**Trolley Dilemmas Fail to Predict Ethical Judgment in a Hypothetical Vaccination Context**

**Abstract**
We investigated whether the responses of 68 ethics committee members and staff to trolley dilemmas could predict their responses to research ethics problems concerning vaccine trials. Trolley dilemmas deal with the issue of sacrificing some for the benefit of many, which is also a core issue in the vaccination trial dilemmas. The subjects’ responses to trolley dilemmas showed no statistically significant correlation with their responses to our vaccination trial dilemmas. We concluded that, if there is a component of transferrable intuition between the contexts, it must be small and dominated by other factors. Furthermore, the willingness to sacrifice some for many was larger in the trolley context, despite a more favorable risk/reward ratio and the voluntary participation of the subjects at risk in the vaccination situations. We conclude that one’s general willingness to trade lives in the trolley context may be an artifact that is due to its unrealistic setting.

**Introduction**

Trolley dilemmas are iconic philosophical thought experiments, often assumed to track ethical orientation on a deontological-utilitarian axis(Foot, 1967; Thomson, 1976, 1985, 1990; Unger, 1996; Kamm, 2007, 2015; Bruers & Braeckman, 2013). In the classical version, a trolley moves towards five people on a track. If nobody intervenes, these individuals face a certain death. However, you are standing by a switch and can move the trolley onto a side track by turning the switch. This move will save the five people on the main track, but one person on the side track will die. Should you turn the switch? In various surveys, most people (around 90%) reply that they would turn the switch and kill one person instead of five. This choice aligns with utilitarian ethics maximizing utility or good. However, when presented with a modified trolley problem where a fat man is standing on a bridge above the track, most people do not find it permissible to push the man onto the track to stop the trolley and, thereby, save five people by sacrificing one. These results are often interpreted to mean that people are not guided by mere consequentialism in moral choices, but that deontological ethics also play a decisive role (Kamm, 1989).

Opposing the idea that trolley dilemmas reveal some basic, universal morality (Hauser, Young, & Cushman, 2008), the dilemmas are often criticized for being artificial and not involving the same psychological processes and moral reasoning as more realistic moral situations (Pincoffs, 1986; Singer,1999; Bauman, McGraw, Bartels, & Warren, 2014). Nevertheless, trolley dilemmas play significant roles in medical ethics and medical research ethics discussions and guide enquiries into sham surgery, vaccine trials, and challenge studies of human volunteers (Fritz, 2015; FitzPatrick, 2003; Hope & McMillan, 2004; Albin, 2005; Spier, 2011; Rosenbaum, 2018). For example, trolley analogies played a role when Spier (2011) argued that phase III studies in vaccine trials are unnecessary and when Rosenbaum (2018) discussed whether a large dengue vaccine program should be initiated.

If trolley dilemmas reveal some basic ethical intuitions, one would expect a link between how people reply to them and how they reply to more realistic problems of harming a few for the benefit of many. In the present paper, we report on an empirical study that was designed to test such a link in a medical research ethics context. We used a questionnaire posing traditional trolley dilemmas and questions regarding the ethical justifiability of hypothetical clinical trials, where a potentially harmful vaccine would be tested on a group of people, in order to verify its safety for subsequent population wide use. We expected that a high (or low) willingness to trade some lives for the benefit of many in the trolley cases would correspond to a similar tendency for the same respondents to have a high (or low) willingness to accept the vaccine trials. Our study was conducted at a Norwegian conference for medical research ethics committees with subjects who are used to considering, and even deciding, such issues.

Vaccination contexts include research trials, where new vaccines come with unknown risks, and vaccine programs, where risks are better (although not fully) known. In ethics and psychology literature, vaccination contexts are widely regarded as mirroring some key characteristics of trolley dilemmas (Young and Koenigs, 2007; Bartels, 2008; Wiss, Andersson, Slovic, Västfjäll, & Tinghög, 2015; Bialek & De Neys, 2016; Rosenbaum, 2018). Vaccination contexts and trolley dilemmas both present a conflict between the utilitarian principle of “taking actions that give the best result over all” and the deontological moral rule of “doing no harm.” In both types of dilemmas, some people will be harmed to benefit a higher number of people. Although trolley problems and vaccination contexts are similar regarding this core characteristic, there are also many differences between them that may influence people’s moral reasoning. Relevant differences include variation with regard to risk-reward ratio, consent, immediacy, moral status of participants, incentives, uncertainty, and the doctrine of the double effect, all of which we will address in more detail in the discussion.

Even though trolley dilemmas can be viewed as representing a utilitarian-deontological axis, some studies in moral psychology have indicated that the unwillingness to push the fat man off the bridge is more an emotional and instinctive aversion towards the use of direct physical violence than a result of ethical reasoning (Greene, Sommerville, Nystrom, Darley, & Cohen, 2001; Royzman & Baron, 2002; Lanteri, Chelini, & Rizzello, 2008). An alternative account by Kahane (2012; 2015) suggested that commonsensical moral notions rather than a utilitarian outlook on life are the basis for the choices people make in some of the trolley dilemmas. Thus, relevant literature disagrees on whether trolley dilemmas are informative about the respondents’ ethical reasoning with regard to utilitarian and deontological thinking. We will take this disagreement into account in the discussion of our results.

In the following sections, we present our methods and results and discuss our findings, the main finding being a negative one: The subjects’ replies to the trolley dilemmas did not contain information that could predict or explain their responses to the vaccine trial problems.

**Methods**

The study was conducted during the 2016 conference for the Norwegian regional (and national) ethics committees for health research. The administrative staff also participated, as well as some speakers and guests. This conference (“Storfellesmøte” in Norwegian) is the main annual event for discussing applied medical research ethics in Norway.

The questionnaire was designed with six items, grouped into a Trolley Index and a Vaccination Index. Each item was a yes/no question, and the index scores were defined as the number of yes-responses in the respective item groups. The indexes were designed with the goals of validity, reliability, and responsiveness in mind.

The following three standard trolley dilemmas were chosen (Bruers & Braeckman, 2013):

*Dilemma 1: The switch*. A trolley is moving towards five people on the main track. You are standing at a switch. If you turn the switch, the trolley will be diverted to a side track, but there is one person on this side track. Turning the switch will result in that person’s death, and the five people on the main track will be saved.
**Should you turn the switch?**

 *Dilemma 2: The bridge*. A fat man is standing on a bridge above the track. You can save the five people on the track below by pushing the fat man from the bridge in front of the trolley, so that the trolley will be stopped by his heavy weight. The fat man will die, and the five people will be saved.
**Should you push the man?**

*Dilemma 3: The loop.* As in the first dilemma, you are standing at a switch. But this time, the side track turns back onto the main track. If there is no one on the side track, the trolley will still move onto the main track and will kill the five people. But on the side track is a fat man. So if you turn the switch, the fat man will block the trolley, save the five people on the main track, but die himself.
**Should you turn the switch?**

Since a subject’s Trolley Score is calculated as the number of yes-responses to these dilemmas, it has a value range of 0 to 3. The validity of a questionnaire index is defined as the degree to which it “measures the right thing.” The purpose of our index is to measure respondents’ willingness to trade lives in a trolley-problem context; we expect it has high face validity, since we have chosen commonly used trolley problems.

The responsiveness of an index is its ability to capture differences among respondents. Care was taken in choosing dilemmas known to have a certain response variation individually while also having different rates of positive responses, increasing the likelihood that the scores will spread out well.

The research ethics dilemmas were designed to have similarities with research projects encountered by the committees, and are actually (substantially) modified versions of a real project evaluated by one of the committees. The dilemmas were formulated as follows:

*Dilemma 4:* A research group has developed a vaccine against the contagious disease Alobe, which occasionally has large outbreaks in the central African country Manigua. The effect of the vaccine (the immunity) is short lived, but mass-vaccination around an outbreak is expected to save thousands of lives in the future. In experiments, the vaccine triggers fatal heart failure in 3% of laboratory mice, but this side effect is unlikely for humans, who will receive a smaller dosage in relation to body weight. However, the Manigian authorities refuse to give approval of the vaccine, until its safety is established. The project will therefore test the safety of the vaccine on 1000 consenting adults, but discontinue the testing in case of deaths. The participants will receive no compensation, and will most likely have no benefit from the vaccine, since the effect is short lived. One still expects subjects to volunteer, since it is inherent in the Manigian culture to take personal risk in order to save others.
**Is the study ethically justified?**

*Dilemma 5:* The same setting as above, except that the assumption about Manigian culture turns out to be incorrect. In order to recruit volunteers, the project will offer the Alobe vaccine bundled with a new vaccine against Malaria, which is otherwise not accessible. The Malaria vaccine is known to have a positive health effect that outweighs the risk of the Alobe vaccine.
**Is the study ethically justified?**

*Dilemma 6:* The same setting as above, but the research group argues that the subjects are autonomous individuals, and prefer to compensate them with two Manigian annual salaries, which is easier to administer than bundled vaccination. The subjects will be informed that this amount is sufficient to finance Malaria vaccination for their entire nuclear family at the local health station, if they choose to do so.
**Is the study ethically justified?**

The Vaccination Score was defined as the number of yes-responses to Dilemmas 4 to 6, and also has a range of 0 to 3. For this index, validity is more of an issue. The index is supposed to capture willingness to sacrifice some people’s welfare for the good of others in a medical research context. We can claim that the dilemmas have acceptable face validity because they are (loosely) based on an actual project evaluated by one of the ethics committees. Ideally, we would prefer to include more than three dilemmas and to cover more diverse problem situations. However, to achieve a high response rate, we needed to restrict the questionnaire to one page. To ensure responsiveness, the versions were designed, based on informal piloting among colleagues, to have different levels of acceptability. Dilemma 4 was intended to give around 50% yes-responses. In Dilemma 5, the bundling with a Malaria vaccine was intended to increase the percentage of yes-responses by ensuring a positive net health effect, while Dilemma 6 was expected to give lower yes-rates due to ethical issues with monetary incentives.

In addition to the six dilemmas, the questionnaire included the demographic items of age, sex, role in the meeting, and number of ethics committee meetings attended. Reducing risk of identification, response options for age and number of meetings were grouped.

The main hypothesis was that the Trolley Score would show statistically significant correlation with the Vaccination Score, with a two-sided test and significance level of 5%. The experiment had 80 % power to detect a correlation of 0.33, computed by the R function pwr.r.test. A correlation of 0.33 would imply that approximately 10% of the Vaccination Score variability could be explained by the Trolley Score. As secondary outcomes, we compared the trolley response distributions to previously published findings and investigated possible correlations between the demographics variables and the Trolley and Vaccination Scores.

**Results**

**Descriptive statistics**

There were 75 respondents, one of whom filled in demographics only. There were few missing values, and 68 questionnaires were fully completed. Tables 1 to 3 give the distributions of the demographics variables, dilemma responses, and Trolley/Vaccination Scores, respectively.

*Table 1*

*Table 2*

*Table 3*

**Psychometric properties**

We tested the reliability of our two indexes with the Cronbach’s alpha statistic (DeVellis, 2017). It measures the internal correlation among items of an index on a scale from 0 to 1. The Trolley Index received an alpha score of 0.76, and the Vaccination Index scored 0.61. A score above 0.7 is sometimes considered desirable, and our Vaccination index does not quite meet this standard. However, the alpha statistic is very sensitive to the number of questions, and with only three items, 0.61 is reasonable. Furthermore, there is a tradeoff between internal correlation and responsiveness, as an item that correlates strongly with the others contributes with less independent information. Therefore, a moderately high alpha value may be desirable.

We evaluated responsiveness with the entropy-based measure of information efficiency (Dahl & Østerås, 2010), which quantifies the information content on a scale from 0 to 1. Its formula is given as ${-\sum\_{i=1}^{k}p\_{i}log\_{2}(p\_{i})}/{log\_{2}(k)}$, where *k* is the number of response options and $p\_{i}$ is the probability of response *i*. The numerator represents the number of bits (binary information units) that is collected per subject in the given data set, while the denominator is the maximum number attainable. Dilemmas 1 to 6 have individual information efficiencies of 0.62, 0.99, 0.82, 0.97, 0.94, and 0.71, respectively. For Dilemma 2, the response distribution is close to 50-50, which gives an information efficiency score close to 1. Dilemma 1 displays the lowest information efficiency because it has the most skewed distribution.

As stated previously, each scale has a value range of 0 to 3. With four values, optimal information efficiency corresponds to a distribution of 25% for each. Intuitively, one might think that an ideal scale should have a maximum variance among the subjects. This is incorrect, however, as the maximum variance would imply that the subjects’ scores were evenly split between the end points 0 and 3. This would mean that each subject responded identically to all questions in the scale, rendering two out of three questions redundant. The response distributions of the indexes, as given in Table 2, give an information efficiency of 0.90 for the Trolley Index and 0.91 for the Vaccination Index. This means that the amounts of information collected by the two indexes are 90% and 91% of the theoretical maximum, which implies high levels of responsiveness.

There is no universally accepted procedure for evaluating the validity of indexes, without comparing them to external gold standards, which are not available in the present case. However, the high response rate indicates that most subjects accepted the dilemmas as meaningful. In addition, a subsequent discussion session indicated that the subjects had interpreted the dilemmas correctly, and that they did not dismiss them as irrelevant to the context of medical research ethics.

**Hypothesis test**

The correlation test gave a p-value of 0.087, so there was no statistically significant Pearson correlation. The correlation was estimated to 0.209 with the confidence interval (-0.047, 0.436). The point estimate would imply that 4.4% of the variance in the Vaccination Score could be explained by the Trolley Score (and given the lack of significance, it may actually be 0%).

The hypothesis of correlation between the scores showed no statistical significance for relevant subgroups of the dataset, such as males or females, old or young, members of committees or staff, etc. We also investigated linear regression models with the Vaccination Score as the dependent variable, the Trolley Score as the main predictor, and the different demographic variables as possible confounding factors. None of the demographic variables were statistically significant, so the preferred model included only the Trolley Score, which is mathematically equivalent to the original correlation analysis.

Among the trolley dilemmas, it can be argued that Dilemma 2 stands out as different, since the emotional barriers to pushing people may confuse the utilitarian-deontological response we are trying to measure. Therefore, we also tested a variation of the hypothesis where this dilemma was excluded from the Trolley Index. This pushed the p-value up to 0.14, which shows that the non-significance finding was not due to the inclusion of Dilemma 2.

Table 4 gives the correlation matrix for all six dilemmas with p-values in parentheses.

*Table 4*

We see a strong correlation between trolley dilemmas and a slightly less strong correlation between vaccination dilemmas, which is related to the higher Cronbach’s alpha score for the Trolley Index. From a strict hypothesis testing perspective, the significance level should be adjusted for the number of correlations that are estimated. Such a Bonferroni correction reduces the p-value limit from 0.05 to 0.003, which means that among the vaccination dilemmas, only 5 and 6 show statistically significant correlation. However, the only formal hypothesis test in our experiment was for the index level correlation. We therefore regard the empirical correlations in Table 4 as supplementary descriptive statistics only and leave the p-value interpretation to the reader. This also applies to the estimated correlation between Dilemmas 2 and 6 (0.27), with a p-value of 0.024.

**Discussion**

As the two scales constitute the foundation of our experiment, their psychometric properties are vital. Since they were accepted as relevant by the subjects, and most of them responded to all items, we conclude that they have at least reasonable face validity. The scales also scored acceptably on a standard test of reliability. Since we had only three items in each scale, it was particularly important to achieve a high level of responsiveness in order to collect as much information as possible. An appropriate responsiveness measure is information efficiency, according to which both scales scored at least 90%. Hence, the scales had favorable psychometric properties and were suitable for the purpose of the experiment.

The lack of a statistically significant correlation between the Trolley Scores and Vaccination Scores is our main finding. Accordingly, subjects with a relatively high Trolley Score did not have a consistent tendency toward a higher (or lower) Vaccination Score. Strictly speaking, a negative result like this does not rule out a possible correlation, as it merely fails to demonstrate a positive one. However, even interpreted at face value, the point estimate of the correlation implies that the Trolley Score explains only 4.4% of the variance in the Vaccination Score.

There are several morally and psychologically relevant differences between the dilemma settings, which may play a role:

* *Risk-reward ratio*. In the trolley dilemmas, five people can be saved by sacrificing one, while in the vaccination dilemmas, thousands of people are likely to be saved, with few expected casualties (arguably none in Dilemma 5). Hence, there is a large difference in the expected risk-reward ratio in favor of the vaccination dilemmas.
* *Consent.* In the trolley cases, people are sacrificed without being asked, while in the vaccination trials, participation is voluntary and consent is given. This difference renders the vaccine trials less problematic than the trolley cases on a crucial dimension.
* *Immediacy*. Trolley dilemmas force the subject to choose between one and five here and now, while choosing to approve or reject the vaccine trials regards something further away in space and time. Since vaccine participants are more distant than the possible victims in the trolley cases, we would expect the vaccine cases to be psychologically easier to accept along this dimension.
* *Moral status of research participants.* Research participants are prone to exploitation and, therefore, considered in need of protection. Extensive literature and international regulations exist to protect research participants, which may increase the barrier towards accepting the vaccine trials. Anonymous random people on a track do not have this particular right to protection.
* *Incentives.* In Dilemma 5, and even more so in Dilemma 6, the research subjects are given incentives for participating in a study that they might otherwise refuse. This dimension is not present in the trolley dilemmas, as the “participants” do not have a choice. Ethics committees are typically cautious when it comes to incentives, as any type of coercion should be avoided in recruiting research participants.
* *Uncertainty.* The trolley dilemmas present certain outcomes, while the vaccination dilemmas include uncertainty regarding the harmful side effects over all, as well as randomness in the outcomes for each included participant.
* *Doctrine of the Double Effect (DDE)*. DDE considers the moral difference between killing and letting die (Foot, 1967). The claim is that it is more permissible to let people die as a side effect, as a foreseeable but not intended consequence, than it is to kill someone as a mere means to save someone else. In Dilemma 1, one can argue that the action is performed as a means to save the five, while the death of the single person is an unintended side effect. In Dilemmas 2 and 3, however, the fat man’s body is used as a direct means to stop the trolley, so his death cannot be considered a side effect. The vaccination dilemmas are less clear-cut DDE-wise, as the research subjects’ bodies are used more indirectly by providing information that is required in order to save the lives of many.

When formulating our hypothesis of a correlation between the settings, the idea was that, although these differences are likely to shift the response distributions up or down, this should be largely independent of the deontological-utilitarian dimension. For instance, if the immediacy of the trolley situation increases the number of yes-responses in Dilemmas 1 to 3, we would still expect the subjects with a higher than average Trolley Score, due to a utilitarian view, to also have a higher than average Vaccination Score. Conversely, the fact that the vaccination dilemmas included informed consent should increase Vaccination Scores over all, keeping utilitarian responders above deontological ones in both settings.

We see two alternative explanations to the observed lack of correlation:

1. *The utilitarian-deontological axis is not a valid construct, as people approach each new setting differently.*
2. *The utilitarian-deontological axis is a valid and transferrable construct, but its impact is small compared to other factors*.

Explanation 1 implies that “orientation on the utilitarian-deontological axis” is not a valid conceptualization and, therefore, fails to represent any moral intuition transferrable from one context to another. Although our findings are compatible with this view, we cannot draw such a general conclusion from a single moderately sized experiment. However, studies by Kahane (2012; 2015) also appear to support Explanation 1. He argued that subjects approach trolley dilemmas by weighing more commonsensical norms against each other, rather than weighing utilitarian against deontological reasoning.

Explanation 2 assumes that even though the axis is valid, and utilitarian-deontological orientation is transferable from one context to the other, this effect is dominated by other factors. From a statistical point of view, we might consider the relative weight that a randomly chosen subject assigns to consent, immediacy, moral status, etc., as noise, which masks the effect of his/her utilitarian-deontological orientation. Another source of random noise is the decision-making process itself, as most people find the dilemmas difficult and their thought process may end up with unpredictable results. A person might place a higher or lower importance on consent or risk-reward ratios in the spur of the moment. In the language of the signal processing theory, the signal-to-noise ratio may be too low for a statistically significant measurement of the transferability of a subject’s utilitarian-deontological orientation.

From a scientific point of view, Explanation 2 is problematic because it cannot be falsified, as no finite experiment can prove that a correlation is exactly zero. The present experiment was powered to detect a correlation level that was considered a minimum for practical relevance (explaining 10% of the variability). Therefore, the fact that it failed to produce statistical significance supports the conclusion that the effect is either nonexistent or too small to be important. The trolley dilemmas turned out to be essentially useless for predicting or explaining people’s responses to the given vaccination dilemmas. The important question, then, is how far we can generalize this result of non-generalizability. The most striking differences between the trolley and vaccination settings appear to be those related to risk-reward ratios, urgency, and consent. We would therefore expect our finding of low predictive value from the trolley responses to generalize well to other dilemmas relating to vaccination research, since any reasonable vaccination program will be intrinsically non-urgent, based on consent, and have high risk-reward ratio.

We also believe that this conclusion would be relevant for a broad range of medical research dilemmas, as few medical interventions have the characteristics of being urgent, performed without consent, and pose life-threatening risks to others than those who are expected to benefit from the intervention. We might add the following item to our list of morally or psychologically relevant differences or regard it as a summary of the ones presented:

* *Realism.* The trolley dilemmas describe scenarios that responders are extremely unlikely to experience, while the vaccination dilemmas are much more realistic (particularly for our sample of research ethics committee members).

The difference in realism may trigger different modes of reasoning by the subjects, particularly if the trolley context is perceived as a “toy problem” in which the task is to figure out a “hypothetically correct answer,” as opposed to exercising authentic moral judgment. Hence, a part of the explanation may be that trolley scenarios are too unrealistic and artificial to give a significant bearing on moral judgment in the more realistic vaccination dilemmas. If this effect were complete, it would fit into our Explanation 1 above. Nevertheless, it is a valid alternative that the effect of (lack of) realism reduces the transferrable component of the subjects’ moral judgment to the extent that it becomes dominated by the random noise discussed under Explanation 2. If this is correct, it means that trolley dilemmas mostly generalize to unrealistic problems. This line of criticism is not new, as several studies suggest that trolley dilemmas are too artificial and de-contextualized to be of much use for real-life ethics (O’Connor, 2012; Bauman, McGraw, Bartels, & Warren, 2014; Gold, Pulford, & Colman, 2014). According to Bauman, McGraw, Bartels, & Warren (2014, p. 536-7), there are good reasons to worry that “judgment and decision-making processes people use in these unusual situations may not accurately reflect moral functioning in a broader set of situations.” They suggested three main problems for the external validity of trolley problems: (1) Many find trolley problems morbidly funny, which may inflict on moral judgment and indicates a suboptimal experimental setting. (2) Respondents often do not find the trolley dilemmas believable or acceptable, and question several premises: Can the fat man really stop the trolley with his body, and how come the people on the track cannot escape? (3) People have very different psychological responses to more realistic moral dilemmas compared to the trolley dilemmas. These general threats to external validity for the trolley problems may add to the explanation of why we cannot extract any information from people’s responses to trolley problems regarding their responses to medical ethics dilemmas.

Several studies of medical research ethics have used trolley problems to guide their discussion (Albin, 2005; FitzPatrick, 2003; Hope & McMillan, 2004; Spier, 2011; Rosenbaum, 2018). For example, Rosenbaum (2018) used trolley problems as analogues to the distribution of the dengue vaccine. If distributed in a massive vaccine program for all children over 9 years old, this vaccine would decrease severe cases of dengue with 80%. The problem is that some children, who would otherwise not be afflicted, would become severely ill because of the vaccine. Rosenbaum discussed, from a public health perspective, whether the right thing to do would be to start a massive vaccine program, analogous to how most people judge the first trolley problem. She ended up arguing that this option probably is not the ethically best choice because of other possible negative consequences, such as increased skepticism to vaccines.

Trolley dilemmas are similarly used in other areas of medical ethics. For example, Albin (2005) argued that it is morally permissible to perform sham surgery on control subjects in research on operation procedures. Sham surgery is, by many, regarded to be unethical due to its relatively high risk and invasive nature (Macklin, 1999). According to Albin, it is permissible to put the controls at risk in order to save future patients as long as the core problem is analogous to the classical trolley problem and the participants consent and know the risk. Use of trolley analogies can also be found in, for example, FitzPatrick (2003) and Hope and McMillan(2004). Opposing such extrapolations of trolley-based reasoning, our study shows that even though the risk is small, and consent is in place, the context of research may elicit different moral intuitions and/or reasoning than trolley problems. Therefore, one should be cautious in using trolley problems as tools for reasoning in medical research ethics.

This normative suggestion may be criticized for relying on the assumption that moral responses to more real-life ethics dilemmas are more likely to be accurate than responses to idealized thought experiments. It could be the case that more realistic and, thereby, messy scenarios may engage irrational reactions compared to responses to clean-cut trolley dilemmas. Even if this were the case, our discussion demonstrates that the existing morally relevant differences between the problem sets need to be considered to decide the ethical justifiability of the vaccine trials. In the trolley dilemmas, however, there are no resources that can help with evaluating the moral importance of these differences. Thus, even though we may not conclude that the responses to the vaccine trials were “correct moral responses,” we propose that trolley dilemmas strongly underdetermine the moral reasoning required concerning the vaccine trials.

**Other findings**

The response distribution for the trolley dilemmas corresponds well with previously published results with one particular diversion: there is a tendency to sacrifice the fat man on the bridge more readily compared to other surveys. In our sample, 43% replied that the fat man should be pushed compared to 10 to 25% in other surveys (Hauser, Young, & Cushman, 2008). However, the pattern is that a large majority sacrifices the one person in Dilemma 1, far fewer push the heavy man in Dilemma 2, and Dilemma 3 is in between 1 and 2. We may speculate whether the consequentialist response to the bridge dilemma could result from experience of the respondents in deliberating over ethical problems, perhaps increasing the ability to detach from the emotional response normally evoked by the bridge dilemma (Greene, Sommerville, Nystrom, Darley, & Cohen, 2001).

A secondary finding is that, although they were relatively willing to act in the trolley dilemmas, the participants were quite reluctant to approve of the proposed vaccine trials. On a scale from 0 to 3, the average Trolley Score was 2.04, while the average Vaccination Score was only 0.94. This means that the responders gave a yes-response in slightly more than two out of three trolley dilemmas and slightly less than one out of three vaccination dilemmas. We were surprised that our responders found it considerably more acceptable to sacrifice one non-consenting person to save five than to let consenting adults take a relatively small risk to save thousands. In particular, even the most acceptable vaccination dilemma (Dilemma 4) was approved by fewer than the least acceptable trolley dilemma (Dilemma 2). It seems the respondents find it more acceptable to kill a bystander by pushing him in front of a trolley to save five than allow people to voluntarily take a relatively small risk to save a large number of fellow citizens. According to our list of relevant differences above, we find that Dilemma 4 scores better on items *Risk-reward ratio* and *Consent*. The items of *Incentives* and *DDE* appear to be irrelevant or neutral, while items that favor Dilemma 2 are *Immediacy*, *Moral status of research participants*, and *Uncertainty*. Of these, uncertainty is the least relevant, since the possibility of a bad outcome in Dilemma 4 is already accounted for in the risk-reward ratio. Furthermore, although moral status is relevant, we will suggest that it is dominated by considerations of consent, since the subjects in the vaccination dilemma are aware of their status and are free to leave, as opposed to the trolley victims. The bottom line, then, is that immediacy dominates risk-reward ratio considerations. It must be noted that the psychological aspect of immediacy that might weigh in favor of accepting Dilemma 2 only has to do with time, as the spatial immediacy in pushing the man should work against it. Although psychologically relevant, it is hard to accept the idea that the timing of people’s deaths can outweigh a risk-reward ratio that is at least 100 times more favorable for the vaccination dilemma (1:5 versus (a few):thousands).

Instead, we think that the summary item of *Realism* may be a more correct explanation. If our subjects regard the trolley dilemma as a theoretical toy problem and relate the vaccination dilemma to a real decision problem with actual human beings, this would provide a better explanation of our findings.

**Concluding Remarks**

Many studies of moral dilemmas aim at isolating and identifying factors that influence how one distinguishes right from wrong. We used a somewhat different approach by investigating whether replies to trolley dilemmas by ethics committee members could predict how they reply to real life-like dilemmas, with the same moral core issue of trading some lives for the benefit of many.

We found that a subject’s replies to trolley dilemmas did not correlate with his/her replies to questions about the ethical justifiability of vaccination trials. We find it surprising that the Trolley Score failed to distinguish between subjects that were more or less willing to accept the vaccination project. Although the two contexts are different in many ways, a core issue in both is the sacrificing of a few persons’ welfare for the benefit of many. We expected that the subjects’ orientation on the deontological-utilitarian axis would play an important role in both contexts, so that subjects with a higher Trolley Score would also accept the vaccination project more readily than those with a lower Trolley Score. The importance of the findings is strengthened by the fact that the population studied is highly relevant for the given medical dilemmas.

Therefore, even if trolley dilemmas and vaccination contexts share a core dimension of sacrificing some for the benefit of many, the context sensitivity of moral judgements seems to be so high that barely any “core intuitions” carry over from the simplified trolley cases to the more contextual vaccine cases (in support of, for example, Bartels, 2008). These results indicate that one should be cautious in using trolley dilemmas to guide reasoning in vaccine trials, in particular, and in medical research ethics, in general. The medical context under study is different from the trolley dilemmas in several ways, which is surely a reason why moral intuitions did not carry over. However, given the simplified and unrealistic nature of the trolley dilemmas, this will be the case for most realistic medical ethics contexts.

**Best Practices**

Based on the results of our study, we recommend not to use analogies to trolley dilemmas as a basis for ethical evaluation of medical research scenarios that involve the question of trading some lives or well-being for the benefit of many.

**Research Agenda**

Trolley dilemmas have an iconic status in ethics and psychology. As discussed in this paper, they have also had an influence on human research ethics. It is of great value to obtain some empirical data on connections between how people judge trolley dilemmas and how they judge research ethics dilemmas considered analogous to trolley dilemmas on some crucial dimensions. Although we provide data from members of research ethics committees in this paper, it would be of further interest to investigate whether similar results would be obtained in other demographics as well as to address connections between ethical judgements in trolley dilemmas and other types of research ethics dilemmas than the vaccination scenarios chosen here.

**Educational Implications**

Trolley problems are common parts of any ethics curriculum. In teaching medical ethics, our results are educational regarding the connection between such theoretical ethics problems and real-life research ethics dilemmas.

**References**

Albin, R.L. (2005). Sham surgery controls are mitigated trolleys. *Journal of Medical Ethics* 31:149-152 doi:10.1136/jme.2003.006155.

Bartels, D.M. (2008). Principled moral sentiment and the flexibility of moral judgment and decision making. *Cognition* 108: 381-417.

Bauman, C.W., McGraw A.P., Bartels D.M., and Warren, C. (2014). Revisiting External Validity: Concerns About Trolley Problems and Other Sacrificial Dilemmas in Moral Psychology. *Personality and Social Psychology Compass* 8(9): 536-554, doi: 10.1111/spc3.12131

Bialek, M. and De Neys. W. (2016). Conflict detection during moral decision-making: evidence for deontic reasoners’ utilitarian sensitivity. *Journal of Cognitive Psychology* 28(5): 631-639.

Bruers, S. and J. Braeckman (2013). A Review and Systematization of the Trolley Problem. *Philosophia* 42: 251-269.

Dahl, F. A., Østerås N., (2010): Quantifying Information Content in Survey Data by Entropy, *Entropy*, 12 (2): 161-163.

DeVellis, R. F. (2017). *Scale Development: Theory and Applications (Applied Social Research Methods)* 4th Edition. Sage, Thousand oaks, USA.

Foot, P. (1967). The problem of abortion and the doctrine of the double effect, *Oxford Review* 5.

FitzPatrick, W. (2003). Surplus Embryos, Nonreproductive cloning and the Intend/Foresee Distinction, *The Hastings Center Report* 33: 29-36.

Fritz, Z. (2015). Can ‘Best Interests’ derail the trolley? Examining withdrawal of clinically assisted nutrition and hydration in patients in the permanent vegetative state. *Journal of Medical Ethics* doi:10.1136/medethics-2015-103045

Gold, N., B.D, Pulford, & Colman, A.M. (2014). The outlandish, the realistic, and the real: contextual manipulation and agent role effects in trolley problems. *Front Psychol* 5: 35.

Greene, J., Sommerville, B. R., Nystrom, L. E., Darley, J. M., and Cohen, J. D. (2001). An fMRI investigation of emotional engagement in moral judgment. Science, 293, 2105–2108.

Hauser, M., Young, L., and Cushman, F. (2008). Reviving Rawls’ linguistic analogy: operative principles and the causal structure of moral actions. In *W. Sinnott-Armstrong (Ed.), Moral psychology and biology*. NY: Oxford U. Press.

Hope, T. and McMillan, J. (2004). Challenge studies of human volunteers. *Journal of Medical Ethics* 30: 110-116, doi: 10.1136/jme.2003.004440.

Kahane, G. (2012). On the wrong track: process and content in moral psychology. *Mind & Language* 27(5): 519-545.

Kahane, G. (2015). Sidetracked by trolleys: Why sacrificial moral dilemmas tell us little (or nothing) about utilitarian judgment. *Social Neuroscience* 10(5): 551-560.

Kamm, F. M. (1989). Harming some to save others. *Philosophical Studies*, 57, 227–260.

Kamm, F. (2007). *Intricate Ethics*. Oxford University Press.

Kamm, F. (2015). *The Trolley Problem Mysteries.* Oxford University Press.

Lanteri, A., Chelini, C., and Rizzello, S. 2008. An Experimental Investigation of Emotions and Reasoning in the Trolley Problem. *Journal of Business Ethics* 83: 789-804.

Macklin, R. (1999). The Ethical Problems of Sham Surgery in Clinical Research. *The New England Journal of Medicine* 341: 992-996.

O’Connor, J. (2012). The Trolley Method of Moral Philosophy. *Essays in Philosophy* 13(1): 242-255.

Pincoffs, E.L. (1986). *Quandaries and Virtues: Against Reductivism in Ethics*. Lawrence, Kansas, University of Kansas.

Rosenbaum, L. (2018). Trolleyology and the Dengue Vaccine Dilemma. *The New England Journal of Medicine* 379: 305-307.

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

Singer, P. (1999). Living high and letting die. *Philosophy and Phenomenological Research* 59: 183-187

Spier, R.E. (2011). Vaccine Safety: An examination of the value and necessity of Phase III trials*. Procedia in Vaccinology* 4:1-8.

Thomson, J.J. (1976). Killing, letting die, and the trolley problem. *The Monist* 59(2): 204-217, doi: 10.5840/monist197659224.

Thomson, J.J. (1985), The trolley problem, *The Yale Law journal* 94: 1395-1415.

Thomson JJ. (1990). The trolley problem. In: Thomson JJ. *The Realm of Rights.* Cambridge: Harvard University Press, 1990:176–204.

Unger, P.K. (1996). *Living High and Letting Die: Our illusion of Innocence.* Oxford: Oxford University Press.

Young, L. and M. Koenigs 2007, Investigating emotion in moral cognition: a review of evidence from functional neuroimaging and neuropsychology. *British Medical Bulletin* 84(1): 69-79

Wiss, J., Andersson, D., Slovic, P., Västfjäll, D., and Tinghög G. (2015). The influence of identifiability and singularity in moral decision making. *Judgment and Decision Making* 10(5): 492-502.