Judgment and Decision Making, Vol. 15, No. 3, May 2020, pp. 421–442

Consequences, norms, and inaction: A critical analysis

Jonathan Baron* Geoffrey P. Goodwin†

Abstract

Gawronski, Armstrong, Conway, Friesdorf and 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 (which causes some harm in order to prevent greater harm), 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 dilemmas 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 process. It thus should be included as an explanation of deontological responding, not an alternative response type. 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 course of action is consistent with a moral norm. This problem is almost inherent in the attempt to switch act and omission while maintaining equivalent norms. We support this argument with data from experiments with new and old scenarios, in which we asked subjects to judge both norms and consequences. We also find that GACFH's results are not as consistent as they appear to be in the paper.

Keywords: utilitarianism, deontology, omission bias, process dissociation

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.[1] 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 refers to a category of moral theories that emphasize properties of action other than their consequences, such as whether an action violates basic rights, or whether an action 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. The model is similar to, and builds upon, an earlier model based on the idea of process dissociation (PD; Conway & Gawronski, 2013). In this paper, we discuss largely the single paper that introduces the CNI model, but some of our comments apply to other papers using the CNI model or the PD model as applied to moral judgments. We do not attempt to discuss or even enumerate all these papers (many of which have appeared while the present paper was under review.). Instead, we intend this paper as an expression of

We thanks Andreas Glöcker, the reviewers, and Ilana Ritov for helpful comments.

*Department of Psychology, University of Pennsylvania, 3720 Walnut St., Philadelphia, PA, 19104. Email: baron@upenn.edu. ORCID: 0000-0003-1001-9947.

†Department of Psychology, University of Pennsylvania.

[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.

https://doi.org/10.1017/S193029750000721X Published online by Cambridge University Press

general concern about the use of both methods to study moral judgment.

In a typical sacrificial 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 responses 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 pairing 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 a sacrificial 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 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 resulting cases and incongruent (conflicting) for the other two. The pattern of subjects' responses to the four dilemmas indicates how much each principle is driving their 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.

The PD model, on which the CNI model builds, uses only one type of congruent case, in which both norms and consequences prescribe inaction. For example, the act of killing is not only proscribed by a norm against killing but actually leads to more deaths than doing nothing. This case is compared to a standard sacrificial dilemma. Thus, the PD model contrasts two types of cases instead of four. It does not try to reverse the usual consequences of acts and omissions. This reversal is what generates the two additional cases. In the CNI model, "perverse" responses (those based on neither norms nor consequences) to congruent cases can arise from a preference for action, for one congruent case (in which both norms and consequences proscribe action), or from a preference for inaction, for the other (in which both norms and consequences prescribe action).

In this paper, we discuss two general concerns with this approach. We begin with the main philosophical assumption made by GACFH, which is that a bias against action is not deontological. We argue that whether an option involves action or omission, in itself, is a feature of the sort that deontology can take into account but utilitarianism cannot. We then discuss the difficulty of constructing items that meet the new requirements, and other problems with assessing the relationship between model parameters and other variables. We then present evidence that the apparent results of the CNI model are affected by subjects' lack of agreement with the experimenters about which norms and consequences are relevant for the choice, as well as by inconsistencies in the data.

However, we agree that the CNI model raises an important question about the kind of deontological judgments that lead to apparent rejections of utilitarian options, specifically whether they arise from norms that concern action vs. omission as well as other features. We review other literature that tries to answer this question along with related questions about the nature of apparent deontological responses.

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 or principles about choices (acts or omissions) other than those concerned with consequences. Most characteristically, it concerns properties of acts that make them proscribed or forbidden. It can also concern properties of acts that make them prescribed (required), as in the case of duties. The rules may be, but 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.

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 action/omission dilemmas, in which both options lead to some harm, would therefore appear as a bias against action. Such a bias would count as a deontological principle,

https://doi.org/10.1017/S193029750000721X Published online by Cambridge University Press

according to what we take to be the standard definition, because it affects choices for reasons other than consequences.

GACFH ask whether subjects in an experiment consider each type of principle (deontological or utilitarian) in isolation, or neither. When neither principle applies, GACFH suppose that the decision is based on a bias toward action or inaction. A major stumbling block for this effort is that it turns out to be much more difficult than GACFH suppose successfully to disentangle the alleged action/inaction bias from a deontological response tendency.

One underlying reason for this is that GACFH's claim that action biases are conceptually dissociable from a deontological response tendency is questionable. Indeed, we are not sure that such a dissociation is possible. To illustrate why, several considerations are relevant. Consider first that there are some well-known contexts in which a deontological response is definitionally identical with a bias against action. Indeed, deontological and utilitarian theories fundamentally dispute the moral relevance of acts and omissions in many morally pertinent situations.

Deontological rules are mostly 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, according to which 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). In this context, there is no sensible way to conceive of deontology and consequentialism as independent of one another, nor is it possible to dissociate deontological theorizing from its view of the difference between action and omission.[2]

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, GACFH are mistaken to argue that, in these contexts, the action-omission distinction can be conceived of as a third factor, entirely separate from deontological ethics.

People often prefer an option with worse consequences, and this finding, 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. But, if it should turn out that the effect is entirely due to a bias against action, then 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), although, as we explain later, it didn't turn out to be that simple.

Thus, in our view, it is a mistake to regard deontology and utilitarianism as independent of one another, as GACFH propose. 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). Deontology, in contrast, is by definition opposed 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), for instance, that one option is a distinct action whereas the other is an omission.

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, as we define them, will lead to outcomes that are relatively worse. This happens when deontology prohibits an action that would lead to better consequences. The attraction of deontology for decision makers could be one reason why things are sometimes not as good as they could be. We define deontology as anything systematically opposing utilitarian choices because one purpose of research on moral judgment is to look for reasons why things sometimes do not turn out so well. This definition derives from that purpose and is thus not arbitrary.

3 Separating norms and consequences

We turn now to the sacrificial dilemma context to illustrate a corresponding problem in the context of the switched cases that GACFH and others rely on.

In one of the most well-known sacrificial dilemmas, a speeding trolley is hurtling towards five workers on a track. A target individual can choose to switch the trolley to a side track, thereby sparing the five workers originally in danger, but killing a lone worker on the side track. Subjects must

[2] In footnote 3, GACFH suggest that the "doctrine of doing and allowing" would lead to a bias against harmful action, as distinct from a general bias against action. But we argue here that a general bias against action is also deontological, since it is a property of options other than their consequences.

https://doi.org/10.1017/S193029750000721X Published online by Cambridge University Press

decide what is the morally right thing for the target individual to do. In this problem, the utilitarian option involves action: divert the trolley onto the side-track so it kills only one person, whereas the deontological option – motivated by an injunction against actively causing harm – involves omission: do nothing and let the trolley kill five people.

An action that causes less harm is therefore contrasted with an omission that causes greater harm. It is this detail that inspires GACFH's critique. They and others (Crone & Laham, 2017) argue that this sort of dilemma involves an inherent "action confound" – actions, but not omissions are systematically associated with better consequences in these vignettes. Accordingly, it is therefore alleged that it is impossible to know whether subjects' responses might ultimately be driven by this action confound.

To sidestep this problem, these researchers attempt to create a corresponding, reversed case, which switches the association between actions-omissions and utilitarian-deontological options. One way to achieve this is to make the target individual no longer in control of the switch, but rather, a bystander watching someone else control the switch (see Crone & Laham, 2017, for a particularly clear illustration). The switch controller is about to switch the trolley (thus taking the utilitarian option), and the bystander must decide whether to intervene (act) in order to stop them, or instead, to do nothing. In this reversed case, doing nothing now causes better consequences, whereas acting causes worse consequences – thereby reversing the usual association between action (versus omission) and better (versus worse) consequences.

This maneuver, though clever, faces an inherent problem. The overall goal is to keep the consequences identical and just switch the association between action-omission and utilitarian-deontological options. But it is also critical that the researchers are also able to keep the relative strength of the two (deontological) moral norms equivalent across the standard and switched cases. This is where things go awry. There is a much stronger moral norm prohibiting the active killing of a person, than there is a moral norm prescribing acting to prevent someone else from actively killing someone. Thus, in making this switch, the comparison is not quite "fair" because the two moral norms being compared differ substantially in their perceived strength. The proscriptive norm is inherently stronger than the prescriptive norm – the supposed "action bias" is built into deontological norms in the first place. As a consequence, there is no way to detect an "action bias" that is supposedly independent of the deontological norms (and consequences) in consideration here. (The only way to do so would be to show by some independent means that the norms are equivalent in perceived strength, and yet there is still an action bias. But this is not attempted.)

In essence, while one can keep consequences constant in these switched dilemmas, it is not clear that it is feasible to keep norm strength constant. This problem simply reflects the fact that deontology is characteristically defined over properties of actions, and much less characteristically focused on properties of omissions. Because of this, we are somewhat skeptical about the likely success of using such switched dilemmas. At a minimum, it would be important to measure both perceived consequences and perceived norm strength, and hold both constant in order to detect a separate "action bias" that is allegedly independent. This is not done in any of the research we are aware of. There are also additional, related problems, which we highlight in the following examples.

GACFH do not focus on the trolley problem, but instead use somewhat more realistic dilemmas (as do Crone & Laham, 2017). Like Crone and Laham, one way in which GACFH attempt to separate the effects of norms and consequences from those of response bias is by constructing vignettes in which norms and consequences are supposed to conflict, but the usual association of norms with inaction and consequences with action is reversed. These two sorts of cases are called "incongruent" cases, and they are contrasted with two sorts of cases in which norms and consequences point in the same direction – "congruent" cases. We discuss the congruent cases below, but first illustrate some additional problems with the switched incongruent cases.

Consistent with Crone and Laham's strategy, 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 GACFH's 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 reversing the benefits of action relative to inaction. But, this switching has a problem. Consistent with the above analysis, it is unlikely that the prescriptive norm to block another surgeon from killing innocent patients is anywhere near equal in perceived strength to the proscriptive norm prohibiting a surgeon from killing of innocent patients. GACFH do not check on how strong subjects perceive these respective norms to be.

Furthermore, 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

https://doi.org/10.1017/S193029750000721X Published online by Cambridge University Press

fact that someone else has made a decision provides reason to think that he or she knew something that we did not know. As these various considerations suggest, choices attributed to norms could instead be based on consequences.

Thus, relying on the blocking of someone else's decision not only fails to hold norm strength constant, it also does not ensure that the consequences of action are held constant. In fact, it is quite 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.[3]

Moreover, GACFH do not check on which norms subjects think are relevant to each dilemma. There may 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 GACFH's 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 scenario, 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 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. In fact, there is a strong contrary norm not to give in to such requests, as doing so 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.

[3]One early attempt comes from Ritov and Baron (1994) who 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.

In one version of GACFH's police torture scenario, 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." In this case, action to intervene is supposedly (exclusively) prescribed by a norm. However, consideration of the consequences might also dictate this action. 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, but 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.

Similarly, in one version of GACFH's immune-deficiency dilemma (which we call VIRUS, below), 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." In this version the norm is supposed to support omission but the consequences are supposed to support action. However, it seems somewhat unclear what the available choices (and their 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 one in which norms and consequences are congruent — both moral norms and consideration of consequences dictate not using the medication. Also relevant to our concerns, the congruent version in which both the norm and consequences are supposed to 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." Subjects are then asked, "Is it acceptable to take the student out of quarantine to return her to her home country for treatment?" Although norms and consequences are both supposed to support action in this case, a careless subject might see the last sentence before the question as a hint that the action — taking her out of quarantine — might

https://doi.org/10.1017/S193029750000721X Published online by Cambridge University Press

not have better consequences overall, and the term "acceptable" in the question might suggest that the director should have discretion, 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 scenarios, as well as similar scenarios used in process-dissociation (PD) 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., because of the consequences! And some subjects might disagree about what the relevant norms and consequences actually are, which would convert apparently conflicting vignettes into congruent ones. The fact that subjects on the average are sensitive to the norm manipulations does not imply that they interpreted the norms as intended, nor that subjects who respond very differently are interpreting cases in the same way.

As in previous PD work (Conway & Gawronski, 2013), GACFH use dilemmas in which norms and consequences point in the same direction, such that they both dictate either action or inaction (the "congruent" cases). What could 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 (see, e.g., Conway & Gawronski, 2013). But 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, a second possible account of these perverse responses is that subjects disagree with the experimenters about either the norms or the consequences (or both), and they respond on the basis of their disagreement. Indeed, we suspect that the frequency of these perverse responses has been over-estimated in past PD research precisely because subjects disagree with the intended categorization of the vignettes. (We investigate this further below.)

To summarize, while we understand the impetus for GACFH's (and others') attempt to reverse the normal link between actions and consequences in sacrificial dilemmas, their attempt to do so suffers from a mix of conceptual and empirical problems: (1) The manipulation of norms often appears to have inadvertently also manipulated consequences. (2) It is not clear whether subjects perceived the relevant norms in the same way that the experimenters intended. (3) Nor is it clear how subjects perceived the consequences of each course of action. For both (2) and (3), there are plausible reasons to worry that these perceptions may not match GACFH's intentions. But perhaps most fundamentally, (4) It is not clear that switched cases could ever succeed as intended, because the strength of proscriptive norms against particular actions is almost always likely to be stronger than the strength of prescriptive norms to block the performance of those same actions. In this way, the alleged action confound is an explicit feature of deontological ethics, not a separate confounding factor.

4 Correlations with other variables

Despite the problems just listed, GACFH make inferences about other variables, such as gender and cognitive load, on the basis of correlations with model parameters. For example, they ask whether males and females differ in their reliance on norms, consequences, or both. GACFH did not check to see whether such correlations (with gender, and other variables) were consistent across scenarios; this should be done routinely.

As first pointed out by McGuire et al. (2009), conclusions about types of items in moral judgment should be (and often are not) tested across items as well as subjects, so that it is possible to generalize to members of the same class of items. The relevance of testing across items is noticeable in GACFH's Studies 2a and 2b, where the apparent effect of cognitive load (another variable of interest) seems to be due largely to a single item (d4incon, in the paper), which was affected by cognitive load much more than any other item. This item was also an outlier with respect to the small number of action choices, as compared to five other items in the same category (incon). In general, the correlations between external variables, such as gender and cognitive load, and responses to items in the same category are highly variable.[4]

The CNI model can be fit to each set of four cases for each subject, thus creating six potential estimates of each parameter. In looking for correlations between parameters of the model and external factors such as gender or cognitive load, we need to ask whether these correlations are consistent across the six sets. It is possible that some correlations are specific to one particular set, or possibly even go in opposite directions for different sets. In the GACFH article and others that use the CNI approach, data are analyzed without apparent attention to subject variation or to variation from set to set. 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.[5] The p-

[4]We did not test all the studies systematically. Our point is that the failure to examine generality across items renders all conclusions suspect unless similar tests are done.

[5]Moshagen (2010, p. 52) says, "multiTree offers no means to diagnose or handle heterogeneity across items and/or participants. This is considered

https://doi.org/10.1017/S193029750000721X Published online by Cambridge University Press

values in GACFH thus do not allow their usual interpretation, and this concern seems to apply most of the published work using the CNI model to date.

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 for women to pay more attention than men to norms, but the same or more attention to consequences? As it turns out, this result could arise for 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. Or men may tend to perceive the relevant norms and consequences less in accordance with the researchers' intentions. One of these alternatives must be true.

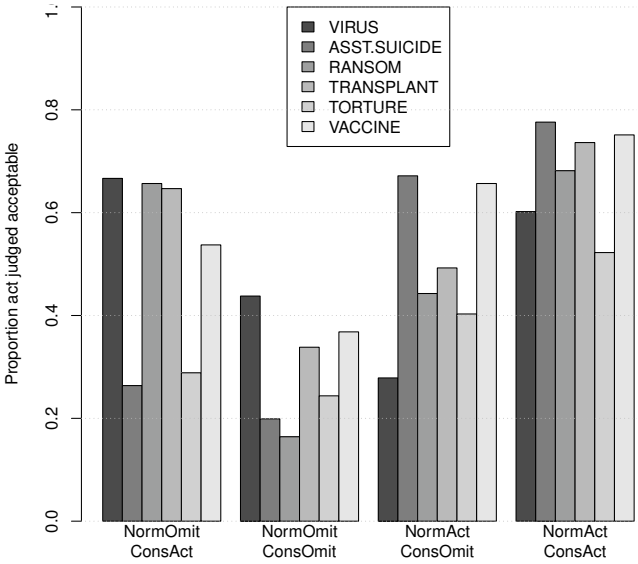
The CNI model also makes what we think are unrealistic assumptions about the ordering of the two processes represented by the *N* and *C* parameters, those that represent the probability of basing the choice on norms or consequences, respectively.[6] First, it assumes that consequences are evaluated before norms. This assumption seems to conflict with the spirit of the corrective dual-process model inspired by Greene et al. (2009), in which a judgment based on moral norms is intuitive and thus immediate, but can be overridden by subsequent reflective judgments based on consequences. As we show in Appendix A, the order of the two processes matters in drawing conclusions about the relative correlations of the two parameters with external variables such as gender.

The second possibly unrealistic assumption is that the two processes are ordered at all, rather than occurring in parallel. Subjectively, subjects report being immediately aware of the conflict between two considerations, when conflict is present. This impression is supported by findings showing effects of conflict on response time, regardless of the eventual response (Białek & De Neys, 2016, comparing incongruent and congruent dilemmas; Baron & Gürçay, 2017), as well as by research showing no evidence of a shift in relative favorability of the two response options over time (e.g., Bago & De Neys, 2019; Gürçay & Baron, 2017; Koop, 2013).

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 limitation.

[6]In principle, the processes could execute simultaneously even if one is dominant in determining the response, but the point of a conditional (tree) information flow is to reduce the use of resources by avoiding one process when possible.

FIGURE 1: Proportion of responses in which the act was judged acceptable, for each of the six cases used by GACFH in their Study 1a. The bars are grouped by the GACFH'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 GACFH.



5 Other concerns about the congruent/conflicting distinction

As we have discussed, it is not clear that the GACFH dilemmas succeeded in manipulating subjects' perception of norms and consequences as intended. Examination of GACFH's data provides additional support for this concern. Figure 1 shows the proportion of "acceptable" responses for the four versions of each case for Study 1a, based on their public data. (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 (NormOmit-ConsOmit), and 32% for the version in which both imply action (NormAct-ConsAct). 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.

Disagreement between experimenters' intention and subjects' understanding is, of course, not limited to the dilemmas used here (e.g., Shou & Song, 2017). We suspect, however, that the problem is most serious in the often-implausible scenarios based on examples made up by philosophers, as compared to scenarios based more on real cases such as vaccination. And the problems may be more serious still when researchers attempt to modify dilemmas in order to fit them into a pre-determined design scheme, as is typically done in applications of the CNI method.

https://doi.org/10.1017/S193029750000721X Published online by Cambridge University Press

Perhaps more disturbing is the large overlap among different versions, across cases. We might expect endorsement of action for all of the NormOmit-ConsOmit versions to be consistently lower than for all other versions, and the action endorsement for NormAct-ConsAct versions to be consistently higher, but this is not observed. There is even one reversal within a vignette: Action endorsement in the NormOmit-ConsOmit version of VIRUS is higher than in the NormAct-ConsOmit version of VIRUS. These puzzling results underscore the potential discrepancies between experimenter classifications and subject perceptions.

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 (as well as for the PD 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 action and omission. GACFH argue that these perverse responses arise from biases toward action or inaction. We 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. This allowed us to examine whether judgments were still perverse, both when subjects agreed with the experimenters' classification, and when they did not (in which case their own classifications were the relevant benchmark).

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.[7]

We used four versions each of 5 scenarios. Three scenarios were new[8], and two were taken verbatim from GACFH, except for the final questions. In all cases we asked, "What should you/he/she do?" rather than asking what was "acceptable".[9] All cases are in Appendix B. Each scenario had four versions (cases), which we label according to whether the norm favors action or omission (by design) and whether the consequences favor action or omission, e.g., "NormAct-ConsOmit", which represents what we intended to be a conflict between norm and consequences. The first three scenarios (EUTHANASIA, DATA, TRAIN) were intended to be as clear as possible about both norms and consequences.[10] The last two scenarios (RANSOM, VIRUS) were taken verbatim from GACFH, except for the change in the question asked, as noted. The four versions of each scenario were presented sequentially in a block. The order of the versions was randomized separately within each block. The four scenarios 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 (depicting a congruent case in which action was favored by both norms and consequences) 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).
>
> 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

[7] See http://www.sas.upenn.edu/~baron/q.html for the nature of the panel. Members of this panel have usually done several other studies on different topics. They are removed if they show signs of non-serious responding, as indicated by response times — collected without their knowledge — that are outliers on the low side. In the two studies reported here, we did not eliminate anyone and found no reason to do so.

[8] Two of these were modified versions of cases used by Baron, Scott, Fincher & Metz, 2015.

[9] "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 and should be taken.

[10] DATA 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.

https://doi.org/10.1017/S193029750000721X Published online by Cambridge University Press

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 GACFH, 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 experimenter judgments, perverse responses were defined as responses to the Should measure in which subjects indicated that the protagonist should do something that neither norms nor consequences (as defined by the experimenters, a priori) favored. For the subject judgments, we counted a response as perverse if a subject's own judgment of either consequences or norms conflicted with their response to the Should measure, and if neither one of their judgments agreed with their Should response (thus counting "not sure" as if it were "disagree" for this purpose).

It is apparent that, especially for Ransom and Virus (the items from GACFH) 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 GACFH because they did not agree with the intended classification of the cases. Comments from several subjects indicated specific points of disagreement, or the introduction of new choice options, e.g., for Virus: "I would have specialists from her home country flown in."; "Quaranteen [sic] 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 with our classification,

Table 1: Proportions of subjects in Experiment 1 (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.

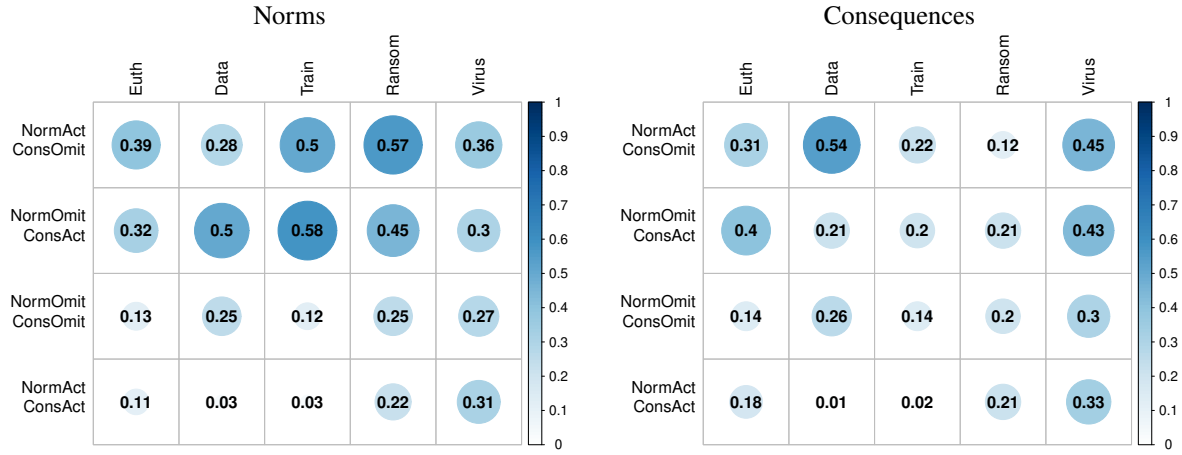
	Our cases			GACFH cases	
	Euth	Data	Train	Ransom	Virus
Experimenter judgments:					
NormOmit-ConsOmit	0.06	0.23	0.14	0.26	0.26
NormAct-ConsAct	0.14	0.02	0.04	0.32	0.40
Subject judgments:					
NormOmit-ConsOmit	0.05	0.04	0.05	0.07	0.07
NormAct-ConsAct	0.02	0.01	0.01	0.04	0.09

and a few subjects explained it in their comments, e.g., for Euthanasia: "I believe the patient has a right to not want to die but . . . I believe the paitent [sic] does not have the right to make someone kill them."; and for Train, "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." But in terms of the subjects' own classifications, there were, again, very few perverse responses. We thus conclude that GACFH were, as suspected, substantially overestimating the number of perverse responses, thus also overestimating the possible "biases" that could produce them, and mis-estimating differential attention to norms and consequences (given that this comparison is affected by the number of perverse responses).

We next examined subjects' classification data directly (i.e., their responses to the questions asking about consequences and norms for each item). 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 (Baron, 2009). 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.

https://doi.org/10.1017/S193029750000721X Published online by Cambridge University Press

FIGURE 2: Proportion of subjects in Experiment 1 who disagreed with the experimenter's classification in terms of norms and consequences.



The other explanation is that subjects were careless. To test the role of carelessness we measured the response times 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 −.36 ($p \sim .000$) across subjects with the number of errors (disagreements with the experimenter) using all cases — faster responding subjects tended to make more perverse responses. Lrt also correlated −.21 ($p = .036$) with the number of responses that were inconsistent with both of the subjects' own judgments of norms and consequences. And Lrt correlated −.50 ($p \sim .000$) with the total number of responses inconsistent with the experimenters' classification, across all cases. (By definition, these latter two analyses included only congruent cases.) These correlations for the five cases were, respectively (in the order of Figure 2): −.45, −.31, −.38, −.23, and −.15; note that they were lower for the two cases from GACFH, 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 CNI design, but asking this question does not imply any agreement with GACFH's assumption that such bias is independent of deontological responding. We measured consistency with (experimenter-defined) norms, with consequences, and with action for the four versions of each of the five scenarios, thus obtaining five measures for each subject of the consistency with which judgments were based on norms, consequences, or action (therefore, 15 measures in total per subject). Each subject's score ranged from 1 to −1 — note that these three measures are not independent, as a score of 1 or −1 for one attribute forces the other two to be 0). Then we computed coefficient α for each of the attributes: .56 for consequences, .30 for norms, and .28 for inaction; 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 −.02). Thus, it appears that there are individual differences in action/inaction bias, which are somewhat consistent across items. However, the α for the incongruent items only was .03, compared to .35 for the congruent items. Thus, any individual consistency in bias seems to depend largely on perverse responses to the congruent items. We argue that these responses are largely due to the rich opportunities that these items provide for interpretations that differ from those intended by the experimenters, and, by this criterion, action/inaction bias plays no consistent role in responses to incongruent items.

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. The bias also varied substantially with items: −0.13, 0.24, 0.05, −0.10, −0.12, for the five items (in the order shown in Table 1).

Of additional interest, the sums of the two consistency measures for Norms and Consequences were 0.80, 0.75, 0.82, 0.42, and 0.34 (out of a maximum of 1.00) for the five scenarios in the order shown in Table 1. Thus, our sce-

https://doi.org/10.1017/S193029750000721X Published online by Cambridge University Press

narios seem to have succeeded in reducing the number of perverse ("neither norms nor consequences") responses.

In sum, our results overall are consistent with our concern that GACFH are vastly overestimating the number of perverse responses to 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 GACFH's experiments as well as ours.)

7 Experiment 2

A major problem with constructing cases both for GACFH and for us in Experiment 1 was that of switching the norm between action 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 across the four cases within each scenario. 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 action or omission. We thought that most of the perverse responses were the result of disagreement with the experimental design regarding whether the cases were in fact congruent.

In Experiment 2, we constructed 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 (as in PD experiments, except that half of the congruent cases in our experiment favor action). Even this 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 actions (likewise held constant).

Such pairs of cases are still difficult to construct. General norms rarely if ever require completely unconditional action, so our cases refer to specific conditions or roles. Moreover, as we argued earlier, by switching the consequences that are usually associated with an action, we risk changing the norms as well because in order to accommodate this switch, the underlying circumstances of the action must also change. Nonetheless, we attempted as best we could to construct such cases.

We also changed the questions following each case, 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 four new scenarios (each with two items rather than four), Jury, Police, Jet and Strokes, which are described in Appendix C, we included two four-item scenarios from Experiment 1, Data and Virus, with the new wording of questions just described.[11] The two scenarios (with four items each) were split into four groups of two. Now, each of the eight 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 eight pairs was randomized for each subject, and the order of the two members of each pair was randomized for each pair but otherwise presented in adjacent order.

Again, we 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 on 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 scenario retained from GACFH), but was greatly reduced

[11] The error in the NormAct-ConsOmit Data case was corrected.

https://doi.org/10.1017/S193029750000721X Published online by Cambridge University Press

FIGURE 3: Proportion of subjects in Experiment 2 who disagreed with the experimenter's classification in terms of norms and consequences.

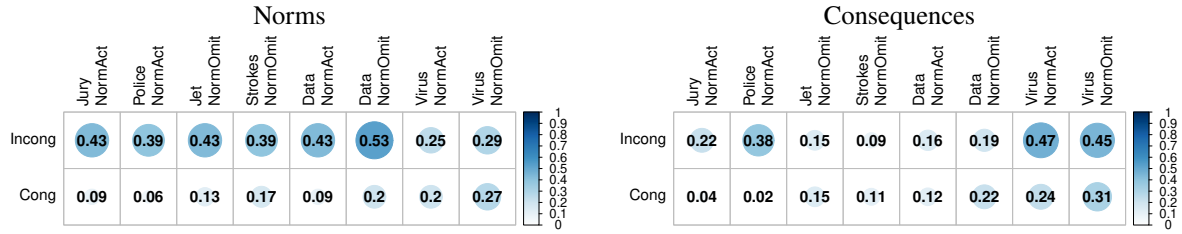


TABLE 2: Proportions of subjects in Experiment 2 (n=100) whose choice of action/omission disagreed with the classification of items defined as congruent by the experimenters, in the top two rows. The bottom two rows use the subjects' own classifications of whether the items are congruent. Note that the four new scenarios (the first four) were pairs, each with only one norm.

	Jury	Police	Jet	Strokes	Data	Virus
Experimenter judgments:						
NormOmit-ConsOmit			0.11	0.04	0.21	0.29
NormAct-ConsAct	0.05	0.05			0.08	0.28
Subject judgments:						
NormOmit-ConsOmit			0.03	0.04	0.07	0.04
NormAct-ConsAct	0.02	0.04			0.02	0.09

when we used the subjects' classifications rather than the experimenters'. Note that the results shown for Data and Virus show four numbers in each column rather than two because each of these scenarios involved two pairs. Our Data scenario was designed to be like those used by GACFH in that it uses all four variants on the same basic scenario, thus requiring an unusually artificial switch between act and omission.

Perverse responding on congruent trials was again associated with high speed. The correlation with Lrt (based on the fastest half of the trials for each subject) was −.63 ($P \sim .000$), suggesting that at least some of the perverse responses arose from carelessness. The Lrt correlation was −.344 ($p \sim .000$) when responses were defined in terms of the subjects' own judgments 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 −.201 ($p = .045$) for the total number of disagreements. The correlation was especially great (−.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. That is, they manipulated their perception of norms or consequences so that the two judgments pointed to the same response option. Across the eight 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).[12]

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

Subjects also again had a different concept of "omission" than what we had in mind. On average, they were "correct" only 70% of the time in classifying the second response as an omission. Some of this error 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

[12]The correlation for congruent cases, which should be 1.00, had a mean of .645, but there were far fewer cases of disagreement here, which were mostly for subjects who responded quickly.

https://doi.org/10.1017/S193029750000721X Published online by Cambridge University Press

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 and 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 eight 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; $t_{99} = 2.34$, $p = .022$). Once again, as in Experiment 1, consistent individual differences in action/inaction bias are limited to the congruent items ($\alpha = .41$). The incongruent items had an α of only .18 (95% confidence interval −0.06 to 0.42).

Experiment 2 provides further evidence that subjects' disagreement with the experimenters' classification of norms and consequences, mostly based on the difficult design of the dilemmas, led GACFH to considerably overestimate the number of perverse responses. Moreover, the correlations we found with response times 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 apparent deontological responses in sacrificial dilemmas are the result of a bias against action. In principle, the CNI method could help to answer this question, although our concern is that the bias would have to be extreme in order to yield a sufficient number of perverse responses on congruent items that the subjects clearly understood as congruent. Perhaps a useful modification of the method would be to include cases with options that are equivalent in terms of their consequences, and with a relevant norm that requires neutrality between act and omission, such as the "equipoise" norm for clinical trials of new medical treatments.

As it happens, other research provides some useful hints about the possible role of bias toward omission or action. Several experiments (Spranca et al., 1991, Experiment 3; Ritov & Baron, 1990; Baron & Ritov, 1994; Baron, 1995; Ritov & Baron, 1995; Baron & Ritov, 2004) were done to ask 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 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 often 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 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 much of 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. Again, a bias against action cannot account for much of the original result, which involved direct causality as well.

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.

https://doi.org/10.1017/S193029750000721X Published online by Cambridge University Press

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 that subjects (on the average) strongly preferred omission to direct action, but not much to indirect action. 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, Gürçay & 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) probably play some role in explaining the existence of non-utilitarian responses, but it has now been clearly demonstrated that other factors are highly 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., 1991, Experiment 3; Baron & Ritov, 1994). These two factors work together when an action produces a perceived loss, but they oppose each other when an action produces a perceived gain. Sometimes the amplification effect is stronger, so there is a small bias toward beneficial action for gains (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.[13]

In sum, we have evidence that 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.[14]

9 Conclusions

We have argued that the application of the CNI model to questions about the determinants of utilitarian and deontological responses to sacrificial dilemmas may yield misleading conclusions. Several problems were specific to the initial paper (GACFH) and could be easily fixed in later work. These include the failure to ask whether relations with various external variables (gender, cognitive load, etc.) are consistent across subjects and across item groups.

Other methodological problems are more difficult to resolve. Most seriously, the application of the CNI model requires the use of congruent items that must yield enough "perverse" responses (those that both violate norms and produce worse consequences) so that the model provides results that differ from the standard analysis. We have argued that the construction of such items has the side effect of permitting extensive re-interpretation by subjects, so that they are not as congruent as intended. This problem is also present in the PD model.

The CNI model does overcome one problem with the PD model. The PD model relies on congruent-incongruent pairs (as does our Experiment 2), but all the congruent items are designed so that both consequences and norms favor inaction. Thus, perverse responses to these congruent items can indicate a bias toward harmful and norm-violating action, but any bias toward inaction cannot be detected. The CNI model overcomes this problem by including item pairs in which the norms and consequences for congruent items both favor action. Thus, perverse responses can arise from biases toward inaction as well as toward action.

Although the inclusion of the latter congruent cases solves one problem, it introduces others when it attempts to replace

[13] In 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.

[14] Still, a method like the CNI could be useful in quantifying the role of action/inaction bias, as a function of various conditions.

https://doi.org/10.1017/S193029750000721X Published online by Cambridge University Press

a norm favoring inaction with one favoring action, e.g., by making "action" the prevention of someone else's action. The supposedly matched norms may then differ in their perceived strength, and the apparent consequences may differ as well. This is arguably the most intractable problem faced by the CNI model.

Inaction biases do seem to play a small role in accounting for observed neglect of consequences in sacrificial dilemmas and other action/omission dilemmas, as we have noted. While the role of such biases is an interesting empirical question, it does not challenge the fact that observed non-utilitarian responses are truly non-utilitarian. A heuristic of "do no harm" (actively) is in fact a deontological principle, as we have argued, because it tells us that the consequences of inaction are less relevant. Thus, a finding that people are sometimes biased toward inaction (or action) does not impugn the conclusion that they are sometimes not utilitarian.

References

Alexander, L., & Moore, M. (2016) Deontological ethics. *The Stanford Encyclopedia of Philosophy* (Winter 2016 Edition), Edward N. Zalta (ed.), https://plato.stanford.edu/archives/win2016/entries/ethics-deontological/.

Bago, B., & De Neys, W. (2019). The intuitive greater good: Testing the corrective dual process model of moral cognition. *Journal of Experimental Psychology: General, 148*(10), 1782–1801. https://doi.org/10.1037/xge0000533.

Baron, J. (1995). Blind justice: Fairness to groups and the do-no-harm principle. *Journal of Behavioral Decision Making, 8*, 71–83.

Baron, J. (1996). Do no harm. In D. M. Messick & A. E. Tenbrunsel (Eds.), *Codes of conduct: Behavioral research into business ethics*, pp. 197–213. New York: Russell Sage Foundation.

Baron, J. (2009). Belief overkill in political judgments. *Informal Logic, 29*, 368–378.

Baron, J. & Gürçay, B. (2017). A meta-analysis of response-time tests of the sequential two-systems model of moral judgment. *Memory and Cognition, 45*(4), 566–575.

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., & Ritov, I. (2009). Protected values and omission bias as deontological judgments. In D. M. Bartels, C. W. Bauman, L. J. Skitka, & D. L. Medin (Eds.), *Moral Judgment and decision making*, Vol. 50 in B. H. Ross (series editor), *The Psychology of Learning and Motivation*, pp. 133–167. San Diego, CA: Academic Press.

Baron, J., Scott, S., Fincher, K., & Metz, S. E. (2015). Why does the Cognitive Reflection Test (sometimes) predict utilitarian moral judgment (and other things)? *Journal of Applied Research in Memory and Cognition, 4*(3), 265–284

Białek, M. & De Neys, W. (2016). Conflict detection during moral decision-making: Evidence for deontic reasoners' utilitarian sensitivity, *Journal of Cognitive Psychology, 28*, 631–639. https://doi.org/10.1080/20445911.2016.1156118.

Gawronski, B., Armstrong, J., Conway, P., Friesdorf, R., & Hütter, M. (2017). Consequences, norms, and generalized inaction in moral dilemmas: The CNI model of moral decision-making. *Journal of Personality and Social Psychology, 113*(3), 343–376.

Gleicher, F., Kost, K. A., Baker, S. M., Strathman, A. J., Richman, S. A., & Sherman, S. J. (1990). The role of counterfactual thinking in judgments of affect. *Personality and Social Psychology Bulletin, 16*, 284–295.

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

Gürçay, B., & Baron, J. (2017). Challenges for the sequential two-systems model of moral judgment. *Thinking and Reasoning, 23*, 49–80.

Hare, R. M. (1981). *Moral thinking: Its levels, method and point*. Oxford: Oxford University Press (Clarendon Press).

Heit, E., & Rotello, C. (2014). Traditional difference-score analyses of reasoning are flawed. *Cognition, 131*, 75–91.

Johnson, E. J., & Goldstein, D. (2003). Do defaults save lives? *Science, 302*, 1338–1339.

Koop, G. J. (2013). An assessment of the temporal dynamics of moral decisions. *Judgment and Decision Making, 8*, 527–539.

Landman, J. (1987). Regret and elation following action and inaction: Affective responses to positive versus negative outcomes. *Personality and Social Psychology Bulletin, 13*, 524–536.

McGuire, J., Langdon, R., Coltheart, M., Mackenzie, C. (2009). A reanalysis of the personal/impersonal distinction in moral psychology research, *Journal of Experimental Social Psychology 45*(3), 577–580

Moshagen, M. (2010). multiTree: A computer program for the analysis of multinomial processing tree models. *Behavioral Research Methods, 42*, 42–54.

Patil, I., Zucchelli, M. M., Kool, W., Campbell, S., Fornasier, F., Calò, M., Silani, G., Cikara, M., Cushman, F. (2018). Reasoning supports utilitarian resolutions to moral dilemmas across diverse measures. Available at: https://psyarxiv.com/q86vx.

Rachels, J. (1979). Killing and starving to death. *Philosophy, 54*(208), 159–171.

Rachels, J. (2001). Killing and letting die. In L. Becker & C. Becker (Eds.), *Encyclopedia of Ethics, 2nd ed.*, pp. 947–950. New York: Routledge.

Rawls, J. (1971). *A theory of justice*. Cambridge, MA: Harvard University 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. (1994). Judgments of compensation for misfortune: the role of expectation. *European Journal of Social Psychology, 24*, 525–539.

Ritov, I., & Baron, J. (1995). Outcome knowledge, regret, and omission bias. *Organizational Behavior and Human Decision Processes, 64*, 119–127.

Ross, W. D. (1930). *The right and the good*. (Reprinted 2002 by Oxford University Press, Oxford.)

Royzman, E. B. & Baron, J. (2002). The preference for indirect harm. *Social Justice Research, 15*, 165–184.

Shou, Y., & Song, F. (2017). Decisions in moral dilemmas: The influence of subjective beliefs in outcome probabilities. *Judgment and Decision Making, 12*, 481–490.

Spranca, M., Minsk, E., & Baron, J. (1991). Omission and commission in judgment and choice. *Journal of Experimental Social Psychology, 27*, 76–105.

Appendix A: Demonstration of inconsistencies as a function of order and case sets.

When the CNI (or PD) model is used to examine the effects of other variables such as gender or cognitive load on N and C parameters (norms and consequences), these effects can depend on the implicit ordering in the model itself as well on the particular sets of items used to estimate the model (discussed in our main text). The CNI model, like the PD model on which it is based, defines the parameters as if the subject first decides whether to make a utilitarian choice (based on "consequences"), with probability C. If that option is rejected, the subject then decides, with probability N, whether to make a deontological choice (based on "norms"), and, finally, if both options are rejected, the subject decides on the basis of response bias toward inaction or action. Thus N, but not C, is a conditional probability. The choice of this implicit ordering is arbitrary, and, if anything, opposite to that implied by corrective (sequential) dual-process views, in which the utilitarian response arises as a correction of an intuitive response based on norms (Baron & Gürçay, 2017).

GACFH base their conclusions on effect sizes, e.g., the effect of gender on the N parameter. But an alternative approach, which we think is reasonable, would be to examine the actual differences in the parameters. If this had been done, the ordering of the two decisions just described would clearly matter. For a simple example, suppose that subjects have an action bias of 1.0, so that they always favor action in the congruent condition in which both norms and consequences favor action (NormAct-ConsAct). Action responses in the other congruent condition (NorOmit-ConsOmit) represent cases in which neither norms nor consequences drove the response. (This is the implicit assumption in the PD model, which omits the former congruent condition). For one experimental condition (or group of subjects), the proportion of "action" responses is: .6 when consequences favor action and norms favor inaction (NormOmit-ConsAct); .6 when consequences favor inaction and norms favor action (NormAct-ConsOmit); and .2 when consequences and norms both favor inaction. For the other condition, the respective three proportions are .8, .8, and .6. (These numbers are chosen so that the CNI model fits exactly.) Then, the CNI model as stated implies that the C parameter is .4 and .2 for the two conditions, respectively, an effect of .2, and the N parameter is .67 and .25, an effect of .42. If the ordering of decisions in the model is reversed, so that we have NCI instead of CNI, the C parameter is .67 and .25 for the two conditions, respectively, an effect of .42, and the N parameter is .4 and .22, an effect of .2. Thus, the CNI model implies that the effect is mainly on C and the NCI model implies that it is mainly on N.

GACFH note that their conclusions do not change if the ordering is reversed, and this claim may be approximately true, because effect sizes (their preferred measure), like correlations, compare effects to standard deviations, and the standard deviations could depend on ordering in the same way as the parameters themselves. Yet it seems strange to say that one effect is larger than another when, in fact, it could be smaller in the parameters of interest. The N and C parameters are not intended as arbitrary scales but as meaningful measures of cognitive processes.

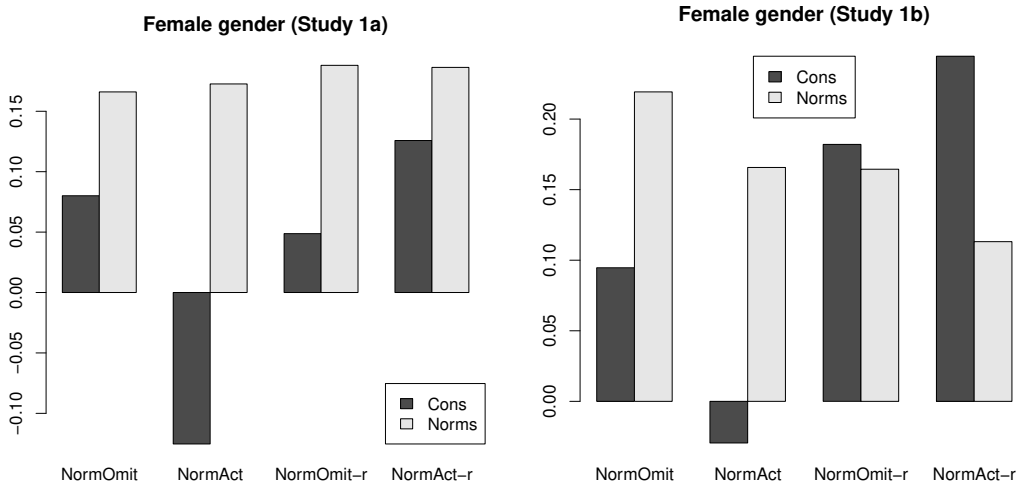
As we noted, a second source of potential inconsistency is the use of different sets of items. We illustrate this here by breaking the GACFH data into two parts, one corresponding to a PD model for NormOmit in which the norm prescribes omission (in both congruent and incongruent cases), and the other to a "reverse" PD model for NormAct in which the norm prescribes action. The NormAct cases allow subjects to display a strong bias toward inaction in the congruent cases, which would not be detectable in the original PD model.

Moreover, also as noted, the CNI model (and the PD model) could be stated in a different order, in which the decision to attend to consequences is dependent on an implicit prior decision not to attend to norms. These two classifications — NormOmit vs. NormAct, and original vs. reversed order — allow four estimates of the two main parameters, Consequences (C) and Norms (N), using analyses analogous to the PD model in each case.

For illustration, Figure 4 shows the correlations of gender, the main variable of interest, in GACFH's Studies 1a

https://doi.org/10.1017/S193029750000721X Published online by Cambridge University Press

FIGURE 4: Correlation of each parameter (Cons, Norms) with the main variable of interest (gender) in GACFH's Studies 1a and 1b. The parameters are derived from four PD models applied to each subject. NormOmit and NormAct represent the cases where the norm prescribes inaction and action, respectively. "-r" represents the models in which the ordering of implied decisions is reversed (so that attention to consequences is contingent on an implicit decision not to attend to norms).



and 1b with the parameters for Norms and Consequences (N and C).[15] It is apparent that the four ways of calculating the relative effects on Norms and Consequences do not yield consistent results. In Study 1a, NormOmit and NormAct yield different conclusions about the direction of the effect of gender on Consequences. (These effects are small, but our point is that this can happen, in principle.) In Study 1b, reversing the ordering of consideration of consequences and norms reverses the relative effects of Consequences and Norms. Consequences show a larger effect with the reversed model. Note that the results reported by GACFH are understandable as a compromise between the NormOmit and NormAct versions of the original ordering. (Application of the PD model to these data yields similar inconsistencies.) The following section provides the details for Figure 4 (which are presented only for completeness).

Calculations for Figure 4.

The CNI model was apparently fit to individual responses (act or omit, coded as 1/0) to each item. To illustrate possible inconsistencies in the model, we simplified the procedure so that we could calculate parameters of interest for each subject. First, we averaged each subject's responses in each of the four conditions (acinc [NormAct-ConsOmit], accon [NormAct-ConsAct], incon [NormOmit-ConsOmit, inc [NormOmit-ConsAct]) so that we got a proportion of action responses for each condition. (Each condition had six scenarios, so the average was always a proportion out of six.) Thus, we ignored variation among the six cases within each condition by collapsing over it.

Second, we split our procedure into two parts, one focusing on the NormOmit conditions, in which the norm prescribed inaction, and the other focusing on the NormAct condition, in which the norm prescribed action. These two calculations need not agree, so GACFH's procedure estimated a best fit, taking all conditions into account. But our point is to illustrate a possible source of inconsistency. Note that the NormOmit focus consists of the cases in which action and omission were reversed, e.g., by making the action the prevention of someone else's action.

Table 3 illustrates the original CNI model, with probabilities of entering each of the four possible states of the model: Cons, Norm, Act (bias) and Omit (bias). The probability of the Cons state is C. The Norm state is entered with probability N only if the Cons state is not entered. Thus, the probability of entering the Norm state is $(1-C)N$. Finally, if neither state is entered the Act state is entered with probability $(1-C)(1-N)A$, where A is the response bias toward action, and the alternative state of Omit is entered with probability $(1-C)(1-N)(1-A)$. The 1's and 0's in the table indicate action or inaction.

The probabilities W through Z are the proportions we calculated in the data. So now we can calculate C and N: $C = Y - Z$, and $N = (Z-X)/(1-C) = (Z-X)/(1-Y+Z)$.[16] This calculation is based largely on the "in" cases, although it was necessary to include X as well.

[15] Studies 2 and 3 did not report consistent correlations with these parameters. The public Study 4 data did not include psychopathy measures and moral judgments in any way that could be matched.

[16] In a few cases the denominators were 0, and these were counted as missing data.

https://doi.org/10.1017/S193029750000721X Published online by Cambridge University Press

TABLE 3: The original CNI model in tabular form, with calculations of the probability of action in each of the four conditions, taking into account the state probabilities.

		NormOmit ConsAct	NormOmit ConsOmit	NormAct ConsAct	NormAct ConsOmit	p(state)
Cons		1	0	1	0	C
	Norm	0	0	1	1	$(1-C)N$
	Act	1	1	1	1	$(1-C)(1-N)A$
	Omit	0	0	0	0	$(1-C)(1-N)(1-A)$

$W = p(action\|NormOmit\text{-}ConsAct) =$	$C+$		$(1-C)(1-N)A$
$X = p(action\|NormOmit\text{-}ConsOmit) =$			$(1-C)(1-N)A$
$Y = p(action\|NormAct\text{-}ConsAct) =$	$C+$	$(1-C)N+$	$(1-C)(1-N)A$
$Z = p(action\|NormAct\text{-}ConsOmit) =$		$(1-C)N+$	$(1-C)(1-N)A$

TABLE 4: The CNI model reversed, with calculations of the probability of action in each of the four conditions.

		NormOmit ConsAct	NormOmit ConsOmit	NormAct ConsAct	NormAct ConsOmit	p(state)
Norm		0	0	1	1	N
	Cons	1	0	1	0	$(1-N)C$
	Act	1	1	1	1	$(1-N)(1-C)A$
	Omit	0	0	0	0	$(1-N)(1-C)(1-A)$

$W = p(action\|NormOmit\text{-}ConsAct) =$		$(1-N)C+$	$(1-C)(1-N)A$
$X = p(action\|NormOmit\text{-}ConsOmit) =$			$(1-C)(1-N)A$
$Y = p(action\|NormAct\text{-}ConsAct) =$	$N+$	$(1-N)C+$	$(1-C)(1-N)A$
$Z = p(action\|NormAct\text{-}ConsOmit) =$	$N+$		$(1-C)(1-N)A$

We did an alternative calculation relying on the "ac" cases, as follows: $C = W - X$, and $N = (Y - W)/(1 - C) = (Y - W)/(1 - W + X)$.

Table 4 shows the corresponding calculations based on the assumption that the steps in the tree are reversed, so that the first choice is whether or not to enter the Norm state, and then, only if that decision is negative, is the decision made whether to enter the Cons state:

The probabilities in Table 4 lead to the same two pairs of estimates: First, $N = Z - X$, and $C = (Y - Z)/(1 - N) = (Y - Z)/(1 - Z + X)$. Second, $N = Y - W$, and $C = (W - X)/(1 - N) = (W - X)/(1 - Y + W)$.

Appendix B: 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

https://doi.org/10.1017/S193029750000721X Published online by Cambridge University Press

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?

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 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.[17]

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 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.

[17]This 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.

https://doi.org/10.1017/S193029750000721X Published online by Cambridge University Press

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).<p>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).<p>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

https://doi.org/10.1017/S193029750000721X Published online by Cambridge University Press

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 C: New 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

https://doi.org/10.1017/S193029750000721X Published online by Cambridge University Press

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.

A. 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.

B. 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.

A. 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.

B. 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.

https://doi.org/10.1017/S193029750000721X Published online by Cambridge University Press