Consequences, norms, and inaction: A comment

Jonathan Baron* Geoffrey P. Goodwin†

Abstract

Gawronski, Armstrong, Conway, Friesdorf & Hütter (2017, GACFH) presented a model of choices in utilitarian moral dilemmas, those in which following a moral principle (the deontological response) leads to worse consequences than violating the principle (the utilitarian response). In standard utilitarian dilemmas, the utilitarian option usually involves action, and the deontological response, omission. GACFH's proposed CNI model holds that deontological responses in such dilemmas arise in three different ways, only the first of which reflects deontological thinking: the activity of a psychological process leading to a deontological choice, the inactivity of a different process leading to a utilitarian choice, or a bias toward inaction. GACFH attempt to separate these three processes by presenting new dilemmas in which action and omission are switched, such that action reflects a deontological choice and omission reflects a utilitarian choice. They also present dilemmas in which these two normally opposing processes lead to the same choice. They conclude that utilitarian and deontological responses are indeed separable and independent, and that this has been obscured in past research which has treated them as naturally opposed to one another. We argue that: 1., a bias toward harmful inaction is best understood as an explanation of deontological responding rather than as an alternative explanation; 2., standard utilitarian dilemmas are designed to assess the relative strength of deontological and utilitarian arguments, and the only ways in which a subject could fail to show either response tendency is through being inattentive, reactive, or antisocial; 3., the CNI model and its implementation do not do a good job of answering the empirical question about how a bias toward inaction should be explained; and, 4., previous research has in fact shed considerable light on this empirical question.

Key words: utilitarianism, deontology, omisison bias

1 Introduction

A great deal of research has now established that many people's moral judgments do not follow utilitarian principles. In one sort of demonstration (among many), subjects are asked to compare two options, one of

*Department of Psychology, University of Pennsylvania, 3720 Walnut St., Philadelphia, PA, 19104. Email: baron@upenn.edu.
†Department of Psychology, University of Pennsylvania.

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 is a category of moral theories that emphasize properties of action other than consequences, such as whether the actions violate basic rights, or whether the act conflicts with a required duty. Deontological theories can justify not pushing the man in a variety of ways, but most of them involve a prohibition on 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.[2] In the typical dilemma, consequences favor acting, e.g., pushing the man, a bias toward inaction opposes action, and moral norms usually also oppose action (e.g., "don't kill people"). GACFH further suppose that deontological and utilitarian responding are not simply poles of a single dimension but, rather, alternative and independent ways of thinking about the dilemmas. In particular, GACFH assume that the basic utilitarian principle is based on consequences and the basic deontological principle is based on norms. In their view, a preference for inaction (or action) is separate, different from either philosophical approach. In standard sacrificial dilemmas, they argue that an apparent utilitarian response could arise either from a focus on consequences or from a preference for action, and an apparent deontological response could arise either from a norm or a preference for inaction.

To assess the role of each of the three components of the model, GACFH use a design in which they manipulate the association of norms and consequences with action or inaction. In this design, the consequences of action are manipulated so that action is sometimes better than inaction, but sometimes worse. Separately, norms either forbid (proscribe) or require (prescribe) action. These manipulations give rise to three new versions of the standard sacrificial dilemma. In the standard version of this dilemma, action has better conse-

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

[2] The model is similar to, and builds upon, an earlier model based on the idea of process dissociation.

quences 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 agree for two of the four cases and disagree for the other two. The pattern of subjects' responses to the four dilemmas indicates which principle is driving the responses. If action is always chosen, or inaction is always chosen, then the responses are driven by a bias toward action or inaction, not by either norms or consequences.

In this comment, we discuss several concerns with this approach and the reported findings. We begin with the main philosophical assumption made by GACFH, which is that a moral distinction between acts and omissions is not itself an indicator of deontological reasoning. We then discuss two aspects of the empirical approach adopted by GACFH, the difficulty of constructing items that meet the new requirements, and the estimation of the model itself. Finally, we argue that the primary empirical question, whether apparently deontological responses arise from a bias against action, has already been largely answered by previous literature, so that this new approach is not needed for this empirical purpose.

2 Omission bias and deontological reasoning

The basic finding that some people prefer an option with worse consequences, by itself, shows that people do not follow utilitarian principles, however it is explained. For example, Ritov and Baron (1990) found that many people oppose vaccination when the side effects of the vaccine cause half as many deaths as the disease that the vaccine prevents. GACFH, however, suggest that the result is an artifact if it is due to a general preference for not acting — as if the response itself cannot be taken as a rejection of utilitarian principles. It is not clear why they think this. If it should turn out that the effect is entirely due to a bias against action, then we would conclude that "people are non-utilitarian in certain cases because they have a bias against action." In other words, the preference for inaction is part of the explanation for non-utilitarian responding, rather than an entirely separate response tendency. This is more or less what Ritov and Baron (1990) had in mind at the outset (likewise Spranca et al., 1991). Hence the poorly-chosen name "omission bias" for the effects reported then. Utilitarianism implies that the morally better option is the one that minimizes harm (or maximizes good — there is no natural zero point, so "harm" and "good" are always relative terms). If some other descriptive principle of decision making leads people away from this option, then that other principle is an explanation of why judgments are sometimes not utilitarian.

Utilitarianism makes no inherent distinction between acts and omissions (Rachels, 1979, 2001), so a bias

toward omissions is inherently opposed to utilitarianism. The view that there is no moral difference between acts and omissions is itself an argument against proposed moral rules that proscribe some action — a category that includes a large proportion of moral rules commonly endorsed. In sum, a finding that people judge acts and omissions differently is itself indicative that they are not following utilitarian principles (assuming that other extraneous differences between acts and omissions are held constant). As a consequence, GACFH's attempt to treat a bias towards inaction as somehow separate from non-utilitarian thinking is, in our view, a conceptual mistake. Of course, it is still worthwhile finding out whether a bias toward omission is the only reason for non-utilitarian judgments in sacrificial dilemmas, an issue we discuss.

3 Separating norms and consequences

GACFH seem to assume that one could be both a consequentialist and a deontologist at the same time, and some of their results give the appearance of supporting this – i.e., some people seem sensitive to both the Consequences and Norms parameters in their model (described below). Yet the very definition of deontology is based on its opposition to consequentialism (including utilitarianism). While consequentialism evaluates options in terms of expected consequences only, deontology adds (or substitutes) additional criteria concerning properties of the behavior in question other than its consequences (Alexander & Moore, 2016). There is a point to this. To the extent that choices do what they are expected to do, then utilitarian choices will bring about the best options on the average, while deontological choices will lead to outcomes that are relatively worse. Thus, the attraction of deontology could be part of the reason why things are not as good as they could be.

Because of the necessary conflict between deontological and utilitarian responses, treating them as independent creates several problems. The idea is to allow responses that are sensitive to *neither* norms nor consequences by using dilemmas in which norms and consequences point in the same direction, such that they both dictate either action or inaction. What could possibly account for a response tendency that seems to conflict with both norms and consequences? The most obvious possibility is that such responses are driven by an antisocial tendency, which causes some subjects to choose the option that is both relatively harmful and contrary to moral norms. This response tendency might also be caused by reactance or misbehavior on the part of some subjects who deliberately give non-serious responses. It is also possible that such responses are the result of inattention. Such "congruent" conditions may therefore be useful for the purpose of excluding inattentive, non-serious, or deeply anti-social subjects (e.g., sociopaths). But, beyond this, it is not clear what else of substance can be learned from such responses.

Because of the necessary conflict between deontological and utilitarian responses, treating them as in-

dependent creates several problems and has little compensating benefit. The attempt to allow sensitivity to both utilitarian and deontological responses (or neither) within the same dilemma by making norms and consequences point in the same direction, such that they both dictate either action or inaction, has no clear conceptual benefit. What could possibly account for a response tendency in the other direction? ***

To compare effects of norms and (better) consequences while removing any effects of bias toward action or inaction, vignettes are constructed in which norms and consequences are supposed to conflict, but the usual association of norms with inaction and consequences with action is reversed. This method, however, creates additional consequences that blur the intended distinction. One way these dilemmas are constructed is by switching an originally harmful action to one that blocks someone else's harmful action. For instance, in the standard transplant scenario dilemma, the subject is asked to imagine themselves as a surgeon, and the target action is to kill a patient in order to harvest his organs for other needy patients. In the switched case, the target action is to intervene to prevent another surgeon from killing a patient for the same purpose. The point of this manipulation is to make action forbidden (proscribed) in the first case, but required (prescribed) in the second, while holding constant the relative benefits of action over inaction. But, this switching has a problem. When the action is to contravene someone else's action, it has additional consequences aside from preventing the consequences of that action. It may hurt the decision maker's feelings, possibly leading him or her to take retaliatory action against the one who contravenes. It may also violate the lines of authority, thus weakening these lines for the future by discouraging those in command from taking their responsibility seriously (Baron, 1996). It may also be illegal or against the rules, and rule following likewise has a value as a precedent for future cases. In addition, the fact that someone else has made a decision provides reason to think that he or she knew something that we did not know. Thus, the use of the idea of blocking someone else's decision does not ensure that the consequences of action are held constant, with the only factor differentiating the two cases being whether the action is proscribed or prescribed. In fact, it is difficult to find a clear way to reverse action and omission that holds everything else constant, i.e., to look for a true framing effect.[3]

Moreover, there is no real way to check on what norms each vignette is supposed to encapsulate, i.e., no manipulation checks on these norms, and no prior pilot testing or conceptual checks of any sort. While it seems likely that there are norms of some sort being invoked, there is no way to test whether people are really attentive to those norms or not. There may also be multiple norms invoked in some scenarios, further compounding this problem (see example below on the norms involved in the abduction dilemma). And, as

[3]Ritov & Baron (1994) examined compensation and penalty judgments in a situation in which action or omission could lead to the same harmful outcome. Specifically, a train was headed down a hill toward a tree that had fallen across the tracks. In one of several conditions, the engineer decides not to stop the train, the train hits the tree, and one passenger is injured. In another condition, the engineer stops the train, but the train stops so quickly that one passenger is injured. (In yet another condition, the engineer tries to stop the train and fails.) Judgments depended on expectations and were affected by the consequences of the rejected option. Modification of this experiment might allow the sort of test that GACFH were attempting, in the context of moral judgment. We thank Ilana Ritov for pointing this out.

the last paragraph suggests, people could be making their decisions entirely based on consequences.

For example, in the abduction dilemma, it seems as though GACFH think there is a relevant norm to approve ransom payments to guerrillas if it means saving a journalist from beheading. The approval of such payments is apparently prescribed, whereas the vetoing of such payments is apparently proscribed. In the basic version of this vignette, in which the norm to approve this payment is opposed by consequences, the consequences include many further deaths caused by the guerrillas in the war they are waging. GACFH therefore want to treat approval of the ransom payment as reflecting sensitivity to a moral norm. 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 caused in the guerrilla war. (They might also think that the journalist's innocence is a factor that amplifies the badness of his death compared to the deaths of combatants in a war. And it's not very clear in this case that the later deaths in the guerrilla war can really be thought of as "caused" by refusing the ransom. This is not so much about norms as it is about the directness of causation.) Additionally, it's not even clear that this norm to make such ransom payments is widely endorsed. There is a strong contrary norm not to give in to such requests. There is just too much latitude in this vignette (and others) to conclude anything firmly about whether norms or consequences (or some combination) are driving people's decision-making. At least, we should have data on what each subject thinks the norm is for this kind of case.

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

Conceptual problems of this sort are not limited to these examples alone, but apply to GACFH's entire set of vignettes. GACFH make the point that such problems are immediately solved by the fact that people showed some sensitivity to the norms factor. But, as our previous examples make clear, it might turn out that people are sensitive to the manipulation of norms for other reasons – e.g., the consequences! And some subjects might disagree about what the norms are, which would convert apparently conflicting vignettes into congruent ones.

More generally, the issue addressed in the literature is not really about the conflict between consequences and *norms* per se. Many norms are fully consistent with (or even directly express) utilitarian views, e.g., "Punishments for offenses should be greater when the harm caused is greater (other things being equal)." Others are based on intuitive (prescriptive) rules that usually maximize utility (Hare, 1981). Some of these rules are best followed even when they appear to lead to worse consequences, because the decision makers may not be capable of recognizing such true exceptions. (Hare's example is the prohibition of adultery, which may seem to serve the greater good by those tempted to engage in it.) The relevant conflict is between utilitarian principles and deontological rules, which have special properties other than being norms. Specifically, deontological rules take the form of duties (positive or negative, i.e., to do something or not to do something) and are concerned with properties of actions other than their consequences, such as whether they are acts or omissions, whether they are direct physical causes of a harmful outcome, or whether they bring about harmful consequences as a causal means to achieve a good consequence or as side effects that do not themselves cause the desired consequence. Thus, the basic contrast between norms and consequences that GACFH explore is, we think, based on a mischaracterization of relevant scholarship.

4 The model

GACFH's model is sequential, with the first step being to decide whether consequences will determine the answer or not. The probability of this step is C (for "consequences"). If not, then the next step is to determine whether norms determine the answer (with probability N, given that the stage is reached) or not. And, if neither consequences nor norms determine the answer, it is chosen according to a response bias (I in the original model). We will use the response-bias parameter A, which equals 1–I, the probability of action rather than inaction, if this node of the tree is reached.

The basic model is this, using the following abbreviations from the data file (which is publicly available): Ac refers to dilemmas in which action is proscribed by the norm; In refers to dilemmas in which action is prescribed (so inaction is proscribed); Con refers to dilemmas in which norms and consequences are consistent; Inc refers to dilemmas in which they are inconsistent. Thus: for AcInc, action is opposed by the norm but leads to better consequences (as in the standard dilemmas used in other research); for AcCon, action is opposed by the norm and leads to worse consequences; for InCon, action is favored by the norm and leads to better consequences; and for InInc, action is favored by the norm but leads to worse consequences. According to the model, the probabilities of choosing action in the four dilemmas are:

Consequences	Norms	p(action)
action	inaction	AcInc = C + (1-C)(1-N)A
inaction	inaction	AcCon = (1-C)(1-N)A
action	action	InCon = C + (1-C)N + (1-C)(1-N)A
inaction	action	InInc = (1-C)N + (1-C)(1-N)A

It is apparent that C, the probability of deciding by consequences, should equal AcInc–AcCon (the difference of the first two cases), and it should also equal InCon–InInc. We can thus use these two measures of C as a preliminary test of the model. (Other checks are redundant with these.) When we carried out this check for each of the 6 scenarios, one of them, the one about the police using torture, showed a very large deviation, largely because InInc (choice of the action of stopping the torture by reporting it) was much higher than it should have been ($t_{200} = 6.08$, $p = 0.000000036$, Bonferroni corrected). However, otherwise the model fit fairly well.

We compared the original CNI model of GACFH, as just described, to two other models by fitting three models to each subject's data from the first study reported. We fit the original model coding V_c, whether consequences led to action or not, as 1, 0, 1, 0 for the four conditions in the order listed. This variable thus had 12 1's and 12 0's for each subject (since each of the four dilemma types had six dilemmas). Likewise, V_n, whether norms predict action or not, was 0, 0, 1, 1, and A was effectively an intercept but multiplied by (1-C)(1-N). The response was coded as 1 for choosing action, 0 otherwise. We used the nls() function (nonlinear least squares) in R.[4]

The second model simply asked whether the response (0,1) was a linear function of V_n and V_c. The intercept in this standard regression model took the place of the A parameter. This model has no clear theoretical interpretation in the present context. It lacks the ordering assumption of the original model and fails to replace it with anything plausible. The point of it is simply to modify the original model as little as possible while removing the assumption that consequences and norms are considered sequentially. The original model assesses the effect of N only when C is 0, while the second model assesses the effect of N and C simultaneously.

We can think of these models as determining hit rates (action responses when action yields the better outcome) and false-alarm rates (action responses when action yields the worse outcome). The difference between hits and false alarms is thus C, regardless of whether the scenarios are consistent or inconsistent (as noted for the CNI model). Both hits and false alarms are affected by A, the bias toward responding action. The ROC curve of signal detection theory is a plot of hits against false alarms as the bias (A) is varied, and it would thus consist of a diagonal line with a slope of 1, if A were varied systematically while holding

[4]The model failed to converge for 11 subjects, and these were removed from all analyses.

other factors constant. Heit and Rotello (2014) have argued that this assumption is implausible and usually false when tested.[5] Moreover, the assumption of linear ROC curves, implicit in the CNI model, can lead to false or misleading conclusions concerning correlations with external variables (such as sex) or effects of manipulations such as cognitive load.

The third model was the same as the second except that it used logistic regression instead of linear regression. This model has a clear theoretical interpretation. Namely, N and C have multiplicative effects on the probability of choosing action, or, in other words, they have additive effects on the log odds of choosing action. This assumption, like that of the second model, maintains symmetry between the norms and consequences. And it is consistent with Luce's (1959) choice axiom, and with the idea that ROC curves are curved (as they usually are) rather than linear. Support for this interpretation comes from the finding that a multilevel model predicting the response from V_n and V_c, using logistic regression, found no significant interaction between norms and consequences as predictors. As Luce's model would predict, their effects are additive on the log odds scale implicit in logistic regression.[6]

The squared residuals were about the same for the first two models, and significantly lower for the third model, the logistic model, than either of the other two models ($p < .025$, t test).[7] Given that this logistic model is empirically no worse than the others and more theoretically sensible, it would make sense to use it in future studies of trade-offs of norms and consequences, when the response is binary. It would also be useful to ask for confidence or strength-of-preference judgments, and then do a linear regression like that of the second model. Such additional data would also permit analysis of ROC curves.

In sum, we have very little reason to accept the sequential assumption of the CNI model. Parsimony favors a model that is consistent with a simpler form of conflict concerning whether deotological considerations are strong enough to override the utilitarian response (e.g., Baron & Gürçay, 2017).

5 Correlations with other variables

GACFH makes inferences about other variables, such as sex and cognitive load, on the basis of correlations with model parameters. GACFH did not check to see whether these correlations were consistent across scenarios; this should be done routinely. We checked the correlation between sex and attention to norms in Study 1a, and it worked for all scenarios.[8]

[5] We cannot plot ROC curves from the present data because bias is not manipulated within subjects.

[6] We used the glmer() function in the lme4 package of R, with crossed random effects for subjects and for the 6 scenarios.

[7] We also analyzed Study 2a. Here, the logistic model was again significantly better than the original model (p=.015), but the linear model was between the two and not significantly different from either. In contrast with the results reported by GACFH, this analysis failed to find a significant correlation between the load condition and the action bias parameter, or any other parameter, for any of the three models. We tested these correlations using subjects as the unit of analysis.

[8] The p-values reported in the article are much lower than those we obtain for tests using subjects alone as the unit of analysis, which, in turn, should give lower expected p-values than those obtained testing concurrently across both subjects and items. It seems that

As we have already noted, the CNI model seeks to distinguish attention to norms and consequences, independently. Women were found to pay more attention to norms than men, but apparently no less attention to consequences. What could it mean to pay more attention to norms, but the same attention to consequences? As it turns out, this result could arise for other reasons aside from the apparent one. In particular, men could pay less attention to everything. They could be more variable, more prone to errors, or less consistent with the model. Or they could be closer to the floor or ceiling (i.e., with a stronger response bias for or against action), leaving less room for other factors to have any effect. Or men could be more antisocial. One of these alternatives must be true. The role of other influences on model parameters — while necessarily present — is especially clear from the fact that the N and C parameters of the original model correlate positively across subjects ($r = .70$).[9] This correlation could arise from individual differences in the influences just mentioned.

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

6 What we already know about omission bias

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

the software used for hypothesis testing, multiTree, ignores both subject and item variance, and thus uses observations as the units of analysis. Moshagen (2010, p. 52) says, "multiTree offers no means to diagnose or handle heterogeneity across items and/or participants. This is considered a major limitation and will be addressed in future versions." The changelog up to the current version (v046) does not mention any correction of this deficiency. The p-values in GACFH thus do not allow their normal interpretation. In particular, we cannot generalize these results to the population from which the subjects are drawn. (And the same may be true of the items.)

[9] The correlation is .08 for the linear model and .64 for the logistic model.

& Ritov, 2009). Of course, the basic result in standard sacrificial dilemmas also relies on a manipulation of consequences: action leads to better consequences than omission, yet omission is chosen.

Here is what we know based on these earlier studies.

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

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

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

Finally, Baron, Scott, Fincher, and Metz (2015) and Baron, 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) plays some role in explaining the existence of non-utilitarian responses, but it is now pretty clearly demonstrated that other factors are relevant too. One major determinant has something to do with the perception of direct causality (as also argued by Greene et al., 2009), and another has something to do with an attachment to particular moral rules (as also argued by GACFH, and by other results in Baron & Ritov, 2009, concerning protected values).

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

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

References

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/.

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. & 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., Gürçay, B., & Luce, M. F. (2017). Correlations of trait and state emotions with utilitarian moral judgments *Cognition and Emotion*. http://dx.doi.org/10.1080/02699931.2017.1295025.

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

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.

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.

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.

Luce, R. D. (1959). *Individual choice behavior: A theoretical analysis*. New York: Wiley.

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

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.

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

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