

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/cphp20

Intuitions about moral relevance—Good news for moral intuitionism

Hossein Dabbagh

PHILOSOPHICAL **PSYCHOLOGY** ·

> To cite this article: Hossein Dabbagh (2021) Intuitions about moral relevance— Good news for moral intuitionism, Philosophical Psychology, 34:7, 1047-1072, DOI: 10.1080/09515089.2021.1960297

To link to this article: https://doi.org/10.1080/09515089.2021.1960297

© 2021 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.



0

Published online: 27 Jul 2021.

	>
Ø	

Submit your article to this journal 🖸

Article views: 1912



View related articles

View Crossmark data 🗹



Citing articles: 2 View citing articles 🕑

ARTICLE

OPEN ACCESS OPEN ACCESS

Routledae

Tavlor & Francis Group

Intuitions about moral relevance—Good news for moral intuitionism

Hossein Dabbagh

Department for Continuing Education, Philosophy Tutor at University of Oxford, Oxford, UK

ABSTRACT

Investigating the reliability of moral intuitions is not only one of the hot topics in moral psychology but also of high importance for philosophical questions. In a recent study, the influence of framing on people's intuitions about the moral relevance of certain properties of moral scenarios was investigated to assess the validity of some crucial assumptions of moral intuitionism. While it was found that people's intuitions about moral relevance were not affected by framing effects, the findings were not fully satisfying news for moral intuitionism since a worrisome proportion of participants disagreed with the purportedly self-evident moral relevance statements. In this paper, I propose a potential alternative explanation for the low agreement rate that would be less or rather not at all - worrisome for moral intuitionism, namely that the problematic results might be due to experimental pragmatics. To test this hypothesis, I rerun two experiments with especially low agreement rates and applied some simple measures to block unintended pragmatic considerations. This resulted in significantly higher agreement rates with the purportedly self-evident moral relevance statements than in the original. Indeed, the agreement rates were at such a high level that the findings of my study provide good news for moral intuitionism.

ARTICLE HISTORY

Received 21 July 2019 Accepted 21 July 2021

KEYWORDS

Moral intuitionism; experimental pragmatics; self-evidence; moral relevance; framing effects

1. Introduction

Moral intuitionism is the view that there are certain self-evident moral propositions that our moral intuitions can give us non-inferential justification for believing. Some critics such as, Sinnott-Armstrong, contend that "some recent research in psychology and brain science undermines moral intuitionism" (2006, 340). The research Sinnott-Armstrong is referring to found that our moral intuitions can be influenced by morally irrelevant factors, i.e. framing effects, such as the order in which moral dilemmas are presented to participants. This entails that we cannot be non-inferentially justified in holding our moral beliefs. The framing effects can cast doubt on any non-inferential justification, and our moral beliefs need to be

CONTACT Hossein Dabbagh Absein.dabbagh@gmail.com Philosophy Tutor at University of Oxford © 2021 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives License (http://creativecommons.org/licenses/by-nc-nd/4.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited, and is not altered, transformed, or built upon in any way. confirmed by an inference to be validated.¹ Most of these empirical psychology findings are concerned with moral intuitions about the permissibility of a particular action, e.g. whether it is right to kill one person in order to save five persons. However, some forms of moral intuitionism do not require that our moral intuitions about the permissibility of actions are immune to the influence of morally irrelevant factors but emphasize the self-evidence of *pro tanto duties*, in the sense that our intuitions about certain moral self-evident propositions are unsusceptible to framing factors.² In these forms of moral intuitionism, principles of pro tanto duties are understood in terms of moral reasons. That is to say, principles of pro tanto duties specify facts that provide moral reasons, i.e. moral reason-giving facts, which count in favor of or against performing actions.³ These principles state which facts provide reasons and explain why certain actions ought or ought not to be done.⁴ For example, I might have a duty to meet you tomorrow. The fact that I promised to meet you tomorrow provides me with a moral reason to do that. The pro tanto duty of fidelity specifies that the fact that I have promised to Φ provides me with a moral reason to Φ .

The difference between moral intuitions about the permissibility of certain *particular* actions (e.g. doing Φ in that situation is morally wrong) and moral intuitions about certain general self-evident moral propositions (e.g. it is generally wrong to Φ) is that since a moral intuition about the permissibility of an action is a moral intuition about the truth of some proposition – a proposition expressing something about a particular action, moral intuitions about certain general self-evident moral propositions are a priori but moral intuitions about the permissibility of actions are not necessarily so. The contrast is between moral intuitions that are selfevident and moral intuitions that are not. Moral intuitionists who believe in certain general self-evident propositions believe that some normative facts are self-evident and known a priori in the sense that one is noninferentially justified in believing some general moral propositions - such as that torturing someone for amusement is pro tanto wrong. These general self-evident moral propositions can be known on the basis of a noninferential moral intuition produced by an adequate understanding of the conceptual constituents.⁵ Since self-evident moral propositions are those propositions one can justifiably believe on the basis of adequately understanding them alone, self-evident propositions must be all *a priori* truths.⁶ Hence, in order to empirically challenge these forms of moral intuitionism, skeptics need to show that moral intuitions about what counts in favor of or against performing certain actions – intuitions about moral relevance – are not reliable. However, if empirical studies indicated that there are such reliable intuitions, then this would be good news for the epistemology of moral intuitionism.

Recently, James Andow (2018) conducted several experiments to investigate whether people's intuitions about the moral relevance of certain properties of cases might be relatively resistant to framing effects. For instance, he tested whether agreement ratings to statements such as "The fact that pushing the bystander will lead to the death of one innocent bystander who would otherwise have survived counts against you pushing the bystander" are prone to order effects or exhibit an actor versus observer bias. Overall, he found that people's intuitions about moral relevance were not affected by framing effects. However, the findings were not fully satisfying news for moral intuitionism since, in some experiments, about onethird of participants disagreed with the purportedly self-evident statements about moral relevance. This poses a problem for moral intuitionists because disagreement about self-evident statements shows that our intuition about self-evident moral propositions is not reliable. Our moral intuitions are reliable if they will remain credible when we rationally reflect on them. However, when there is relatively low agreement about our moral intuitions about self-evident moral propositions, it seems that they are not credible. More importantly, if we have disagreements about self-evident moral propositions, this entails that we are not non-inferentially justified in believing them because disagreements can create the need for confirmation for iustification.⁷

However, there is a potential alternative explanation for the relatively low agreement rate for the purportedly self-evident moral relevance statements that would be less - or rather not at all - worrisome for moral intuitionism, namely that the problematic results might be due to experimental pragmatics. For example, some participants might have interpreted statements like the aforementioned as being about the overall moral status of the corresponding action (e.g. whether the action is permissible or ought to be performed). To test this hypothesis, I reran two experiments by Andow (2018) in their original form and also added conditions in which I applied some measures to block such unintended re-interpretations of the moral relevance statements. While I replicated Andow's findings in the original conditions, the agreement rates in the new, extended conditions were significantly higher (~90%) and at the same level as the proportion of correct answers to a simple transitivity task ("If Peter is taller than Alex, and Alex is taller than Max, who is the smallest among them?"). Hence, the findings of my experiments provide good news for the aforementioned versions of moral intuitionism.

Below, I elaborate that some moral intuitionists are not really concerned with permissibility judgments. I then proceed with describing what kind of judgment is important for such forms of moral intuitionism. Next, I transit to the studies of Andow (2018) and shortly describe what he tested and what he found (especially the somewhat troublesome finding of the relatively low agreement rates for the purportedly self-evident statements about moral relevance). Finally, I describe what could have caused the surprisingly low agreement ratings without spelling trouble for moral intuitionism.

2. Skepticism about Moral Intuitionist Epistemology

The moral intuitionist epistemology claims that there are certain selfevident moral propositions, which are propositions that our moral intuitions can give us non-inferential justification for believing. Almost all moral intuitionists believe in self-evident propositions.

However, there are some empirical findings that cause trouble for the epistemology of moral intuitionism. For instance, in order to reject the epistemology of moral intuitionism through findings derived from empirical psychology, Sinnott-Armstrong attacks the strongest form of epistemological intuitionism (i.e. Audi's intuitionism). Sinnott-Armstrong starts by defining moral intuition "as a strong immediate moral belief" (2008, 47).⁸ He maintains that since moral intuitions are moral beliefs, they might arise after reflection, yet this does not entail that they have positive epistemic status. He writes,

it is better to define moral intuitions neutrally so that calling something a moral intuition does not entail by definition that it has any particular epistemic status, such as being true or probable or justified (2008, 75 in the notes).

Sinnott-Armstrong (2011) argues that moral believers have reason to ascribe a large probability of error to their moral belief because various empirical studies have shown that morally irrelevant factors influence our moral intuitions.⁹ Among those factors are so-called "framing effects", referring to the phenomenon that human judgment and decision making can be affected by how one and the same information is presented. Consider the famous Asian disease problem¹⁰: 600 people are affected by a deadly disease, and participants' task was to choose between two treatments, A and B. If treatment A is applied, 400 people will die, and 200 will live; if treatment B is applied, there is a 33.3% chance that no person will die but a 66.6% chance that everyone will die – hence, both treatments have the same expected value in terms of people dying/surviving. It was then manipulated whether the framing of the two treatments was positive or negative:

Framing	Treatment A	Treatment B
Positive	Saves 200 lives	A 33% chance of saving all 600 people, 66% possibility of saving no one.
Negative	400 people will die	A 33% chance that no people will die, 66% probability that all 600 will die.

When the choice was presented with positive framing, 72% of participants chose Treatment A. In contrast, only 22% of participants made this choice in the negative framing decision. Another framing effect found in moral psychology is the order of presentation in which moral dilemmas are presented to participants. For instance, participants' judgment about whether a certain action should be performed in a moral dilemma was influenced by whether A was presented before or after another moral dilemma B.¹¹ Based on such and similar findings, Sinnott-Armstrong claims that it is not justified to have some moral beliefs without needing them to be inferred from other beliefs.

Most of these empirical psychology findings are concerned with moral intuitions about the permissibility of a particular action, e.g. whether it is right to kill one person in order to save five persons. But what if there are some moral intuitions about the moral relevance of certain properties of cases that might be resistant to framing effects? Let us focus on moral intuitions about certain *pro tanto* duties (as moral reason-giving facts for or against certain actions), "promise-keeping, *ceteris paribus*, is morally right". If there are such intuitions, this would be good news for the epistemology of moral intuitionism according to which there are some self-evident moral propositions. The history of moral intuitionism shows that although different moral intuitionists disagree about what these moral propositions are, it is widely accepted that there are at least a number of self-evident moral propositions, e.g. "it is *pro tanto* wrong to pay an unequal amount to those who do the same job" or "killing innocent people for no reason is absolutely wrong".

Some influential forms of moral intuitionism, e.g. Rossian intuitionism, focus on intuitions about self-evidence of *pro tanto* duties as moral reasongiving facts for or against certain actions, rather than our intuitions about the permissibility of particular actions.¹² Hence, in order to empirically challenge these forms of moral intuitionism, it needs to be shown that intuitions about moral relevance, i.e. what counts in favor of or against performing certain actions, are unreliable to a worrisome degree. For example, consider the intuition that the death of an innocent person counts against performing any action with such a consequence. Could this intuition be subject to framing effects?

3. Rossian Intuitionism as an Alternative

In this section, I present a kind of moral intuitionism that does not require that moral intuitions about which actions are permissible/ought to be done are reliable. The kind of intuitionism I am interested in requires merely that our moral intuitions about some facts that count in favor of or against performing an action are reliable. W. D. Ross's intuitionism is a good example of this kind of moral intuitionism. Like many other intuitionists, there are two kinds of beliefs in Rossian epistemology. One kind consists of beliefs about *pro tanto* duties (as moral reason-giving facts for or against certain actions), which are basic, intuitive and self-evident. The other kind consists of beliefs about all-things-considered duties, which are not intuitive (inferred), not basic, and not self-evident.¹³

Ross believes that self-evident propositions are recognized as true with no *need* of supporting evidence. If one considers and understands self-evident propositions, supposedly one can see their truth without any need of proof.¹⁴ However, this does not entail that self-evident propositions are unprovable. It is possible that one can identify some arguments in favor of them.¹⁵ Although self-evident propositions can be known on the mere basis of an understanding of them, they can also be known in other ways, e.g. proposing an argument or illustration in favor of them.¹⁶ However, since such arguments are *not* necessary to be justified in believing a self-evident proposition, they are "epistemically supererogatory".¹⁷

Again, it is a misunderstanding to think that self-evident propositions can be justified in *only* one way, i.e. by sufficiently understanding them. This misunderstanding leads some philosophers to think that the term "selfevidence" needs to be changed. For example, Hooker prefers to say that moral intuitions are propositions that come with "independent credibility".¹⁸ An independently credible proposition is attractive without reference to evidence beyond itself and yet might turn out to be mistaken. For example, in *The Right and The Good*, Ross himself wrote that the proposition "pleasure is intrinsically good" is self-evident. However, he then, through further reflection and thought experiment, changed his mind and held that the proposition "pleasure is not intrinsically good" is self-evident.¹⁹

Ross thought that the principles of *pro tanto* duties (identifying moral reason-giving facts for or against actions) are known via their selfevidence in the sense that we gain such moral knowledge through a reflective non-inferential procedure, e.g. "intuitive induction".²⁰ Intuitive induction is the exercise of our ability to leap from our knowledge of some particular facts to knowledge of some universal principle. For example, we move from the perception of a single particular truth – such as this particular set of 6 apples added to that particular set of 6 apples adds up to 12 apples – to grasping the universal truth that 6 + 6 = 12. Thus, the knowledge we gain by intuitive induction is a way in which we see a self-evident truth or axiom by reflection. When we encounter a proposition like "The angles of all triangles always add up to 180 degrees", by intuitive induction from the knowledge of a single particular example (this triangle's sum of angles is 180 degrees), we assert that the proposition is true. Our knowledge of particular instances of the universal is simply like a ladder that we use to ascend to reach the knowledge of universals. We can put the ladder away once we have climbed it. Once we have a clear understanding of the axioms, we can know them simply on the basis of understanding them – that is, directly, non-inferentially, intuitively, or self-evidently. Since intuitive induction is not like Humean induction, it uses our ability to leap straight from the knowledge of particular moral cases to the knowledge of strictly universal principles as a necessary truth. So, by the method of intuitive induction, Ross thought, we reflect on particular moral cases to know universal moral principles.²¹ For example, when someone experiences that keeping promises seems right in situations A, B, etc., one is able to form a general principle that keeping promises, *ceteris paribus*, is right.

4. Previous Empirical Study on Moral Relevance

In his paper, "Are Intuitions About Moral Relevance Susceptible to Framing Effects?", Andow (2018) discusses the results of a series of experiments about the susceptibility of moral intuitions to framing effects. In the experiments, he uses variants of the popular trolley dilemma in which an out-ofcontrol train is about to run over and kill five persons. In one variant (henceforth: Switch) the only possibility to save the lives of the five persons is to push a button, which will redirect the threatening train onto another track where it will kill "only" one person; in another variant (henceforth: Push) the only possible intervention is to push a heavy person from a bridge onto the tracks and in front of the train, thereby stopping the train and saving the five persons (but killing the one). To give an impression of the relevance statements that were used, here is one referring to the Push dilemmas that includes a fact that presumably counts clearly against doing the corresponding action (henceforth: Count Against [CA]): "The fact that pushing the bystander will lead to the death of one innocent bystander who would otherwise have survived counts against Abigail pushing the bystander", and one referring to the Switch dilemmas that includes a fact that presumably counts clearly in favor (henceforth: Count in Favor [CF]) of doing the corresponding action: "The fact that pushing the button will prevent the death of five innocent people who would otherwise have died counts in favour of Abigail pushing the button". For these moral relevance statements, Andow investigated whether they exhibit an actor versus observer bias, i.e. whether participants' judgments about moral relevance statements are affected by whether the dilemmas and statements are worded in the third person ("Abigail") or from the perspective of the participant ("you"). Furthermore, he tested whether participants' judgments about moral relevance statements are affected by the order of presentation in which the corresponding moral dilemmas were presented to participants, e.g. first Switch and then Push or the other way around (henceforth: order effects).

The results of Andow's experiments "might be cautiously used to provide some indirect support for the intuitionist resistance" (2018, 140). Andow found that people's intuitions about moral relevance are not or only slightly affected by the framing effect. However, there was a relatively high disagreement rate about self-evident moral propositions. For instance, in one experiment (1b), the agreement rate with "The fact that pushing the bystander will lead to the death of one innocent bystander who would otherwise have survived counts against Abigail pushing the bystander" was overall only 72%. This is evidence that it is not true that all people have shared intuitions about certain moral self-evident propositions. In another experiment (3a) also, the intuitions about moral relevance were subject to order effect in participants' response to CA and CF, and the agreement rates with the statements about moral relevance were low. Andow seems to share this characterization of his results when stating that, while they "can't be used to mount a strident defence of intuitionist moral epistemology", he is inclined to "think that the current results might be cautiously used to provide some indirect support for the intuitionist resistance" (2018, 140). On the one hand, Andow thinks that the results of his experiments can provide some support for moral intuitionism because people's intuitions are only slightly affected by the framing effect. On the other hand, he believes that the results of his experiments are not positive enough to be a confirmation of moral intuitionism because people do not share high levels of agreement about self-evident moral propositions.

My motivation for the following experiments was to investigate whether an even stronger case for the empirical defense of moral intuitionism can be made. More concretely, I aimed to a) replicate the finding that judgments about moral relevance were insusceptible to framing effects and b) investigate whether the high rate of disagreement might be due to experimental pragmatics. However, experimental pragmatics can occur in many forms.²² For my purpose, I concentrate on the point that sometimes participants interpret their task differently from what the experimenter intended. In the case of Andow's experiments and the surprisingly high disagreement rates, I speculated that at least some of the disagreement was based on participants interpreting the test question not as being about whether they agree with the respective moral relevance statements but as being about the overall permissibility of performing the or not performing the described action. If this hypothesis is right, I should be able to significantly decrease the disagreement rate with the moral statements by introducing measures that block unintended interpretations of the test question, e.g. adding a question about what should be done in the moral dilemma to make clear that the question about moral relevance is not about the morally right course of action.

5. Experiments

I chose to replicate and extend Experiment 1b and Experiment 3a by Andow because in these experiments, the agreement rates with the statements about moral relevance were especially low and, thereby, worrisome for moral intuitionists. Andow did not find strong evidence for the influence of framing effects in his studies. But he found surprisingly low agreement rates for purportedly obvious statements about moral relevance.

Experiment 1

In this experiment, I replicate and extend Experiment 1b by Andow in which he assessed whether judgments about moral relevance concerning two moral dilemmas, Push and Switch (see below), are affected by framing a moral dilemma in the second ("You") or third person ("Abigail") perspective. He did not find an effect of this manipulation, but there were two troubling findings for moral intuitionists: First, and most importantly, the overall agreement rate with statements about moral relevance – e.g. "The fact that pushing the bystander will lead to the death of one innocent bystander who would otherwise have survived counts against Abigail pushing the bystander" was overall only 72%. Second, almost a fifth of all participants (20% for CF and 17% for CA) gave inconsistent moral relevance judgments across the two moral dilemmas. To be more specific, they judged that killing an innocent bystander counts against performing the respective action in Push but not in Switch (or vice versa).

In the following experiment, I extended Andow's experiment by including conditions that were modified in several ways to raise agreement rates for judgments of moral relevance. First, I added a question about what should be done in the moral dilemmas. This additional question was introduced to discourage participants not to interpret statements about moral relevance as being about the morally appropriate action. Second, I rephrased and explained the statements about moral relevance to facilitate their understanding. Third, I used binary response options (disagree vs agree) instead of gradual response options because I believed that gradual response options might confuse participants since the answers to the moral relevance questions seem obvious (*cf.* Cullen, 2010).

I. Participants

In this experiment, participants were recruited using Prolific Academic (Palan & Schitter, 2018), a UK-based equivalent to Amazon MTurk, and completed an online survey built using Unipark. All participants were required to be at least 18 years old and native English speakers (also holds for the second experiment). 444 participants started the survey, and the data of 427 were included in the analysis (17 were excluded for not finishing the experiment or finishing it in less than 40 seconds). The mean age was 25 years; 42% were male, and 58% female. Participants received £0.25 for estimated 3 minutes of their time (\pounds 5/h).

li. Design, Procedure and Materials

Participants were randomly assigned to one of four conditions in a 2 (version: Original vs Extended; between) * 2 (perspective: Actor vs Observer; between) * 2 (scenario: Switch vs Push²³; within) * 2 (kind of moral relevance: Count in Favor [CF] vs Count Against [CA]; within) mixed design.

In Extended, participants first read general instructions to familiarize themselves with the task and the rating scale (see Appendix). In Original, participants started the experiment without general instructions (as in Andow, 2018). They were then presented with the two scenarios (order randomized), each on a separate screen and together with the two statements about moral relevance, namely CA and CF. In Extended, participants were additionally, and first, asked whether the agent should perform the available "intervention"²⁴ in the scenario (henceforth: should-question) before they were presented with the two statements about moral relevance. This additional question was introduced to avoid that participants interpret the moral relevance statements as being about the morally appropriate action. Another additional feature in Extended was a short explanation (see below) of the moral relevance questions that was presented after the should-question and before the two questions about moral relevance.

The two scenarios were worded as in the original study (Andow, 2018), with the Observer vs Actor manipulation in brackets:

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/You] 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.

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/You] 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.

In Original, the two questions and answer options about moral relevance were worded as follows, with the Push vs Switch and the Observer vs Actor manipulation in brackets:

CA: The fact that pushing the [bystander/the button] will lead to the death of one innocent bystander who would otherwise have survived counts against [you/Abigail] pushing the bystander.

CF: The fact that pushing the [bystander/the button] will prevent the death of five innocent people who would otherwise have died counts in favour of [you/Abigail] pushing the bystander.

The response option after each statement was a six-point Likert-item ranging from "strongly disagree" to "strongly agree".

In Extended, the two questions and answer options about moral relevance were worded as follows, again with the Push vs Switch and the Observer vs Actor manipulation in brackets:

CA: If [you push/Abigail pushes] the [bystander/the button], one innocent person will be killed who would otherwise have survived. This is a negative aspect of pushing the bystander/the button and, evaluated on its own (without considering the other effects of this action), counts against doing it.

CF: If [you push/Abigail pushes] the [bystander/the button], five innocent people will be saved who would otherwise have died. This is a positive aspect of pushing the bystander and, evaluated on its own (without considering the other effects of this action), counts in favour of doing it.

The response options after each statement were "disagree" and "agree".

On the final page, participants were asked about their gender, age, whether they have already taken part in a study involving trolley dilemmas and were also presented with a simple transitivity task ("If Peter is taller than Alex, and Alex is taller than Max, who is the smallest among them?"). The transitivity task was included to have a rough comparison between the proportions of the right answer to this question and to the questions about moral relevance.

lii. Results and Discussion

The results are summarized in Table 1 and Figure 1. Participants' moral relevance judgments in the Original conditions were transformed into a binary format, with ratings smaller than 4 now classified as "disagree" and ratings equal to or greater than 4 as "agree".

1058 👄 H. DABBAGH

Scenario	Perspective	Relevance	Version	Disagree	%	Agree	%	χ ²	p-value
Scenario	reispective	Relevance	VEISION	Disagree	70	Agree	70		p-value
Push	Actor	CA	Original	31	28%	79	72%	$\chi^{2}_{1,216} = 15.53$	p = .0001
Push	Actor	CA	Extended	8	8%	98	92%		
Push	Actor	CF	Original	39	35%	71	65%	$\chi^{2}_{1,216} = 11.79$	p = .0006
Push	Actor	CF	Extended	16	15%	90	85%		
Push	Observer	CA	Original	30	28%	76	72%	$\chi^{2}_{1,211} = 12.11$	p = .0005
Push	Observer	CA	Extended	10	10%	95	90%		
Push	Observer	CF	Original	36	34%	70	66%	$\chi^{2}_{1,211} = 12.41$	p = .0004
Push	Observer	CF	Extended	14	13%	91	87%		
Switch	Actor	CA	Original	36	33%	74	67%	$\chi^{2}_{1,216} = 15.84$	p = .0001
Switch	Actor	CA	Extended	11	10%	95	90%		
Switch	Actor	CF	Original	24	22%	86	78%	$\chi^{2}_{1,216} = 7.41$	p = .0065
Switch	Actor	CF	Extended	9	8%	97	92%		
Switch	Observer	CA	Original	41	39%	65	61%	$\chi^{2}_{1,211} = 20.83$	p < .0001
Switch	Observer	CA	Extended	12	11%	93	89%		
Switch	Observer	CF	Original	23	22%	83	78%	$\chi^{2}_{1,211} = 17.33$	p < .0001
Switch	Observer	CF	Extended	3	3%	102	97%		
Overall			Original	260	30%	604	70%	$\chi^{2}_{1,1708} = 109.18$	p < .0001
			Extended	83	10%	761	90%		

 Table 1. Moral relevance ratings and the comparison of Original versus Extended conditions.

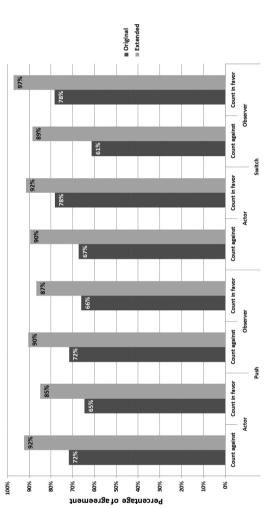
Effect of Actor versus Observer Manipulation

Neither in the Original nor in the Extended conditions did I find an effect of the perspective (Actor vs Observer) manipulation, p > .05 for each of the eight Actor vs Observer comparisons. Hence, I replicated Andow's finding that it did not make a difference for judgments about moral relevance whether the scenario was framed in the second versus third person perspective.

Agreement Rates with Moral Relevance Statements in Original versus Extended Conditions

Overall, the agreement rate with the moral relevance statements was 80%, with 70% in the Original conditions and 90% in the Extended conditions. The agreement rate in the Original condition was close to the one obtained in the original study by Andow (72%) and significantly lower than in the Extended conditions, $\chi^2_{1,1708} = 109.18$, p < .0001. Agreement rates were higher not only in the Extended conditions but also for each of the eight comparisons, all ps < .01 (see Table 1).

The proportion of inconsistent moral relevance judgments across the two scenarios – e.g. judging that the death of one person counts against performing the action in Push but not in Switch – was 23% in the Original conditions (24% for CA und 21% for CF) and 10% in the Extended conditions (8% for CA and 12% for CF). Hence, the proportion in the Original condition was close to the one obtained in the original study by Andow (20% for CF and 17% for CA) and overall significantly higher than in the Extended conditions, $\chi^2_{1,854} = 14.18$, p < .0001.





Furthermore, the agreement rate for the moral relevance statements in the Extended conditions, 90%, was close to the proportion of correct answers to the simple transitivity task (89%; "If Peter is taller than Alex, and Alex is taller than Max, who is the smallest among them?).

To sum up, I found no influence of the perspective (Actor vs Observer) manipulation on moral relevance judgments. The agreement rate for judgments about moral relevance in the Extended conditions was not only higher than in the Original conditions but also, at 90%, high in absolute terms. Hence, although the findings of my first experiment are good news for moral intuitionists since they would make room for relying on some of our moral intuitions, this does not entail that one should accept the intuitionist epistemology entirely in order to explain the findings. Because one might still believe that some of our intuitions are reliable even if one does not necessarily believe that there are some *real* self-evident moral propositions.

Experiment 2

This experiment is based on Experiment 3a by Andow (2018), in which he assessed whether judgments about moral relevance are affected by the order in which moral scenarios are presented. He found order effect for CF judgments for Switch, with higher agreement rates when Switch was presented first, as compared to Switch being presented after Push.²⁵ Moreover, the overall agreement rate with judgments about moral relevance in the Switch scenario was again only about 72%. While in the original study by Andow, a third scenario, Loop, was used, I will here only use Switch and Push since the order effect was strongest when these two scenarios were paired.

In the following experiment, I extended Andow's experiment by including conditions that were modified in the same ways as in my first experiment to (potentially) raise agreement rates for judgments of moral relevance.

I. Participants

406 participants started the survey, and the data of 399 were included in the analysis (7 were excluded for not finishing the experiment or finishing it in less than 40 seconds). The mean age was 36 years; 33% were male, and 67% female. Participants received £0.25 for estimated 3 minutes of their time (\pounds 5/h).

li. Design, Procedure and Materials

Participants were randomly assigned to one of four conditions in a 2 (experiment: Original vs Extended; between) * 2 (order: Switch first vs Push first; between) * 2 (scenario: Switch vs Push; within) * 2 (kind of moral relevance: Count in Favor [CF] vs Count Against [CA]; within) mixed

			J						
Version	Scenario	Position	Relevance	Disagree	%	Agree	%	χ ²	p-value
Original	Push	First	CA	28	28%	72	72%	$\chi^{2}_{1,200} = 15.53$	p = .2866
Original	Push	Second	CA	35	35%	65	65%		
Original	Push	First	CF	52	52%	48	48%	$\chi^{2}_{1,200} = 9.08$	p = .0026
Original	Push	Second	CF	31	31%	69	69%		
Original	Switch	First	CA	44	44%	56	56%	$\chi^{2}_{1,200} = 1.33$	p = .2482
Original	Switch	Second	CA	36	36%	64	64%		
Original	Switch	First	CF	9	9%	91	91%	$\chi^{2}_{1,200} = 23.39$	p < .0001
Original	Switch	Second	CF	38	38%	62	62%		
Extended	Push	First	CA	12	12%	88	88%	$\chi^{2}_{1,199} = 0.04$	p = .8445
Extended	Push	Second	CA	11	11%	88	89%		
Extended	Push	First	CF	18	18%	82	82%	$\chi^{2}_{1,199} = 0.16$	p = .6927
Extended	Push	Second	CF	20	20%	79	80%		
Extended	Switch	First	CA	14	14%	85	86%	$\chi^{2}_{1,199} = 0.06$	p = .8141
Extended	Switch	Second	CA	13	13%	87	87%		
Extended	Switch	First	CF	5	5%	94	95%	$\chi^{2}_{1,199} = 0.33$	p = .5635
Extended	Switch	Second	CF	7	7%	93	93%		

Table 2. Moral relevance ratings and the effect of order of presentation.

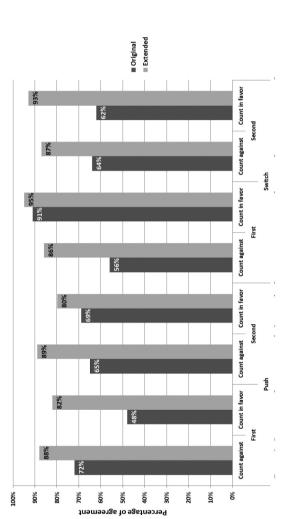
design. The only difference to my first experiment was that the perspective manipulation was replaced by order manipulation.

lii. Results and Discussion

The results are summarized in Table 2 and Figure 2. Participants' moral relevance judgments in the Original conditions were again transformed into a binary format, with ratings smaller than 4 now classified as "disagree" and ratings equal to or greater than 4 as "agree".

Order Manipulation

In the Extended conditions, I did not find an order effect (all ps > .5; see Table 2). In the Original conditions, I found an order effect for the CF statements in Switch and Push. For Push, the agreement rate with the CF statement was lower when Push was presented first (48%), as compared to when Push was preceded by Switch (69%), p < .01. For Switch, the agreement rate with the CF statement was higher when Switch was presented first (91%), as compared to when Switch was preceded by Push (62%), p < .0001. This latter finding also occurred in the original study by Andow (2018). The agreement rates for the CA statements in the original conditions were not affected by the order of presentation of the scenarios.





Agreement Rates with Moral Relevance Statements in Original versus Extended Conditions

Overall, the agreement rate with the moral relevance statements was 77%, with 66% in the Original conditions and 87% in the Extended conditions. The agreement rate in the Original conditions was overall significantly lower than in the Extended conditions, $\chi^2_{1,1596} = 103.58$, p < .0001, and also in six of the eight comparisons, with the following two exceptions. The agreement rate for the CF statement in Push, when presented after Switch, was not significantly higher in the Extended condition (80%), as compared to the Original condition (69%), $\chi^2_{1,199} = 304.58$, p = .08; and neither was the agreement rate for the CF statement in Switch, when presented first, significantly higher in the Extended condition (95%), as compared to the Original condition (91%), $\chi^2_{1,199} = 1.19$.

The proportion of inconsistent moral relevance judgments across the two scenarios – e.g. judging that the death of one person counts against performing the action in Push but not in Switch – was 26% in the Original conditions (24% for CF und 28% for CA) and 12% in the Extended conditions (14% for CF and 10% for CA), with the proportion in the Original conditions significantly higher than in the Extended conditions, $\chi^2_{1,798} = 24.37$, p < .0001.

Furthermore, the agreement rate for the moral relevance statements in the Extended conditions, 87%, was close to the proportion of correct answers to the simple transitivity task (90%).

To sum up, I found no influence of the order manipulation on moral relevance judgments in the Extended conditions, and the agreement rate for judgments about moral relevance in the Extended conditions was not only higher than in the Original conditions but with 87% also high in absolute terms. Hence, the findings of my second experiment also provide good news for moral intuitionists. However, as stated before, we should be aware that although stability in judgments about moral relevance would be good news for moral intuitionists, e.g. Rossian intuitionists, such findings would also be good news for moral intuitionists who, unlike Ross, believe that we can use non-inferentially justified intuitions to support moral principles governing all-things-considered actions.

6. Discussion

Are intuitions about moral relevance susceptible to framing effects? Andow (2018) conducted several experiments to investigate this question and found that people's intuitions about moral relevance were relatively unaffected by framing effects. However, the findings were not fully satisfying news for moral intuitionism since, in some experiments, about one-third of

participants disagreed with the purportedly self-evident moral relevance statements, such as "The fact that pushing the bystander will lead to the death of one innocent bystander who would otherwise have survived counts against you pushing the bystander". I speculated that the relatively low agreement rate for the purportedly self-evident moral relevance statements might be because participants might have interpreted the statements as being about the overall moral status of the corresponding action (e.g. whether the action is permissible or ought to be performed). To test this hypothesis, I replicated two of Andow's experiments in which the agreement rate was especially low and added in each experiment a condition in which I made some modifications to ensure that participants understand the statements about moral relevance in an intended way. For instance, I added a question about whether the proposed action should be performed in order to make clear that the statements about moral relevance were not about the overall moral assessment of the proposed action.

Experiment 1 investigated whether intuitions about moral relevance exhibit an actor versus observer bias. I replicated Andow's finding that it did not make a difference for judgments about moral relevance whether the scenario was framed in the second versus third person perspective. Neither in the original conditions used by Andow nor in my extended conditions did this factor have an influence on participants' moral relevance judgments. Furthermore, I also replicated the relatively low agreement rates for the moral relevance statements in the original conditions: Overall, only about 70% of participants (72% in the original study) agreed with the moral relevance statements. Crucially, in my extended conditions, the agreement rate was significantly higher than in the original condition, and also high in absolute terms (90%) and at roughly the same level as the proportion of correct answers to a simple intransitivity task ("If Peter is taller than Alex, and Alex is taller than Max, who is the smallest among them?"). Hence, the results of my first experiment were good news for moral intuitionism.

Experiment 2 investigated whether intuitions about moral relevance are influenced by the order in which the moral dilemmas were presented to participants. In the original conditions used by Andow, the order of presentation of the moral dilemmas affected the agreement rate for the moral relevance statement in two out of four cases. Moreover, the overall agreement rate with the purportedly self-evident moral relevance statements was relatively low (66%). In contrast, in the extended conditions, the order of presentation had no influence on moral relevance ratings and the overall agreement rate was significantly higher than in the original conditions, high in absolute terms (87%) and at roughly the same level as the proportion of correct answers to a simple intransitivity task. Hence, the results of my second experiment were also good news for moral intuitionism. Let us now bring on a critical reflection on my experiments and findings, including an acknowledgment of the limits of my experiments, a discussion of different possible interpretations of the findings, and a critical reflection on the kind of support that the findings provide for moral intuitionism.

As I have shown in this paper, empirical studies are important for the epistemology of moral intuitionism. There are a number of empirical studies which try to undermine the epistemology of moral intuitionism on the grounds that there are many disagreements about self-evident propositions. For example, Andow found troubling results in his experiments since there is relatively low agreement about self-evident moral propositions. However, what I did in this paper was a kind of *negative support* for moral intuitionism. My experiments indicate that the worrisome conclusion suggested by Andow's findings is not warranted. Although I defended moral intuitionism against one kind of threat, this does not bring out positive support entirely for moral intuitionism. Rather it just shows that holding some minimal pragmatic requirements will change the worrisome conclusion in favor of moral intuitionism.

One might, however, wonder why Andow's wording causes a problem and why my revised wording fixes it? What would be my pragmatic-based explanation of the original findings? What inferences caused the pattern of disagreement, and why did my modifications block those inferences?

In response, I found two worries against Andow's wording. 1) Some participants might understand the moral relevance statements as "action should be done" or "actions shouldn't be done". For example, they might understand "counts in favour" as "action should be done". To block this inference, I added the question in my experiment about what should be done, evaluated on its own, and asked the participants first: if [you push/ Abigail pushes] the [bystander/the button], five innocent people will be saved who would otherwise have died. This is a positive aspect of pushing the bystander and, evaluated on its own (without considering the other effects of this action), counts in favor of doing it. 2) It could also be that "counts against" is understood as "there is a decisive reason not to do the action". That is why I added the question in my experiment, evaluated it on its own, and asked the participants first: if [you push/Abigail pushes] the [bystander/the button], one innocent person will be killed who would otherwise have survived. This is a negative aspect of pushing the bystander/the button and, evaluated on its own (without considering the other effects of this action), counts against doing it.

However, one might object that the statements in question, in my Extended materials, involve a conjunction of two clauses: [i] This is a negative/positive aspect of the action and [ii] evaluated on its own (without considering the other effects) counts against/in favor of doing it. My findings show near-unanimous agreement with such statements. But only the second clause in each case clearly concerns "moral relevance"; the first clause could well be interpreted as a vague evaluative statement, but not necessarily a moral reason-giving one. This can be a potential problem since it could be a unanimous agreement with the first clause that is really driving the high levels of agreement in these studies, weakening my central claim to have uncovered evidence of lack of disagreement about moral relevant judgments.²⁶

To tackle this potential problem, I suggest that further studies should be done to support what my findings indicate: that there is a near-unanimous agreement with moral relevance statements. These studies should be as closely matched as possible to Andow's setup of the experiment while at the same time avoiding that the moral relevance statement is misunderstood. The experiments in my paper put more emphasis on the second point (that the moral relevance statement should not be misunderstood), resulting in an experimental setup that diverges from Andow's in several points. While it does not seem *prima facie* plausible that participants' nearunanimous agreement is mainly based on the positive/negative aspect clause, because it does not seem plausible that participants choose "agree" when they only agree with one part of a conjunctive statement, this possibility is not entirely ruled out with the current experimental setup and should be addressed in future studies.

7. Conclusion

To sum up, my experiments show that participants have relatively high agreement among their intuitions about moral relevance statements. This entails that this group of people has some shared intuitions about self-evident *pro tanto* duties (as moral reason-giving facts for or against certain actions). Moreover, framing effects on intuitions about moral relevance are minor (rather: non-existent). Thus, these experiments do not indicate that moral intuitions are unreliable. However, this does not entail that moral intuitions are reliable because the experiments merely do the former and not the latter.

How about asking whether these experiments show us things about the reliability of *our* shared moral intuitions, e.g. "it is wrong to torture someone for one's own amusement" or "that an act would hurt an innocent person must count morally against it" or "that an act would reduce the pain an innocent being is suffering counts morally in favour of it". Well, strictly speaking, the answer must be no, since *we* were not subjects in the experiments. It is, however, possible to move from findings about the intuitions had by the *subjects* in the experiments to conclusions about the reliability of *our* shared intuitions if we add some premise along the lines of "what is true of the influence of framing effects on the moral relevance intuitions of the

subjects of these experiments is also probably at least roughly true of the influence of framing effects on the moral relevance intuitions *we* have". With this extra premise, we can draw conclusions about the reliability of *our* shared moral intuitions. However, we should acknowledge that the possibility to draw conclusions about the reliability of "our shared intuitions" is even more limited. Because there are more relevant differences, e.g. differences between intuitions about trolley scenarios and intuitions in an experimental setup and having intuitions in the course of everyday life.

Notes

- 1. For an alternative view, see Dabbagh (2017).
- 2. Sinnott-Armstrong's argument is supposed to show that intuitions cannot be non-inferentially justified, whether the non-inferential justification is due to self-evidence or due to something else. Self-evidence is only one way a moral intuition might be non-inferentially justified. For instance, some intuitionists have a reliabilist account of the non-inferential justification of moral intuitions. This kind of account entails that moral intuitions are non-inferentially justified, but it does *not* entail that they are self-evident.
- 3. To qualify this point, a *pro tanto* duty, in fact, is *not a type of duty*. But it is *related to duty in a particular way*. For example, Ross himself writes that such duties "suggest that what we are speaking of is a certain kind of duty, whereas it is in fact not a duty, but something related in a special way to duty" (1930, 2002, 20).
- Ewing (1959, pp. 63, 110); Urmson (1975, pp. 112–3); Stratton-Lake (2011, pp. 147– 150).
- 5. These foundational beliefs (e.g. self-evident propositions) may get extra justification by reference to other beliefs. See Audi (1996).
- 6. However, not all *a priori* propositions are self-evident. For example, a proposition like "all bachelors are unmarried or Obama's eyes are blue" is *a priori* but it is not selfevident in itself.
- 7. Note that the justification of moral intuitions (e.g. intuition that x is wrong) is important here, not intuitions about the justification of moral intuitions (e.g. intuition that moral intuitions are unreliable).
- 8. For an alternative view, see Dabbagh (2018a & 2018b).
- 9. See Sinnott-Armstrong (2011, pp. 15-16).
- 10. Tversky and Kahneman (1981).
- 11. Wiegmann and Waldmann (2014).
- 12. Ross, 2002 [Ross, 2002]); Audi (2004); Stratton-Lake (2002).
- 13. Although Ross's *The Right and the Good* maintains that basic moral beliefs can be justified only non-inferentially, his earlier paper "The Basis of Objective Judgments in Ethics" says that basic moral beliefs can be justified in both ways, i.e. inferentially and non-inferentially. In fact, Ross thinks there are some intuitively true beliefs about *pro tanto* duties that *can* be recognized directly. But this does not commit Ross to holding that *all* basic moral propositions *are* known directly.
- 14. See Ross, 2002, Ross, 2002, pp. 29-30, 40) and (1939, 29-30).
- 15. This is actually what Ross himself seems to think when he writes in his article that: "even if it could be inferred that love or aesthetic experience is good, I feel sure that

our judgment that they are so is intuitive; that something can be inferred does not prove that it cannot be seen intuitively" (1927, 121).

- 16. For more details, see Stratton-Lake (2002).
- 17. See Stratton-Lake and Intuitionism in Ethics (2016). When Ross writes that a selfevident proposition is "evident without any need of proof, or of evidence beyond itself", this means that sufficient understanding of a self-evident proposition is enough to be justified in believing that proposition, but this does not entail that other justifications for believing it cannot exist. See Hooker (2002, fn. 7) and Audi (1999). Self-evident propositions might be derivable from other propositions by deduction.
- 18. See Hooker (2000, p. 12) & (2002, 163).
- 19. Ross (1939, pp. 271-275).
- 20. Ross (1939, p. 170).
- 21. Ross (1939, p. 49).
- 22. See Cullen (2010), Grice (1989), Hagan and Royzman (2017), and Noveck and Reboul (2008), and Samland, J and Waldmann (2016).
- 23. In Andow (2018), this version of the trolley dilemma is labeled "Physical".
- 24. I said intervention becuase the formulation "the available action in the scenario" would be misleading when there is not only one available "action" in the scenario: we have an action (pushing a button in Switch, pushing the person in Footbridge), and an omission (not doing anything).
- 25. In Andow (2018), this version of the trolley dilemma is labeled "Physical".
- 26. I would like to thank the anonymous reviewers for raising this critical point.

Acknowledgments

I would like to express my gratitude to Alex Wiegmann, who kindly helped me with the experiments. I would also like to thank James Andow, Brad Hooker, Regina Rini, and the editor and anonymous reviewers for their sharp and comprehensive critical comments.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Notes on contributor

Hossein Dabbagh is a Philosophy Tutor at the University of Oxford's Department for Continuing Education. He is also an Honorary Research Fellow at the UCL Institute of Education.

References

- Andow, J. (2018). Are Intuitions About Moral Relevance Susceptible to Framing Effects? *Review of Philosophy and Psychology*, 9(1), 115–141. https://doi.org/10.1007/s13164-017-0352-5
- Audi, R. (1996). Intuitionism, Pluralism, and the Foundation of Ethics. In W. Sinnott Armstrong & M. Timmons (Eds.), *Moral Knowledge? New Readings in Moral Epistemology* (pp. 101–136). Oxford University Press.

Audi, R. (1999). Self-Evidence. Philosophical Perspectives, 13, 205-228.

- Audi, R. (2004). *The Good in the Right: A Theory of Intuition and Intrinsic Value*. Princeton University Press.
- Cullen, S. (2010). Survey-driven romanticism. *Review of Philosophy and Psychology*, 1(2), 275–296. https://doi.org/10.1007/s13164-009-0016-1
- Dabbagh, H. (2017). Sinnott-Armstrong Meets Modest Epistemological Intuitionism. The Philosophical Forum, 48(2), 175–199. https://doi.org/10.1111/phil.12151
- Dabbagh, H. (2018a). Feeling Good: Integrating the Psychology and Epistemology of Moral Intuition and Emotion. *Journal of Cognition and Neuroethics*, 5(3), 1–30.
- Dabbagh, H. (2018b). The Seeming Account of Self-Evidence: An Alternative to Audian Account. Logos and Episteme, 9(3), 261-284. https://doi.org/10.5840/logos-episteme20189322
- Ewing, A. C. (1959). Second Thoughts in Moral Philosophy. Routledge & Kegan Paul.
- Grice, H. P. (1989). Studies in the way of words. Harvard University Press.
- Hagan, J. P., & Royzman, E. (2017). The shadow and the tree: Inference and transformation. In J.-F. Bonnefort & B. Trémolière (Eds.), *Moral inferences* (pp. 37–55). Routledge.
- Hooker, B. (2000). Ideal Code, Real World. Oxford University Press.
- Hooker, B. (2002). Intuitions and Moral Theorizing. In P. Stratton-Lake (Ed.), *Ethical Intuitionism: Re-evaluations* (pp. 161–183). Oxford University Press.
- Noveck, I. A., & Reboul, A. (2008). Experimental Pragmatics: A Gricean turn in the study of language. *Trends in Cognitive Sciences*, 12(11), 425–431. https://doi.org/10.1016/j.tics. 2008.07.009
- Palan, S., & Schitter, C. (2018). Prolific.ac—A subject pool for online experiments, Journal of Behavioral and Experimental Finance, Volume 17, Pages 22-27ISSN 2214–6350, https:// doi.org/10.1016/j.jbef.2017.12.004
- Ross, W. D. (1927). The Basis of Objective Judgments in Ethics. International Journal of Ethics, 37(2), 113–127. https://doi.org/10.1086/intejethi.37.2.2377758
- Ross, W. D. (1939). The Foundations of Ethics. Clarendon Press.
- Ross, W. D. (2002). The Right and the Good. (P. Stratton-Lake ed.). Clarendon Press. 1930
- Samland, J, W. A., & Waldmann, M. R. (2016). Lying despite telling the truth. *Cognition*, 150, 37–42. https://doi.org/10.1016/j.cognition.2016.01.017
- Sinnott-Armstrong, W. (2006). Moral Intuitionism Meets Empirical Psychology. In T. Horgan & M. Timmons (Eds.), *Metaethics After Moore* (pp. 339–365). Oxford University Press.
- Sinnott-Armstrong, W. (2008). Framing Moral Intuitions. In W. Sinnott-Armstrong (Ed.), Moral Psychology (Vol. 2, pp. 47–76). MIT Press.
- Sinnott-Armstrong, W. (2011). An Empirical Challenge to Moral Intuitionism. In J. Graper Hernandez (Ed.), and 200-203, *The New Intuitionism* (, 11–28). Continuum.
- Stratton-Lake, P. (2002). Introduction. In P. Stratton-Lake (Ed.), *Ethical Intuitionism: Re-evaluations* (pp. 1–28). Oxford University Press.
- Stratton-Lake, P. (2011). Eliminativism about derivative prima facie duties. In T. Hurka (Ed.), *Underivative Duty* (pp. 146–165). Oxford University Press.
- Stratton-Lake, P., (2016). "Intuitionism in Ethics", *The Stanford Encyclopedia of Philosophy*, E. N. Zalta (ed.), https://plato.stanford.edu/archives/win2016/entries/intuitionism-ethics /
- Tversky, A., & Kahneman, D. (1981). The framing of decisions and the psychology of choice. *Science*, 211(4481), 453–458. https://doi.org/10.1126/science.7455683
- Urmson, J. O., (1975). "A Defence of Intuitionism". *Proceedings of the Aristotelian Society*, 75: 111–119.

Wiegmann, A., & Waldmann, M. R. (2014). "Transfer effects between moral dilemmas: A causal model theory". Cognition, 131(1), 28–43. https://doi.org/10.1016/j.cognition.2013. 12.004

Appendix

Instruction Extended:

Dear participant,

Thank you for taking part in this study!

Please take your time and read the following instruction and scenarios carefully.

On the following pages, we will present you with two situations in which a person has to make a decision. Please try to understand the situations as well as possible. Please assume that in the situations, everything happens as described (even if you think that the scenario – or parts of it – are completely unrealistic). That implies that the mentioned options are the only ones available. You can and should also assume that there are no general rules for these kinds of situations and that the person is not going to be prosecuted for their decision.

After having understood the situation, your task will be to judge what the person in each situation should do (in terms of *morality*) and also to indicate whether you disagree or agree with statements about certain aspects of the scenario.

Below, you see an arbitrary example question together with a rating scale to demonstrate how we will ask you to provide your judgments (unlike the later task, this example is not about morality):

Simon has to buy a new watch and can choose between only two models. Watch A is of good quality, costs 100 pounds and has a color that Simon doesn't like very much. Watch B is of equally high quality, costs 500 pounds and is Simon's favorite color. Should Simon buy Watch B?

[definitely not 1-2-3-4-5-6 definitely]

In this case, you should tick box 1 if you think that Simon should definitely not buy watch B. Tick box 6 if you think that Simon should definitely buy watch B. If your judgment lies between these extremes, tick the box between 1 and 6 that best reflects your judgment.

When making a difficult decision like the one above, one has to assess the positive and negative aspects of each option that is available. Maybe the positive aspects of one alternative seemed better to you compared to the positive aspects of the other alternative. Or maybe the negative aspects of one alternative seemed less bad to you than the negative aspects of the other option.

No matter how strongly you weighed each option's positive and negative aspects and to which conclusion you ultimately came, we are interested in the general direction (positive or negative) in which you evaluated aspects of the situation.

Please answer the following questions about your reasoning when you thought about the hypothetical situation.

If Simon buys watch B, he will spend more money than necessary considering the quality of the watch. This is a negative aspect of buying watch B and, evaluated on its own (without considering the other effects of this action), counts against doing it.

[disagree/agree]

If Simon buys watch B, he will have a watch that is his favorite color. This is a positive aspect of buying watch B and, evaluated on its own (without considering the other aspects of this action), counts in favor of doing it.

[disagree/agree]

In both cases, you should choose the option "agree" even if you only mildly or somewhat agree to the claim, and, likewise, choose "disagree" even if you only mildly or somewhat disagree.

Please read and answer each of the questions carefully. You will not be allowed to omit questions, and you will not be able to return to previous pages.

The situations might appear somewhat unrealistic to you. Still, please assume that everything takes place exactly as described and take your judgment task seriously. Furthermore, the answer may seem obvious, but this fact should not disturb you.

You are free to leave the study at any time, and your participation is completely voluntarily. We guarantee you that the collected data will be made anonymous and only be used in the study at hand. If you agree with the stated terms, click on the Continue button below. If you have any further questions regarding this experiment, please feel free to contact us at awiegmann@aol.com.

For the reader's convenience, here is a comparison of the test page in the Original versus the Extended version for Push dilemma in the Actor version.

Vignettes

Original:

A runaway trolley is headed toward five innocent people who are on the track and who will be killed unless something is done. You 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.

The fact that pushing the bystander will lead to the death of one innocent bystander who would otherwise have survived counts against you pushing the bystander. [strongly disagree 1-2-3-4-5-6 strongly agree 6]

The fact that pushing the bystander will prevent the death of five innocent people who would otherwise have died counts in favor of you pushing the bystander. [strongly disagree 1-2-3-4-5-6 strongly agree 6]

Extended:

A runaway trolley is headed toward five innocent people who are on the track and who will be killed unless something is done. You 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.

Please answer the following questions about the described situation:

Should you push the bystander?

Please choose the option that best expresses your moral intuition:

[definitely not 1-2-3-4-5-6 definitely]

When making a difficult decision like the one above, one has to assess the positive and negative aspects of each option that is available. Maybe the positive aspects of one action alternative seemed more desirable to you compared to the positive aspects of the other alternative. Or maybe the negative aspects of one alternative seemed less bad to you than the negative aspects of the other option.

No matter how strongly you weighed each option's positive and negative aspects and to which conclusion you ultimately came, we are interested in the general direction (positive or

1072 👄 H. DABBAGH

<u>negative</u>) in which you evaluated aspects of the situation. Please answer the following questions about your reasoning when you thought about the hypothetical situation.

If you push the bystander, one innocent person will be killed who would otherwise have survived. This is a negative aspect of pushing the bystander and, evaluated on its own (without considering the other effects of this action), counts against doing it.

[disagree/agree]

If you push the bystander, five innocent people will be saved who would otherwise have died. This is a positive aspect of pushing the bystander and, evaluated on its own (without considering the other effects of this action), counts in favor of doing it.

[disagree/agree]