Are intuitions about moral relevance susceptible to framing effects?*

James Andow

April 23, 2019

Abstract

Various studies have reported that moral intuitions about the permissibility of acts are subject to framing effects. This paper reports the results of a series of experiments which further examine the susceptibility of moral intuitions to framing effects. The main aim was to test recent speculation that intuitions about the moral relevance of certain properties of cases might be relatively resistent to framing effects. If correct, this would provide a certain type of moral intuitionist with the resources to resist challenges to the reliability of moral intuitions based on such framing effects. And, fortunately for such intuitionists, although the results can't be used to mount a strident defence of intuitionism, the results do serve to shift the burden of proof onto those who would claim that intuitions about moral relevance are problematically sensitive to framing effects.

1 Experimental Ethics and the Reliability of Intuitions

In recent decades, psychologists, economists and experimental philosophers have examined the factors underlying moral intuitions (see, e.g., Cushman et al., 2006; Gold et al., 2015; Greene, 2007; Greene et al., 2009, 2004, 2001; Hauser et al., 2007; Mikhail, 2000, 2007; Nichols and Mallon, 2006; Petrinovich and O'Neill, 1996; Waldmann and Dieterich, 2007). Typically the focus has been on intuitions about the moral *permissibility* of actions. Such intuitions have been shown to vary between different types of trolley cases (Cushman et al., 2006; Greene et al., 2001; Hauser et al., 2007) and progress has been made in understanding the mechanisms underlying such differences (Cushman and Greene, 2012; Greene et al., 2009). In this literature, there has been particular focus on the extent to which participants' intuitions about permissibility are driven by 'consequentialist' or 'deontological' considerations. Accordingly, much of the focus

*This is an author post-print copy produced for self-archiving. See the published version for the final text and page references.

has been on intuitions about cases which aim to tease apart such considerations. In particular, various types of trolley case have been used in which a protagonist can by various means prevent a runaway train from causing five deaths and bring about the death of an innocent bystander.

Here are the four trolley cases which I will be using in this paper (these are taken from Liao et al. 2011 with minor changes).

Switch A runaway trolley is headed toward five innocent people who are on the track and who will be killed unless something is done. Abigail can push a button, which will redirect the trolley onto a second track, saving the five people. However, on this second track is an innocent bystander, who will be killed if the trolley is turned onto this track.

Push A runaway trolley is headed toward five innocent people who are on the track and who will be killed unless something is done. Abigail can push a button, which will activate a moveable platform that will move an innocent bystander in front of the trolley. The runaway trolley would be stopped by hitting the innocent bystander, thereby saving the five but killing the innocent bystander.

Physical A runaway trolley is headed toward five innocent people who are on the track and who will be killed unless something is done. Abigail can run to a nearby bridge on which a heavy bystander is standing and push this bystander from the bridge. The runaway trolley would be stopped by hitting the innocent bystander, thereby saving the five but killing the innocent bystander.

Loop A runaway trolley is headed toward five innocent people who are on the track and who will be killed unless something is done. Abigail can push a button, which will redirect the train onto a second track, where there is an innocent bystander. The runaway trolley would be stopped by hitting the innocent bystander, thereby saving the five but killing the innocent bystander. The second track loops back towards the five people. Hence, if it were not the case that the trolley would hit the innocent bystander and grind to a halt, the train would go around and kill the five people.

The empirical investigation of the causal factors underpinning our moral intuitions potentially has important implications for moral philosophy. Intuitions about cases are often taken as the primary source of evidence for and against theories in normative ethics. If empirical investigation were to show intuitions to be sensitive to factors which are clearly irrelevant to whether an act is in fact moral permissible, this would raise serious worries about the reliability of intuitions and thus about the epistemic foundation of moral philosophy.

This is not just a hypothetical challenge. Various researchers have reported results which suggest that moral intuitions are sensitive to such irrelevant factors. One such

factor is that moral intuitions appear to be sensitive to order of presentation (Lanteri et al., 2008; Lombrozo, 2009; Nichols and Mallon, 2006; Petrinovich and O'Neill, 1996; Schwitzgebel and Cushman, 2015; Wiegmann et al., 2012).[1] By this point, quite a number of studies have reported such results (for a survey, see, Wiegmann et al., 2012). In the following, I focus on two particular types of order effects which have been reported to influence participants' judgements about moral responsibility.[2]

Switch effects Acting in Switch receives lower permissibility ratings when it follows Push or Physical (relative to when Switch is presented first). Such an effect has been found in a number of studies (Lombrozo, 2009; Nichols and Mallon, 2006; Wiegmann et al., 2012).

Loop effects Acting in Loop receives lower permissibility ratings when it follows Push or Physical (compared to when Loop follows Switch). This effect is reported in Liao et al. (2011). Indeed, Liao et al. (2011) find that a greater proportion of participants agree that pushing the button is permissible in Loop when it follows Switch rather than Push.

There is also evidence that intuitions about permissibility are subject to two other effects of framing.[3]

Actor-Observer effects Acting in a case, e.g., Switch, receives lower permissibility ratings when the case concerns the actions of the participant themselves than when the actions are those of a third party (Nadelhoffer and Feltz, 2008).

Asian-Disease effects Tversky and Kahneman (1981) find a very robust framing effect whereby whether the outcomes of potential government programmes are described in terms of numbers of people being 'saved' or 'dying' strongly influences which of the two programmes participants prefer. Participants prefer Program A (described as saving 200 people) to Program B (described as involving a 1/3 probability all 600 will be saved and 2/3 that no one will be saved), but prefer Program B when these same outcomes are described in terms of death.

I should also note that here are various types of responses which one might give to the challenge that evidence of framing effects presents to the methods and epistemology of moral philosophy. For some discussion of some of them, see Andow (2015,

[1] See Wiegmann and Waldmann (2014) for one explanation of these (which they call 'transfer' effects).

[2] I will call both Switch and Loop effects 'order' effects. Some might want to reserve the expression 'order effect' for effects like Switch effect where the order of scenarios is manipulated rather than those like Loop effects where what is manipulated is which scenario is presented directly before a target scenario. Nothing turns on this terminological choice.

[3] There are various other framing effects which have also been reported (see, e.g., Bartels, 2008; Gonnerman et al., 2012; Nahmias et al., 2007; Nichols and Knobe, 2007; Tobia et al., 2013; Weinberg et al., 2012). I willl only in detail those which play a role in the experimental studies reported later in the paper.

2016); Demaree-Cotton (2015); Nado (2014). However, my focus today is to consider one particular type of underexplored response.

2 The Intuitionist Resistance

The epistemic foundation of moral philosophy is only threatened by the results that irrelevant framings influence moral intuitions given certain assumptions about moral epistemology. The threat depends on the idea that intuitions play an important role in moral epistemology, yes, but it also depends on the idea that a certain kind of intuitions play an important role. It is very easy to state the empirical challenge such that it assumes an overly simplistic view of intuition-based understandings of the methods and epistemology of moral philosophy.

Importantly, some of the most influential *intuitionists* such as Ross (2002) do not place any significant weight on intuitive judgments about the *permissibility* of certain actions in certain types of scenario. Instead, the emphasis is on the self-evidence of prima facie duties. Here's how Stratton-Lake (2014) explains that notion

> ... principles of prima facie duty are, roughly, principles stating that certain facts count in favour of an act and others count against. So these principles state, for instance, that the fact that one's act would produce some good, or the fact that it would be the keeping of a promise, or the expression of gratitude, etc., counts in favour of it, and the fact that, for example, it would involve the infliction of harm on someone counts against it. What we ought to do is determined by all of these facts, and how they weigh up against each other. Ross denied that we can ever know what we ought to do, and rejected the view that there could be strictly universal, self-evident principles specifying what we ought to do.

Indeed, Stratton-Lake (2014) goes so far as to suggest that intuitionists such as Ross have the resources to resist the extant challenges to the epistemology of moral philosophy discussed in the previous section. These challenges are based on evidence of the unreliability of intuitions about the moral *permissibility* of actions. But, for all such evidence shows, intuitions about *prima facie* or *pro tanto* duties, or as I shall call them *intuitions about moral relevance*, may be reliable sources of evidence about what counts in favour of and counts against actions.[4]

As Stratton-Lake (2014) puts it

[4]It is worth noting that the same lessons apply mutatis mutandis for particularists such as Dancy (1983). While particularists accept the idea that the very same feature which counts in favour of acting in one case may count against acting in another case, no sensible particularist should be open to the idea, for example, that the same feature might count against acting in case A when preceded by case B, yet in favour of acting in that very same case A when preceded by a third case C. But I will not discuss particularism any further here.

> . . . it is hard to imagine someone thinking that the fact that one would have to kill an innocent person in order to save five didn't count against it, or that the fact that their act would save five innocent people didn't count in favour of it, regardless of their overall verdict about whether they should kill the one or let the five die. That their act involved physically pushing someone in front of the trolley, or pulling a lever that would release a trap door dropping them onto the track would plausibly make no difference to such intuitions. Nor would framing effects introduced by the order of presentation of the cases. If such a priori expectations are correct—and they would need to be empirically tested—then empirical psychology would raise no problems for a Rossian intuitionism that claims only that principles of prima facie duty are self-evident.

Stratton-Lake's comments specifically concern Ross's picture of the epistemic foundations of moral philosophy. However, the point can be a more general one. Intuitions about moral permissibility are plausibly not the only moral intuitions which play an epistemic role in moral philosophy. As well as intuitions about what is permissible and what is not, we are plausibly guided by intuitions about what counts in favour of and against acting in certain ways, about what factors are of moral significance, about what the morally relevant considerations are. Any intuitionist who places such intuitions at the foundation of their moral epistemology rather than intuitions about permissibility seems to be able to make use of the line of resistance Stratton-Lake outlines.

So, can such intuitionists weather the empirical challenge to the epistemic foundation of moral philosophy? There is a debate to be had here about where the burden of proof lies. Is it on the moral intuitionist to demonstrate the reliability of intuitions about moral relevance? Or is it on the skeptic to demonstrate such intuitions' unreliability? There are considerations that speak each way. In this respect, their position is akin to that of those who attempt to mount an 'expertise defence' of intuitions in the face of evidence of unreliability. There's an important question about whether one is entitled to assume that the experts are more reliable than ordinary people in the absence of further relevant evidence.[5] Debates like these are tricky ones. One can, however, bypass such debates by taking up the burden of proof for oneself. This is what I have done in the following. In the remainder of this paper, I report the results of a series of experiments which aim to investigate whether intuitions about moral relevance are susceptible to each of the four framing effects mentioned in the previous section.

[5] I have argued elsewhere that, although an 'expertise defence' of philosophers' intuitions may fail for other reasons, such an assumption is in fact justified (Andow, 2015)

3 Actor-Observer Effects

Experiment 1a

Experiment 1a examines whether intuitions about moral relevance are subject to the 'actor-observer' effect whereby participants seem to hold themselves to different moral standards than they do other people.

Participants In this and the rest of the experiments reported in the paper, participants were recruited using Prolific Academic, a UK-based equivalent to Amazon MTurk, and completed an online survey built using Qualtrics. Participants were not permitted to take part in more than one of the studies presented in this paper (with one exception).[6] In this, and the other studies, all participants were resident in the UK, native English speakers, and at least 18 years old. 199 participants were recruited.

Mean age was 31.07. 110 were Male (55.3%), and the rest female. 30 had studied some philosophy at university level (15.1%). In all the studies in this paper, participants were rewarded for their participation (£0.17 @ £5.10/h based on a predicted completion time of 2 minutes).

Materials Participants were randomly allocated to receive questions about one of four cases. Two of these cases were the standard Push and Switch cases, which we'll call Observer versions. The remaining two were Actor versions of Push and Switch in which 'Abigail' is replaced with 'you'. Following each case, participants were asked two questions to gauge their intuitions about moral relevance (order randomized). The statements used were CA and CF. The wording of these statements is inspired by the prediction of Stratton-Lake (quoted in section 2).

> CA – The fact that pushing the [button/bystander] will lead to the death of one innocent bystander who would otherwise have survived counts against [Abigail/you] pushing the [button/bystander].
>
> CF – The fact that pushing the [button/bystander] will prevent the death of five innocent people who would otherwise have died counts in favour of [Abigail/you] pushing the [button/bystander].

Responses to all statements were given on a 6-point scale from 'strongly disagree' to 'strongly agree'. Finally basic demographic information was requested.

Results The mean ratings of relevance by condition are in Table 1 and displayed in Figure 1. A 2x2x2 ANOVA was conducted with condition (Actor, Observer) and

[6] The exception is Experiment 3c which was run more than 8 months after the next most recent study in this paper. Participants who took part in previous studies were allowed to take part in Experiment 3c.

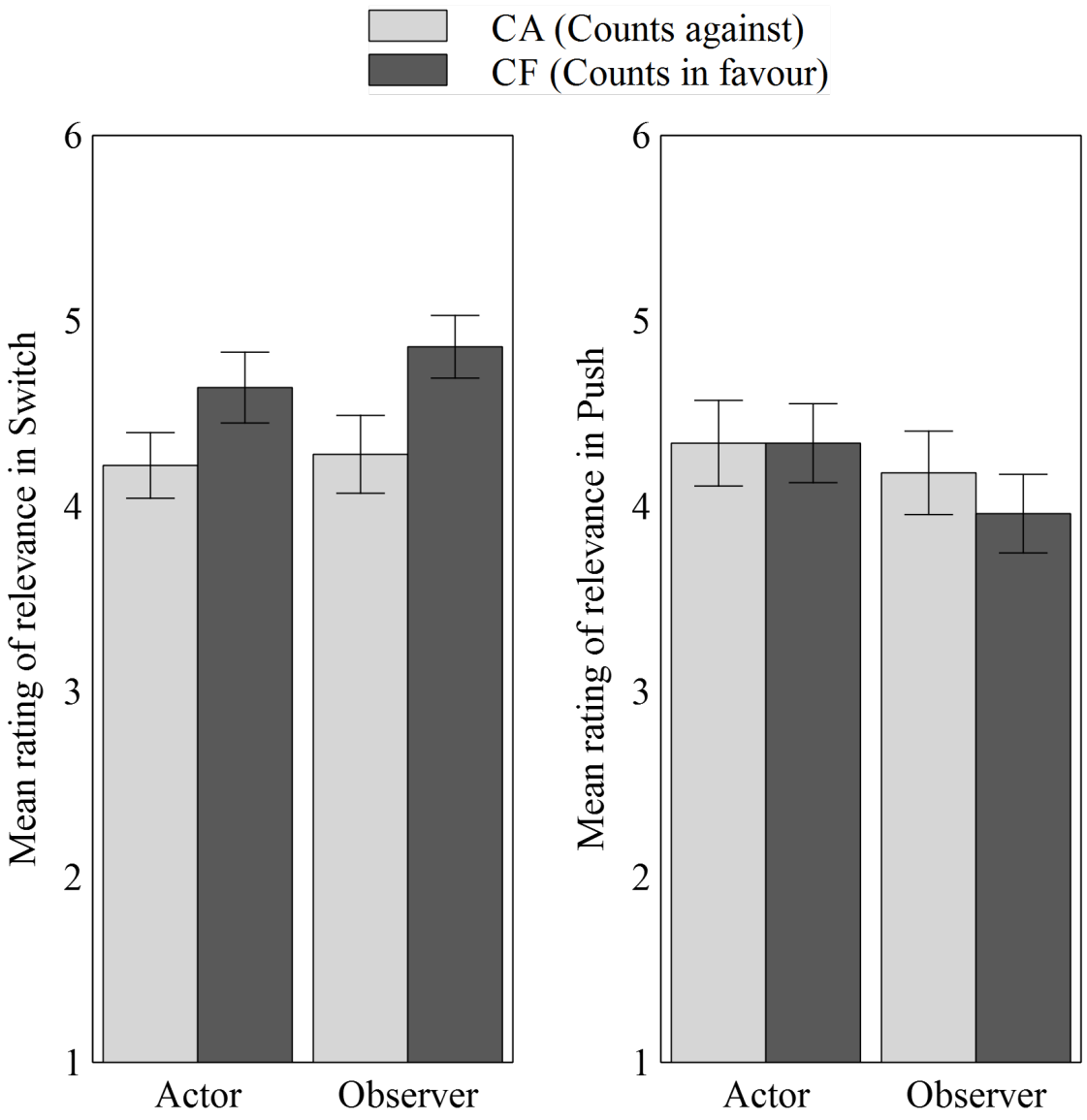

Figure 1: Mean ratings of agreement with the relevance items (CA and CF) for both scenarios (Push and Switch) by condition (Actor, Observer). Error bars indicate one standard error. (Experiment 1a)

Table 1: Mean responses and standard deviations to relevance items (Experiment 1a)

Condition	Scenario	Statement	N	Mean	SD	Percent Agree
Actor	Switch	CA	50	4.22	1.25	76%
		CF	50	4.64	1.35	84%
	Push	CA	50	4.34	1.64	74%
		CF	50	4.34	1.51	76%
Observer	Switch	CA	50	4.28	1.49	74%
		CF	50	4.86	1.20	90%
	Push	CA	49	4.18	1.58	74%
		CF	49	3.96	1.49	69%

scenario (Switch, Push) as between-subjects factors and question-type (CA, CF) as a within-subjects factor. The three-way interaction was not significant ($p = .480$) neither were the two-way interactions between question-type and condition ($p = .906$), or condition and scenario ($p = .183$). There was a significant interaction between question-type and scenario ($F(1, 195) = 5.07, p = .025, \eta_p^2 = .025$).[7] There was borderline significant main effect of scenario ($F(1, 195) = 3.71, p = .056, \eta_p^2 = .019$), but not of condition ($p = .675$).

To consider the nature of the interaction between question-type and scenario, two independent samples t-tests were conducted. For CF, participants gave lower ratings for Push than Switch ($t(191.331) = 3.03, p = .003, d = 0.34$).[8] There was no significant difference for ratings of CA ($p = .952$). Pearson's chi-squared tests reveal a significant difference between the proportion of participants in the different scenarios who agree with CF ($\chi^2(1, 199) = 6.31, p = .012, \phi = .178$) but not CA ($p = .838$).[9]

Discussion Experiment 1a finds no evidence that intuitions about moral relevance are subject to Actor-Observer effects. This is somewhat surprising as differences between actor and observer framings are well documented and widespread in psychology. Moral intuitionists might take some comfort from this result.

However, there are two observations we can make about these results which should perhaps trouble the intuitionist. The first potentially worrying observation is that Experiment 1a finds some perhaps surprisingly high levels of disagreement with the relevance items. Stratton-Lake claimed that it is hard to imagine someone disagreeing with claims such as "The fact that . . . will lead to the death of one innocent bystander

[7]The standardly used rules of thumb for interpreting partial eta-squared (η_p^2) are that a small effect = 0.01, medium = 0.06, and large = 0.14 .

[8]The standardly used rules of thumb for interpreting Cohen's d are that a small effect = 0.2, medium = 0.5, and large = 0.8.

[9]The standardly used rules of thumb for interpreting phi (ϕ) are that a small effect = 0.1, medium = 0.3, and large = 0.5.

who would otherwise have survived counts against . . . ". Indeed, the Rossian intuitionist might be inclined to count such a claim among their 'self-evident principles' of prima facie duty. However, in this experiment, around 25% of participants fail to agree that the fact an action will cause a death counts against it (across all four groups).

The second is that levels of agreement with CF in particular seem to vary significantly between Switch and Push. So, even if there is no evidence of an Actor-Observer framing effect, one might see some epistemically worrying aspects to these results. If participants say that a certain feature counts against Push but not against Switch then they are either not interpreting the question as one about the moral relevance of certain factors or else they are exhibiting an epistemically troubling kind of inconsistency (for the difference between Push and Switch is irrelevant to the facts about whether an action prevents deaths counts in favour of perfoming it). Naturally, given this between-subject design, no participant is recorded giving such an 'inconsistent' response pattern. However, the pattern of results is nonetheless notable.

Either of these observations might also be used to raise doubts about whether participants are responding to the relevance items in the intended fashion. This issue will be discussed in detail in the general discussion.

Experiment 1b

This experiment represents a slight variation on Experiment 1a. There are two main changes. First, the Physical scenario was not used in Experiment 1a, but might be thought to be more likely to exhibit an Actor-Observer effect. Consequently, it is used here. Second, the between-subjects design in Experiment 1a, didn't allow for a direct measurement of how many participants give seemingly inconsistent responses to CA and CF items across different scenarios. Consequently, in Experiment 1b, scenario (Physical, Switch) is a within-subjects factor rather than between-subjects factor.

Participants 204 participants were recruited. Mean age was 33.93. 82 were Male, 120 Female, and the rest Other. 12 had studied some philosophy at university level (5.90%).

Materials Each participant answered both CA and CF questions (order randomized) about both Switch and Physical (order randomized). Participants were randomly assigned to receive questions about Actor or Observer versions of scenarios. Basic demographic information was collected.

Results The mean ratings of relevance by condition are in Table 2 and displayed in Figure 2.

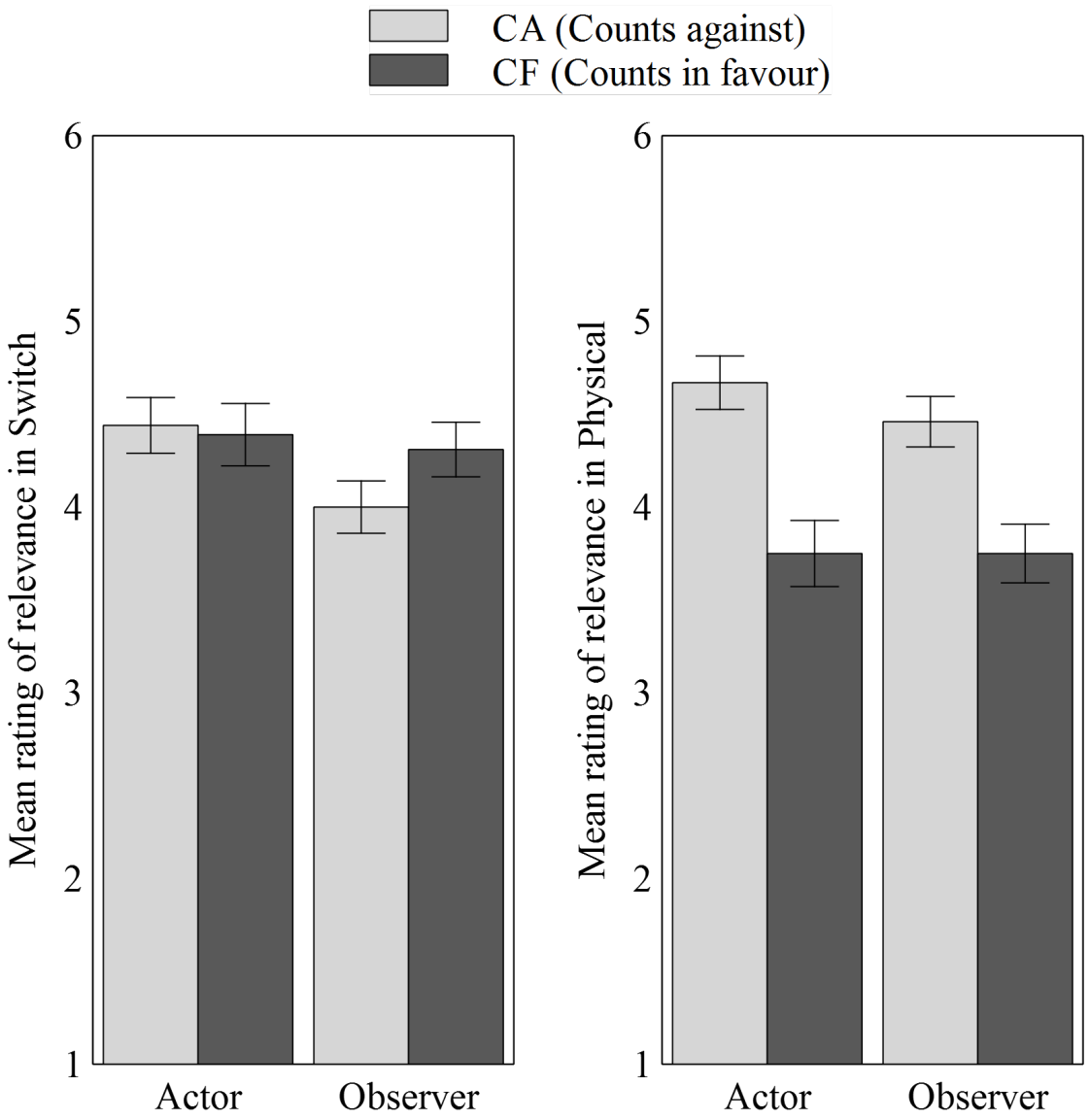

Figure 2: Mean ratings of agreement with the relevance items (CA and CF) for both scenarios (Physical and Switch) by condition (Actor, Observer). Error bars indicate one standard error. (Experiment 1b)

Table 2: Mean responses and standard deviations to relevance items (Experiment 1b)

Condition	Scenario	Statement	N	Mean	SD	Percent Agree
Actor	Switch	CA	102	4.44	1.51	74%
		CF	102	4.39	1.70	77%
	Push	CA	102	4.67	1.45	79%
		CF	102	3.75	1.79	63%
Observer	Switch	CA	102	4.00	1.43	68%
		CF	102	4.31	1.48	77%
	Push	CA	102	4.46	1.38	78%
		CF	102	3.75	1.64	60%

A 2 x 2 x 2 ANOVA was conducted with condition (Actor, Observer) as a between-subjects factor and scenario (Physical, Switch) and question-type (CA,CF) as within-subjects factors. The three-way interaction was not significant ($p = .549$). Neither were the two-way interactions between question-type and condition ($p = .357$) and scenario and condition ($p = .169$). There was a significant interaction between question-type and scenario ($F(1, 202) = 46.395, p < .0005, \eta_p^2 = .187$). There was no significant main effect of condition ($p = .151$). There were significant main effects of question-type ($F(1, 202) = 5.07, p = .025, \eta_p^2 = .024$) and scenario ($F(1, 202) = 5.56, p = .019, \eta_p^2 = .027$).

To consider the nature of the interaction between question-type (CA vs CF) and scenario (Physical vs Switch), two paired samples t-tests were conducted. For CF, participants gave lower ratings for Physical than Switch ($t(203) = 6.32, p < .0005, d = .44$). For CA, participants gave higher ratings for Physical than Switch ($t(203) = 4.22, p < .0005, d = .30$). Pearson's chi-squared tests reveal a significant difference in between the proportion of participants in the different Scenarios agreeing with CF ($\chi^2(1, 204) = 68.42, p < .0005, \phi = .58$) and CA ($\chi^2(1, 204) = 64.71, p < .0005, \phi = .56$). The proportion of participants giving inconsistent answers across the two scenarios was 20% for CF and 17% for CA.

Discussion Like Experiment 1a, Experiment 1b also fails to find any evidence that intuitions about moral relevance are subject to Actor-Observer effects. This is a notable result as differences between actor and observer framings are very well documented and widespread in psychology. Again, it may be that intuitions about moral relevance are subject to actor-observer effects which the current studies did not have power to detect. However, if that is so, then the distorting effects are small.

However, following up on the issues about apparently inconsistent patterns of response flagged up in the discussion for Experiment 1a, the results of Experiment 1b are not entirely comforting for intuitionist. Around one in five participants indicates attitudes about moral relevance which an intuitionist would deem inconsistent, e.g.,

they think that the fact that acting will lead to a death counts against acting in Physical but not in Switch (despite the fact that both cases have this feature).

Again, the low levels of agreement with the relevance items doesn't fit particularly well with Stratton-Lake's predictions (see section 2).

Again, one might think either of these findings might be taken to raise a question of whether participants are responding to the items in the intended fashion—concerns of this kind will be followed up later in the general discussion.

4 Asian Disease effects

Experiment 2

Experiment 2 examines whether intuitions about moral relevance are subject to the 'Asian-Disease' type of framing effect introduced in §1.

Participants 98 participants were recruited. A small number of participants were excluded prior to this count who either did not complete the survey or else gave an incorrect answer to a simple comprehension question. Mean age was 31.38. 40 were Male and the rest Female. 17 had previously studied some philosophy or psychology at university level. 3 participants indicated a prior familiarity with the 'Asian disease' type of cases.

Materials Participants read a case slightly adapted from Tversky and Kahneman

> Imagine that the UK is preparing for the outbreak of an unusual disease, which is expected to kill 600 people. Two alternative programs to combat the disease have been proposed. Assume that the exact scientific estimate of the consequences of the programs are as follows:

Participants were asked a simple comprehension question. Half of participants then were asked which of the following programes they favoured:

- If Program A is adopted, 200 people will be saved.
- If Program B is adopted, there is 1/3 probability that 600 people will be saved, and 2/3 probability that no people will be saved.

While the other half were asked about the following two programmes (which are equivalent but framed in terms of death rather than saving).

- If Program A is adopted 400 people will die.
- If Program B is adopted there is 1/3 probability that nobody will die, and 2/3 probability that 600 people will die.

Table 3: Means and standard deviations for all relevance items (Experiment 2)

Condition	Statement	N	Mean	SD	Percent Agree
Save	CFA	47	4.74	1.11	87%
	CAA	47	4.34	1.27	74%
	CFB	47	4.74	1.01	94%
	CAB	47	4.79	1.12	91%
Death	CFA	51	4.57	0.94	90%
	CAA	51	5.18	1.20	92%
	CFB	51	4.67	1.21	84%
	CAB	51	4.51	1.24	84%

Both groups were then asked about four items concerning moral relevance (order randomized). The relevance items were as follows:

> **Regardless of which program you ultimately favour**, please indicate the extent to which you agree with the following statements.
>
> - CFA – The fact that adopting Program A will lead to 200 people [being saved/not dying] from the disease counts in favour of Program A.
> - CAA – The fact that adopting Program A will lead to 400 people [not being saved/dying] from the disease counts against Program A.
> - CFB – The fact that adopting Program B might lead to [all 600 people being saved/none of the 600 people dying] from the disease counts in favour of Program B.
> - CAB – The fact that adopting Program B might lead to [none of the 600 people being saved/all 600 people dying] from the disease counts against Program B.

Answers were provided on a 6-point scale from 1 (strongly disagree) to 6 (strongly agree). Finally demographic information was collected.

Results The effect of framing on intuitions about which programme is favourable was significant. There was a difference between conditions with respect to which programme was favourable ($\chi^2(1, N = 98) = 21.59, p < .001, \phi = .47$): 66% of participants who received the 'save' framing chose program A, compared to only 19.6% of participants who received the 'death' framing.

Means and standard deviations for all relevance items by condition are in Table 3, see also Figure 3. To analyze the data concerning intuitions about moral relevance, I treat intuitions about what are morally relevant considerations in the case of Programme A and what are morally relevant considerations in the case of Programme B

separately. The rationale for this is that participants are asked what counts for and against two distinct options (unlike in the trolley cases where participants were just asked what counts for and against pushing the bystander/button).

First, consider intuitions about relevance concerning Programme A. A 2 x 2 ANOVA was conducted with condition (Save, Death) as a between-subjects factor and question-type (CAA,CFA) as a within-subjects factor. There was a significant interaction ($F(1, 96) = 10.67, p = .002, \eta_p^2 = .100$). The main effect of condition was trending but not significant ($F(1, 96) = 3.80, p = .054, \eta_p^2 = .038$). There was no significant main effect of question-type ($p = .513$). To consider the nature of the interaction, two independent-samples t-tests were conducted. CAA ratings were found to be significantly higher for the Death framing than the Switch ($t(96) = 3.35, p = .001, d = .68$), but there was no difference between framings for CFA ($p = .399$). Pearson's chi-squared tests reveal a significant difference between the proportion of participants in the different framings agreeing with CAA ($\chi^2(1, 98) = 5.60, p = .018, \phi = .24$) but not CFA ($p = .643$).

Now, consider intuitions about relevance concerning Programme B. A 2 x 2 ANOVA was conducted with condition (Save, Death) as a between-subjects factor and question-type (CAB,CFB) as a within-subjects factor. Neither the interaction ($p = .483$) nor the main effects of question-type ($p = .688$) or condition ($p = .339$) were significant.

Discussion It seems that framing reasons in terms of 'saving' and 'death' may have some influence on judgments about what is morally relevant. For example, Program A results in 400 people no longer being alive, which might be taken to count against Program A. One can frame this in terms of (a) the 400 people dying or (b) the same people not being saved from the disease. When framed in terms of death, participants seem to think this is a more relevant consideration. However, the effects of framing do not seem to extend to intuitions about relevance concerning Program B.

Tversky and Kahneman articulate the effect of framing on judgments about which program is favourable in terms of asymmetrical attitudes towards risk taking: people favour the 2/3 risk that 600 will die over the certainty of 400 deaths; but people favour the certainty of saving 200 lives over the risk that no one will be saved. However, the results of the current study suggests that the framing affects intuitions about moral relevance not with respect to the risky option – Program B – but with respect to the option involving certainties. Taken in isolation, then, the current results might suggest that favouring Program A over B given the 'save' framing is more the result of (a) a lower inclination to see 'leading to 400 people [not being saved/dying]' as a consideration against Program A, than (b) a differential attitude to risk. However, more research would need to be done to establish this, and more work be done to consider this result within the wider context of research on attitudes to risk.

Note that in this study, the levels of disagreement with the relevance items are notably lower than in Experiments 1a and 1b. This may perhaps be due to the precise

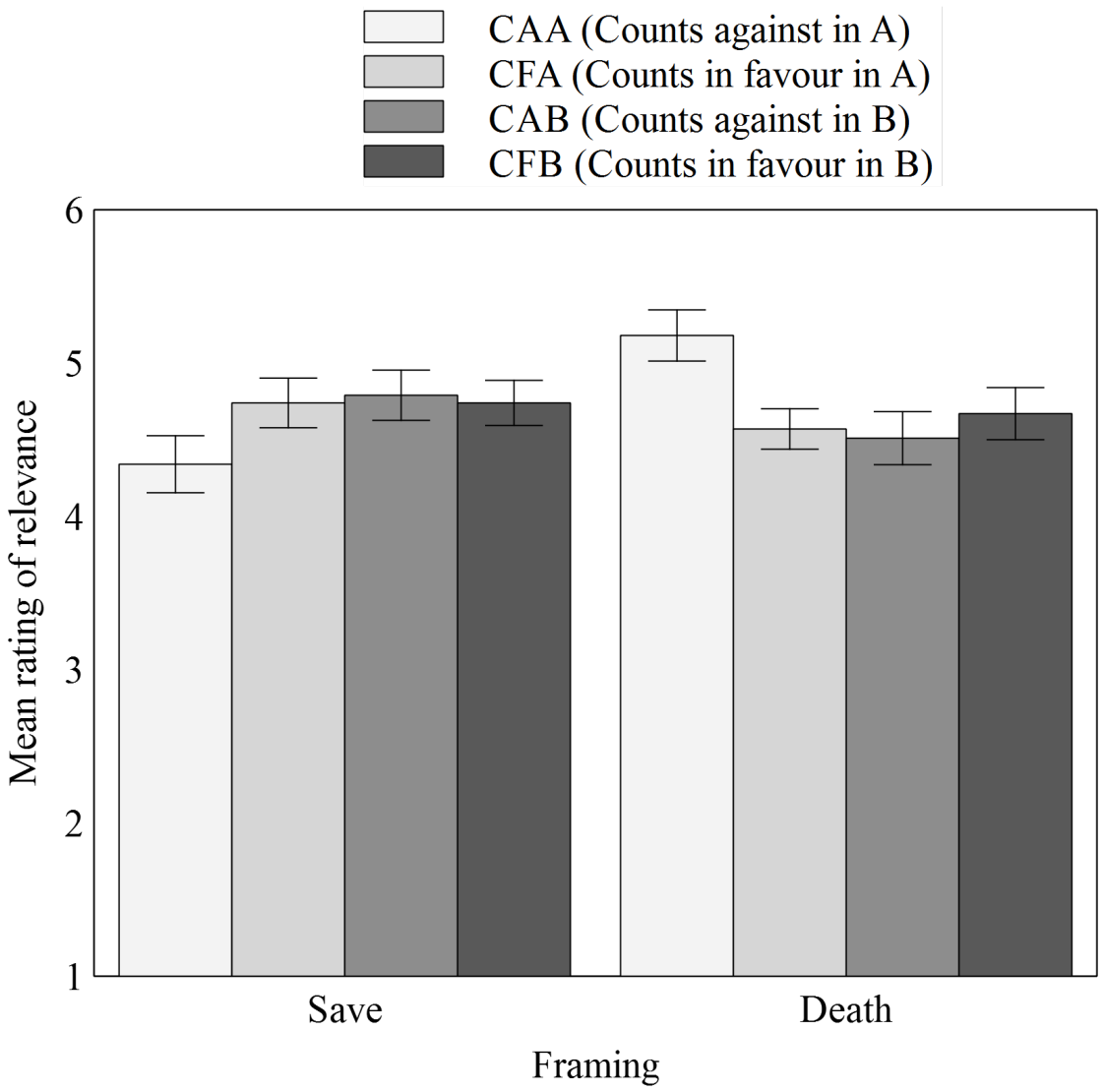

Figure 3: Mean ratings of agreement with the relevance items (CAA, CFA, CAB, CFB) by framing (Save, Death). Error bars indicate one standard error. (Experiment 2)

Table 4: Mean ratings and standard deviations in Switch by order (Experiment 3a)

Condition	Statement	N	Mean	SD	Percent Agree
Physical-Switch	CA	103	3.88	1.21	66%
	CF	103	4.11	1.43	69%
Switch-Loop	CA	107	3.96	1.41	71%
	CF	107	4.64	1.29	82%

way in which the statements were introduced, i.e., including "Regardless of which programme you ultimately favour", which may clarify the nature of the question for participants. Note that the previously discussed issue of participants giving apparently inconsistent response patterns doesn't so clearly arise as different questions about relevance are asked about Programme A and Programme B.

5 Order Effects

Experiments 3a, 3b, and 3c examine the extent to which intuitions about relevance are subject to Switch and Loop order effects (as described in §1).

Experiment 3a

Participants 320 participants were recruited. In this study, mean age was 29.41, 176 participants were Male (55%), 143 Female (44.7%), and 1 Other (0.3%). Fifty-seven had studied some philosophy at university level (17.8%).

Methods Each participant saw two cases. The cases used were Switch, Physical and Loop (see section 1 above). Participants were randomly allocated to one of three orders: (1) Physical, Loop; (2) Physical, Switch. (3) Switch, Loop. Participants were asked to rate their agreement with CA and CF (order randomized). Finally basic demographic information was requested.

Switch effect results The mean ratings of relevance for Switch by condition are in Table 4 and displayed in Figure 4. A 2 x 2 ANOVA was conducted with order (Physical-Switch, Switch-Loop) as a between-subjects factor and question-type (CA, CF) as a within-subjects factor. There was no significant interaction ($p = .101$). The main effects of order ($F(1, 208) = 5.93, p = .016, \eta_p^2 = .028$) and question-type ($F(1, 208) = 10.78, p = .001, \eta_p^2 = .049$) were significant. Coding responses as 'agree' or 'disagree' allows us to examine the proportion of participants who agree with CA/CF for each order of presentation (also in Table 4). Pearson's chi-squared tests indicated

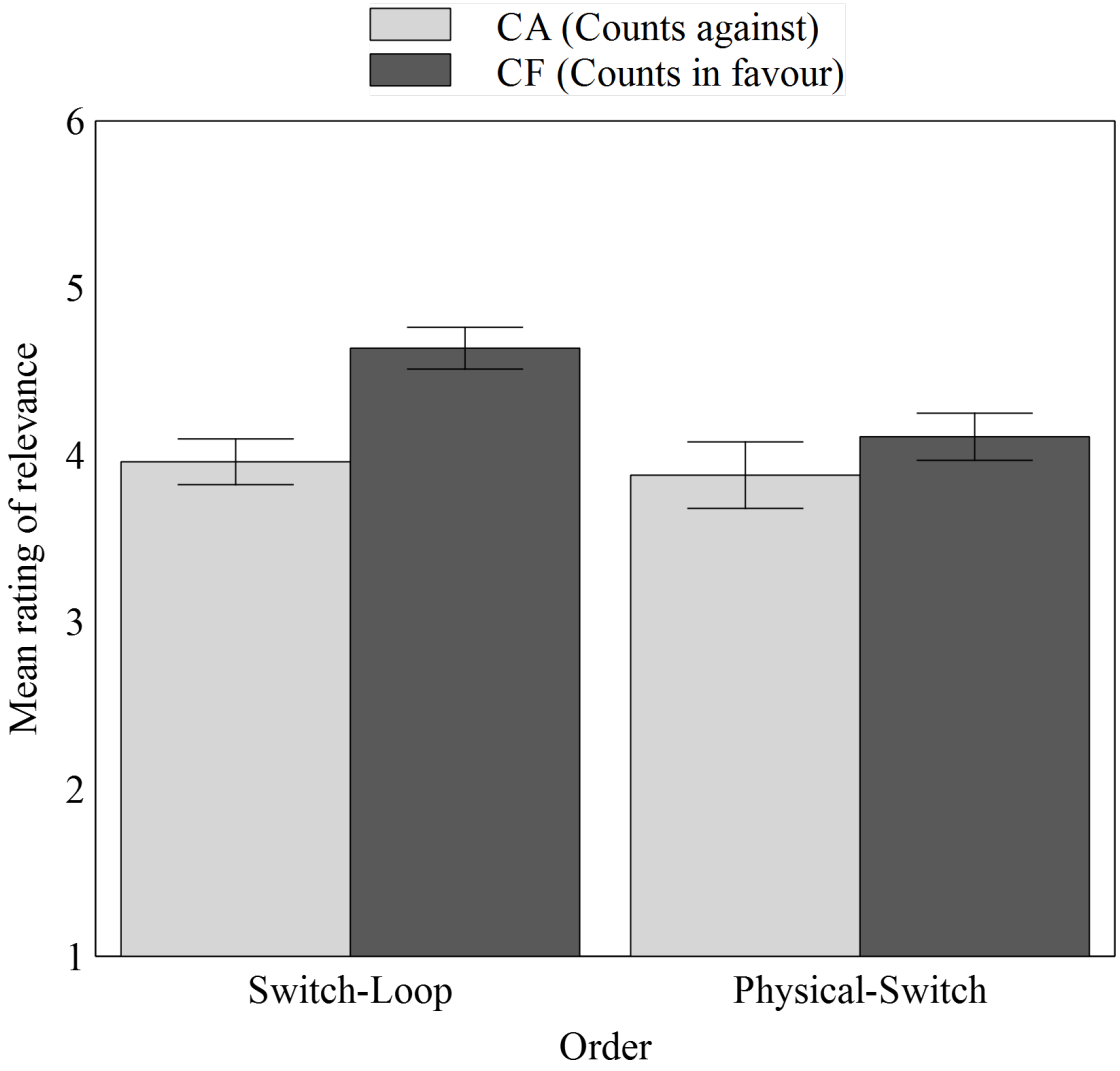

Figure 4: Mean ratings of agreement with the relevance items (CA, CF) for Switch by Order (Switch-Loop, Physical-Switch). Error bars indicate one standard error. (Experiment 3a)

Table 5: Mean ratings and standard deviations in Loop by order (Experiment 3a)

Condition	Statement	N	Mean	SD	Percent Agree
Physical-Loop	CA	110	4.27	1.31	72%
	CF	110	4.06	1.45	71%
Switch-Loop	CA	107	3.91	1.47	65%
	CF	107	4.31	1.43	76%

no effect of order on the proportion of participants who agree with CA ($p = .43$). However, a significantly higher proportion of participants agreed with CF when Switch was presented first ($p = .03, \phi = .16$).

Loop effect results The mean ratings of relevance for Loop by condition are in Table 5 and displayed in Figure 5. A 2 x 2 ANOVA was conducted with order (Physical-Loop, Switch-Loop) as a between-subjects factor and question-type (CA, CF) as a within-subjects factor. There was a significant interaction ($F(1, 215) = 4.48, p = .035, \eta_p^2 = .02$). There was no significant main effect of order ($p = .633$) or question-type ($p = .505$). To consider the nature of the interaction, two independent-samples t-tests were conducted. There was a borderline significant effect of order on CA ($t(215) = 1.94, p = .054, d = .29$) but not CF ($p = .212$). These responses were coded as 'agree' or 'disagree' as before. Pearson's chi-squared tests indicate no effect of order on the proportion of participants who agree with CA ($p = .25$) or CF ($p = .43$).

Discussion As in Experiments 1a and 1b, the apparent levels of the disagreement with the relevance items is surprisingly high. In the switch case, 45% of participants seem to indicate some level of disagreement with CA. As in the previous studies, this raises a potential worry that participants' responses to CA and CF do not reflect their intuitions about what counts in favour of and against acting. This feature of the results thus provides some motivation for using a different question design, e.g., similar to that used in Experiment 2.

However, let's consider what the findings of Experiment 3a mean under the assumption that the results do reflect participants' intuitions about moral relevance. Experiment 3a reveals some signs of order effects in participants' responses to CA and CF. Let's consider loop effects first. Here, there is a small interaction between order and question-type. This indicates that order affects what participants' think to be of most relevance. Seeing Physical first seems to lead to participants leaning towards CA and, more importantly, away from CF. The most important difference seems to be the shift in levels of agreement with CF from 82% when Switch is presented first to 69% when Physical is presented first. The same pattern is also reflected slightly in the results concerning switch effects. Although the interaction is not significant in this case, the pattern of results suggests that seeing Physical first leads to participants be-

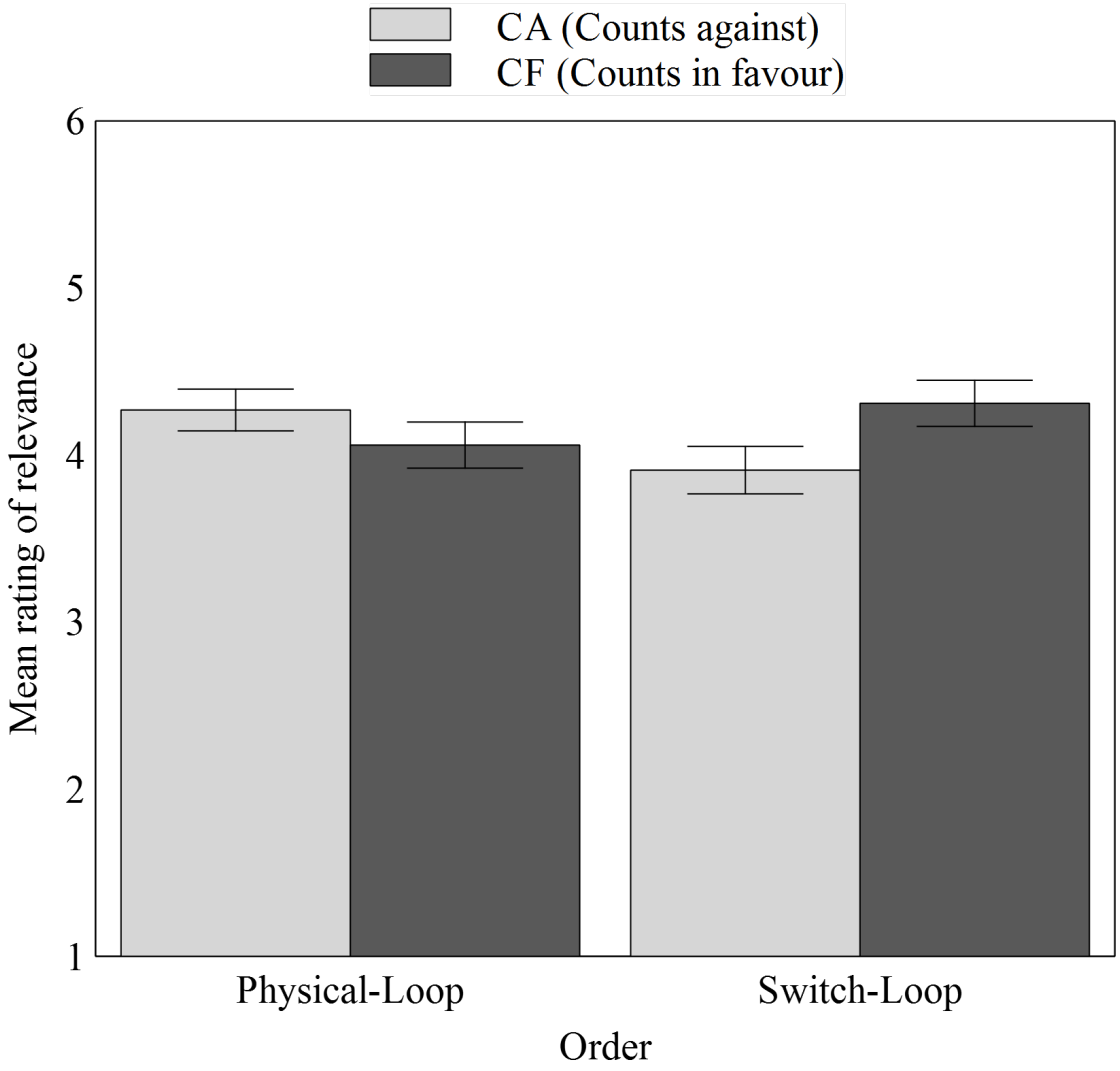

Figure 5: Mean ratings of agreement with the relevance items (CA, CF) for Loop by Order (Physical-Loop, Switch-Loop). Error bars indicate one standard error. (Experiment 3a)

ing more inclined to agree with CA and less inclined to agree with CF. Note that it is not implausible to interpret this shift in responses as a shift in intuitions about moral relevance, as seeing Physical first may incline participants to realise that preventing deaths is actually not so very important to them.

So, if these results can be taken to reflect participants' intuitions about moral relevance, then they suggest that such intuitions may not be as stable as the intuitionist might have hoped: there are some signs that they are subject to irrelevant effects.

Experiment 3b

Like Experiment 3a, Experiment 3b aims to examine the extent to which intuitions about moral relevance are subject to Switch and Loop effects. However, Experiment 3b makes use of a slightly different question design which aims to guard against the worry highlighted in the discussion of Experiment 1a that participants' responses to CA and CF may not reflect their intuitions about moral relevance. There are two main changes: (1) participants receive either only CA or only CF statements; (2) the relevance question in presented in a way which aims to more clearly distinguish it from the issue of whether it is all things considered okay to push the button.

Participants 246 participants were recruited. Mean age was 33.1 87 were Male (35.4%), 156 Female (63.4%), and 3 other (1.2%). 21 had studied some philosophy at university level (8.5%).

Materials The cases used were Physical, Switch and Loop. All participants saw two cases. Participants were randomly allocated to received one of three orders: (1) Physical-Switch; (2) Physical-Loop; (3) Switch-Loop. For each case seen, participants were first asked about moral permissibility using statement MP:

> It is morally permissible for Abigail to [push the button to redirect the train onto the second track / push the bystander in front of the trolley].

Participants were then presented with the same case again, and invited to express their intuitions about moral relevance of using the following wording similar to that used in Experiment 2:

> This time, **regardless of whether or not you think it is morally permissible for Abigail to push the [button/bystander]**, please indicate the extent to which you agree with the following statement.

Participants were randomly allocated to receive either CA or CF statements. Responses to all statements were given on a 6-point scale from 'strongly disagree' to 'strongly agree'.

Table 6: Mean ratings and standard deviations in Switch (Experiment 3b)

Condition	Statement	N	Mean	SD	Percent Agree
Physical-Switch	CA	46	4.20	1.34	70%
	CF	36	4.03	1.30	78%
Switch-Loop	CA	33	4.24	1.39	73%
	CF	49	4.76	1.15	90%

Table 7: Mean ratings and standard deviations in Loop (Experiment 3b)

Condition	Statement	N	Mean	SD	Percent Agree
Physical-Loop	CA	42	4.12	1.37	69%
	CF	40	4.25	1.34	78%
Switch-Loop	CA	33	4.00	1.46	64%
	CF	49	4.49	1.28	82%

Switch effect results The mean ratings of relevance for Switch by condition are in Table 6 and displayed in Figure 6. A 2 x 2 ANOVA was conducted with order (Physical-Switch, Switch-Loop) and question-type (CA, CF) as between-subjects factors. Neither the interaction effect ($p = .097$), nor main effects of order ($p = .059$) or question-type ($p = .399$) were significant. Pearson's chi-squared tests reveal no significant differences in the proportion of participants in the different framings agreeing with CA ($p = .76$) or CF ($p = .14$).

Loop effect results The mean ratings of relevance for Loop by condition are in Table 7 and displayed in Figure 7. A 2 x 2 ANOVA was conducted with order (physical-loop, switch-loop) and question-type (CA, CF) as between-subjects factors. Neither the interaction effect ($p = .401$), nor main effects of order ($p = .777$) or question-type ($p = .147$) were significant. Pearson's chi-squared tests reveal no significant differences in the proportion of participants in the different framings agreeing with CA ($p = .62$) or CF ($p = .63$).

Discussion These results still reveal a perhaps surprising degree of dissent from the statements CA and CF—claims one might well have supposed to be obviously true given the intended meaning—despite the efforts to ensure that the issue of relevance is distinguished from that of permissibility. However, this experiment finds little evidence of either Switch or Loop effects. It is possible that this experiment did not have sufficient power to detect any small effects which are there. In both cases, there are similar trends in the data to those in the previous experiment.

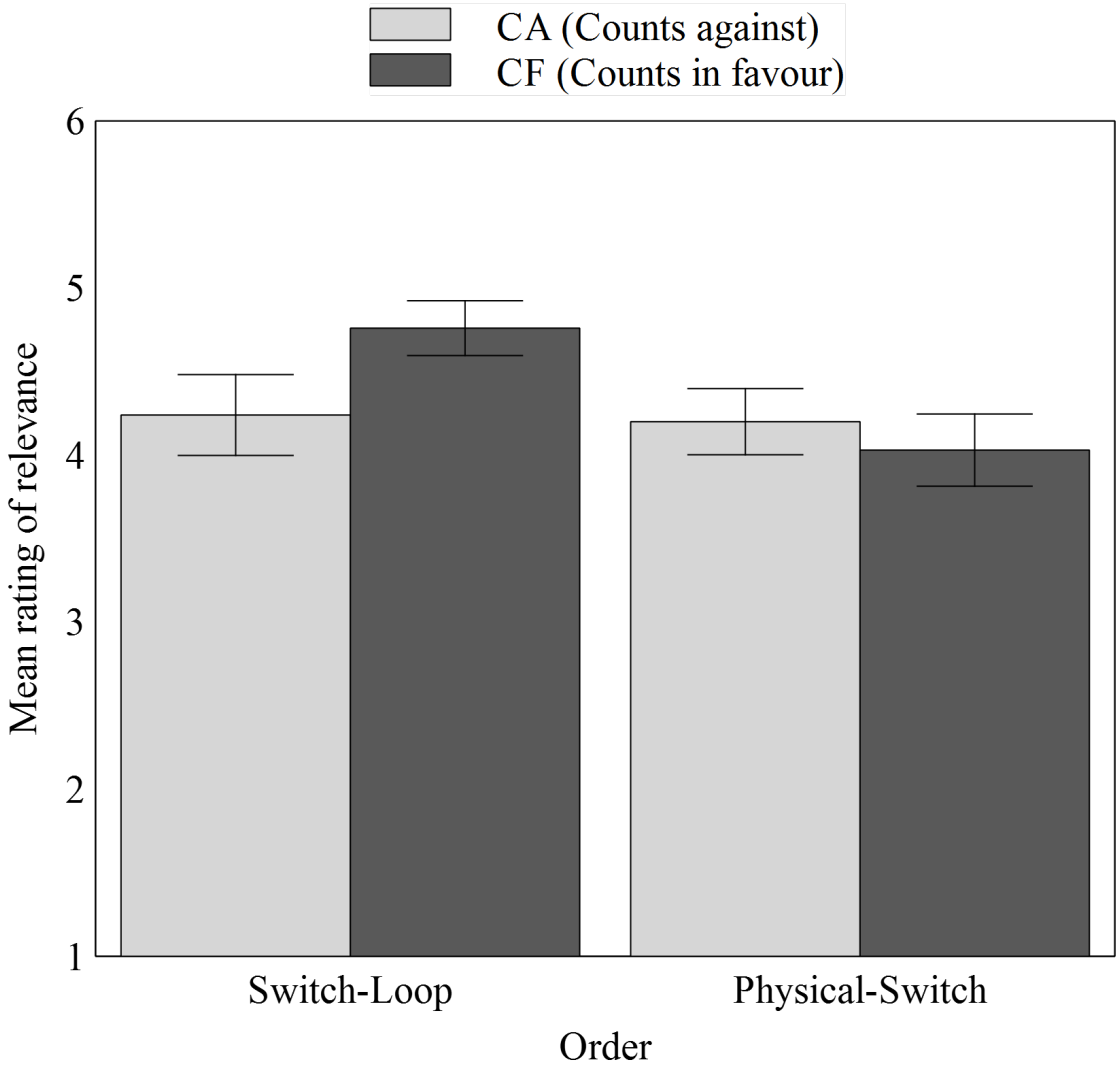

Figure 6: Mean ratings of agreement with the relevance items (CA, CF) for Switch by Order (Switch-Loop, Physical-Switch). Error bars indicate one standard error. (Experiment 3b)

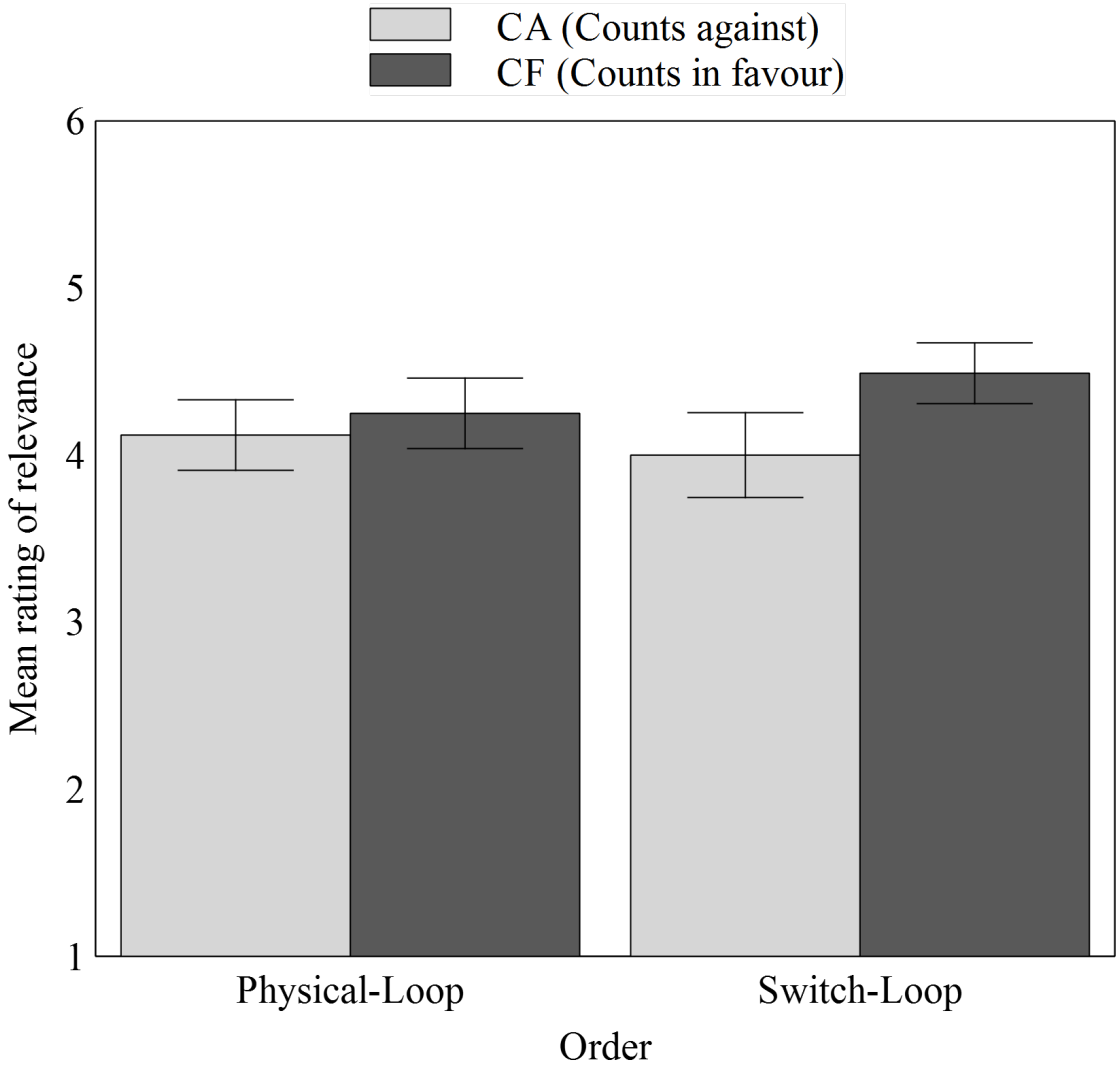

Figure 7: Mean ratings of agreement with the relevance items (CA, CF) for Loop by Order (Physical-Loop, Switch-Loop). Error bars indicate one standard error. (Experiment 3b)

Experiment 3c

Like Experiment 3a and 3b, 3c aims to examine the extent to which intuitions about moral relevance are subject to Switch and Loop effects. However, Experiment 3c again makes use of a slightly different question design. This aims to further guard against the worry that participants' responses to CA and CF may not reflect their intuitions about moral relevance. One reason to worry about the results from Experiments 3a and 3b is the surprising level of dissent from CA and CF. The relevance question is now presented in a way which aims to even more clearly distinguish it from the issue of whether it is all things considered okay to push the button.

Participants 300 participants were recruited. Mean age was 36.54. 116 were Male (28.7%), and 184 Female (61.3%). 16 had studied some philosophy at university level (5.3%).

Materials The cases used were Physical, Switch and Loop. All participants saw two cases. Participants were randomly allocated to received one of three orders: (1) Physical-Switch; (2) Physical-Loop; (3) Switch-Loop. For each case, participants were asked about new versions of CA and CF. The precise wording of the question and items was as follows.

> Sometimes in a difficult moral choice a person can have moral reasons to do something, but also moral reasons not to do it. **Regardless of what you think about the best final decision in these cases**, we would like to know whether you agree about each of the following claims:
>
> **CA** The fact that pushing the [button/bystander] will lead to the death of one innocent bystander who would otherwise have survived is at least one moral reason against Abigail pushing the [button/bystander].
>
> **CF** The fact that pushing the [button/bystander] will prevent the death of five innocent people who would otherwise have died is at least one moral reason in favour of Abigail pushing the [button/bystander].

This alternative wording, e.g., 'at least one moral reason', aims to make very clear to participants the nature of the intuitions of interest. Participants received both CA and CF in a randomised order. Responses to all statements were given on a 6-point scale from 'strongly disagree' to 'strongly agree'.

Switch effect results The mean ratings of relevance for Switch by condition are in Table 8 and displayed in Figure 8. A 2 x 2 ANOVA was conducted with order (Physical-Switch, Switch-Loop) as a between-subject factor and question-type (CA, CF) as within-subjects factors. The interaction was not significant ($p = .099$). Neither

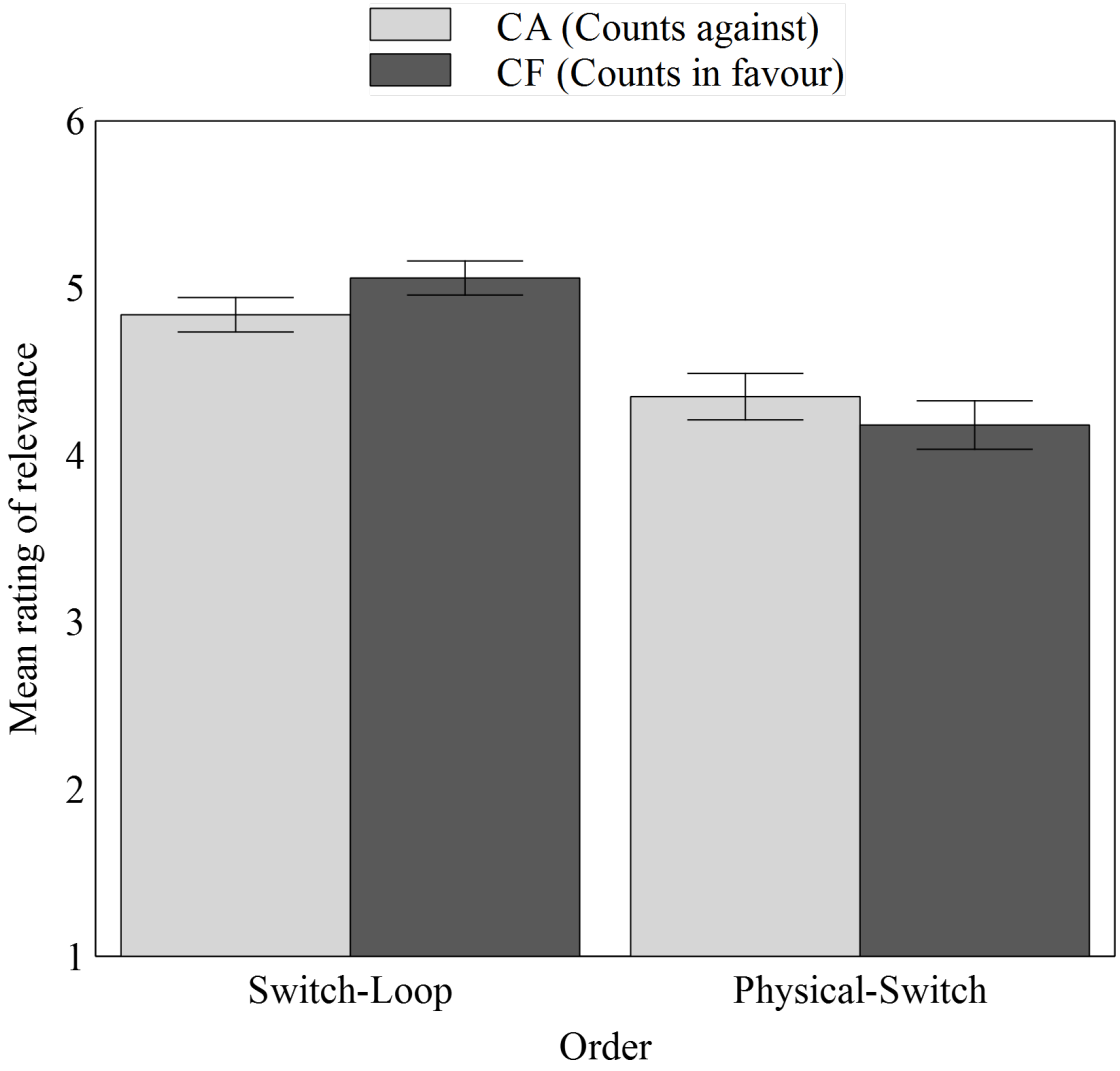

Figure 8: Mean ratings of agreement with the relevance items (CA, CF) for Switch by Order (Switch-Loop, Physical-Switch). Error bars indicate one standard error. (Experiment 3c)

Table 8: Mean ratings and standard deviations in Switch (Experiment 3c)

Condition	Statement	N	Mean	SD	Percent Agree
Physical-Switch	CA	100	4.35	1.39	78%
	CF	100	4.18	1.45	75%
Switch-Loop	CA	101	4.84	1.04	90%
	CF	101	5.06	1.03	93%

Table 9: Mean ratings and standard deviations in Loop (Experiment 3c)

Condition	Statement	N	Mean	SD	Percent Agree
Physical-Loop	CA	99	4.49	1.21	82%
	CF	99	4.53	1.30	84%
Switch-Loop	CA	101	4.69	1.09	88%
	CF	101	5.00	1.05	93%

was there a significant main effect of question-type ($p = .838$). However, there was a significant main effect of order ($F(1, 199) = 27.89, p < .0005, \eta_p^2 = .12$). Pearson's chi-squared tests reveal higher levels of agreement in Switch-Loop for CA ($p = .019, \phi = .165$) and CF ($p < .0005, \phi = .247$).

Loop effect results The mean ratings of relevance for Loop by condition are in Table 9 and displayed in Figure 9. A 2 x 2 ANOVA was conducted with order (Physical-Loop, Switch-Loop) as a between-subject factor and question-type (CA, CF) as within-subjects factors. The interaction was not significant ($p = .246$). Neither was there a significant main effect of question-type ($p = .157$). However, there was a significant main effect of order ($F(1, 198) = 8.70, p = .004, \eta_p^2 = .04$). Pearson's chi-squared tests reveal no significant differences in the proportion of participants in the different framings agreeing with CA ($p = .288$). However, the proportion in Switch-Loop who agreed with CF was higher ($p = .041, \phi = .145$).

Discussion Experiment 3c seems to show slighly lower levels of dissent from statements CA and CF than that observed in Experiments 3a and 3b. That said, the observed level might still trouble some intuitionists. For example, 25% of participants, in one condition, appear do not agree with idea that the fact that acting will save five people who would otherwise have lived is at least one moral reason in favour of acting. This issue will be taken up again in the general discussion.

Experiment 3c finds some evidence of both Switch and Loop order effects. Moreover, the pattern of results is highly consistent with that seen across Experiments 3a and 3b. The main pattern for loop cases seems to be that seeing Physical first makes participants less likely to agree with CF. The main pattern for switch effects seems to

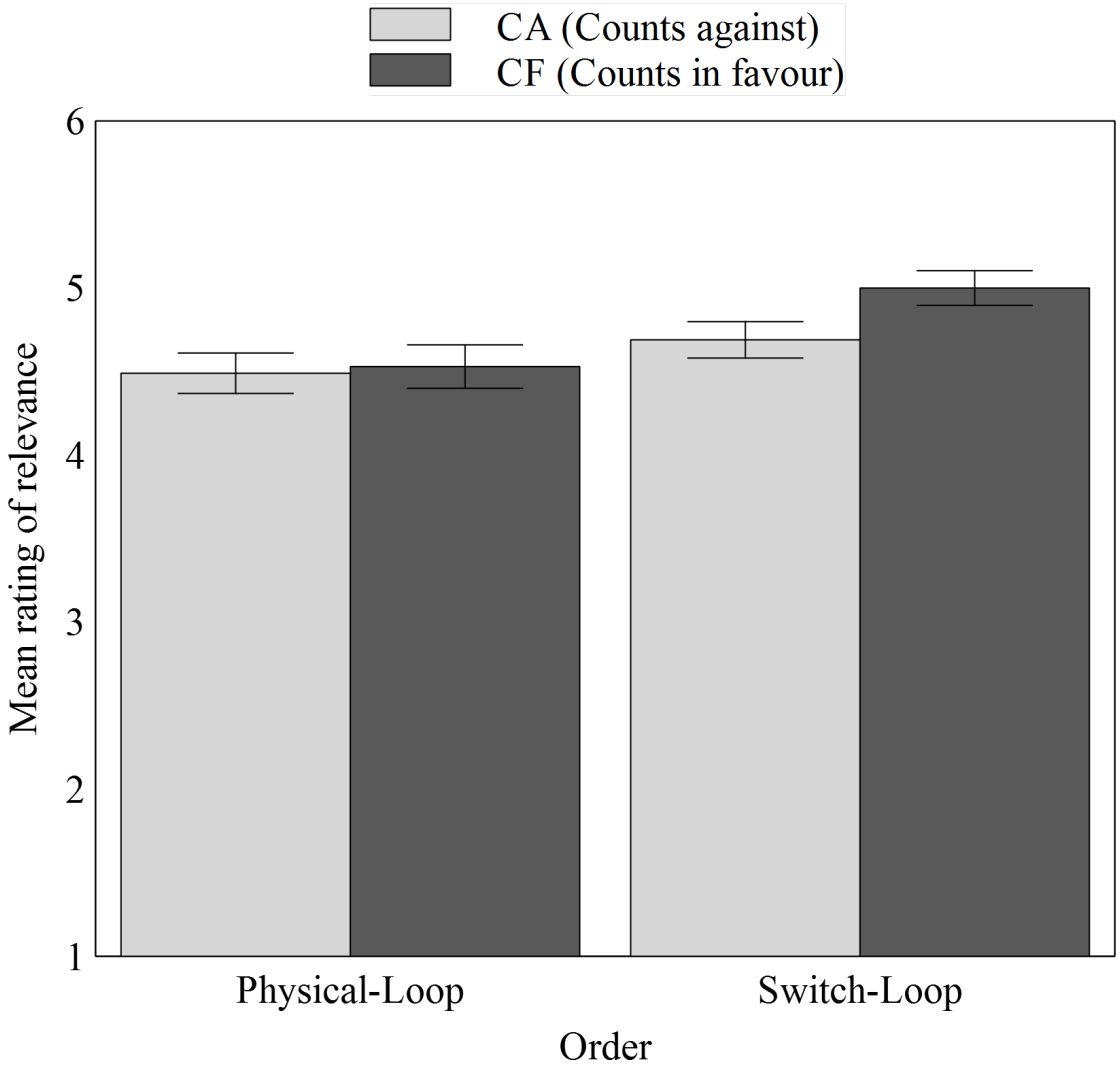

Figure 9: Mean ratings of agreement with the relevance items (CA, CF) for Loop by Order (Physical-Loop, Switch-Loop). Error bars indicate one standard error. (Experiment 3c)

be that seeing Physical first makes participants less likely to agree with both CA and CF, and, although the interaction is not significant in this case, this reduction seems to be more pronounced for CF.

Despite minor variations in which main effects and interactions were found to be significant, the results of Experiments 3a–c are relatively consistent. On the one hand, this internal consistency should bolster our trust in each of the three alternative ways of probing intuitions about moral relevance used in Experiments 3a–c. On the other hand, this evidence cumulatively suggests that such intuitions may be prone to some small effects of order whereby prior exposure to Physical depresses participants' intuitions concerning moral relevance.

6 General Discussion and Conclusions

Here is a brief summary of the main results of the studies reported above. Experiments 1a and 1b finds no evidence of any actor-observer effect. Experiment 2 does find some evidence that framing the scenario and responses in terms of 'saving' rather than 'dying' has some effect on participants' responses. However, the influence of framing was limited: it was observed only for questions about one of the programs, and in all cases and for all statements a large majority of participants indicate some level of agreement. Experiments 3a, 3b, and 3c find no robust evidence of large order effects of either the Switch or Loop variety. There does seem to be a consistent pattern whereby seeing Physical first leads to lower ratings of the idea that the fact that acting will save five people counts in favour of acting. However, such differences do not consistently reach significance, and when they do, tend to be associated with small effect sizes.[10]

What should we conclude on the basis of these results? My tentative answer is going to be that they provide some indirect support for the intuitionist resistance and that the burden is upon those who would undermine the epistemic standing of intuitions about moral relevance to provide some convincing evidence. However, the results in this paper can't be used to mount an all out defence of intuitionist's moral epistemology. Certainly, they can't be used to defend the conclusion that intuitions about moral relevance are not subject to troubling framing effects. So before I get to cautiously arguing that the results do provide some indirect support for the intuitionist's case, I'll first take some time to explain the various reasons why a more strident defence isn't possible on the basis of these results.

There are various aspects of the results reported that one might take to suggest that the results do not genuinely reflect participants' intuitions about moral relevance. I will consider this possiblity shortly. However, even assuming that the current findings do genuinely reflect participants' intuitions about moral relevance, there are a

[10]There are two exceptions which shouldn't be overlooked. These are the main effect of order in Experiment 3b and the effect of framing on responses to statement CAA in Experiment 2.

number of reasons why a strident defence is not possible. One straightforward reason is that there are other types of framing effect, not directly investigated in this paper, to which intuitions about relevance might be subject. A second is that one shouldn't overinterpret a null result. These studies find no convincing evidence of troublingly large Switch, Loop, Actor-Observer or Asian-Disease effects on intuitions of moral relevance. But this is, of course, not the same as convincing evidence that there are no such effects on intuition about moral relevance.

Another reason that a strident defence of intuitionism is not possible on the basis of the results reported here are is that there are actually various aspects of the results which should trouble the intuitionist.

(1) The results were not completely absent of evidence of the specific types of framing effect which the studies were trying to investigate; they did find some small framing effects.

(2) The current studies found some convincing evidence of a different type of sensitivity to irrelevant factors (albeit not the type of sensitivity they were trying to find). In Experiment 1, around one in five participants seem to agree that a factor counts in favour of (or against) acting in Push (or Physical) but not Switch (or vice versa) when it is present in both.[11] This suggests a form of framing effect which might be troubling for an intuitionist (although not, note, particularists). For, such a response pattern, if it truly reflects participants' intuitions about whether certain features count in favour of or against actions, indicates some degree of unreliability.

(3) The relevatively high levels of disagreement with the relevance items, e.g., CA and CF, found across all the experiments reported in this paper (ranging from around 6% up to 40%), don't sit easily within an intuitionist's worldview. Intuitionists are inclined to agree with Stratton-Lake that is hard to imagine someone disagreeing with claims such as "The fact that ... will lead to the death of one innocent bystander who would otherwise have survived counts against ..." and to think that claims such as these are among the 'self-evident principles' of prima facie duty. If the current results are to be believed, then the numbers of participants who do not find such principles self-evident should perhaps trouble the intuitionist. To the extent that intuitionists *are* surprised by the observed levels of dissent, the current results cannot be used to mount a stident defence of standard intuitionist approaches in moral philosophy. Why? Because surprising results concerning intuitions about moral relevance present a prima facie challenge to the use of *armchair* intuitionist methodology. The philosophical methods typically adopted by intuitionists don't typically involve large scale surveys of folk intuitions but rather depends on personal careful consideration of cases.

(4) The main results of these experiments may underestimate the extent to which intuitions about moral relevance of intuitions about moral relevance are subject to

[11] Revisiting the data for Experiment 3, a similar trend can be seen there too. To illustrate, in Experiment 3c, it is as high as 25% of participants giving inconsistent response patterns.

Switch, Loop, Actor-Observer, and Asian-Disease cases. The cases used in the current studies are, for the most part, cases in which one might think it is *obvious* what the morally relevant considerations are—deaths and saving lives—even taking into account the surprising degree of dissent shown by participants in the studies reported in this paper. But, of course, the intuitionist epistemology is supposed to be a general moral epistemology. The intuitionist epistemology and methods are also supposed to apply in more complex and subtle cases: cases in which it may not be so very obvious what the relevant considerations are. One might suspect that intuitions about moral relevance in such cases would be less secure and more prone to framing effects. Consequently, a natural next step in the empirical investigation into intuitions about moral relevance would be to examine whether intuitions about moral relevance were subject to framing effects in more complex and subtle cases—cases in which it is not immediately obvious what the morally relevant considerations might be nor how strong they are.

Let's turn to the possibility that the current results do not reflect participants' intuitions about moral relevance. There are two aspects of the results that, one might argue, would be so surprising under the assumption that the reflect participants' genuine intuitions about moral relevance that we should reject the assumption. The two aspects in question are the following: the apparently inconsistent patterns of response (agreeing with CA for Push but not Switch, for instance); and the relatively high levels of dissent from the relevance items (ranging from around 6% up to 40%).

I don't think we should be so skeptical that participants' responses should be interpreted as reporting their intuitions about moral relevance. One first reason for resisting such a response is that I don't think we should overplay how surprising these results would be. Take the finding of some apparent inconsistency between agreement with CA and CF between Push and Switch for example. In most of the above studies, the number of participants giving such responses is around one in five. This result is consistent with Demaree-Cotton (2015)'s analysis of the typical magnitude of effects of framing on the polarity of intuitions across numerous studies on philosophical intuitions. Importantly, Demaree-Cotton herself argues that this figure is in fact reassuringly low for proponents of intuition-based methods in philosophy.[12] Now, take the finding of relatively high levels of dissent from relevance items. There are a number of reasons why we shouldn't overestimate the scale of this finding. Note that one never expects to find 100% agreement to survey questions, as there will always be some noise in the sample due to a small number of participants who either aren't paying attention, or make a mistake, or misunderstand the question.[13] Note also, one should expect some small number of participants to genuinely disagree (e.g., because

[12] Although, I have myself argued that smaller effects might nonetheless be troubling Andow (2016).

[13] Regrettably, no attention checks were used in these experiments. However, participation was restricted to those with a high rate of acceptance for their submissions in Prolific ($> 90\%$).

they endorse a form of nihilism about moral reasons).[14] So, while the results might be surprising if they reflect intuitions about moral relevance, one shouldn't overestimate the scale of the surprise.

A second reason is the fact that across Experiments 3a, 3b, and 3c the general pattern of responses was fairly consistent. These three experiments used different question designs. The alternative wordings used in Experiments 3b and 3c are closely based on alternative forms of the question suggested by commenters who were initially concerned that the wording used in 3a was not effectively tapping intuitions about moral relevance. So, absent further evidence, I think participants should be interpreted as genuinely reporting intuitions about which factors are morally relevant. One might well remain sceptical that participants' *fully considered* judgments about what factors are morally relevant—given appropriately lengthy amounts of consideration or whatever conditions you deem necessary to reach a fully considered judgment—would retain these high levels of dissent. However, I think it would be unduly sceptical to doubt that participants are reporting their intial intuitions about whether the stated factors are relevant.

Finally consider that, in Experiment 3c—the experiment with the form of question which I take to be least ambiguous and most likely to elicit genuine intuitions about moral relevance—the apparent level of dissent in the target cases is roughly 15% across all conditions and statements. Once we factor in the impact of noise, e.g., participants who fail to pay attention, and understandable disagreement due to beliefs such as moral nihilism, it may be that the level of apparent dissent is not so surprisingly high in the first place.[15]

In light of these considerations, I am inclined to think that the current results should be interpreted as indicating participants' intuitions about moral relevance. Of course, I can't rule out that (some) participants are giving responses which indicate, for example, (a) whether they think the stated reason is *the most important reason* against action, (b) whether they think the bystander's *innocence* is beside the point, or even (c) simply whether they think the action is permissible. And future empirical work, using different ways of probing the relevant intuitions, or perhaps using protocol analysis to investigate how participants are responding to questions of the form used in the current studies, may be able to speak to these issues.

Indeed, this is far from the end of the road for empirical research into intuitions about moral relevance.[16] For example, in order to properly assess the standing of intu-

[14] Although, this is likely to be pretty small. To illustrate, in Experiment 3a, only 4% of participants disagreed with both relevance items for both cases they received, and nihilism isn't the only coherent motivation for giving such a response pattern.

[15] Although, note, the level of apparently inconsistent responses goes up to 25% in this experiment.

[16] One reviewer suggests that participants recruited through services such as Prolific and Mturk might be familiar with trolley problems and questions like 'Should one do x?'. In future work, it would be worth taking steps to recruit participants who are not familiar with the task of providing normative

itionist's epistemology, questions concerning *expert* intuitions about moral relevance and intuitions formed *after a greater opportunity to reflect* will need to be addressed empirically. One potential problem for the intuitionist resistance would be if there were reason to expect that more reflection and greater expertise would lead to intuitions about moral relevance being more susceptible to framing effects. I'll admit that I can't easily anticipate a compelling argument to that effect. However, the intuitionist resistance would stand to be bolstered by concrete evidence concerning such intuitions (or, of course, undermined, depending on the results). So, another natural next step for this research project would be to conduct the relevant studies.

In sum: the current results certainly can't be used to mount a strident defence of intuitionist moral epistemology; nonetheless, I am inclined to think that the current results might be cautiously used to provide some indirect support for the intuitionist resistance. This indirect support is to shift the burden of proof onto those who would empirically undermine intuitionism in moral philosophy on the basis that intuitions about moral relevance are problematically sensitive to framing effects (in similar ways to intuitions about permissibility).[17] The main take-home message of the current studies is that any such framing effects on intuitions about moral relevance are minor. Intuitionists have predicted that intuitions about moral relevance would rather stable (and, in particular, more stable than intuitions about moral permissibility) (Stratton-Lake, 2014). The burden is now upon their critics to produce data showing that intuitions about moral relevance are problematically sensitive to effects of framing, even if intuitionists are not entitled to assume that the epistemic standing of intuitions about moral relevance is impeccable or that such intuitions are completely immune to effects of framing.

Acknowledgements Thanks to the reviewers whose comments greatly improved this paper, as well as to Sarah Fisher, Aimie Hope, Joshua Knobe, Robin Scaife, Kelly Schmitdke, Philip Stratton-Lake, and attendees at the 1st Conference of Experimental Philosophy Group Germany for helpful comments and discussion.

References

Andow, J. (2015). Expecting moral philosophers to be reliable, *Dialectica* 69(2): 205–220.
URL: *http://dx.doi.org/10.1111/1746-8361.12092*

judgments about such cases.

[17] Bear in mind that the framing effects which Experiments 1–3 were looking for—Switch Effects, Loop Effects, Actor-Observer Effects and Asian Disease Effects—are generally considered to be robust and are widely replicated for intuitions about moral permissibility. Although, note, that there are interesting discussions to be had about the magnitude and methodological significance of the effects observed by this previous work (see Demaree-Cotton, 2015).

Andow, J. (2016). Reliable but not home free? What framing effects mean for moral intuitions, *Philosophical Psychology* 29(1):904–911.
URL: *http://dx.doi.org/10.1080/09515089.2016.1168794*

Bartels, D. M. (2008). Principled moral sentiment and the flexibility of moral judgment and decision making, *Cognition* 108(2): 381–417.

Cushman, F. and Greene, J. D. (2012). Finding faults: How moral dilemmas illuminate cognitive structure, *Social Neuroscience* 7(3): 269–279. PMID: 21942995.
URL: *http://dx.doi.org/10.1080/17470919.2011.614000*

Cushman, F., Young, L. and Hauser, M. (2006). The role of conscious reasoning and intuition in moral judgment testing three principles of harm, *Psychological science* 17(12): 1082–1089.

Dancy, J. (1983). Ethical particularism and morally relevant properties, *Mind* 92(368): 530–547.
URL: *http://www.jstor.org/stable/2254092*

Demaree-Cotton, J. (2015). Do framing effects make moral intuitions unreliable?, *Philosophical Psychology* online first: 1–22.
URL: *http://dx.doi.org/10.1080/09515089.2014.989967*

Gold, N., Pulford, B. D. and Colman, A. M. (2015). Do as I say, don't do as I do: Differences in moral judgments do not translate into differences in decisions in real-life trolley problems, *Journal of Economic Psychology* 47: 50 – 61.
URL: *https://doi.org/10.1016/j.joep.2015.01.001*

Gonnerman, C., Reuter, S. and Weinberg, J. (2011). More oversensitive intuitions: Print fonts and could choose otherwise. Unpublished manuscript, Indiana University, Bloomington, Indiana.

Greene, J. D. (2008). The secret joke of Kant's soul. In W. Sinnott-Armstrong (Ed.), *Moral psychology, Vol. 3: The neuroscience of morality: Emotion, brain disorders, and development*, Cambridge, MA: MIT Press, pp. 35–80.

Greene, J. D., Cushman, F. A., Stewart, L. E., Lowenberg, K., Nystrom, L. E. and Cohen, J. D. (2009). Pushing moral buttons: The interaction between personal force and intention in moral judgment, *Cognition* 111(3): 364–371.

Greene, J. D., Nystrom, L. E., Engell, A. D., Darley, J. M. and Cohen, J. D. (2004). The neural bases of cognitive conflict and control in moral judgment, *Neuron* 44(2): 389–400.

Greene, J. D., Sommerville, R. B., Nystrom, L. E., Darley, J. M. and Cohen, J. D. (2001). An fmri investigation of emotional engagement in moral judgment, *Science* 293(5537): 2105–2108.

Hauser, M., Cushman, F., Young, L., Kang-Xing Jin, R. and Mikhail, J. (2007). A dissociation between moral judgments and justifications, *Mind & Language* 22(1): 1–21.

Lanteri, A., Chelini, C. and Rizzello, S. (2008). An experimental investigation of emotions and reasoning in the trolley problem, *Journal of Business Ethics* 83(4): 789–804.

Liao, S. M., Wiegmann, A., Alexander, J. and Vong, G. (2011). Putting the trolley in order: Experimental philosophy and the loop case, *Philosophical Psychology* 25(5): 661–671.

Lombrozo, T. (2009). The role of moral commitments in moral judgment, *Cognitive Science* 33(2): 273–286.

Mikhail, J. (2000). Rawls' linguistic analogy: A study of the 'generative grammar' model of moral theory described by John Rawls in 'a theory of justice. (PhD dissertation, Cornell University, 2000).

Mikhail, J. (2007). Universal moral grammar: Theory, evidence and the future, *Trends in Cognitive Sciences* 11(4): 143–152.

Nadelhoffer, T. and Feltz, A. (2008). The actor-observer bias and moral intuitions: adding fuel to Sinnott-Armstrong's fire, *Neuroethics* 1(2): 133–144.

Nado, J. (2014). Philosophical expertise, *Philosophy Compass* 9(9): 631–641.

Nahmias, E., Coates, D. J. and Kvaran, T. (2007). Free will, moral responsibility, and mechanism: Experiments on folk intuitions, *Midwest Studies in Philosophy* 31.

Nichols, S. and Knobe, J. (2007). Moral responsibility and determinism: The cognitive science of folk intuitions, *Nous* 41: 663–685.

Nichols, S. and Mallon, R. (2006). Moral dilemmas and moral rules, *Cognition* 100(3): 530–542.

Petrinovich, L. and O'Neill, P. (1996). Influence of wording and framing effects on moral intuitions, *Ethology and Sociobiology* 17(3): 145–171.

Ross, W. D. (2002). *The Right and the Good*, Clarendon Press.

Schwitzgebel, E. and Cushman, F. (2015). Philosopher's biased judgments persist despite training, expertise and reflection, *Cognition* 141: 127–137.

Stratton-Lake, P. (2014). Intuitionism in ethics, in E. N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy*, Winter 2014 edition.

Tobia, K., Buckwalter, W. and Stich, S. (2013). Moral intuitions: Are philosophers experts?, *Philosophical Psychology* 26: 629–638.

Tversky, A. and Kahneman, D. (1981). The framing of decisions and the psychology of choice, *Science* 211(4481): 453–458.

Waldmann, M. R. and Dieterich, J. H. (2007). Throwing a bomb on a person versus throwing a person on a bomb intervention myopia in moral intuitions, *Psychological Science* 18(3): 247–253.

Weinberg, J., Alexander, J., Gonnerman, C. and Reuter, S. (2012). Restrictionism and reflection: Challenge deflected or simply redirected?, *The Monist* 95(2): 200–222.

Wiegmann, A., Okan, Y. and Nagel, J. (2012). Order effects in moral judgment, *Philosophical Psychology* 25(6): 813–836.
URL: *http://dx.doi.org/10.1080/09515089.2011.631995*

Wiegmann, A. and Waldmann, M. R. (2014) Transfer effects between moral dilemmas: A causal model theory, *Cognition* 131(1): 28–43.
URL: *https://doi.org/10.1016/j.cognition.2013.12.004*