

A. The cashier puts \$150 in cash on the counter. Jane picks up all the money after seeing that it is \$150.

B. Jane sees the cashier credit her account for \$150. Jane says nothing and receives the extra \$100.

Lie: Bill's knows that a co-worker and friend has embezzled \$1,000 from their company. Bill plans to tell the investigators that his friend was out of town at the time, which was not true.

A. Bill tells the investigators that the friend was away. Bill's friend is not punished.

B. The investigators misread some memos and think that the friend was out of town, and they tell this to Bill. Bill says nothing to correct them. His friend is not punished.

Check: Susan gets an appeal to send money to help victims of an earthquake in a poor country. Susan plans to send a \$100 check. She has exactly \$100 in her checking account, but she has

several thousand dollars in a savings account, and she can transfer money to her checking account in two days.

A. Susan writes the check, then decides to go out to an expensive restaurant that takes checks only. Susan tears up the check to help the victims.

B. Susan decides to go out to an expensive restaurant that takes checks only. Susan does not write the check to help the victims.

Mall: Jill has been stalked by a crazy man, who shows up with a gun in a crowded mall where Jill is shopping and points the gun at her.

A. Jill moves 3 feet to the left, so that she is behind another person. The crazy man shoots but hits the other person. Jill is unharmed.

B. Jill moves 3 feet to the left, thinking that the crazy man would not change the position of the gun before he shot. Jill knows there is another person behind her. The crazy man does not change the position of the gun, so he hits the person who was standing behind Jill. Jill is unharmed.

Plane: A missile has just been mistakenly fired at a large commercial airliner from a military base. If nothing is done, 200 passengers will die. A military air traffic controller can change the course of airplanes.

A. The controller alters the course of a smaller commercial airliner with 20 passengers so that it is placed in the path of the missile. The large airliner is safe, but the missile destroys the smaller airliner. The controller knew this would happen.

B. The controller alters the course of the commercial airliner. The airliner is safe, but the missile destroys a smaller commercial airliner with 20 passengers flying right behind the large airliner. The controller knew this would happen.

Species: 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. One of the scientists must decide what to do. If nothing is done, 10 species will become extinct.

A. The scientist sprays the forest with a chemical that will destroy the 3 plant species in which the insects make their nests, thus killing the insects and saving the other plants.

B. The scientist sprays the forest with a chemical that will destroy the insects. The same chemical

will kill 3 of the plant species as a side effect.

The items were presented in a random order chosen for each subject. After each pair of items, the subjects answered the following questions.

Is option A morally wrong? [yes, no, equal or hard to say]

Is option B morally wrong? [yes, no, equal or hard to say]

Which option is better or less wrong? [A, B, hard to say]

Rate the strength of your negative emotion when you think about option A. [Four point scale from “none” to “the maximum.”]

Rate the strength of your negative emotion when you think about option B.

Does choosing option A cause harm? [yes, no, equal or hard to say]

Does choosing option B cause harm? [yes, no, equal or hard to say]

In which option is it easier for an outsider to determine the decision maker's intent? [A, B, hard to say]

Results

Table 2 shows the mean differences for each of the 14 cases. Except for “push,” subjects tended to judge the utilitarian response for the conflict items to be less wrong, and, except for “cash” and “species,” they tended to show omission bias for the no-conflict items. (The species item was generally not judged to be wrong for both options.) The correlation between the mean wrongness score for the conflict items and no-conflict items was 0.38 ($t_{84} = 3.73$, $p = .0003$); this is to our knowledge the first demonstration of this kind of correlation. (It is worthy of note that the last three no-conflict items were not properly about acts vs. omissions, since both options were acts.)

[Table 2 about here.]

Judgments of causality correlated as hypothesized with judgments of whether omissions were better than acts, but judgments of detectability did not. (If anything, their correlations were negative.) In particular the omission-better judgment 2) correlated with the causality difference across the 14 cases ($r = .73$, $p = .0034$), as shown in Figure 3, and the subject means correlated across subjects ($r = .48$, $p = .0000$) showing individual differences as well as differences across cases. Similar results were found for the correlations between the

wrongness difference and the causality difference ($r = .81$, $p = .0005$, and $r = .46$, $p = .0000$, respectively). The corresponding correlations for wrongness and detection were $-.10$ and $-.03$, respectively, and neither came close to significance.⁶

[Figure 3 about here.]

Some of the individual items were included as mini-experiments to test various hypotheses. The difference between flu10 and flu25 concerns the effect of the magnitude of benefit, when items are presented on different pages. The wrongness difference was significant one-tailed ($t_{85} = 0.63$, $p = 0.0280$) but neither the causality nor emotion difference (nor the detection difference) came close to significance. The wrongness difference apparently reflects the influence utilitarian reasoning, but this influence must have occurred in subjects who were conflicted. If they had been unconflicted utilitarians, they would have strongly favored the lesser harm and would have shown no differences. This result thus supports Greene's (2007) view that responses to moral dilemmas involve conflict. Note, though, that emotion may not suffice as a source of conflict, since the negative emotion was generally for inaction.

Attention to numbers could also account for the fact that many subjects endorsed the action in the gun item, which is based on a classic dilemma that, with smaller numbers than 9, usually elicits more endorsement of omissions (e.g., Baron, 1992). The contrast between the trolley and push items, and that between the gun and button items, were tests of the effects of proximity or contact (e.g., Cushman, Young, & Hauser, 2006). The difference between trolley and push in omission-better judgments was significant ($t_{85} = 3.97$, $p = 0.0002$), but that between gun and button was not ($t_{85} = 1.02$, $p = 0.3100$). For trolley/push, the difference in causality was also significant ($t_{85} = 2.19$, $p = 0.0313$), as was the difference in emotion ($t_{85} = 5.29$, $p = 0.0000$), but not the difference in detectability ($t_{85} = -0.89$). Thus, contact (and proximity) may operate because they affect perceived causality.

⁶A mixed-model analysis with subjects and items as random effects (Baayen, Davidson, & Bates, 2008) yielded a significant effect of detectability opposite to the hypothesis ($p < .01$, using several methods of testing the hypothesis). A possible explanation of this result is that it came from those cases in which action is favored; if action is the better option, then it may be good, not bad, for it to be detectable. This explanation is supported by the following analysis. We measured the correlation between omission-better judgment and detectability, across subjects, for each of the 14 cases. This correlation was itself correlated with the mean of the omission-better judgments, across the 14 cases ($r = .66$, $p = .0102$). But the highest of these across-subject correlations, for Brick, was still only .03, so we still have no evidence for the hypothesized effect of detectability when omissions are considered better. The results do suggest, though, the utilitarian actions are considered better if they are public: "Tu Gutes und sprich darüber!"

In sum, our results provide additional support for a role of causality. It seems likely that omission bias is not one simple phenomenon, however, and other factors such as movement, knowledge, and presence remain to be explored in scenarios designed for this purpose.

Conclusion

We have focused here on two related biases, protected values (PVs) and omission bias. These biases are related to many others, and, indeed, overlap with them in many experiments. For example, omission bias is typically confounded with status-quo bias.

We argued that PVs for omission are difficult to maintain, so people have relatively few of them. On the other hand, we suspect that PVs are not all that serious. Many may be constructed on the spur of the moment, unreflectively, in order to answer a question in an experiment. Thus, we do find some PVs for omissions. People do not always think about the problem of proliferating obligations. We also suspect, though, that some PVs are the result of prior thought. Although this prior thought may be unreflective, its results may be more resistant to challenges of the sort used by Baron and Leshner (2000). Such challenges might work better with PVs for omissions, a topic for future research.

When people have PVs for action, and when they are asked whether they will undertake some action that causes the negative outcome in question, they will logically, we argue, be very reluctant to undertake such action. Thus, measures of omission bias, will show a relatively strong bias. In principle, if people really held to the principle of the PV, they would be unwilling to act at all, and their threshold for action would be zero. People are not always so consistent. We have thus found, repeatedly, correlations between PVs and omission bias, and we have discussed possible reasons why our results conflict with the study of Bartels and Medin (2007), which failed to replicate our basic result in one condition.

Note that PVs are not the only cause of omission bias, so we can expect some omission bias even in their absence. We have discussed other possible causes of omission bias. A major source of the bias is a judgment that it is wrong to cause harm. If the harm has some other cause, then the responsibility for it is diffused between the decision maker and the other cause, so omissions that cause harm are less bad. We have reported additional evidence for a role of causality.

Omission bias seems likely to have several different determinants. We have argued that PVs can exacerbate it, and perceived causality is relevant too. But its confounding with other types of bias leads to other explanations. For example, the default usually is better than alternatives to it, so it is a reasonable

heuristic, other things being equal, to prefer the default. Likewise for the status-quo, and for what is natural. In experiments it is not the case that other things are equal, so the biases found seem real, yet they make sense in terms of generally useful heuristics. The problem is that people often apply these heuristics blindly without thinking about why they are generally useful and whether those reasons apply in the case at hand. It is possible that some biases like these have very serious effects (Baron, 1998).

Other determinants of omission bias could be analogous to thinking about causality. For example, people may reason that they are not responsible for something if it would have happened anyway if they knew nothing about it or if they were in no position to stop it. In judgments of moral responsibility for the purpose of blame and punishment, such an argument might be reasonable. It is less reasonable for someone who is a decision maker, in the situation, who knows what the options are. In such cases, it may amount to a kind of self-deception.

We have also examined the role of emotion. It has been proposed (e.g., Greene et al., 2004; Haidt, 2001) that emotion is a cause of non-utilitarian biases, and there is some evidence for this claim. Yet it is also possible that non-utilitarian judgments can cause the emotion, rather than the reverse. A moral rule may be elicited before any emotion is experienced (Bucciarelli, Khemlani, & Johnson-Laird, 2008). We doubt that emotion is required for biases, since biases similar to those we discuss are found in relatively “cold” domains, such as logic and probability.

In two experiments, we find that reported emotion can be manipulated without any effect on PVs and omission bias. The same studies find the usual correlation between reported emotion and magnitude of bias. These findings together suggest that a causal role for emotion is not universal. One might question whether the emotion reported in scenario studies is “real,” or, alternatively, whether subjects are simply telling us what they think they would feel if the situation were real. Some evidence for real emotion is provided by studies of brain activity during thinking about scenarios (Greene et al., 2004; Greene, 2007).

Conclusions about the nature of omission bias are relevant to prescriptive questions, practical questions about how to reduce biases that might cause harm, or how to work around them. If emotion is a major causal determinant, then prescriptive remedies might involve trying to get people to “cool off” before making relevant judgments (as suggested in a different context by Camerer, Issacharoff, Loewenstein, O’Donoghue, & Rabin, 2003). We have suggested that the role of emotion might be overemphasized, although we note that emotional states may still affect biases at least because they change the way a situation is perceived or the tendency to take action.

To the extent to which PVs affect omission bias, as they seem to do, then it might be worthwhile to encourage reflection on apparent absolute values, as done experimentally by Baron and Leshner (2000). Recently, Bartels (2008) has provided additional evidence for the lability of PVs.

The role of causality is of interest because the kind of causality that seems relevant is direct, physical, causality, in which the agent's behavior starts a chain of events that lead to an outcome through known principles, possibly with other enabling conditions. An alternative view of causality is embodied in tort law, often called "but for" causality. By this view, you "cause" something if it would not have happened but for some choice you made, regardless of whether that choice is an act or omission. If people were to think of this kind of causality as relevant to more decisions, then omission bias might be reduced, and, with it, its harmful effects.

References

- Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effect modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, in press.
- Baron, J. (1992). The effect of normative beliefs on anticipated emotions. *Journal of Personality and Social Psychology*, *63*, 320–330.
- Baron, J. (1993). *Morality and rational choice*. Dordrecht: Kluwer.
- Baron, J. (1998). *Judgment misguided: Intuition and error in public decision making*. New York: Oxford University Press.
- Baron, J. (2008). *Thinking and deciding*. New York: Cambridge University Press (4th edition).
- Baron, J., & Leshner, S. (2000). How serious are expressions of protected values. *Journal of Experimental Psychology: Applied*, *6*, 183–194.
- Baron, J., & Miller, J. G. (1999). Limiting the scope of moral obligations to help: A cross-cultural investigation. Manuscript.
- Baron, J. & Ritov, I. (1994). Reference points and omission bias. *Organizational Behavior and Human Decision Processes*, *59*, 475–498.
- Baron, J. & Ritov, I. (2004). Omission bias, individual differences, and normality. *Organizational Behavior and Human Decision Processes*, *94*, 74–85.
- Baron, J., & Spranca, M. (1997). Protected values. *Organizational Behavior and Human Decision Processes*, *70*, 1–16.
- Bartels, D. M. (2008). Principled moral sentiment and the flexibility of moral judgment and decision making. *Cognition*,
- Bartels, D. M., & Medin, D. L. (2007). Are morally-motivated decision makers insensitive to the consequences of their choices? *Psychological Science*, *18*, 24–28.
- Bennett, J. (1981). Morality and consequences. In S. M. McMurrin (Ed.), *The Tanner Lectures on human values* (Vol. 2, pp. 45–116). Salt Lake City: University of Utah Press.
- Bicchieri, C. (2006). *The grammar of society: The nature and dynamics of social norms*. New York: Cambridge University Press.
- Broome, J. (1991). *Weighing goods: Equality, uncertainty and time*. Oxford: Basil Blackwell.
- Bucciarelli, M., Khemlani, S., & Johnson-Laird, P. N. (2008). The psychology of moral reasoning. *Judgment and Decision Making*, *3*, 121–139.

- Camerer, C. F., Issacharoff, S., Loewenstein, G., O'Donoghue, T., & Rabin, M. (2003). Regulation for conservatives: Behavioral economics and the case for "asymmetric paternalism." *University of Pennsylvania Law Review*, *151*, 1211–1254.
- Connolly, T., & Reb. J. (2003). Omission bias in vaccination decisions: Where's the "omission"? Where's the "bias"? *Organizational Behavior and Human Decision Processes*, *91*, 186–202.
- Cushman, F., Young, L., & Hauser, M. (2006). The role of conscious reasoning and intuition in moral judgments: Three principles of harm. *Psychological Science*, *12*, 1082–1089.
- Fiske, A. P., & Tetlock, P. E. (1997). Taboo trade-offs: Reactions to transactions that transgress the spheres of justice. *Political Psychology*, *18*, 255–297.
- Goodwin, G. P., & Darley, J. M. (2008). The psychology of meta-ethics: Exploring objectivism. *Cognition*, *106*, 1339–1366.
- Greene, J. D., Nystrom, L. E., Engell, A. D., Darley, J. M., & Cohen, J. D. (2004). The neural bases of cognitive conflict and control in moral judgment. *Neuron*, *2*, 389–400.
- Greene, J. D. (2007). The secret joke of Kant's soul. In W. Sinnott-Armstrong (Ed.), *Moral Psychology, Vol. 3: The Neuroscience of Morality: Emotion, Disease, and Development*, pp. 35–117 (with comments and reply). MIT Press, Cambridge, MA.
- Greene, J. D., & Baron, J. (2001). Intuitions about declining marginal utility. *Journal of Behavioral Decision Making*, *14*, 243–255.
- Haidt, J. (2001). The emotional dog and its rational tail: A social intuitionist approach to moral judgment. *Psychological Review*, *108*, 814–834.
- Hare, R. M. (1981). *Moral thinking: Its levels, method and point*. Oxford: Oxford University Press (Clarendon Press).
- Kahneman, D., Knetsch, J. L., & Thaler, R. H. (1990). Experimental tests of the endowment effect and the Coase theorem. *Journal of Political Economy*, *98*, 1325–1348.
- Kahneman, D., & Ritov, I. (1994). Determinants of stated willingness to pay for public goods: A study of the headline method. *Journal of Risk and Uncertainty*, *9*, 5–38.
- Kuhse, H. (1987). *The sanctity of life doctrine in medicine: A critique*. Oxford: Oxford University Press.
- McNaughton, D., & Rawling, P. (1991). Agent relativity and the doing-happening distinction. *Philosophical Studies*, *63*, 167–185.
- Nucci, L. (2001). *Education in the moral domain*. New York: Cambridge University Press

- Parfit, D. (1984). *Reasons and persons*. Oxford: Oxford University Press (Clarendon Press).
- Piaget, J. (1965). *The moral judgment of the child*. New York: Free Press.
- Ritov, I., & Baron, J. (1990). Reluctance to vaccinate: omission bias and ambiguity. *Journal of Behavioral Decision Making*, 3, 263–277.
- Ritov, I., & Baron, J. (1992). Status-quo and omission bias. *Journal of Risk and Uncertainty*, 5, 49–61.
- Ritov, I., & Baron, J. (1999). Protected values and omission bias. *Organizational Behavior and Human Decision Processes*, 79, 79–94.
- Ritov, I., & Kahneman, D. (1997). How people value the environment. In M. H. Bazerman, D. M. Messick, A. E. Tenbrunsel, & K. A. Wade-Benzoni (Eds.), *Environment, ethics and behavior*, pp. 33–51. San Francisco: New Lexington Press.
- Rozin, P., Spranca, M., Krieger, Z., Neuhaus, R., Surillo, D., Swerdlin, A., & Wood, K. (2004). Natural preference: instrumental and ideational/moral motivations, and the contrast between foods and medicines. *Appetite*, 43, 147–154.
- Royzman, E. B. & Baron, J. (2002). The preference for indirect harm. *Social Justice Research*, 15, 165–184.
- Samuelson, W., & Zeckhauser, R. (1988). Status-quo bias in decision making. *Journal of Risk and Uncertainty*, 1, 7–59.
- Schweitzer, M. (1994). Disentangling status quo and omission effects: Experimental evidence. *Organizational Behavior and Human Decision Processes*, 58, 457–476.
- Singer, P. (1993). *Practical ethics* (2nd ed.). Cambridge: Cambridge University Press.
- Tetlock, P. E., Lerner, J. & Peterson, R. (1996). Revising the value pluralism model: Incorporating social content and context postulates. In C. Seligman, J. Olson, & M. Zanna (Eds.), *The psychology of values: The Ontario symposium, Volume 8*. Hillsdale, NJ: Erlbaum.
- Turiel, E. (1983). *The development of social knowledge: Morality and convention*. Cambridge University Press.

List of Figures

1	RB threshold and PV-act by item, Study 1.	41
2	Emotion and PVs by item (each scale ranging from 0 to 1), Study 2. Arrows represent the effect of the cause manipulation.	42
3	Causality judgment and “omission better” by item, Study 3.	43

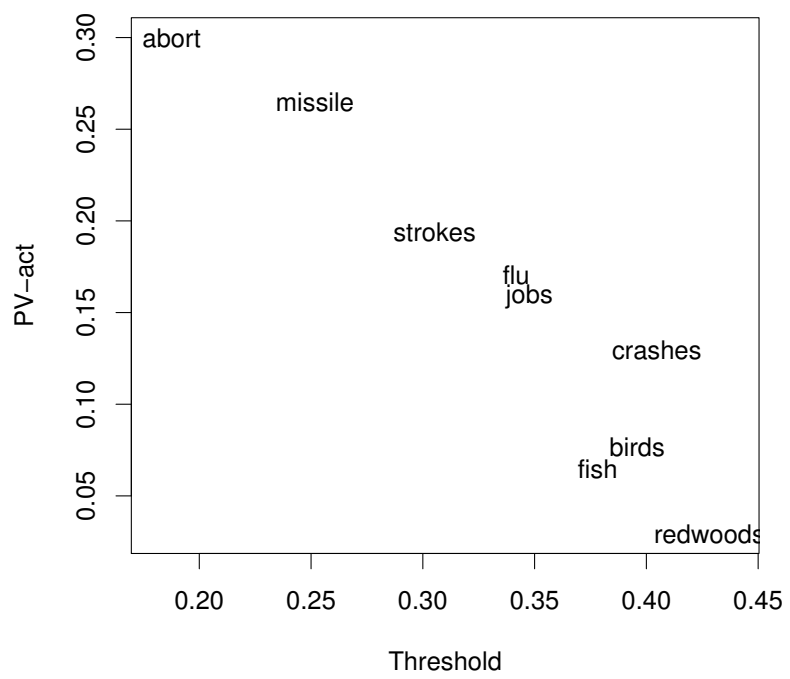


Figure 1: RB threshold and PV-act by item, Study 1.

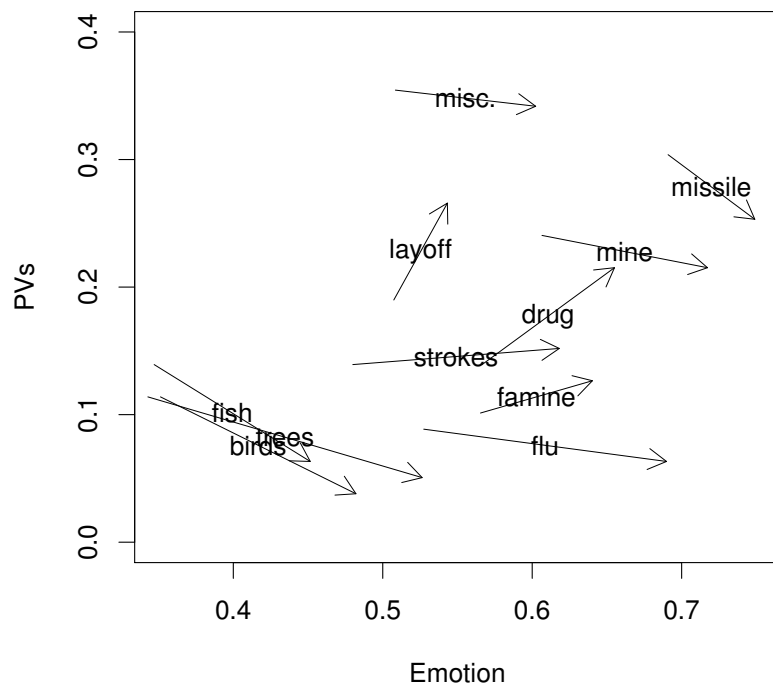


Figure 2: Emotion and PVs by item (each scale ranging from 0 to 1), Study 2. Arrows represent the effect of the cause manipulation.

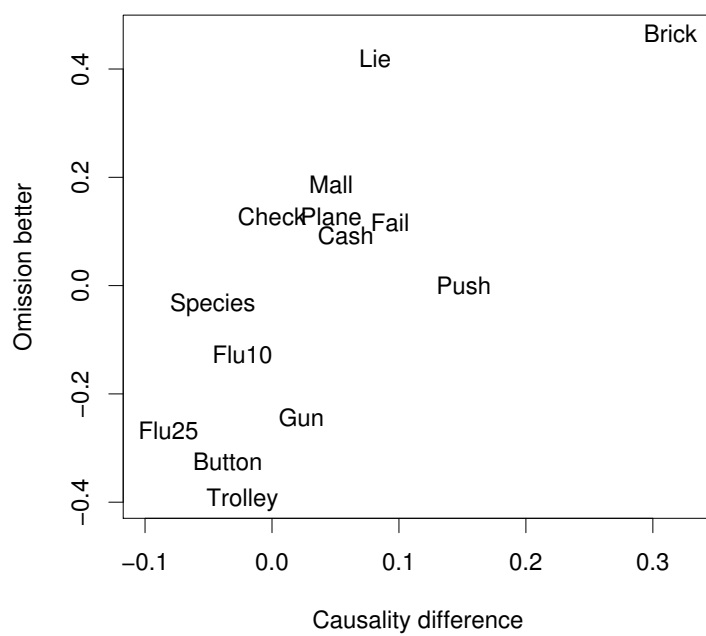


Figure 3: Causality judgment and “omission better” by item, Study 3.

Table 1: Coefficients for prediction of PV-act and PV-omit from logistic mixed-effect model, Study 1. PV values are 1 with PVs, 0 without. RB and CV values are thresholds on a scale of 1–11, with 1 meaning the lowest possible threshold. “RB first” refers to the order of RB and CR within the session. “Act first” refers to the order of the PV-act and PV-omit questions within a page. “.sm” refers to subject means; “.vm” refers to case means. Significance codes: *** 0.001, ** 0.01, * 0.05 + 0.1.

Threshold					
Predictors	RB	RB.sm	RB.vm	RB first	Act first
PV-Act	−0.20***	−0.21*	−1.03***	0.90*	0.53**
PV-Omit	−0.06*	−0.17	−0.56***	0.92+	0.26
Predictors	CR	CR.sm	CR.vm	RB first	Act first
PV-Act	−0.27***	−0.26**	−1.01***	0.67+	0.53**
PV-Omit	−0.08*	−0.30*	−0.52***	0.53	0.23
Zeros					
Predictors	RB	RB.sm	RB.vm	RB first	Act first
PV-Act	1.23***	1.88**	4.61***	0.66+	0.30
PV-Omit	0.68***	0.78	2.23**	0.82	0.10
Predictors	CR	CR.sm	CR.vm	RB first	Act first
PV-Act	1.40***	1.38*	4.62***	0.82*	0.47*
PV-Omit	0.40*	1.28+	2.42**	0.76	0.21

Table 2: Mean response to omission bias item, mean differences — act minus omission (or direct minus indirect) — for wrongness, emotion, and cause judgments, and responses that acts are more detectable, for the 14 cases in Study 3. All values are re-scaled so that the possible range is from -1 (omission greater) to 1 (action greater).

	Omission	Wrong	Emotion	Cause	Detection
Flu25	-0.27	-0.20	-0.04	-0.04	0.02
Flu10	-0.13	-0.09	-0.03	-0.01	0.12
Fail	0.12	0.02	0.01	0.05	-0.07
Trolley	-0.40	-0.21	-0.12	-0.01	0.38
Push	0.00	0.11	0.07	0.08	0.29
Gun	-0.24	-0.01	-0.09	0.01	0.36
Button	-0.33	-0.02	-0.09	-0.02	0.31
Brick	0.47	0.23	0.13	0.16	0.24
Cash	0.09	-0.01	0.01	0.03	0.12
Lie	0.42	0.12	0.08	0.04	0.13
Check	0.13	0.05	0.02	0.00	0.06
Mall	0.19	0.10	0.02	0.02	-0.06
Plane	0.13	0.03	0.03	0.02	-0.02
Species	-0.03	-0.01	-0.00	-0.02	0.06