Consequences, norms, and inaction: A comment

Jonathan Baron* Geoffrey P. Goodwin†

Abstract

Gawronski, Armstrong, Conway, Friesdorf & Hütter (2017, GACFH) presented a model of choices in utilitarian moral dilemmas, those in which following a moral principle or norm (the deontological response) leads to worse consequences than violating the principle (the utilitarian response). In standard utilitarian dilemmas, the utilitarian option involves action, and the deontological response, omission. GACFH propose that responses in such dilemmas arise in three different ways: a psychological process leading to a deontological choice, a different process leading to a utilitarian choice, or a bias toward inaction or action. GACFH attempt to separate these three processes with new dilemmas in which action and omission are switched, and in which the utilitarian and deontological processes lead to the same choice. They conclude that utilitarian and deontological responses are indeed separable, and that past research has missed this fact by treating them as naturally opposed. We argue that a bias toward harmful inaction is best understood as an explanation of deontological responding rather than as an alternative explanation, and we briefly review the literature concerning the incomplete sufficiency of this explanation. We also argue that GACFH’s results can be largely explained in terms of subjects’ unwillingness to accept the researchers’ assumptions about which consequence is worse and which is consistent with a moral norm. We support our argument with data from experiments with new and old scenarios, in which we asked subjects to judge these classifications.

Key words: utilitarianism, deontology, omission bias

1 Introduction

A great deal of research has now established that many people’s moral judgments do not follow utilitarian principles. In one sort of demonstration (among many), subjects are asked to compare two options, one of which leads to a better result than the other, e.g., fewer deaths, but many subjects choose the other option, thus violating the utilitarian principle of doing the most good, or the least harm, aggregated across those affected.
In order to get this result, the more harmful option must be made attractive in some way that is irrelevant to the utilitarian calculation. Usually this involves telling subjects that the harm from the utilitarian option must be actively and directly caused by the decision maker, e.g., pushing a man off of a bridge, to his death, in order to stop a trolley that will otherwise kill several other people. For some individuals, refraining from directly causing this harm thereby becomes more attractive than causing it, even though the overall consequences are worse. These cases are called “sacrificial dilemmas.”

The usual analysis of these dilemmas is that they pit utilitarian responding (responding in terms of the greater good) against deontological responding, where deontology is a category of moral theories that emphasize properties of action other than consequences, such as whether the actions violate basic rights, or whether the act conflicts with a required duty. Deontological theories can justify not pushing the man in a variety of ways, but most of them involve a prohibition of active killing, whatever the consequences.

Gawronski, Armstrong, Conway, Friesdorf & Hütter (2017, henceforth GACFH) report an experimental analysis of sacrificial dilemmas, using a new method, which they call the CNI model because it considers three possible determinants of responses: Consequences, Norms, and Inaction. In the typical dilemma, consequences favor acting, e.g., pushing the man, a bias toward inaction opposes action, and moral norms usually also oppose action (e.g., “don’t kill people”). GACFH further suppose that deontological and utilitarian responding are not simply poles of a single dimension but, rather, alternative and independent ways of thinking about the dilemmas. In particular, GACFH assume that the basic utilitarian principle is based on consequences and the basic deontological principle is based on norms. In their view, a preference for inaction (or action) is separate, different from either approach. In standard sacrificial dilemmas, they argue that an apparent utilitarian response could arise either from a focus on consequences or from a bias toward action, and an apparent deontological response could arise either from a norm or a bias toward inaction.

To assess the role of each of the three components of the model, GACFH use a design in which they manipulate the association of norms and consequences with action or inaction. In this design, the consequences of action are manipulated so that action is better than inaction in half of the items, and worse in the other half. Orthogonally, norms either forbid (proscribe) or require (prescribe) action. These manipulations give rise to three new versions of the standard sacrificial dilemma. In the standard version of this dilemma, action has better consequences than inaction, but it is forbidden by a norm. In a new, reversed version, action is worse than inaction, but it is required by a norm. In cases of this sort, the action in question is one of

---

1 We use the term “utilitarian” rather than “consequentialist” because the latter is a broader class of principles, some of which ignore the number of people affected (e.g., Rawls’s [1971] “difference principle”). For most of the dilemmas at issue here, numbers are highly relevant to the conflict.

2 The model is similar to, and builds upon, an earlier model based on the idea of process dissociation. In this paper, we discuss largely the single paper that introduces the CNI model, but most of our comments apply to other papers using the CNI model and process dissociation as applied to moral judgments.
preventing someone else’s action, or reversing an action already chosen, or reversing its effects. In two further versions, norms and consequences align — such that both dictate either action, or alternatively, inaction. Thus, consequences and norms are congruent for two of the four cases and incongruent (conflicting) for the other two. The pattern of subjects’ responses to the four dilemmas indicates which principle is driving the responses. If action is always chosen, or inaction is always chosen, then the responses are driven by a bias toward action or inaction, not by either norms or consequences.

In this comment, we discuss several concerns with this approach and the reported findings. We begin with the main philosophical assumption made by GACFH, which is that a moral distinction between acts and omissions is not itself an indicator of deontological judgments, i.e., judgments that depend on features of options other than their consequences. We then discuss the difficulty of constructing items that meet the new requirements, and we present evidence that major results of the CNI model are affected by subjects’ lack of agreement with the experimenters about what norms and consequences imply. Finally, we argue that the primary empirical question, whether apparently deontological responses arise from a bias against action, has already been largely answered by previous literature, so that this new approach is not needed for this empirical purpose.

2 Omission bias and deontological reasoning

Utilitarianism, as a principle of choice, says that we should choose the option with the best overall consequences. A minimal definition of deontology, consistent with the literature in both philosophy and psychology, is that it consists of moral rules about choices (acts or omissions) other than those concerned with consequences, particularly properties of acts. Many of these properties are taken as sufficient to make certain actions prescribed as matters of duty, and other properties make actions proscribed or forbidden. The rules need not be absolute; some of them may be overridden by other rules or by consequences (Ross, 1930). Some philosophers add additional specifications to what they count as deontology. But, in general, we can think of deontology as consisting of rules that must be considered in addition to, or instead of, utilitarian outcomes. If the outcomes of utilitarian choices are indeed best, as expected, then the outcomes of deontological principles, consistently applied, will at least sometimes be relatively worse. Thus, the attraction of deontology could be part of the reason why things are not as good as they could be.

An example of a well-known deontological rule is “primum non nocere” (“first, do no harm”). The usual interpretation is that it is wrong to act in a way that causes harm, presumably even if the action is likely to prevent greater harm. This principle, even if not absolute, leads to a bias against action in any conflict situation in which the act could be harmful. Consistent application of this principle in a series of act/omission
dilemmas, in most of which both options lead to some harm, would appear as a bias against action.

GAFCH ask whether subjects in an experiment consider each type of principle, or both, or neither. In their analysis, the subject decides sequentially, first whether consequences apply (and stops there if they do), then whether norms apply, and, if neither principle applies, decide by response bias. When neither consideration drives the decision, GAFCH suppose that the decision is based on a bias toward action or inaction. Such a bias must be considered as another (non-absolute) rule, e.g., “When you don’t know what to do, don’t do anything,” or “Don’t just sit there, do something.” Such rules are, by definition, deontological, because they concern inherent properties of the choices, and not their consequences. So the assertion that biases of this sort are separate from both utilitarian and deontological considerations is inconsistent with how these theories are defined.

There is in fact a fundamental tension between deontological and utilitarian theories concerning the role of acts and omissions in many morally pertinent situations. Deontological rules are normally prohibitions of actions, while utilitarianism, focusing as it does on consequences, makes no fundamental distinction between acts and omissions. This tension is evident in the much-discussed deontological principle of doing versus allowing: it is morally impermissible to cause the death of someone (or to cause some other significant harm), but it may be permissible to allow the same person’s death. For deontologists, the action of killing has intrinsically undesirable properties, regardless of its consequences. From a utilitarian perspective, however, this distinction is irrelevant – doing and allowing are morally equivalent (see, e.g., Rachels, 1975), if all other things are equal (which frequently they are not). Accordingly, the two theories clash in a fundamental way concerning this distinction, which has ramifications in several practical situations (e.g., the morality of active versus passive euthanasia; Rachels, 1975).

This sort of fundamental opposition is what motivated the philosophical, and later the psychological, use of trolley dilemmas, and other sacrificial dilemmas as a tool for inquiry. Though such dilemmas may have other problems, GAFCH are mistaken to argue that the action-omission distinction can be conceived of as a third factor, entirely separate from deontological ethics.

The basic finding that some people prefer an option with worse consequences, by itself, shows that people do not follow utilitarian principles, however it is explained. For example, Ritov and Baron (1990) found that many people oppose vaccination when the side effects of the vaccine cause half as many deaths as the disease that the vaccine prevents. GACFH, however, suggest that the result is an artifact if it is due to a general preference for not acting — as if the response itself cannot be taken as a rejection of utilitarian principles. It is not clear why they think this. If it should turn out that the effect is entirely due to a bias against action, then

---

3 All but one of the dilemmas used by GAFCH clearly have this property. The one that is not clear is when payment of a ransom will buy food for the kidnappers’ needy families. Yet a subject might imagine that any payment of ransom would encourage other kidnappings for less benign purposes.
we would conclude that “people are non-utilitarian in certain cases because they have a bias against action.” In other words, the preference for inaction is part of the explanation for non-utilitarian responding, rather than an entirely separate response tendency. This is more or less what Ritov and Baron (1990) had in mind at the outset (likewise Spranca et al., 1991). Hence the poorly-chosen name “omission bias” for the effects reported then. (See our later discussion concerning the drivers of the effect of interest.) Utilitarianism implies that the morally better option is the one that minimizes harm (or maximizes good — there is no natural zero point, so “harm” and “good” are always relative terms). If some other descriptive principle of decision making leads people away from this option, then that other principle is an explanation of why judgments are sometimes not utilitarian.

The very definition of deontology is based on its opposition to consequentialism (including utilitarianism). While consequentialism evaluates options in terms of expected consequences only, deontology adds (or substitutes) additional criteria concerning properties of the behavior in question other than its consequences (Alexander & Moore, 2016). There is a point to this. To the extent that choices do what they are expected to do, then utilitarian choices will bring about the best options on the average, while deontological choices will lead to outcomes that are relatively worse. Thus, the attraction of deontology could be part of the reason why things are not as good as they could be.

3 Separating norms and consequences

To distinguish the separate effects of norms, consequences and response bias, GAFCH use dilemmas in which norms and consequences point in the same direction, such that they both dictate either action or inaction. What could possibly account for a response tendency that seems to conflict with both norms and consequences? The most obvious possibility is that such responses are driven by an antisocial tendency, which causes some subjects to choose the option that is both relatively harmful and contrary to moral norms. This response tendency might also be caused by reactance or misbehavior on the part of some subjects who deliberately give non-serious responses. It is also possible that such responses are the result of inattention. Such “congruent” conditions may therefore be useful for the purpose of excluding inattentive, non-serious, or deeply anti-social subjects (e.g., sociopaths). But, beyond this, it is not clear what else of substance can be learned from such responses.

Another way in which GAFCH attempt to separate the effects of norms and consequences from those of
response bias is to construct vignettes in which norms and consequences are supposed to conflict, but the usual association of norms with inaction and consequences with action is reversed. This method, however, creates additional consequences that blur the intended distinction. One way these dilemmas are constructed is by switching an originally harmful action to one that blocks someone else’s harmful action. For instance, in the standard transplant scenario dilemma, the subjects are asked to imagine themselves as a surgeon, and the target action is to kill a patient in order to harvest his organs for other needy patients. In the switched case, the target action is to intervene to prevent another surgeon from killing a patient for the same purpose. The point of this manipulation is to make action forbidden (proscribed) in the first case, but required (prescribed) in the second, while holding constant the relative benefits of action over inaction. But, this switching has a problem. When the action is to contravene someone else’s action, it has additional consequences aside from preventing the consequences of that action. It may hurt the decision maker’s feelings, possibly leading him or her to take retaliatory action against the one who contravenes. It may also violate the lines of authority, thus weakening these lines for the future by discouraging those in command from taking their responsibility seriously (Baron, 1996). It may also be illegal or against the rules, and rule following likewise has a value as a precedent for future cases. In addition, the fact that someone else has made a decision provides reason to think that he or she knew something that we did not know. Thus, the use of the idea of blocking someone else’s decision does not ensure that the consequences of action are held constant, with the only factor differentiating the two cases being whether the action is proscribed or prescribed. In fact, it is difficult to find a clear way to reverse action and omission that holds everything else constant, i.e., to look for a true framing effect.4

Moreover, GAFCH do not check on which norms subjects think are relevant to each dilemma. There may also be multiple norms invoked in some scenarios, further compounding this problem. And, as the last paragraph suggests, choices attributed to norms could be based on consequences.

For example, in the abduction dilemma, it seems as though GACFH think there is a relevant norm to approve ransom payments to guerrillas if it means saving a journalist from beheading. The approval of such payments is apparently prescribed, whereas the vetoing of such payments is apparently proscribed. In the basic version of this vignette, in which the norm to approve this payment is opposed by consequences, the consequences include many further deaths caused by the guerrillas in the war they are waging (because the payment will be used to buy weapons). GACFH therefore want to treat approval of the ransom payment as reflecting sensitivity to a moral norm, because it will save the immediate victim, a journalist. However, any

---

4Ritov & Baron (1994) examined compensation and penalty judgments in a situation in which action or omission could lead to the same harmful outcome. Specifically, a train was headed down a hill toward a tree that had fallen across the tracks. In one of several conditions, the engineer decides not to stop the train, the train hits the tree, and one passenger is injured. In another condition, the engineer stops the train, but the train stops so quickly that one passenger is injured. (In yet another condition, the engineer tries to stop the train and fails.) Judgments depended on expectations and were affected by the consequences of the rejected option. We thank Ilana Ritov for pointing this out. Later in this paper, we modify the train dilemma for use in an experiment.
apparent sensitivity to the alleged norm could be explained entirely in terms of the consequences. People might generally approve payment of the ransom because they think that the beheading of the journalist, and the attendant publicity it would bring, would be worse overall than the deaths of combatants in the guerrilla war (which are also less certain to occur). Additionally, it’s not even clear that this norm to make such ransom payments is widely endorsed. There is a strong contrary norm not to give in to such requests, as it involves collaboration with evil. Thus, a refusal to make the payment could arise from a norm rather than (as assumed) a consideration of consequences. There is just too much latitude in this vignette (and others) to conclude anything firmly about whether norms or consequences (or some combination) are driving people’s decision-making. At least, we should have data on what each subject thinks the norm is for this kind of case.

In the prescribed norm version of the police torture vignette, you must decide whether to stop your partner from using torture to interrogate a suspect who is accused either of having kidnapped children or of having stolen paintings. Stopping the torture requires “stopping [your partner] by reporting him to your supervisor.” Evidently, in this case, “you” have already decided not to participate in illegal torture, which is why your partner is doing it. (You have tried all other interrogation methods.) The combination of this fact plus the mention of “your supervisor” may bring to mind the possibility of being legally liable if you do not report the torture to your supervisor. This concern would increase the likelihood of the “action” response in this case, and it would do so not because it enhances the strength of the relevant norm, but rather, because the reversal changes the consequences that can be anticipated.

In the immune-deficiency dilemma (which we call VIRUS, below), the conflict version in which the norm supports omission but the consequences are supposed to support action, the text reads: “You are the director in a hospital of a developing country. A foreign student who is volunteering in the country got infected with a rare virus. The virus is highly contagious and deadly to seniors and children. The only medication that can effectively stop the virus from spreading has severe side-effects. Although the virus will not kill her, the student suffers from a chronic immune deficiency that will make her die from these side-effects.” Here it seems somewhat unclear what the consequences actually are; subjects might well imagine that there is some other way to prevent the virus from spreading aside from medication, e.g., quarantine. Subjects who considered this possibility would thus see this dilemma as congruent. More relevant to our concerns here, the congruent version in which both the norm and consequences support action reads: “You are the director in a hospital of a developing country. A foreign student who is volunteering in the country got infected with a rare virus. The virus is highly contagious and can cause severe stomach cramps. The student suffers from a chronic immune deficiency that will make her die from the virus if she is not returned to her home country for special treatment. However, taking her out of quarantine involves a considerable risk that the virus will spread.” The version then asks, “Is it acceptable to take the student out of quarantine to return her to her
home country for treatment?” Here, a careless subject might see the last sentence before the question as a hint that the action might not have better consequences overall, and the term “acceptable” in the question might suggest that the director should have discretion in this case, to balance the relative risks, which are not clearly described. Perhaps others could die from the virus aside from this student.

Conceptual problems of this sort are not limited to these examples alone, but apply to GACFH’s entire set of vignettes, as well as similar vignettes used in process-dissociation studies. GACFH make the point that such problems are immediately solved by the fact that people showed some sensitivity to the norms factor. But, as our previous examples make clear, it might turn out that people are sensitive to the manipulation of norms for other reasons – e.g., the consequences! And some subjects might disagree about what the norms and consequences are, which would convert apparently conflicting vignettes into congruent ones. The fact that subjects on the average are sensitive to norms as intended does not imply that subjects who respond very differently are interpreting cases in the same way.

4 Correlations with other variables

GACFH makes inferences about other variables, such as sex and cognitive load, on the basis of correlations with model parameters. GACFH did not check to see whether these correlations were consistent across scenarios; this should be done routinely. The relevance of testing across items might be apparent in Studies 2a and 2b, where the apparent effect of time pressure seems to be due to a single item (d4incon, in the paper) which was affected by time pressure much more than any other item, and which was also an outlier for the small number of action choices, despite 5 other items in the same category (incon). In general, correlations between manipulations, such as sex and time pressure, and responses to items in the same category are highly variable, with considerable overlap across types.

Moreover, the p-values reported in the article are much lower than those we obtain for tests using subjects alone as the unit of analysis, which, in turn, should give lower expected p-values than those obtained testing concurrently across both subjects and items. It seems that the software used for hypothesis testing, multiTree, ignores both subject and item variance, and thus uses observations as the units of analysis. The p-values in GACFH thus do not allow their usual interpretation.

As we have already noted, the CNI model seeks to distinguish attention to norms and consequences, independently. Women were found to pay more attention to norms than men, but apparently no less attention to consequences. What could it mean to pay more attention to norms than men, but apparently no less attention to consequences?

---

5 Moshagen (2010, p. 52) says, “multiTree offers no means to diagnose or handle heterogeneity across items and/or participants. This is considered a major limitation and will be addressed in future versions.” The changelog up to the current version at the time of publication (v046) does not mention any correction of this deficiency.
As it turns out, this result could arise for other reasons aside from the apparent one. In particular, men could pay less attention to everything. They could be more variable, more prone to errors, or less consistent with the model. Or they could be closer to the floor or ceiling (i.e., with a stronger response bias for or against action), leaving less room for other factors to have any effect. Or men could be more antisocial. One of these alternatives must be true. The role of other influences on model parameters — while necessarily present — is especially clear from the fact that the N and C parameters of the original model correlate positively across subjects \((r = .70)\). This correlation could arise from individual differences in the influences just mentioned.

The obvious way to avoid such problems is to compare “norms” and consequences directly, when they conflict. But this is exactly what the standard, inconsistent, scenarios do. We gain no insight from the consistent cases, except that they may be useful for excluding inattentive subjects. The cases in which acts and omission are switched could in theory provide some insight because there is no essential reason why utilitarian choices must involve action — the greatest good might in some cases result from doing nothing. And deontological rules do sometimes prescribe actions (though they more commonly proscribe actions). However, as we have previously noted, the construction of scenarios in which acts and omission are switched can introduce other consequences which complicate any inferences that can be drawn.

5 Concerns about the congruent/conflicting distinction

As we have discussed, it is not clear that the GAFCH dilemmas succeeded in manipulating norms and consequences, as perceived by the subjects, as GAFCH intended. Examination of their data provided additional support for this concern. Figure 1 shows the proportion of “acceptable” responses for the four versions of each case for Study 1a. (Other studies show similar results.) One surprising feature of the data is the large proportion of congruent versions in which the response disagrees with the expected outcome: 29% overall for the version in which norms and consequences both imply omission, and 32% for the version in which both imply action. Of course, it is these responses that the CNI model is meant to explain, but this many of them still seems surprising if the subjects agreed with the experimenters’ classification.

Perhaps more disturbing is the large overlap between different versions, across cases. We might expect the NormOmit-ConsOmit version to be consistently lower than all other versions, and the NormAct-ConsAct versions to be consistently higher, but this is not observed. There is even one reversal within a case: The NormOmit-ConsOmit version of VIRUS is higher than the NormAct-ConsOmit version of VIRUS. It would seem more sensible to view the items in terms of continuous variation in the pull of norms and the pull of consequences, rather than in terms of four discrete categories.

Moreover several factor analyses of the responses (with oblimin rotation and either two or three factors)
Figure 1: Proportion of responses in which the act was judged acceptable, for each of the six cases used by GAFCH. The bars are grouped by the GAFCH’s classification of whether the norm implied action or omission and whether the consequences implied action or omission. Names of the cases differ slightly from those used by GAFCH.

consistently showed that factors were distinguished by the norm, not the consequences and not whether norms and consequences were consistent or not.

6 Experiment 1

Of interest is the origin of “perverse” responses, that is, opposition to action when both consequences and norms are supposed to favor it, and support of action when both considerations oppose it. These responses are crucial for deriving conclusions from the CNI model; if they did not exist, the experiment would be identical to the standard experiment that pits consequences against a norm, except for the attempt to switch act and omission. GAFCH argue that these perverse responses arise from extreme bias toward action or inaction. We

---

6 This is also true for the process-dissociation model.
think they are largely due to disagreement with the experimenter about what norms or consequences imply.

We tried to test this in two ways. One was to make up new scenarios, with greater clarity about both norms and consequences. In this effort, we were only partially successful. The other way was to ask subjects directly about norms and consequences in each case. Of interest was whether perverse judgments still occurred when subjects agreed with the experimenters’ classification, or, alternatively, whether judgments were perverse with respect to the subjects’ own classifications.

6.1 Method

We aimed for, and got, exactly 100 subjects from an internet panel of mostly Americans that has done many other experiments, although none exactly like this one. Payment was $5. Ages ranged from 19 to 79, with a median of 51; 72 were women.

We used four version each of 5 cases. Three cases were new, and two were taken verbatim from GAFCH, except for the final questions. In all cases we asked, “What should you/he/she do?” rather than asking what was “acceptable”. The cases are in Appendix A. Each case had four versions, which we label according to whether the norm favors act or omission (by design) and whether the consequences favor act or omission, e.g., “NormAct-ConsOmit”, which represents what we intended to be a conflict between norm and consequences. The first three cases (EUTHANASIA, DATA, TRAIN) were intended to be as clear as possible about both norms and consequences. The last two cases (RANSOM, VIRUS) were taken verbatim from GAFCH, except for the change in the question asked, as noted. The four versions of each case were presented sequentially in a block. The order of the versions was randomized separately within each block. The four cases were also presented in a random order chosen for each subject.

Subjects were also asked about norms, consequences, and which option was an omission. “Not sure” options were included for all questions after the first. For example, a typical page read:

A passenger train is heading for a tree that has fallen on the tracks. The engineer is required by law to stop the train immediately when this happens (the law is designed to ensure consistent safety practices and to prevent the train from derailing).

7In order to construct all four cases, especially the reversed act and omission, we had to avoid the clearest examples in which norms and consequences agree, e.g., “A cruel dictator has 10 hostages. He plans to kill two of them, picked at random. He will spare these two if you kill five, picked at random. Should you kill 5 to save 2?”

8See http://www.sas.upenn.edu/~baron/q.html for the nature of the panel.

9Two of these were modified versions of cases used by Baron et al., 2015.

10“Acceptable” is a natural deontological concept, hence possibly biasing subjects against utilitarian responding. Even intuitive utilitarians may not think in terms of what is permissible/acceptable or forbidden/unacceptable. They may have no special difficulty in deciding between two unacceptable options (any more than between two acceptable options that presented similar conflicts). In addition, tolerant subjects, whether utilitarian or not, may be willing to say that something is acceptable even when they think that the other option is morally better.

11DATA contained a potentially confusing error in the text of the question for the NormAct-ConsOmit version, but our focus here is on the congruent cases. In the reported graphs, we treat this version as we intended it.
Table 1: Proportions of subjects (n=100) whose choice of action/omission disagreed with the classification of items defined as congruent by the experimenters, in the top two rows. The bottom two rows use the subjects’ own classification of whether the items are congruent.

<table>
<thead>
<tr>
<th></th>
<th>Euth</th>
<th>Data</th>
<th>Train</th>
<th>Ransom</th>
<th>Virus</th>
</tr>
</thead>
<tbody>
<tr>
<td>Experimenter judgments:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NormOmit-ConsOmit</td>
<td>0.06</td>
<td>0.23</td>
<td>0.14</td>
<td>0.26</td>
<td>0.26</td>
</tr>
<tr>
<td>NormAct-ConsAct</td>
<td>0.14</td>
<td>0.02</td>
<td>0.04</td>
<td>0.32</td>
<td>0.40</td>
</tr>
<tr>
<td>Subject judgments:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NormOmit-ConsOmit</td>
<td>0.05</td>
<td>0.04</td>
<td>0.05</td>
<td>0.07</td>
<td>0.07</td>
</tr>
<tr>
<td>NormAct-ConsAct</td>
<td>0.02</td>
<td>0.01</td>
<td>0.01</td>
<td>0.04</td>
<td>0.09</td>
</tr>
</tbody>
</table>

The engineer does not think that any passengers would be injured by a sudden stop, but he is sure that hitting the tree would cause some passengers to be thrown from their seats, or get whiplash.

What should he do?
A. Stop the train, following the law. No passengers would be injured.
B. Do nothing. The train would hit the tree. Some passengers would be injured.

Which option produces the better outcome. (Consider the outcome alone, not how it came to happen.)
Clearly option A  Probably A  Not sure  Probably B  Clearly B

Is there a moral rule that favors one option over the other? (Do not count "produces the better outcome" as a rule for this question.)
The rule favors A.
Conflicting rules but mostly favoring A.
Not sure, or no such rule.
Conflicting rules but mostly favoring B.
The rule favors B.
Is one of the options an omission, as distinct from an act?
A is an omission.  Not sure.  B is an omission.

6.2 Results and Discussion

Our main interest is in the responses to the congruent items, which, for GAFCH, often seem to be perverse, disagreeing with both norms and consequences as defined by the experimenters. Table 1 shows the raw rate of such perverse responses for each pair of congruent versions, using both the experimenters’ and the subjects’ classification of norms and consequences. For the subjects, we counted a response as perverse if
either consequences or norms disagreed with the subject’s choice and if neither criterion agreed with it (thus counting cases where one criterion was “not sure”).

It is apparent that, especially for Ransom and Virus (the items from GAFCH) many subjects gave perverse responses when these were defined in terms of the experimenters’ classification, but very few gave such responses according to the subjects’ own classification. This result supports our worry that subjects were misclassified in GAFCH because they did not agree with the intended classification of the versions. Comments from several subjects indicated specific points of disagreement, e.g., for Virus: “I would have specialists from her home country flown in.”; “Quaranteen student.”; “I’m not sure if you are saying the student would die from the side-effects of the drug.”; “You should be able to keep the student from others which will prevent spreading.”. And, for Ransom: “Paying the ransom so the journalist won’t be beheaded is the only choice to make even if the money is used to purchase guns that could cause more deaths. The deaths would probably happen if more guns are purchased or not.”, “I’m not sure there are any rules for this but I believe ransoms should not be paid regardless the situation.”

Our own cases yielded fewer perverse responses in terms of our own classification, thus providing some support for our effort to clarify the norms and consequences. But there was still considerable disagreement, and a few subjects explained it in comments, e.g.: “I believe the patient has a right to not want to die but . . . I believe the patient does not have the right to make someone kill them.”; “The engineer can’t know that hitting the tree will not derail the train and can’t predict the amount of injuries that may be caused in either scenario.” In terms of the subjects’ own classifications, there were, again, very few perverse responses. We thus conclude that GAFCH were, as suspected, drastically overestimating the number of perverse responses, thus also overestimating the possible “biases” that could produce them.
Subjects frequently disagreed with our classification in the incongruent cases as well as the congruent ones, as shown in Figure 2. (Disagreement requires an answer on the “wrong” side of “not sure”.) We have two related explanations for this high rate of disagreement, aside from the possibility that our items were not clear. One, which applies only to incongruent cases (the top two rows of each table in Figure 2), is that subjects engaged in “belief overkill”, that is, convincing themselves that there was no conflict by bringing their judgments into line with their choice. This explanation implies that, in most of these cases, the two judgments would agree with each other and with the choice. Indeed, in 94% of the 644 incongruent cases in which consequence and norm judgments agreed with each other (and were not both wrong, which occurred in only 19 cases), the two judgments supported the choice. Of course, it is also possible that the “error” in judging the norm or consequence occurred before the choice and caused it.

The other explanation is that subjects were careless. To test the role of carelessness we measured the response time for completing each page (to the nearest 0.1 sec) and took the mean of the logs of the fastest 10 pages (i.e., the fastest half). This value, which we call Lrt, correlated $-0.36$ ($p \sim 0.000$) across subjects with the number of errors (disagreements with the experimenter) using all cases. Lrt also correlated $-0.21$ ($p = 0.036$) with the number of responses that were inconsistent with both of the subjects own judgments of norms and consequences. Roughly putting these effects together led to a correlation of $-0.50$ ($p \sim 0.000$) between Lrt and the number of disagreements with the experimenters’ classification, across all cases. (By definition, these included only congruent cases.) These correlations for the five cases were, respectively (in the order of Figure 2): $-0.45$, $-0.31$, $-0.38$, $-0.23$, and $-0.15$; note that they were lower for the two cases from GAFCH, suggesting a role for other factors aside from carelessness.

A surprising result is that performance on the final question on each page, which option was an omission, was poor: overall 37% correct, 14% incorrect, and 49% unsure. The answer was always option B, and its wording always began with “Do nothing”. Several subjects said they were confused by the question, e.g., “I chose not sure on many of the omissions as in not doing anything is still a choice or action.” That said, correct answers correlated .52 with Lrt.

A secondary issue, which this study allows us to address, is the extent to which the scenarios we used can measure individual differences in attention to consequences and norms, and, in particular, in bias toward action or inaction. Note that we can ask about action/inaction bias because of the counterbalanced association of action with the two other attributes, a unique feature of the GAFCH design, but asking this question does not imply any agreement with GAFCH’s assumption that such bias is consistent with utilitarian responding. We measured consistency with norms, with consequences, and with action for the four versions of each of the five scenarios, thus obtaining five measures for each subject of whether judgments were based on norms, consequences, or action. Then we computed coefficient $\alpha$ for each of the attributes: .56 for consequences, .30
for norms, and .28 for action; all of these had 95% confidence intervals that clearly excluded zero, and all but one item-total correlation (dropping the item from the total) was positive (and the negative one was $-0.02$). Importantly, it does appear that there are individual differences in action/inaction bias. Although the overall mean of action bias was negative, indicating a preference for inaction, it was not significantly negative, and some subjects had a positive bias toward action. Such a bias could be misinterpreted as a utilitarian response in studies in which the utilitarian response is always to act.

In sum, our results overall are consistent with our concern that GAFCH are vastly overestimating the number of perverse responses to what they congruent dilemmas. Part of the problem is that the scenarios themselves are not as clear as they could be about both norms and consequences. And part of it is that some subjects are careless. (This may happen in their experiments as well as ours.)

However, it also seems that a bias toward action may exist in some subjects (as well as a bias toward omission in others), so that GAFCH are correct to point out that such an action bias could be misinterpreted as a utilitarian response in commonly used sacrificial dilemmas.

7 Experiment 2

A major problem with constructing cases for both GAFCH and us, in Experiment 1, was that of switching the norm between act and omission. This switch typically required some additional step in the chain between choice and consequence, such as acting to prevent someone else from acting. In principle, it is possible to estimate the CNI model from cases without this switch applied within the same scenario. The idea of matching four cases within a single scenario, so that the model could be estimated for each group of four cases, apparently did not succeed very well anyway, for us or them. Such matching would be helpful if scenarios contributed substantial variance, which could then be controlled, but this variance appeared to be small compared to the variance caused by differences in how the four conditions were constructed. Moreover, our primary concern here is not to fit the CNI model but to question the assumption that the congruent conditions were useful in measuring any bias toward act or omission. We thought that most of the perverse responses were the result of disagreement with the experimental design about whether the cases were in fact congruent.

In Experiment 2, we construct pairs of cases, rather than quadruples, attempting to hold constant whether the norm favors action or omission, and manipulating only the consequences, so as to create a single congruent and a single incongruent case in each pair. Even manipulation of consequences is not necessary for our central point, which is that the high rate of perverse responses to congruent cases is implausible if the cases were interpreted as intended. However, we did try to construct some cases involving omissions (held constant
within each pair) and others involving acts (likewise held constant).

Such pairs of cases are still difficult to construct. General norms rarely if ever require unconditional action, so our cases refer to specific conditions or roles. Moreover, by changing the consequences, we risk changing the norm as well, since it is hard avoid the thought that one norm is to do what yields the best consequences.

We also changed the questions, in an attempt to clarify them:

For which option is the final outcome better. (Consider the outcome alone, not how it came to happen.)

Clearly option A  Probably A  Not sure  Probably B  Clearly B

Are there any moral principles, aside from those that refer to outcomes, that are supposed to guide decisions like this?

At least one moral principle favors A, and no such principles favor B.
At least one moral principle favors B, and no such principles favor A.
No moral principles (other than those that refer to outcomes) are relevant here.

Is one of the options an omission (not doing something), as distinct from an act (doing something)?

A is an omission.  Not sure.  B is an omission.

In addition to the 4 new cases describe in Appendix B, we included two cases from Experiment 1, but with the new wording of questions used here: Data and Virus.\textsuperscript{12} The two cases (with 4 items each) were split into 4 groups of 2. Now, each of the 8 pairs of items matched on the same norm but varied only in which option led to the better outcome; as a result, one item in each pair was congruent and one was incongruent. The order of the 8 pairs was randomized for each subject, and the order of the two members of each pair was randomized for each paper but otherwise presented in adjacent order.

Again, aimed for, and got, exactly 100 different subjects from the same panel as Experiment 1. Payment was $5. Ages ranged from 20 to 75, with a median of 49.5; 70 were women.

7.1 Results and discussion

Table 2 shows the results for the congruent cases in the same form as Table 1. Perverse responding — that is, conflict with the responses we expected, based our definition of consequences and norms — was fairly low for the new cases, and lower still when we used the subjects’ classifications of norms and consequences rather than our own (lower half of Table 2). For the Data and Virus scenarios, perverse responding was substantially higher, especially for Virus (the one from GAFCH), but was greatly reduced when we used the subjects’ classification rather than that of the experimenters. Note that the Data scenario was designed to be like those

\textsuperscript{12}The error in the NormAct-ConsOmit Data case was corrected.
Table 2: Proportions of subjects (n=100) whose choice of action/omission disagreed with the classification of items defined as congruent by the experimenters, in the top two rows. The bottom two rows use the subjects’ own classification of whether the items are congruent. Note that the four new cases (the first four) were pairs, each with only one norm.

<table>
<thead>
<tr>
<th></th>
<th>Jury</th>
<th>Police</th>
<th>Jet</th>
<th>Strokes</th>
<th>Data</th>
<th>Virus</th>
</tr>
</thead>
<tbody>
<tr>
<td>Experimenter judgments:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NormOmit-ConsOmit</td>
<td>0.11</td>
<td>0.04</td>
<td>0.21</td>
<td>0.29</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NormAct-ConsAct</td>
<td>0.05</td>
<td>0.05</td>
<td></td>
<td></td>
<td>0.08</td>
<td>0.28</td>
</tr>
<tr>
<td>Subject judgments:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NormOmit-ConsOmit</td>
<td>0.03</td>
<td>0.04</td>
<td>0.07</td>
<td>0.04</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NormAct-ConsAct</td>
<td>0.02</td>
<td>0.04</td>
<td></td>
<td></td>
<td>0.02</td>
<td>0.09</td>
</tr>
</tbody>
</table>

Figure 3: Proportion of subjects who disagreed with the experimenter’s classification in terms of norms and consequences.

Figure 3 shows the disagreements with the intended classification of items in terms of norms and consequences. Some of these disagreements are likely due to carelessness. The correlation with Lrt is $-0.201 (p = 0.045)$ for the total number of disagreements. The correlation was especially great ($-0.476$) for judgments of consequences for congruent items.

However, it appears that there were two other major sources of disagreements. First, as indicated from comments, subjects did not distinguish norms and consequences in the same way that philosophers do, or had trouble knowing what we had in mind. For example: “The duty of a police officer is to enforce the law. But if doing so you get an innocent person convicted then it goes against what is morally right.”; “Any example of a moral principle other than those that refer to moral outcomes would be nice since I do not know how you define these in the context of the questions.”; “Are there any moral principles aside from those that refer to outcomes that are supposed to guide decisions like this? — This is a difficult question to answer.” In other
cases, subjects added information not stated in the scenario: “Drug might have other side effects.”; “[I]f I turn in the evidence there will be possible repercussions for me.”; “Move her with as many possible safety measures for others as possible.”.

Second, subjects clearly engaged in what looked like “belief overkill” (Baron, 2009). That is, they manipulated their perception of norms or consequences so that the two judgments pointed to the same response option. Across the 8 incongruent items, the intended correlation of norms and consequences was −1.00. This follows from what we mean by “incongruent” and from the fact that we manipulated both intended norms and consequences. However, the mean within-subject correlation between the judged norms and consequences was positive (.365, $p \sim .000$ by t test across subjects, against the null hypothesis of 0).13

Subsequent studies would do better to define “norms” (principles, etc.) and “consequences” with examples.

Subjects also again had a different concept of “omission”. On average, they were “correct” 70% of the time in classifying the second response as an omission. Some of this may be due to carelessness; the correlation between correct classifications and Lrt was .277 ($p = .005$). However, it is also clear from subjects’ comments that there were conceptual issues too: “For part 4 of this question no option is an omission here. Why is that not a choice?”; “Not giving out the medication is not an omission since it is not medically prudent. An omission has to be theoretically considered possible and no one would want to cause an increase of 700 strokes!” “Neither option is an omission! Why is there no option for me to choose that? The best thing to do is isolate the volunteer with no medication so she will live.” We should not have been surprised by these difficulties, since Spranca, Minsk & Baron (1991, Expt. 6) found substantial variety in what subjects took to define an omission.

As in Experiment 1, we also computed the reliability of reliance on norms, consequences and action across the 8 scenarios. These were close to the values found in Experiment 1: consequences .59, norms .32, action .42. All but one item-total correlation (dropping the item from the total) was positive (norms for Virus when action was the norm, −.01). This time, the mean bias toward action was positive tested across subjects (mean of .07, where 1 indicates always choosing action and −1 always choosing omission; $t_{99} = 2.34, p = .022$).

Thus, individual differences in preferences for acts or omissions, as proposed by GAFCH, do seem to exist, and a bias toward action could be misinterpreted as a utilitarian response in the usual sacrificial dilemmas. Again, these results do not contradict our claim that biases toward action or omission can be one cause of non-utilitarian responding.

Experiment 2 provides further evidence that subjects’ disagreement with the experimenter’s classification of norms and consequences, mostly based on the difficult design of the dilemmas, led GAFCH to considerably
overestimate the number of perverse responses. Moreover, the correlations we found with response time suggest that the major cause of perverse responses, at least in our sample of subjects, is carelessness rather than malevolence. That said, carelessness itself could result from a negative attitude toward the experimenters, which itself is a form of malevolence, so the line is hard to draw.

8 What we already know about omission bias

GACFH address the question of whether worse consequence responses in sacrificial dilemmas are the result of a bias against action. They write as if little attention had been paid to this problem until very recently. For example, GACFH list a few recent attempts to manipulate consequences, claiming that such manipulations have been rarely attempted. They thus ignore the fact that manipulations of this sort appear in papers that they cited (Spranca, Minsk, & Baron, 1991, Experiment 3) as well as older papers that they do not cite (Ritov & Baron, 1990; Baron & Ritov, 1994; Baron, 1995; Ritov & Baron, 1995; Baron & Ritov, 2004). Many of the experiments reported in these papers were done to address the main issue raised by GACFH, whether the original omission bias result was due to a bias against action or an unwillingness to cause harm for the sake of the greater good. And still other earlier studies bear on this question as well (Royzman & Baron, 2002; Baron & Ritov, 2009). Of course, the basic result in standard sacrificial dilemmas also relies on a manipulation of consequences: action leads to better consequences than omission, yet omission is chosen.

Here is what we know based on these earlier studies.

1. Non-utilitarian responses are affected by other factors when the act-omission distinction cannot by itself cause them. Ritov and Baron (1990, Experiment 1) observed a reluctance to vaccinate children against a disease when the vaccine itself would cause the death of some children (not necessarily the the same children who would have been killed by the disease). However, this omission bias was greatly reduced when the children who would be killed by the vaccine were also the ones who were susceptible to being killed by the disease in the first place. Both this condition and the standard condition were matched in terms of the number harmed by action and the number of harms prevented by action. Thus, a bias toward inaction, by itself, could not account for the original result. This comparison has an additional point, of course: it shows that much of the non-utilitarian bias is due to causing harm that would not be caused anyway.

Similarly, Baron and Ritov (1994, Experiment 4) compared the original vaccination case with a “vaccine failure” case, in which the deaths that result if the vaccination is chosen are not caused by the vaccine itself but rather by its not being fully effective (thereby failing to prevent some harm). Again, the numbers harmed under the action option and under the omission option were matched, but the bias against vaccination (action) was much stronger in the original condition in which the harm was caused by the vaccination itself. This shows
that the causal role of the action is important in non-utilitarian choices, holding constant the consequences of acts and omissions.

Royzman and Baron (2002) compared cases in which an action caused direct harm with those in which an action caused harm only indirectly, with the harm actually caused by a side effect. For example, in one case, a runaway missile is heading for a large commercial airliner. A military commander can prevent it from hitting the airliner either by interposing a small plane between the missile and the large plane or by asking the large plane to turn, in which case the missile would hit a small plane now behind the large one. The indirect case (the latter) was preferred. In Study 3, subjects compared indirect action, direct action, and omission (i.e., doing nothing to prevent the missile from striking the airliner). We found substantial omission bias for direct action, but very little omission bias when the action was indirect. Once again, the causal role of the act causing harm was important, holding the act-omission distinction constant. Many of the results just described were replicated using somewhat different methods by Baron and Ritov (2009, Study 3); in this study, judged causality of the action was the main determinant of omission bias.

Finally, Baron, Scott, Fincher, and Metz (2015) and Baron, Gürcay, & Luce (2017) used dilemmas that pitted two actions against each other. Each dilemma pitted an action with the better outcome against an action that violated a moral rule (often a legal rule), but they did not involve any manipulation of numbers. For example, one case involved a person deciding whether to testify for the prosecution at an insider trading trial. The person knows for sure that the defendant is innocent, but also that if he says what he knows, the defendant will be wrongly convicted (based on the incorrect testimony of other witnesses). The person must decide whether to obey the law and tell the truth, as he swore he would do, thus leading to the conviction of the defendant (the deontological option), or instead, to break the law and remain silent (the utilitarian option). Dilemmas of this sort contrast a deontological rule with a utilitarian calculation, but they differ from standard sacrificial dilemmas in that sympathy aligns with the utilitarian choice. Choice of the non-utilitarian options in these dilemmas can therefore not be explained by sympathy. However, non-utilitarian choices in these dilemmas correlated positively with choice of the non-utilitarian options in standard sacrificial dilemmas and with a scale that measured general utilitarian beliefs. This result therefore suggests that there is a particular attachment to deontological rules that guides some subjects’ judgments.

In sum, a bias toward the default ( omission) plays some role in explaining the existence of non-utilitarian responses, but it has now been clearly demonstrated that other factors are relevant too. One major determinant has something to do with the perception of direct causality (as also argued by Greene et al., 2009), and another has something to do with an attachment to particular moral rules (as also argued by GACFH, and by other results in Baron & Ritov, 2009, concerning protected values).

2. The bias toward omissions, such as it is, can also be analyzed in terms of two factors. One is in fact
a bias toward omissions, which has been studied by itself in other contexts, where it is called (or should be called) the default bias, a bias toward whatever you get if you don’t do anything (e.g., Johnson & Goldstein, 2003). The other is an amplification effect: the consequences of action are weighed more heavily than the consequences of omission (Landman, 1987; Gleicher et al., 1992; Spranca et al., 1992, Experiment 3; Baron & Ritov, 1994). These two factors work together when the outcomes are perceived as losses, but they oppose each other when the outcomes are perceived as gains. Sometimes the amplification effect is stronger, so there is a small bias toward beneficial action (Baron & Ritov, 1994; Ritov & Baron, 1995). GACFH tend to conflate these two factors, and speak of a general bias toward omissions as if that had a single cause.14

In sum, we do not need the CNI model in order to assess the role of action/omission bias in findings of non-utilitarian outcomes. A bias toward omissions (a default bias) is sometimes part of the story, but certainly not the whole story. And, even if it were the whole story, the basic claim established by past research, that people sometimes follow deontological rules even when they lead to worse consequences, would still stand.

References


---

14In the present experiments, the default bias appears essentially nonexistent. This may be a consequence of presenting both options as alternatives, rather than asking about only one of them, leaving the other implicit, as is usually done.


Appendix A: Cases used in Experiment 1

**EUTHANASIA-NormAct-ConsOmit:**

You are a doctor. An elderly patient has been in constant pain for months, and nothing seems to help. **The patient has been begging you to help him die**, for several weeks. Assisted suicide and euthanasia are illegal in your state. There are two pain-killing drugs that the patient could take. He has been taking one that is fairly safe but provides little pain relief. The other drug would relieve the pain but cause the patient to drift off and die within a few hours. You see that **the instructions for the night nurse are, by mistake, to give the drug that would cause death**. What should you do?

A. Change the instructions, thus making sure that the law was followed. The patient would continue to suffer.

B. Do nothing. The patient would die a painless death. You would not be prosecuted, because the prosecutor could not prove that you noticed the mistake.

**EUTHANASIA-NormOmit-ConsAct:**

[Same as EUTHANASIA-NormAct-ConsOmit up to here:]

You see that **the instructions for the night nurse are to give the safe drug**. What should you do?

Change the instructions, thus breaking the law. The patient would die a painless death. You could say that the drug was chosen because it was a better pain reliever, thus escaping prosecution of you or the nurse, even though this was not your reason and in fact you are breaking the law.

B. Do nothing. The patient would get the safe drug and continue to suffer.

**EUTHANASIA-NormOmit-ConsOmit:**

You are a doctor. An elderly patient has been in constant pain for months, and nothing seems to help. The patient’s wife has been begging you to help him die, for several weeks, but **the patient himself does not want to die**, despite the pain. Assisted suicide and euthanasia are illegal in your state. There are two pain-killing drugs that the patient could take. He has been taking one that is fairly safe but provides little pain relief. The other drug would relieve the pain but cause the patient to drift off and die within a few hours. You see that **the instructions for the night nurse are to give the safe drug**.
What should you do?

A. Change the instructions, thus breaking the law. The patient would die a painless death. You could say that the drug was chosen because it was a better pain reliever, thus escaping prosecution of you or the nurse, even though this was not your reason and in fact you are breaking the law.

B. Do nothing. The patient would get the safe drug. The law and the patient’s wishes would be honored.

EUTHANASIA-NormAct-ConsAct:

[Same as EUTHANASIA-NormOmit-ConsOmit up to here:]

You see that the instructions for the night nurse are, by mistake, to give the drug that would cause death.

What should you do?

A. Change the instructions, thus making sure that the law and the patient’s wishes were honored.

B. Do nothing. The patient would die a painless death. You would not be prosecuted, because the prosecutor could not prove that you noticed the mistake.

DATA-NormAct-ConsOmit:

With funding from a government agency, Nandita has collected data that are relevant to changes in a government regulation concerning health. The law requires that she make the data available on the Web (with subject identifiers removed), and she has promised to do this. The agency in question must refer to data when it makes any changes, and it cannot use data that are private, so they require all data collection they have funded to be publicly available.

But Nandita learns that the change, if made, would prevent many people from getting a life-saving and cost-effective treatment. She has not told anyone else about the data, and she could just do nothing at this point. If Nandita makes the data public, the change will be made, and many people will be unable to get the treatment.

What should she do?

A. Make the data public, as she has promised to do (and as required by law). The change will be made, and many people will be able to get a life-saving and cost-effective treatment.15

B. Do nothing. The change will not be made.

DATA-NormOmit-ConsAct:

With funding from a government agency, Nandita has collected data that are relevant to changes in a government regulation concerning health. The law requires that she keep the data private (because of privacy concerns, even with subject identifiers of removed), and she has promised to do this. The agency in question must

---

15This item should have said “will not be able”. Some subjects pointed out our error. The error does not affect our analysis of congruent cases. In reporting other results, we did not delete responses to this item, because we assume that most subjects were unaffected by our mistake.
refer to data when it makes any changes, but it cannot use data that have violated this rule by becoming publicly available.

But Nandita learns that the change in question, if made, would **prevent many people from getting a life-saving and cost-effective treatment.** If Nandita makes the data public, the change will not be made.

What should she do?

A. Make the data public, thus breaking the law and her promise to leave them private. The change will be made, and many people will be able to get a life-saving and cost-effective treatment.

B. Do nothing. The change will be made.

**DATA-NormOmit-ConsOmit:**

With funding from a government agency, Nandita has collected data that are relevant to changes in a government regulation concerning health. **The law requires that she keep the data private (because of privacy concerns, even with subject identifiers of removed), and she has promised to do this.** The agency in question must refer to data when it makes any changes, but it cannot use data that have violated this rule by becoming publicly available.

Nandita learns that the change in question, if made, would **allow many people to get a life-saving and cost-effective treatment.** If Nandita makes the data public, the change will not be made.

What should she do?

A. Make the data public, thus breaking the law and her promise to leave them private. The change will not be made, and many people will be unable to get a life-saving and cost-effective treatment.

B. Do nothing. The change will be made.

**DATA-NormAct-ConsAct:**

With funding from a government agency, Nandita has collected data that are relevant to changes in a government regulation concerning health. **The law requires that she make the data available on the Web (with subject identifiers removed), and she has promised to do this.** The agency in question must refer to data when it makes any changes, and it cannot use data that are private, so they require all data collection they have funded to be publicly available.

Nandita learns that the change in question, if made, would **allow many people to get a life-saving and cost-effective treatment.** She has not told anyone else about the data, and she could just do nothing at this point. If Nandita makes the data public, the change will be made, and many people will be able to get the treatment.

What should she do?
A. Make the data public, as she has promised to do (and as required by law). The change will be made, and many people will be able to get a life-saving and cost-effective treatment.

B. Do nothing. The change will not be made.

**TRAIN-NormAct-ConsOmit:**

A passenger train is heading for a tree that has fallen on the tracks. The engineer is required by law to stop the train immediately when this happens (the law is designed to ensure consistent safety practices and to prevent the train from derailing). But the engineer thinks that if he stops the train suddenly it will cause several passengers to be thrown out of their seats, or get whiplash. The engineer also thinks that hitting the tree will not itself derail the train, harm any passengers, or damage the train beyond a few small dents, so that doing nothing and allowing the train to hit the tree would be less risky. The engineer himself would not get in any trouble for having caused the dents.

What should he do?

- A. Stop the train, following the law. Some passengers would be injured.
- B. Do nothing, thus breaking the law. The train would hit the tree. No passengers would be injured

**TRAIN-NormOmit, ConsAct:**

A passenger train is heading for a tree that has fallen on the tracks. An automatic system will stop the train immediately when this happens. It is possible to override the automatic system manually, although that is against the law (the law is designed to ensure consistent safety practices and to prevent the train from derailing).

But the engineer thinks that if the automatic system stops the train suddenly it will cause several passengers to be thrown out of their seats, or get whiplash. The engineer also thinks that hitting the tree will not itself derail the train, harm any passengers, or damage the train beyond a few small dents, so that overriding the automatic system so that the train hits the tree would be less risky. The engineer himself would not get in any trouble for having caused the dents.

What should he do?

- A. Override the automatic system. The train would hit the tree. No passengers would be injured.
- B. Do nothing, thus following the law. The train would stop suddenly. Some passengers would be injured.

**TRAIN-NormOmit-ConsOmit:**

A passenger train is heading for a tree that has fallen on the tracks. An automatic system will stop the train immediately when this happens. It is possible to override the automatic system manually, although that is against the law (the law is designed to ensure consistent safety practices and to prevent the train from derailing).
The engineer does not think that any passengers would be injured by a sudden stop, but he is sure that **hitting the tree would cause some passengers to be thrown from their seats, or get whiplash**.

What should he do?

A. Override the automatic system. The train would hit the tree. Some passengers would be injured.
B. Do nothing. The train would stop and not hit the tree. No passengers would be injured.

**TRAIN-NormAct-ConsAct:**

A passenger train is heading for a tree that has fallen on the tracks. **The engineer is required by law to stop the train immediately** when this happens (the law is designed to ensure consistent safety practices and to prevent the train from derailing).

The engineer does not think that any passengers would be injured by a sudden stop, but he is sure that **hitting the tree would cause some passengers to be thrown from their seats, or get whiplash**.

What should he do?

A. Stop the train, following the law. No passengers would be injured.
B. Do nothing. The train would hit the tree. Some passengers would be injured.

**RANSOM-NormAct-ConsOmit:**

You are the president of your country. A guerilla group operating in a conflict zone has abducted a journalist from your country and threatens to behead him if your government does not pay a ransom of one million dollars. The group will use the ransom money to buy weapons for their guerilla war, which will cause the deaths of many people. As the president, you have the power to approve payment of the ransom.

What should you do?

A. Approve the ransom payment.
B. Do nothing.

**RANSOM-NormOmit-ConsAct:**

You are the president of your country. A guerilla group operating in a conflict zone has abducted a journalist from your country and threatens to behead him if your government does not pay a ransom of one million dollars. The group will use the ransom money to buy weapons for their guerilla war, which will cause the deaths of many people. Congress has approved payment of the ransom, but you have the power to veto the payment.

What should you do?

A. Veto the ransom payment.
B. Do nothing.
RANSOM-NormOmit-ConsOmit:
You are the president of your country. A guerilla group operating in a conflict zone has abducted a journalist from your country and threatens to behead him if your government does not pay a ransom of one million dollars. The group will use the ransom money to buy food for their families, who live in an area that has been plagued by several draughts. Congress has approved payment of the ransom, but you have the power to veto the payment.

What should you do?
A. Veto the ransom payment.
B. Do nothing.

RANSOM-NormAct-ConsAct:
You are the president of your country. A guerilla group operating in a conflict zone has abducted a journalist from your country and threatens to behead him if your government does not pay a ransom of one million dollars. The group will use the ransom money to buy food for their families, who live in an area that has been plagued by several draughts. As the president, you have the power to approve payment of the ransom.

What should you do?
A. Approve the ransom payment.
B. Do nothing.

VIRUS-NormAct-ConsOmit:
You are the director of a hospital in a developing country. A foreign student who is volunteering in the country got infected with a rare virus. The virus is highly contagious and deadly to seniors and children. The student suffers from a chronic immune deficiency that will make her die from the virus if she is not returned to her home country for special treatment. However, taking her out of quarantine involves a considerable risk that the virus will spread.

What should you do?
A. Take the student out of quarantine to return her to her home country for treatment?
B. Do nothing.

VIRUS-NormOmit-ConsAct:
You are the director of a hospital in a developing country. A foreign student who is volunteering in the country got infected with a rare virus. The virus is highly contagious and deadly to seniors and children. The only medication that can effectively stop the virus from spreading has severe side-effects. Although the virus will not kill her, the student suffers from a chronic immune deficiency that will make her die from these side-effects.
What should you do?
A. Give the student the medication.
B. Do nothing.

VIRUS-NormOmit-ConsOmit:
You are the director of a hospital in a developing country. A foreign student who is volunteering in the country got infected with a rare virus. The virus is highly contagious and can cause severe stomach cramps. The only medication that can effectively stop the virus from spreading has severe side-effects. Although the virus will not kill her, the student suffers from a chronic immune deficiency that will make her die from these side-effects.
What should you do?
A. Give the student the medication.
B. Do nothing.

VIRUS-NormAct-ConsAct:
You are the director of a hospital in a developing country. A foreign student who is volunteering in the country got infected with a rare virus. The virus is highly contagious and can cause severe stomach cramps. The student suffers from a chronic immune deficiency that will make her die from the virus if she is not returned to her home country for special treatment. However, taking her out of quarantine involves a considerable risk that the virus will spread. What should you do?
A. Take the student out of quarantine to return her to her home country for treatment.
B. Do nothing.

Appendix B: Cases used in Experiment 2

JURY-NormAct-ConsAct:
You have been called for jury duty and you sent in the form agreeing to show up on a certain date. As it happens, it was difficult to find 12 jurors on that date. Thus, if you don't show up, 11 others would be sent home after setting aside the whole day for this obligation. The day arrives and you don't feel like going. You would rather do what you usually do when you stay home during the day. What should you do?
A. Show up anyway. You and the 11 others would do your duty.
B. Do not show up. 11 others will be sent home after setting aside the whole day.

JURY-NormAct-ConsOmit:
You have been called for jury duty and you sent in the form agreeing to show up on a certain date. As it happens, plenty of extra jurors were found for that date, so everything would proceed normally if a few jurors did not show up at all. The day arrives and you realize that you must stay home to wait for an important phone call concerning a medical decision for an incapacitated relative, for whom you have the power to make decisions. You expected the call on the previous day, but it has not come through yet. What should you do?

A. Show up for jury duty anyway. You will probably miss the phone call, and your relative could suffer harm as a result.

B. Do not show up. The court would proceed as normal, and you will be home if the phone call comes in.

POLICE-NormAct-ConsAct:

You are a police officer. You, acting alone, have just confiscated bomb-making equipment from the home of a suspected terrorist. The equipment is now in your home. You are supposed to turn in the equipment so that it can be used as evidence to convict the suspect. What should you do?

A. Turn in the equipment. The suspect will surely be convicted.

B. Do nothing. The suspect will probably get away with the crime.

POLICE-NormAct-ConsOmit:

You are a police officer. You, acting alone, have just confiscated bomb-making equipment from the home of a suspected terrorist. The equipment is now in your home. You are supposed to turn in the equipment so that it can be used as evidence to convict the suspect, but you know for sure that the suspect is innocent and would probably be falsely convicted if you turn in the equipment. (You know this in a way that cannot be used in court or shared with your superiors.) What should you do?

A. Turn in the equipment. The suspect will probably be falsely convicted.

B. Do nothing and say nothing.

JET-NormOmit-ConsAct:

You are an air traffic controller. A guided missile has been accidentally launched that will kill 100 airplane passengers in a large jet. Directing a second jet into the path of the missile, to take the hit, would save the 100 passengers, but the second jet has 25 passengers, who would die from the missile. What should you do.

"Direct the second jet into the missile’s path, thus killing all 25 in the second jet but saving the 100 in the first jet.”.

"Do nothing. The missile will hit the first jet and 100 will die.”

JET-NormOmit-ConsOmit:

You are an air traffic controller. A guided missile has been accidentally launched that will kill 25 airplane passengers in a small jet. Directing a second jet into the path of the missile, to take the hit, would save the 25
passengers, but the second jet has 100 passengers, who would die from the missile. What should you do.

"Direct the second jet into the missile’s path, thus killing all 100 in the second jet but saving the 25 in the first jet."

"Do nothing, the missile will hit the first jet and 25 will die."

**STROKES-NormOmit-ConsAct:** You are the administrator of a government hospital system (like the VA in the U.S.). 1000 emergency room patients in government hospitals suffer debilitating strokes each year. Giving a new drug to all emergency room patients with stroke symptoms would prevent these 1000 strokes, but the drug itself would cause debilitating strokes in 300 patients, probably not the same ones whose strokes would be prevented. What should you do?

A. Order that the new drug be used. 1000 strokes would be prevented but 300 would be caused by the drug.

B. Do nothing. 1000 strokes would occur.

**STROKES-NormOmit-ConsOmit:** You are the administrator of a government hospital system (like the VA in the U.S.). 300 emergency room patients in government hospitals suffer debilitating strokes each year. Giving a new drug to all emergency room patients with stroke symptoms would prevent these 300 strokes, but the drug itself would cause debilitating strokes in 1000 patients, probably not the same ones whose strokes would be prevented. What should you do?

A. Order that the new drug be used. 300 strokes would be prevented but 1000 would be caused by the drug.

B. Do nothing. 300 strokes would occur.