Routledge
Taylor & Francis Group

Philosophical Psychology

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

Why do people seem to be more utilitarian in VR than in questionnaires?

Bartosz Maćkiewicz, Jan Wodowski & Joanna Andrusiewicz

To cite this article: Bartosz Maćkiewicz, Jan Wodowski & Joanna Andrusiewicz (06 Dec 2023): Why do people seem to be more utilitarian in VR than in questionnaires?, Philosophical Psychology, DOI: 10.1080/09515089.2023.2282060

To link to this article: https://doi.org/10.1080/09515089.2023.2282060

Published online: 06 Dec 2023.

PHILOSOPHICAL PSYCHOLOGY
https://doi.org/10.1080/09515089.2023.2282060

Routledge
Taylor & Francis Group

Why do people seem to be more utilitarian in VR than in questionnaires?

Bartosz Maćkiewicz[a], Jan Wodowski[a] and Joanna Andrusiewicz[b]

[a]Faculty of Philosophy, University of Warsaw, Warsaw, Poland; [b]Center for Bioethics and Biolaw, Faculty of Philosophy, University of Warsaw, Warsaw, Poland

ABSTRACT

Several experimental studies on moral judgment and moral decision-making show that in virtual reality people tend to make more "characteristically utilitarian" decisions than when responding to standard questionnaires. An explanation of this phenomenon that has been considered by many authors states that the feature of VR studies that is responsible for this effect is the visual salience of the harmful consequences of "deontological" decisions. The present paper makes three points, the first of which is theoretical: we argue that this explanation, which draws from Cushman's dual-process account of moral judgment, is in fact not coherent with this account's predictions with respect to behavior in VR. The second point is that this explanation does not sufficiently explain the existing empirical findings concerning the footbridge dilemma because these studies differ in important aspects of experimental design from studies on the switch dilemma. The third point is empirical: we present two original VR studies that were designed to check the robustness of the increased "utilitarian" tendency and directly test the explanation that is based on the visual salience of harmful consequences. The results of the experiments provide evidence that the effect is quite robust but the proposed explanation is inadequate.

ARTICLE HISTORY
Received 2 January 2023
Accepted 2 November 2023

KEYWORDS
Moral psychology; dual-process theory; virtual reality; moral dilemmas

1. Introduction

People seem to make different decisions in studies where moral dilemmas are presented using Virtual Reality technology than when answering questionnaires. The main finding of several VR studies (Francis et al., 2016, 2017, 2018, 2019; Patil et al., 2014) is that in VR people tend to make "characteristically utilitarian" decisions more often than in studies that use more traditional modes of stimuli presentation; however, it is not clear why this is the case. The goal of this paper is thus twofold. The first aim is to review the emerging literature that reports such differences (Section 1) and discuss

CONTACT Bartosz Maćkiewicz b.mackiewicz@uw.edu.pl Faculty of Philosophy, University of Warsaw, Krakowskie Przedmieście 3, Warsaw 00-047, Poland

the existing account of this phenomenon (Section 2). We will then show that the emerging literature is inadequate on theoretical and empirical grounds (Section 3). The second aim is to present original experimental results from two studies that, on the one hand, show the robustness of this phenomenon, and, on the other hand, call into doubt the existing explanations (Section 4 and Section 5). Since individual differences among the participants could be a potential confounding factor, in addition to moral dilemmas in the experiment we included two personality inventories (HEXACO-PI-R and TriPM-41) to control for variation in personality traits. Finally, we will present two exploratory meta-analyses conducted to test the robustness of the phenomenon under considerations and to assess its magnitude (Section 6).

2. Moral judgments, moral decisions, and virtual reality

There is a rich tradition of empirical research on moral judgments and moral decisions. Psychologists and experimental philosophers have investigated these issues using scenarios borrowed from classical philosophical literature. One of the most discussed and used moral dilemmas is the so-called "trolley problem" (Foot, 1967; Thomson, 1976). In the first version of this dilemma, the "switch" scenario (Foot, 1967), an agent is presented with a situation in which they can save five workers on trolley tracks by pulling a lever, which results in redirecting the trolley to another track; however, on the second track, there is one worker who will inevitably be killed by the redirected trolley. In the "footbridge" version introduced by Thomson (1976), the action that can be performed to save the five workers consists in pushing a fat man off a bridge to stop the trolley. The outcome is the same as in the "switch" scenario: five people are saved and one person is killed.

In the psychological (e.g., Mikhail, 2002; Nakamura, 2012; Waldmann & Dieterich, 2007) and (to a lesser extent) the philosophical (Edmonds, 2013; Lanteri et al., 2008; Singer, 2005) literature, the decision to act in these situations is associated with the "characteristically utilitarian" ("utilitarian", for short) mode of thinking (the lives of five people have greater value than the life of one person), and refraining from acting is considered to be the "characteristically deontological" ("deontological", for short) choice (actively killing a person in these situations is wrong, regardless of the other outcomes). Although we are not committed to this simplistic view of utilitarianism and deontology, we will follow the existing literature in our use of this distinction.

Scenarios depicting these two versions of the trolley dilemma are typically presented in verbal form to participants of experimental studies, whose responses are collected using questionnaires in which they are typically asked to judge the rightness or wrongness of acting in such situations

(Ahleniusa & Tännsjö, 2012; Côté et al., 2013; Cushman & Young, 2011; Cushman et al., 2006; Hauser et al., 2007; Lanteri et al., 2008; Lombrozo, 2009; Mikhail, 2002; Moore et al., 2008; Nakamura, 2012; Schwitzgebel & Cushman, 2012) or to declare what they would do if they faced them (Gold et al., 2014; Greene et al., 2001, 2004, 2008; Koenigs et al., 2007; O'Neill & Petrinovich, 1998; Petrinovich et al., 1993). The most striking tendency that has been observed in many studies is that people are inclined to make the "utilitarian" decision in the "switch" scenario and the "deontological" one in the "footbridge" version of the dilemma (e.g., Côté et al., 2013; Greene et al., 2001; Hauser et al., 2007; Lanteri et al., 2008; Lombrozo, 2009). Many accounts of this phenomenon have been proposed (Cushman, 2013; Greene et al., 2001, 2004, 2008; Singer, 2005; Waldmann & Dieterich, 2007). The most prominent line of thinking consists in explaining the difference in judgments about the "switch" and the "footbridge" cases by postulating two processes that are independent to some extent and play a role in forming moral judgments (Edmonds, 2013; Haidt, 2001; Huebner et al., 2009; Kahneman, 2011; Orsi, 2012). The gist of such explanations is that two psychological mechanisms are responsible for evaluating the rightness or wrongness of actions, and these mechanisms are sensitive to different aspects of a situation. On these accounts, certain features that are present in the "footbridge" version but not in the "switch" version trigger or boost the response from the system (resulting, for instance, in strong emotional aversion to pushing the man off the bridge) that produces a negative assessment of the action. The details differ depending on the specific theory. These two systems are sometimes characterized in terms of an opposition between emotional and rational processing (Greene et al., 2001, 2004, 2008; Haidt, 2001), or between representing the intrinsic value of action and the value of its consequences (Cushman, 2013; Cushman & Greene, 2012; Cushman et al., 2012).

Three factors encouraged researchers to employ Virtual Reality technology in their investigations into moral judgments and moral decision-making. The first was the inherent shortcomings of questionnaire methodology. The verbal mode of stimuli presentation (dilemmas presented in text form) and of response collection (usage of self-report measures) could poorly reflect the real-life contexts in which moral judgments and moral decisions are actually made. VR technology is expected to improve the ecological validity of studies (Parsons, 2015; Rovira et al., 2009).[1] In VR, participants experience two types of *presence*: the illusion of "being there" (also called "place illusion" – PI) that emerges as a result of the movement-induced, real-time updating of sensory perception (Slater, 2009; Slater & Sanchez-Vives, 2016); and the illusion that events are really happening, called "Plausibility" (Psi) (Sanchez-Vives & Slater, 2005, Slater et al., 2010; Slater & Sanchez-Vives, 2016), which is related to the realistic manner of

presentation. These two types of illusions of presence bring making moral decisions and moral judgments in experimental settings closer to their real-life counterparts in a manner that is not possible in questionnaire studies, thus greatly improving generalizability.

The second problem with questionnaire studies is that it is questionable to what extent they correctly operationalize the distinction between moral judgments and moral decisions. It is trivial to point out that moral decisions – understood as choosing one of two or more possible actions in ethically challenging situations and performing it in the real world – do not necessarily follow from moral judgments. Among the most common examples of moral inconsistency are akrasia, accidie, and hypocrisy. Akrasia, generally identified with weakness of will (Mele, 2022; Tenenbaum, 2010), is defined as a state when a person deliberately takes an action that goes against what he believes is best or right (Tenenbaum, 2010) or he acts intentionally counter to his own best evaluative judgment (Davidson, 2001). Accidie, in turn, involves the situation when a person recognizes something that could and should be done by her, and yet she does not take any action to achieve it (Tenenbaum, 2010). The third obvious example of a gap between judgment and action is hypocrisy, where a person can make a moral judgment condemning some action but still does it. Each of these states can result from different causes, and each of them can manifest itself both in everyday life, and during research. From the theoretical point of view, it is important to point out that the mechanism responsible for moral judgments might be to some extent separate from the mechanism that controls moral actions. It has been suggested that those separate mechanisms differ with regard to the points of reference they use: moral judgments are to be approached from the *allocentric frame of reference* and moral actions from the *egocentric frame of reference*, one is more focused on the cultural norms and the other on the *self-relevant consequences* (Tassy et al., 2013). The two mechanisms are also said to be separate not only functionally, but neurologically as different brain structures are involved in moral judgments and moral decisions (Tassy et al., 2012). The overall discrepancy between moral judgments and decisions is nicely illustrated by the fact that psychopaths, who can often correctly judge the rightness and the wrongness of actions, nevertheless engage in immoral behavior (Cima et al., 2010). Although some studies have reported differences in responses to "normatively" (e.g., "Is it morally acceptable to . . . ") and "behaviorally" (e.g., "What would you do in this situation . . . ") worded prompts (Tassy et al., 2013), which indicates that to some extent it might be possible to operationalize moral decisions using self-reported measures, there are important limitations to questionnaire methodology, especially in the context of placing the subject in the appropriate frame of reference – in case of studying moral actions: the egocentric frame of reference. To put it simply, there is

a fundamental difference between doing something and declaring that one would do something (e.g., "In this situation I would do so-and-so"). There is also vast evidence that self-reported measures are sensitive to social desirability bias (e.g., Krumpal, 2013; Randall & Fernandes, 1991).

Both the "switch" and the "footbridge" scenarios have been investigated using VR technology by other researchers. Here, we will focus only on studies that compared decisions in VR-based simulations of moral dilemmas with those made when confronted with a description of the dilemma.[2]

As far as the "footbridge" version of the trolley dilemma is concerned, Francis et al. (2016) conducted a series of studies in which they confronted participants with the dilemma in two forms. The first was a standard verbal description of the case; the second was the situation modeled in VR. In Study 1, a sample was recruited from a university participant pool, whereas Study 2 employed a more diverse sample from the general population. In the questionnaire condition in both studies, the minority (20% and 10%, respectively) of the participants decided to act, whereas in the VR condition most of the subjects (70% and 63.3%, respectively) pushed the person from the footbridge. The difference between the VR and the questionnaire condition was statistically significant in both studies.

Several follow-up studies, utilizing the same or slightly modified method and procedure, were run, each addressing novel research questions. One of them investigated the role of personal force in the moral decisions by introducing simulated movement and haptic feedback into the VR condition (Francis et al., 2017). One examined the role of real-life experience in making difficult moral decisions by employing trained paramedics and firefighters as participants (Francis et al., 2018). Another examined the effect of consumption of alcohol on moral decisions (Francis et al., 2019). In each of those studies, regardless of other experimental manipulations, a discrepancy between VR and questionnaire conditions, comparable to that in the original study, has been observed. In VR around 60% of participants decided to push the person from the footbridge, while in questionnaires this proportion oscillated around 20%.

Patil et al. (2014) ran a study using the "switch" version of the dilemma. Again, they employed an experimental design in which they varied the mode of presentation (verbal vs VR). The mode of presentation was manipulated within-subject. The same participants completed both parts of the experiments in two sessions that took place 102 days apart, on average. The order of the sessions (questionnaire and VR) was counterbalanced. In both versions of the experiment, the overall percentage of "utilitarian" decisions was very high, but compared to Francis et al. (2016) the difference was less pronounced (76% vs. 95%).

The stability of the previously established "utilitarian" tendencies in the "switch" version of the dilemma has also been shown. Skulmowski et al.

(2014) ran a VR study employing a repeated confrontation paradigm. In their study, each participant faced the same trolley dilemma 10 times from the perspective of the train driver. They could leave the trolley on the current track and kill 10 people, or they could divert it onto the other track using the arrow keys and kill one person instead. The overall percentage of "utilitarian" decisions was 96% and no significant effect of trial number was detected, meaning the high "utilitarian" rate was consistent across multiple trials.

3. Explanation of the "utilitarian tendency" in VR studies

The most striking finding of the VR-based experiments compared to the questionnaire-based studies is the increased tendency to make "utilitarian" decisions in both the "switch" and the "footbridge" versions of the trolley dilemma. It might seem that the reality of VR, and the visual saliency, will make the subjects much less likely to decide to push the person from the footbridge. Watching and hearing the person falling and dying can be far more drastic than reading about the result of one's hypothetical action. And yet, most of the participants of the VR conditions decide to do that anyway. The explanation of this phenomenon that is put forward in the literature is based on Cushman's (2013) dual-processing theory of moral judgment and, by extension, moral action.[3]

According to his model, two processes are responsible for moral judgments; each of these processes is based on an algorithm that evaluates an action. The first process (*action-based*) assigns value directly to an action based on its typical evaluation. For example, pushing someone off the bridge in the "footbridge" dilemma is assigned a negative value because of the typically harmful nature of pushing a person. On the basis of a causal model of the world, the second process (*outcome-based*) assigns value to the outcome of an action in a specific situation. For instance, pushing someone off the bridge is assigned positive value because its positive consequence is that it saves lives, and this outweighs its harmful consequence, which is killing the person who is pushed.

Cushman claims that in general these two processes are complementary, but in certain specific situations they compete. Moral dilemmas, such as the "footbridge" version of the trolley problem, are examples of these kinds of situations. Cushman's version of the dual-process theory is able to explain regularly observed differences in judgments between the "footbridge" and the "switch" versions of the dilemma. People who refrain from pushing the person in the "footbridge" case but decide to pull the lever in the "switch" case do so due to the negative value representation of pushing a person, which is absent in the case of pulling the lever. As a consequence, in the "footbridge" case, the negative value assigned to the action by the action-based process is greater than the negative value assigned to the inaction by

the outcome-based process; this results in a different final moral judgment than in the "switch" case.

An explanation of the "utilitarian tendency" in VR studies that was put forward by both Francis et al. (2016) and Patil et al. (2014) states that "visual saliency highlights the negative outcome associated with inaction and this begins to outweigh the negativity associated with the action itself" (Francis et al., 2016, p. 7) and that "in VR participants could have been more sensitive to outcomes because they witnessed distressing consequences (gory deaths of virtual humans) of their actions" (Patil et al., 2014, p. 104). In accordance with Cushman's model, the increased salience of the negative outcome that is associated with refraining from action causes the outcome-based process of moral evaluation to assign greater negative value to the inaction compared to questionnaires. This results in an increased tendency to make "utilitarian" decisions compared to the verbal mode of presentation in both versions of the dilemma.

4. Objections to the dual-processing explanation of the "utilitarian tendency"

The dual-processing account of the "utilitarian tendency" is in our view unsatisfactory. Three possible objections will be discussed. The first two concern the generalizability of the explanations; the last one concerns its specificity.

Firstly, notice that the difference in the proportion of utilitarian choices between the questionnaire and the VR conditions obtained by Francis et al. (2016) is significantly larger than the difference in Patil et al. (2014) study. In the former, it is about 50%, whereas in the latter it is only 20%. At first glance, this seems problematic because one may expect that the same phenomenon would not produce wildly different results across studies. There are, however, important factors that may be invoked to explain this variability. The studies differ in many respects, the most important being the moral dilemma that was used: the "switch" scenario in Patil et al. (2014) study, and the "footbridge" scenario in the experiments described by Francis et al. (2016). This difference is related to the overall pattern of the results. As we know from previous research on these dilemmas, in the "switch" variant people tend to see pulling a lever as morally acceptable, whereas in the "footbridge" version they are strongly morally opposed to pushing a person off the bridge. This pattern is reflected in the discussed studies. Consider the results of the questionnaire part of both studies. In Patil et al.'s experiment, 75% of the participants decided to pull the lever, whereas in Francis et al.'s study only 10% of the subjects declared that they would push the man off the bridge. The variability of the differences between the VR mode of presentation and the questionnaires could be seen as a ceiling effect: there was no room for Patil et al.'s participants to be more utilitarian in VR than in the

questionnaires because they had already strongly manifested a utilitarian mode of thinking in the questionnaires. That being said, when looking at the results it is not obvious that the same phenomenon is at play in both studies, and we think that a more thorough empirical investigation is required.

Secondly, let us consider the dual-processing explanation in the context of Patil et al. (2014) study, where this explanation was put forward in the first place. The design of the study was such that the participants were presented with more than one moral dilemma (8 to be exact). The important part of the experiment was that the subjects had to witness graphically disturbing consequences of their decisions: people getting hit by the train and bleeding on the tracks afterward. It is natural to speculate that this view could have influenced the participants' subsequent decisions by increasing the negative value assigned to refraining from action on the basis of its consequences and making this negative value greater than the negative value assigned to the action itself, thus promoting the outcome-based process of moral evaluation. This mechanism does not work so well as an explanation in the case of the studies conducted by Francis et al. (2016) (and the following studies that employ the same VR materials and procedure; Francis et al., 2017, 2018, 2019). In these experiments, the subjects were confronted with the dilemma only once, therefore they were unable to modify their behavior for the sake of avoiding a very brutal and visually salient outcome. One can, however, argue that explicit visuals are not necessary for this effect to occur because the VR mode of presentation is sufficient in itself. This dispute cannot be settled on purely theoretical grounds. There is, however, an important lesson to be drawn from this discussion. If the explanation that relates an increased tendency to take "utilitarian" decisions to the salience (visual and nonvisual, see Section 4) of the outcome is correct, it should be possible to influence the magnitude of this phenomenon by changing how the actual or possible negative consequences of inaction are presented to participants.

The final objection relies on the fact that the dual-process account provides strong reasons to think that the VR mode of presentation should affect not only the outcome-based process but also the action-based process of moral evaluation. The reason for this is the immersive capabilities of VR. Recall that the realistic manner of presentation combined with tracking of head movements induces two types of illusions: place illusion and plausibility illusion. The possible action of pushing a man off a bridge should seem more concrete and "real" than an action that is only described verbally. The immersion should make the participants more sensitive to the harmfulness of the pushing (cf. Cushman 2012 for a discussion of aversion to simulated harmful actions), consequently increasing the negative value assigned to the action itself (pushing a person off a bridge) by the action-based process. As a result, in accordance with Cushman's account, an increased number of "deontological" decisions should

be observed. This is clearly not the pattern that was observed in previous studies (Francis et al., 2016), which raises worries about the specificity of the explanation. In other words, the opposite results, namely fewer utilitarian decisions in the VR conditions compared to the questionnaires, would also be perfectly compatible with Cushman's theory on a very general level. This is why the dual-processing explanation seems *ad hoc*. It must be emphasized that this objection does not aim to undermine Cushman's dual-process account of moral judgment in its entirety. Of course, it is possible that the "utilitarian" tendency has nothing to do with the two processes responsible for moral judgments and moral decisions. In our view, it is simply not completely clear what prediction this account makes as far as the mode of presentation is concerned, so caution must be taken when trying to link this phenomenon to existing accounts of moral judgment.

To sum up, the existing account of the "utilitarian" tendency in VR cannot sufficiently explain the available empirical data, especially when one considers the details of, on the one hand, the experimental designs of the studies that it tries to account for, and, on the other hand, the important differences in the findings of these studies. In addition to this, from the theoretical point of view, it is concerning that the findings directly contradict the arguably straightforward predictions of the dual-process theory of moral judgment.

5. Study I

To empirically test the proposed explanations, we conducted an experimental study. The starting point was the design used by Francis et al. (2016), but we introduced several changes in order to investigate the explanations of the utilitarian tendency. In this section, we will focus on the VR part of the study, but appropriate modifications were also made in the questionnaire vignettes.

The basic idea of the explanation is that, due to its visual (and in Francis et al., 2016 case also nonvisual) salience, the VR mode of presentation emphasizes the harmful consequences of refraining from action, thereby causing the outcome-based process of moral evaluation to override the negative value assigned to the action by the action-based process. If this is true, we have two potential ways of influencing the "utilitarian tendency". The first is to manipulate the visual salience of the harmful outcome. Reducing it should result in more "deontological" decisions due to the lower negative value assigned to the inaction by the outcome-based system. The second is to decrease the negative value assigned to the action by the action-based system by changing the type of action that is required to be performed in order to sacrifice the person standing on the bridge. Such a change should be followed by an even more pronounced "utilitarian" trend.

In Francis et al. (2016) experiment, several details related to the audio-visual salience of the VR simulation gave prominence to the harmful

consequences of refraining from action. The first was the fact that the five people on the tracks were visible from the beginning of the main part of the experiment. The second was their behavior (calling for help and shouting), which was aimed directly at the participant.[4] It is natural to think that this would result in making the cost of not acting more salient, especially when the demeanor of the people on the track is compared to that of the man on the footbridge, who was completely silent. In the present study, we thus decided to change the location of the five workers from in front of the footbridge to behind it. In addition, we decided to eliminate the workers' verbal and non-verbal behavior. The expected change in the results in comparison to Francis et al.'s studies is that the "utilitarian" tendency would be decreased.

The second strategy that we employed in the study was to manipulate the type of action that is required to push the person off the bridge. Recall that the basic idea of the dual-processing explanation of the difference in utilitarian tendencies in the "switch" and the "footbridge" cases is that certain aspects of the latter are responsible for the different outcome of the process that produces deontological judgments. In Cushman's version of this account, the crucial difference is the type of action that is required to save the five people: pushing a person typically leads to more harmful consequences than pulling a lever does.

Note that in Francis et al. (2016) studies, the action that was *really* required to save the lives of five people was not *actual* pushing (understood as physically pushing a person) but rather pressing a button or pushing a joystick on a controller that resulted in the person being pushed off the bridge in the virtual environment. This is important from the point of view of Cushman's theory because it is not clear which action's value would be taken into consideration by the action-based process. Is it the action in real life (doing something with a controller) or the proxy action in the simulation (pushing a person)?

This dissociation between real life and proxy actions might be problematic, as previous studies by Cushman et al. (2012) associated aversion to certain actions (which is crucial in the workings of Cushman's action-based process) to not only the perceptual but also the motoric properties of said actions. It has been established that aversion, measured with total peripheral resistance (TPR), is greater when participants enact harmful actions than when they merely witness them or perform motorically similar actions within different contexts. In this view, the aversive responses to the actions in question in the studies by Francis et al. (2016) would be distorted and possibly weakened, as the bodily aspect of the action (using the controller) does not match the perceptual aspect of said action (pushing a person).

A follow-up study by Francis et al. (2017) addressed this issue by alleviating the dissociation between the real life and proxy action in their study design. In Study 2, participants who wanted to sacrifice the person on the footbridge and save the lives of five had to actually push a mannequin that was standing in front of them in real life. This change did not have an effect

on the proportion of "utilitarian" decisions, as it was comparable to the original results. This is surprising, considering the supposed role of action aversion in Cushman's model. The current characterization of the action-based process does not account for why the change of the *actual* action would not produce a different response of said process.

This calls for a closer examination of the workings of the action-based process. According to Cushman's account, it could be speculated that the amount of negative value assigned to the action by the action-based process can be manipulated by changing the real-life and proxy activity that needs to be performed to achieve the person falling off the footbridge. Testing this would enable us to assess the applicability of the dual-process theory of moral judgment to VR studies on moral decisions. We thus decided to include two actions in the experiment. The first was the pushing of a person that as closely as possible resembled the pushing of a real person, given the limitations of the apparatus (motion-tracking controllers). The second was a neutral gesture that is not associated with any harmful consequence: touching the top of their own head. If Cushman's theory correctly describes the factors at play in the VR version of the "footbridge" dilemma, we should expect an increased "utilitarian" tendency toward the neutral gesture compared to the pushing.

5.1. Method

5.1.1. Participants

Participants were volunteers recruited using social media, university announcement boards, and mailing lists. All information about the prerequisites was specified in the announcement: 18+ years old, native Polish speakers, no uncorrected visual impairment, and no history of epilepsy. The study was reviewed by the appropriate Research Ethics Board[5]; 82 subjects participated in the study (50 females, age: $M = 26.45$, $SD = 8.62$); 38 people were assigned to the VR condition and 44 to the questionnaire condition.

5.1.2. Materials

The VR version of the "footbridge" dilemma was closely modeled after the classical formulation of the scenario described by Thomson (1976), but several changes were introduced to improve the plausibility of the scenario. In the virtual environment, the participant stands on a bridge, under which there are railway tracks. In order to enhance the realism of the presentation, a few buildings and vegetation were also included. In the main session of the experiment, a woman is located in front of the virtual avatar.[7] On the tracks behind, five railway workers are placed. In the training session, they are not present and the woman is replaced with a white block (see Figure 1).

The participant is informed by voice narration that a railcar is approaching and that the only way to stop it is to push the woman off

Figure 1. The virtual environment that was used in the experiment, viewed from four perspectives. The subject was placed in the location of the camera in the top-right picture.

the bridge, which results in pressing a special pressure-sensitive plate located under the footbridge.[8] Pressing the plate activates the emergency brake. They are also informed that it is impossible for the woman to survive the fall. This is a substantial change compared to the standard formulation of this dilemma. The modification was motivated by the implausibility of the original scenario: it is impossible for even the largest man to stop the train, and the absurdity of this assumption would be additionally bolstered by the realistic manner of presentation. The simulation was programmed using the Godot game engine, version 3.2 (cf. Santucci et al., 2020).

Depending on the experimental condition, the action of pushing the woman off the footbridge was performed differently. In the "pushing" condition, the participant was able to push the woman using their own hands. Their position in space was tracked, and simple collision detection with the woman's body was implemented. In the "gesture" condition, the collision detection was disabled, and the only way to push the woman off the bridge was for the participant to touch the top of their own head with both hands, which triggered an invisible pushing force.[9]

The textual vignettes that were used in the questionnaire condition resembled the VR version as closely as possible. The translations of the original Polish vignettes can be found in Appendix A. The wording of the prompt emphasized the performative aspect of moral judgment ("What would you do in this situation?").

5.1.3. Procedure

The study was run in a 2 (mode of presentation: VR vs. verbal) x 2 (type of action: pushing vs. gesture) mixed design. The mode of presentation was manipulated between subjects, and the type of action was manipulated within-subject. The order of presentation of two within-subject conditions was randomized in both (VR and questionnaire) versions of the experiment[10]. The questionnaire part of the experiment was run as an online survey. The VR part of the study was conducted in the lab using an Oculus Rift S head-mounted display.[11]

Before the main session of the experiment in the VR condition, the participants had to complete a training session which consisted of several trials in which the participants, depending on the verbal instructions, had to either perform the target actions (pushing and doing a gesture) in the training block, or refrain from doing so. The participant completed the training after six correct trials, which took 7 minutes 15 seconds or more in case of unsuccessfully navigating the training (+75 seconds per additional trial). The main objectives of this lengthy training session were to familiarize the participants with the causal chains that are effective in the simulation (e.g., pushing an object off the bridge > pressing the plate > stopping the train) and to make sure that participants with no prior exposure to VR technology were able to perform the target actions using the controllers. Following Francis et al. (2016), we also recorded participants' heart rate using a Polar OH1+ optical forearm heart rate monitor.

Both of the main trials lasted 75 seconds each, giving a total VR session time of 10 minutes under normal conditions. The time course of an individual trial was as follows. After 5 seconds in the virtual environment, the verbal instruction started playing. In both conditions, the duration of the instruction was 35 seconds. 12 seconds after the instruction ended, the running train became fully visible. The participants had 18 seconds to make a decision, counted from the end of instruction. After that period, the train hit the five people on the tracks, or, if the participant decided to push the woman off the bridge, stopped right before hitting the workers.

Individual personality differences between the participants may be a potential confounding factor in the study. To control for this, Francis et al. (2016) used the Levenson Psychopathy Scale (a two-dimensional model). Due to the lack of a Polish adaptation of this tool, we have decided to use two related measurement tools in the study. Before the experiment, participants were asked to complete two self-report personality inventories. The first was HEXACO-PI-R (Ashton & Lee, 2009; Skimina et al., 2020), which is a 60-item inventory designed to measure six personality factors: Honesty-Humility, Emotionality, Extraversion, Agreeableness, Conscientiousness, and Openness to Experience. The second was TriPM-41 (Patrick, 2010; Pilch et al., 2015), an inventory developed to measure psychopathy in the triarchic model, which identifies three dimensions: boldness, meanness, and disinhibition. The results from these two questionnaires allowed us to check whether the research group from

condition Q is comparable to the VR group and to trace the correlation between personality factors and moral decisions made in VR and in the questionnaire.

After the experiment, participants in the VR condition completed an additional short questionnaire designed to assess their familiarity with computer games and VR technology. In Francis et al. (2016) study, the participants were only asked to assess their gaming experience. In the present study, the participants were questioned about their VR experience separately. They were also asked to evaluate the moral rightness of the decisions made in VR and to indicate whether they would change them after the experiment.

We also introduced an original measure designed to serve as a proxy of the visual saliency of the consequences of the action that results in the woman's death. The basic idea is that for something to be visually salient to the participant, it must be actually viewed by them. If a subject does not look at the five workers on the tracks, it is difficult to argue that they are visually salient. Recall that, in our version of the experiment, the workers are located behind the footbridge, therefore it is up to the participant to turn around and look at them. In the virtual environment, it is possible to check whether the viewport (i.e., the virtual camera) contains some specific object. Thus, we also recorded the cumulative time for which the participant looked at the five workers on the tracks.

5.2. Results

The results concerning moral decisions are presented in Figure 2. In the questionnaire condition, independently of the type of action required, the small minority of participants claimed that they would sacrifice one person to save the lives of five (pushing: 18.18%; gesture: 27.27%). In contrast, almost a half of the participants in the VR condition made the same decision and pushed the woman off the bridge in order to save the workers (pushing: 52.63%; gesture: 55.26%).[12] The analysis where only the decision in the first trial was taken into considerations results in almost the same proportion of utilitarian choices (pushing: 55.0%, gesture: 55.5%).

No effect of the order of presentation was observed in the VR condition (pushing: $\chi^2(1) = 0.095$, $p = 0.758$; gesture: $\chi^2(1) = 0.095$, $p = 0.758$). In the questionnaire condition, no effect of the order of presentation was observed for the pushing scenarios ($\chi^2(1) = 0.129$, $p = 0.720$) but participants opted for "deontological" decision in the gesture condition slightly more often when this version of the dilemma was presented second ($\chi^2(1) = 4.73$, $p = 0.030$).

The results were further analyzed using logistic regression with the mode of presentation, the type of action, and their interaction entered as predictors, and decision (action vs. omission) as a dependent variable. A statistically significant effect of the mode of presentation was observed

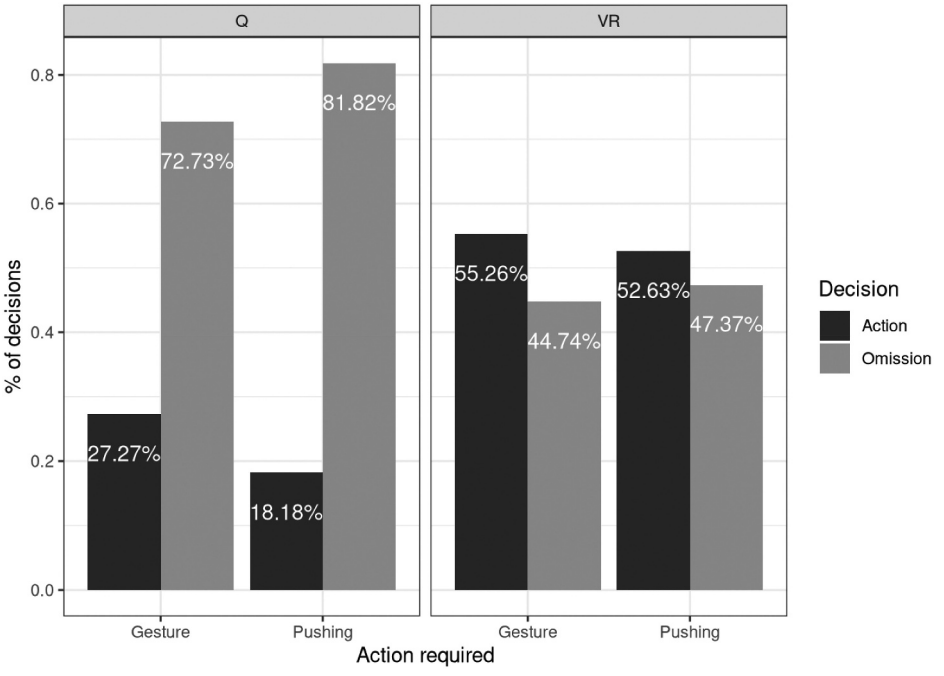

Figure 2. Proportions of "utilitarian" ("action") and "deontological" ("omission") decisions in study I.

($\beta = 1.192$, $p = 0.011$) but not for the type of action ($\beta = -0.523$, $p = 0.312$) and their interaction ($\beta = 0.417$, $p = 0.547$).

We also measured the time between the start of the trial and the moment of decision. Due to the experimental setup, we could only analyze the data for the participants who decided to push the woman off the bridge because the decision to refrain from action was expressed by actually refraining from action and not by pushing a button or choosing an option. No statistically significant differences with regard to the moment of decision were observed between the two types of action in the VR condition (pushing – M = 47.5s, SD = 5.1; gesture – 48.5s, SD = 5.1). We also analyzed the cumulative duration of looking at the five workers. There were no differences between participants who decided to push the woman off the bridge (pushing: $M = 9.1$s., gesture: $M = 10.5$s.) and those who refrained from action (pushing: $M = 14.0$s., gesture: 15.1s.).

Out of the 38 participants in the VR condition, only four declared after the VR session that they would change the decision they had made in the main session of the experiment. No interpretable pattern of responses emerged. There were also no statistically significant differences in *post-hoc* moral judgments between participants who decided to perform an action (pushing: $M = 5.0$, gesture: $M = 4.95$) and those who refrained from acting (pushing: $M = 5.39$, gesture: $M = 5.65$).

In order to investigate the relationships between moral decisions and psychopathic traits, two separate regression models were fitted: one for each within-subject condition. As predictors, we entered the scores from the TriPM-41 inventory for each dimension (boldness, meanness, and disinhibition) and the variable containing information on the between-subjects condition (VR vs. questionnaire). The results are presented in Table 1.

An additional analysis was performed to assess the correlation between prior exposure to video games and VR technology. The logistic regression model did not reveal any statistically significant relationships.[13]

5.3. Discussion

The results of the study do not support the dual-processing explanation of the "utilitarian tendency" that has been observed in VR studies. The main finding is that the proportion of utilitarian decisions is comparable to both studies reported by Francis et al. (2016), even though the salience of the harmful consequences of inaction was supposedly diminished by changes to the experimental materials. The majority of the participants still chose to act in the VR condition, whereas in the questionnaires only a small minority of the subjects decided to sacrifice the person standing on the footbridge. It thus seems that the phenomenon under consideration is yet to be explained.

We also tested the hypothesis that the VR mode of presentation might be responsible for some form of "action bias", especially concerning participants who had had limited exposure to VR technology. The basic idea was that some participants decided to act because they were motivated by curiosity about the capabilities of the new medium. This explanation seemed plausible but turned out to be incorrect as we did not find any statistically significant relationship between the decision that was made and experience with VR or, replicating the results of Francis et al. (2016), video games in general.

Table 1. Two logistic regression models for each condition separately.

	Type of action	
	Pushing (1)	Gesture (2)
Disinhibition	2.215.	−0.104
Boldness	0.984	−0.297
Meanness	−0.368	1.638 *
Mode of presentation (VR)	7.649.	5.551
Disinhibition x VR	0.858	1.895
Boldness x VR	−2.347 *	−1.179
Meanness x VR	−0.868	−2.633 **
Intercept	−7.152	−2.884
No. of observations	82	82
Log Likelihood	−39.511	−44.456
AIC	95.022	104.911

< p< 0.1; *p < 0.05; **p < 0.01; ***p < 0.001

In the study, we also investigated the role of the type of action required to make a "utilitarian" decision in virtual reality; according to Cushman's view, the action-based process of moral evaluation should be sensitive to the difference between neutral actions and typically harmful actions. We did not find any differences in this regard, which indicates that this theory could have limited applicability to results from VR studies.

It is also important to note that in our study hardly anyone changed their decision between trials. According to the discussed explanation of the "utilitarian" tendency, seeing visually salient consequences of the "deontological" decision in the first trial should encourage participants to change their decision in the second trial.[14] This was clearly not the pattern that we observed, regardless of the order of presentation. This finding provides support to the idea that the effect of salience is relatively small and cannot fully account for the "utilitarian" tendency. It is more evidence that the current explanation of this phenomenon is unsatisfactory or at least incomplete.

The present study has several limitations. The first is related to the sample size. If the effect of salience is small, such as that observed in the study conducted by Patil et al. (2014), who obtained a difference of 20pp, then the experimental design that we employed is not likely to detect the effect due to its low power. Note that the proportion of utilitarian decisions was smaller in our study compared to the studies by Francis et al. (2016), but the difference did not reach statistical significance. There are, however, two important points to be made. Firstly, the sample sizes in VR studies are generally low due to organizational and technical difficulties in conducting such experiments on a large scale.[15] Secondly, if the true effect of the visual salience of the consequences is very small and thus undetectable given our experimental design, it is difficult to see how it accounts for the difference between the verbal and the VR modes of stimuli presentation.

The second limitation concerns the possibility that the experimental manipulation of the type of action was ineffective. Recall that we manipulated this factor within-subject: each subject was confronted in random order with two almost identical scenarios that differed in the type of real-world action that was required to save five people. This might be seen as problematic. The action-based process in Cushman's model takes into account the value of an action based on its *typical* evaluation. Here we encounter two issues.

Firstly, how are action types construed in our minds? How general are they, and how are they represented? We assumed that "pushing a person using one's own hands" and "touching the top of one's head" constitute two separate types of action for the needs of the action-based system. We based this assumption on an example given by Cushman himself – pushing a person with your buttocks is different enough from pushing with your hands (Cushman, 2013, p. 282). It is, however, possible that the action-based system operates on a different level of

generality in such a way that both actions fall under the type "pushing a person", which is instantiated by all actions which have similar immediate consequences.

Secondly, under which conditions are the representations of action types triggered? From Cushman and colleagues' work, we know that embodiment plays an important role in this phenomenon. However, it is clear that bodily movements are not necessary for triggering a representation of at least certain types of actions, as verbal descriptions are often sufficient. Now we can see that the actions in the experiment might be treated as "pushing a person" by the action-based process, regardless of how they were physically realized (pushing vs. gesture). The distinction between these two types of actions might be additionally "blurred" by an elaborate training session before the main part of the experiment. It might also be seen as naive to assume that the type of action that is evaluated by the action-based process changes from trial to trial depending on the physical trigger that is required to perform an action in the virtual environment. In other words, if the broad context stays the same (pushing one person off the bridge to save five), there is little to no reason for the action-based process to reevaluate the possibilities that are present. We can partially address this objection by pointing to the fact that when we only analyzed decisions from the first presented dilemma as a between-subjects experimental design (i.e., we completely ignored the potentially invalid second trial), we also did not find any difference in pattern of responses between the "pushing" and "gesture" conditions. We think that both indicated issues (how action types are construed and what triggers their representations) need further empirical investigation.

6. Study 2

One can object to the conclusions we drew from the results of the first study by pointing out that we, in fact, did not directly manipulate the visual (and auditory) saliency of the consequences of the actions. In the second experiment, we followed the original setup of the Francis et al. (2016) study more closely and re-introduced all elements of the experimental simulation that added to the vividity of the scenario. The goal of this follow-up study was to obtain additional evidence that this factor does not substantially increase the tendency to make "utilitarian" decisions in the "footbridge" dilemma.

6.1. Method

6.1.1. Participants

As in Study 1, the participants were volunteers. Pre-screening criteria also remained the same, but those who participated in the first study were unable to take part in the second one. 36 individuals were involved (21 females, 13 males, 2 people chose "Other", age: $M = 24.5$, $SD = 8.32$). The study was reviewed by the Research Ethics Board.[6]

6.1.2. Materials

As pointed out earlier, we re-introduced two elements to the experimental simulation that can be thought to increase the saliency of the harmful consequences of not taking an action. First, the position of the five workers was changed compared to the first study. Instead of behind the footbridge, now they were placed in front of it, and thus they were visible to the participant at all times. Second, following the original experiment, workers were not silent. In the simulation used in this study, after the verbal instruction was finished, the workers started crying for help, directly addressing the subject ("Please help us! You need to do something!").

6.1.3. Procedure

Because the main aim of the experiment was to test whether direct manipulation of salience in VR has an effect on the decision made, this time we decided to omit the questionnaire condition and run the study in a full within-subject design. Only the type of action was manipulated between trials. The order of presentation was randomized. All other details of the experimental procedure remained the same as in the first study, including timing, training session and verbal instructions played to the subjects.

6.2. Results

The results of the follow-up study are presented in Figure 3. In this experiment, about two thirds of the participants chose to push the woman off the bridge (pushing: 63.9%, gesture: 63.9%). An analysis of the responses only from the first trial reveals a similar pattern of response in the pushing

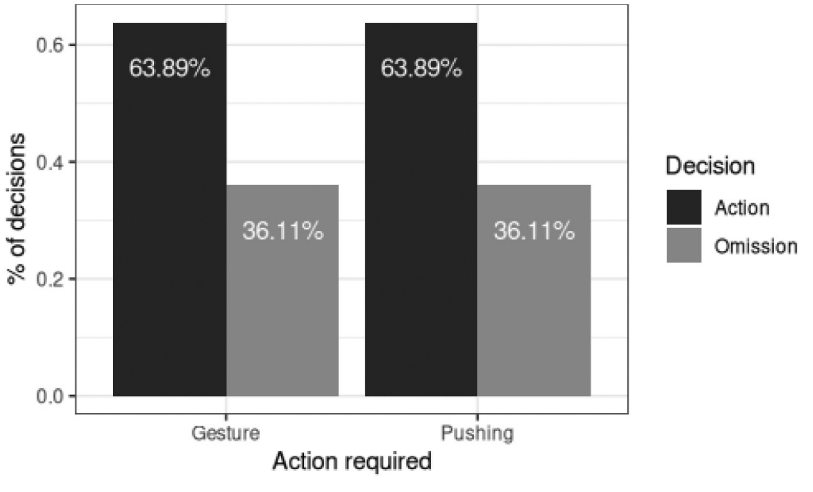

Figure 3. Proportions of "utilitarian" ("action") and "deontological" ("omission") decisions in study 2.

condition (55.5%) but in the gesture condition the proportion of utilitarian responses were considerably higher (83%). The difference, however, was not statistically significant ($\chi^2(1) = 3.273$, $p = 0.070$).

Compared to Study 1, this proportion of "utilitarian" decisions was slightly higher, but the difference did not reach statistical significance for both types of action (pushing: 52.69% vs. 63.9%, vs $\chi^2(1) = 0.962$, $p = 0.327$, gesture: 55.26% vs. 63.9%, $\chi^2(1) = 0.571$, $p = 0.450$). Closer examination of the data with respect to differences between two orders of presentation (pushing-first vs. gesture-first) reveals an interesting pattern in the data. While in the pushing condition, presenting it first resulted in only a slightly higher percentage of the participants deciding to push the woman off the bridge (first: 55.5% vs. second: 72.2%), in the gesture condition the difference was in the other direction and larger (first: 83.3% vs. second: 44.4%). This effect of order was statistically significant ($\chi^2(1) = 5.900$, $p = 0.015$).

6.3. Discussion

The results show that even if we make the consequences of refraining from action as visually and auditorily salient, as in the original Francis et al. (2016) experiment, the proportion of the "utilitarian" decisions does not increase significantly. One of the potential objections to the interpretation of the results of the first study can be thus dismissed. No substantial differences compared to Study 1 provide additional evidence for the inadequacy of the discussed explanation of the disparity of the results obtained using the questionnaire and VR methodology. In contrast to the first study, we have observed a substantial although non-significant effect of order in the pushing condition and significant effect in the gesture condition. This phenomenon calls for further investigation, for example in a more adequately powered replication study. Similarly to Study 1, we did not find any differences related to the type of action required to make a decision in the VR environment.

7. Exploratory meta-analyses

In order to see better how our results fit into the greater picture of research on moral decisions in VR, we conducted two small, exploratory meta-analyses of the existing studies. In the analyses, we only included the studies that are directly comparable with the experiments reported here.[16] The first one concerns the proportion of the participants that made utilitarian decision in the VR versions of the experiments, the second one concerns differences in responses between questionnaire and VR modes of presentation. For both analyses, we used *metafor* R package (Viechtbauer, 2010) and fitted a meta-analytical models to the results of the studies discussed in Section 1 as well as two experiments reported in this article.

For the first meta-analysis (k = 16 studies), we have chosen the raw proportion of people making a utilitarian decision as a measure of effect size, and entered the type of scenario ("switch" vs. "footbridge") as a predictor. The overall meta-analytic effect was estimated to be 0.597 [95% CI: 0.551–0.642]. The "switch" scenario was associated with the increase in tendency to sacrifice a person for a greater number of lives ($b = 0.341$, $p < 0.001$). Figure 4 shows that both our studies sit very close to the estimated effect size. According to the fitted model, the effects display a surprisingly small level of heterogeneity, as indicated by the low I^2 value (18.23%) and non-significant results of the test for residual heterogeneity ($Q(14) = 12.65$, $p = 0.554$).

In the second meta-analysis, we used only the studies where the VR part of the experiment was accompanied by matched questionnaire condition (k = 11 studies). Using Log Odds Ratio as a measure of effect size, we assessed whether the subjects have an increased tendency to make utilitarian decisions in VR compared to the questionnaires. Figure 5 shows the results of the analysis. In all

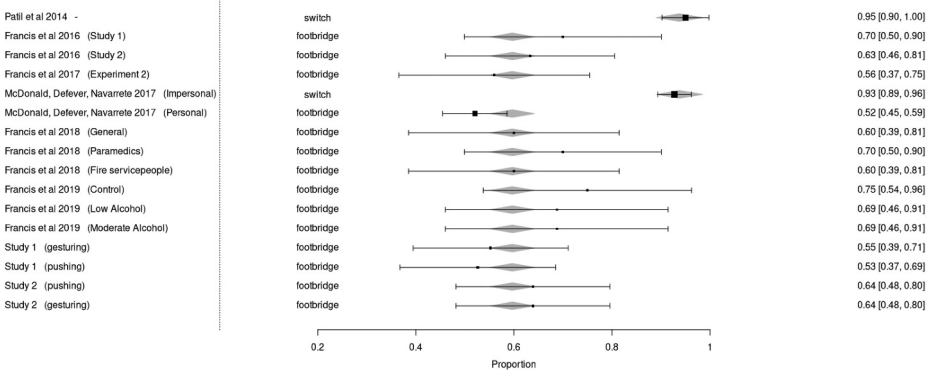

Figure 4. Forest plot of the proportions of utilitarian decisions.

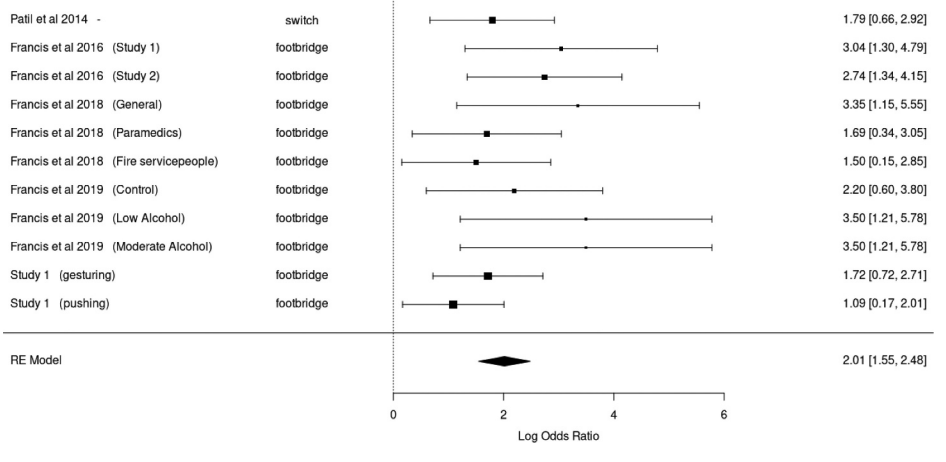

Figure 5. Forest plot of the Log odds Ratios for the effect of mode of presentation (VR vs. questionnaire) on the tendency to make "utilitarian" decisions.

the studies, the effect of mode of presentation reached the level of statistical significance and the effect sizes appear to by rather homogenous ($I^2 = 18.02\%$, Q (10) = 12.13, $p = 0.276$). The overall effect was high according to conventional rules of interpretation (LOR: 2.01 [95% CI: 1.55–2.48], see Chen et al., 2010). All in all, the results of the meta-analysis indicate that the effect under consideration is robust.

8. Conclusions

The main positive contribution of the present studies is that the effect of VR presentation in the "footbridge" dilemma discovered by Francis et al. (2016) was replicated. Both of our experimental studies also strongly suggest that the effect is quite robust: the phenomenon under consideration was present and comparable in size to the original study, despite multiple changes in the experimental setup of the first study that were designed to decrease the magnitude of the effect. In the second study, we intentionally bolstered visual and auditory salience of the harmful consequences of refraining from action and did not obtain a significant change in the subjects' responses. This provides additional evidence against the proposed explanation of "utilitarian" tendency in VR. The observation concerning robustness is further strengthened by the results of two exploratory meta-analyses. Moreover, they have shown how the results of the present studies fit into a broader picture of the phenomenon. The first study also provided evidence that this effect is not influenced by prior gaming or VR experience.

We think that, despite their limitations, the present studies show that existing explanations of the "utilitarian" tendency are unsatisfactory. It is still unclear which features of VR are responsible for this phenomenon. It should be pointed out that even though some studies report differences between textual and VR modes of presentation, in other studies this effect was not detected. For example, in their study on the ethics of autonomous vehicles, where participants were to choose from a car driver perspective who or what to kill on a road in a forced-choice scenario, Sütfeld and his research team (2019) reported that modality (VR vs 2D desktop monitor) did not influence any of the investigated biases and level of abstraction (naturalistic vs text-based) influenced only one aspect of the decision, namely *elderly bias* It is thus reasonable to hypothesize that the effect of VR presentation could, on the one hand, depend on the task; or, on the other hand, on the experimental design. If this is the case, further research should concentrate on varying the details of the experimental setup and the task. Until the "utilitarian" tendency is better understood and the mechanism behind it accounted for, also the issue of generalizability of the results of

VR studies to real-life moral decision-making stands unresolved. While there exist data pointing to an analogous to the text vignettes-VR discrepancy between hypothetical moral judgments and actual real fife moral decision-making (Bostyn et al., 2018), we still do not know enough about the two phenomenons to equate or distinguish them.

Notes

1. This received view was met with criticism in the most recent literature concerning VR research on moral decisions and moral judgment. Ramirez (2019) points out that the types of situations presented in VR studies poorly match real-world contexts in which actual decisions were made. This is an issue for the generalizability of the results but is not crucial for the present studies as they are mainly aimed at testing the explanation of existing experimental results. We do agree with much of Ramirez's argumentation, but for the purpose of the experimental investigation, it was important to stay close to the original scenario. However, some efforts to improve the context-realism of the situation were taken; for instance, we introduced a semi-plausible mechanism of stopping the train.
2. There are studies that use only one method of presentation: text-based questionnaires or VR. While they do match the general pattern of the discrepancy between the methods: fewer decisions to push in questionnaires (e.g., Côté et al., 2013; Greene et al., 2001; Hauser et al., 2007; Lanteri et al., 2008; Lombrozo, 2009), more decisions to push in VR (McDonald et al., 2017), they do not provide the best account of this phenomenon, because of the variation of the scenarios they employ.
3. We do note that this is not the only explanation that has been put forward in the literature. In Francis et al. (2017) the "utilitarian tendency" is also discussed from the point of the frames of reference accounts (see Section 1). According to this explanation, considerations of certain self-relevant consequences associated with the egocentric frame of reference, which is characteristic for choice but not judgment, drive the discussed effect. This explanation, however, is presented as a compliment to rather than a replacement for the explanation discussed in this paper, which is why we focus on the latter.
4. In the discussed paper by Francis et al. they had analyzed the effect of auditory stimuli (calling for help) on the participants' decision by comparing responses made before and after the onset of the voice recording. They did not find statistically significant differences.
5. Rector's Committee for the Ethics of Research Involving Human Participants, University of Warsaw, decision number: 84/2021.
6. Rector's Committee for the Ethics of Research Involving Human Participants, University of Warsaw, decision number: 144/2022.
7. Our choice of the gender of the person to be sacrificed was random. Even though previous studies used males for that role (Francis et al., 2016), we had not expected this difference to have an effect on our results. We presented the dilemmas only twice to each participant, and the gender bias in such scenarios appears only after a couple of trials in a repeated measures design (Skulmowski et al., 2014).
8. Because the experimental setup was slightly different compared to both the standard footbridge dilemma and Francis et al.'s study, we came up with original instructions for the participants. The specific wording (translated into English) used in the study was.

   "In thirty seconds, a train will pass on the tracks under the footbridge you are standing on. Behind you on the tracks is a group of five people. There is a woman in front of you who you can throw off the footbridge by pushing her/touching the

top of your own head with both hands. If the woman is thrown off the footbridge, she will not survive the fall. The falling woman will activate a special lock that will automatically stop the moving train. The only way to save the five people is to throw the woman off the footbridge."

9. The position of the hands was tracked *via* Oculus Rift S controllers that participants held in their hands. A spherical collision shape was assigned to each controller. Then, during the trials, we programmatically checked a) in the "pushing" condition whether one of the spheres collides with the body of the woman on the bridge or b) in the "gesture" condition whether both spheres intersect with the area above the head of the participant. Haptic feedback was not incorporated into the experiment.
10. It was pointed out in the literature that presenting iterated within-subject studies might be susceptible to diminished realism in later trials (Ramirez, 2019). One way to counteract this problem is to vary the order of presentation, which we implemented in our study. Additionally, in Section 4.2 we also present results that take into account only the first experimental trial.
11. Technical specification: 1280×1440(per eye) resolution, 80 Hz refresh rate; head tracking was accomplished using five integrated cameras.
12. It is worth noting that, generally speaking, the participants did not change their decisions between trials. In the pushing-first group, only two people decided to change their decision. One participant decided to sacrifice the woman in the second trial, and one decided to refrain from action. In the gesture-first group, three people changed their behavior: two participants decided not to push the woman in the second trial, and one decided to push her off the bridge.
13. We entered the following variables into the model: how often the participant plays video games (5-point scale); how many hours they play a week; how many games they play in a year; whether they have previous experience with VR; and approximately how many hours they had spent in VR before the experiment.
14. As one of the reviewers pointed out, this result might be caused by consistency bias, which could manifest itself here as a tendency to maintain the first decision in the second trial. We think that this explanation is plausible, and that is the reason we also included the analyses that take into account only the first trial. It is however worth noting that another bias of the opposite direction might be at play here. One of the response biases that might be affecting the results of the study is the desire of the participants to be "good experimental subjects". In this case, the participants might reason that we as experimenters expect different behavior in the second trial with a different way of decision-making (cf Nichols & Maner, 2008). We think that these kinds of effects are potentially dangerous to the generalizability of the results and should be more carefully taken into account in the experimental design.
15. The only exception known to the authors is a study by Navarrete et al. (2012), who recruited 365 participants (data were analyzed for 293 of them). All other relevant studies on moral actions in virtual environments had sample sizes in the range of 30–50 participants.
16. In general, we used a single decision (not a single participant) as a unit of observation in order to achieve consistency across studies. This difference is not relevant in most cases except for the study by Patil et al., who used a repeated measure design where 20 participants were confronted four times with the same dilemma, which results in 80 decisions in total.

Disclosure statement

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

Funding

Funded by the National Science Centre, Poland, under PRELUDIUM [2018/29/N/HS1/02863].

ORCID

Bartosz Maćkiewicz http://orcid.org/0000-0002-9460-5742
Joanna Andrusiewicz http://orcid.org/0000-0002-9036-4769

References

Ahleniusa, H., & Tännsjö, T. (2012). Chinese and Westerners respond differently to the trolley dilemmas. *Journal of Cognition and Culture*, *12*(3–4), 195–201. https://doi.org/10.1163/15685373-12342073

Ashton, M. C., & Lee, K. (2009). An investigation of personality types within the HEXACO personality framework. *Journal of Individual Differences*, *30*(4), 181–187. https://doi.org/10.1027/1614-0001.30.4.181

Bostyn, D. H., Sevenhant, S., & Roets, A. (2018). Of mice, men, and trolleys: Hypothetical judgment versus real-life behavior in trolley-style moral dilemmas. *Psychological Science*, *29*(7), 1084–1093. https://doi.org/10.1177/0956797617752640

Chen, H., Cohen, P., & Chen, S. (2010). How big is a big odds ratio? Interpreting the magnitudes of odds ratios in epidemiological studies. *Communications in Statistics—Simulation and Computation®*, *39*(4), 860–864. https://doi.org/10.1080/03610911003650383

Cima, M., Tonnaer, F., & Hauser, M. D. (2010). Psychopaths know right from wrong but don't care. *Social Cognitive and Affective Neuroscience*, *5*(1), 59–67. https://doi.org/10.1093/scan/nsp051

Côté, S., Piff, P. K., & Willer, R. (2013). For whom do the ends justify the means? Social class and utilitarian moral judgment. *Journal of Personality and Social Psychology*, *104*(3), 490–503. https://doi.org/10.1037/a0030931

Cushman, F. (2013). Action, outcome, and value: A dual-system framework for morality. *Personality and Social Psychology Review*, *17*(3), 273–292. https://doi.org/10.1177/1088868313495594

Cushman, F., Gray, K., Gaffey, A., & Mendes, W. B. (2012). Simulating murder: The aversion to harmful action. *Emotion*, *12*(1), 2. https://doi.org/10.1037/a0025071

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

Cushman, F., & Young, L. (2011). Patterns of moral judgment derive from nonmoral Psychological representations. *Cognitive Science*, *35*(6), 1052–1075. https://doi.org/10.1111/j.1551-6709.2010.01167.x

Cushman, F., Young, L., & Hauser, M. (2006). The role of conscious reasoning and intuition in moral judgment: Testing three principles of harm. *Psychological Science*, *17*(12), 1082–2089. https://doi.org/10.1111/j.1467-9280.2006.01834.x

Davidson, D. (2001). How is weakness of the will possible? In D. Davidson (Ed.), *Essays on actions and events* (2nd ed., first published 1980, pp. 21–42). Oxford University Press.

Edmonds, D. (2013). *Would you kill the fat man? The trolley problem and what your answer tells us about right and wrong*. Princeton University Press.

Foot, P.(1978). Original work published 1967, The problem of abortion and the doctrine of the double effect. In *Virtues and vices: And other essays in moral philosophy* (pp. 19–32). Blackwell. Oxford, UK.

Francis, K. B., Gummerum, M., Ganis, G., Howard, I. S., & Terbeck, S. (2018). Virtual morality in the helping professions: Simulated action and resilience. *British Journal of Psychology*, *109*(3), 442–465. https://doi.org/10.1111/bjop.12276

Francis, K. B., Gummerum, M., Ganis, G., Howard, I. S., & Terbeck, S. (2019). Alcohol, empathy, and morality: Acute effects of alcohol consumption on affective empathy and moral decision-making. *Psychopharmacology (Berl)*, *236*(12), 3477–3496. https://doi.org/10.1007/s00213-019-05314-z

Francis, K. B., Howard, C., Howard, I. S., Gummerum, M., Ganis, G., Anderson, G., Terbeck, S., & Wan, X. (2016). Virtual morality: Transitioning from moral judgment to moral action? *PloS One*, *11*(10), e0164374. https://doi.org/10.1371/journal.pone.0164374

Francis, K. B., Terbeck, S., Briazu, R. A., Haines, A., Gummerum, M., Ganis, G., & Howard, I. S. (2017). Simulating moral actions: An investigation of personal force in virtual moral dilemmas. *Scientific Reports*, *7*(1), 1–11. https://doi.org/10.1038/s41598-017-13909-9

Gold, N., Colman, A. M., & Pulford, B. D. (2014). Cultural differences in responses to real-life and hypothetical trolley problems. *Judgment and Decision Making*, *9*(1), 65–76. https://doi.org/10.1017/S193029750000499X

Greene, J. D., Morelli, S. A., Lowenberg, K., Nystrom, L. E., & Cohen, J. D. (2008). Cognitive load selectively interferes with utilitarian moral judgment. *Cognition*, *107*(3), 1144–1154. https://doi.org/10.1016/j.cognition.2007.11.004

Greene, J. D., Nystrom, L. E., Engell, A. D., Darley, J. M., & Cohen, J. D. (2004). The neural bases of cognitive conflict and control in moral judgment. *Neuron*, *44*(2), 389–400. https://doi.org/10.1016/j.neuron.2004.09.027

Greene, J. D., Sommerville, R. B., Nystrom, L. E., Darley, J. M., & Cohen, J. D. (2001). An fMRI investigation of emotional engagement in moral judgment. *Science*, *293*(5537), 2105–2108. https://doi.org/10.1126/science.1062872

Haidt, J. (2001). The emotional dog and its rational tail: A social intuitionist approach to moral judgment. *Psychological Review*, *108*(4), 814–834. https://doi.org/10.1037/0033-295x.108.4.814

Hauser, M., Cushman, F., Young, L., Kang-Xing Jin, R., & Mikhail, J. (2007). A dissociation between moral judgments and justifications. *Mind & Language*, *22*(1), 1–21. https://doi.org/10.1111/j.1468-0017.2006.00297.x

Huebner, B., Dwyer, S., & Hauser, M. (2009). The role of emotion in moral psychology. *Trends in Cognitive Sciences*, *13*(1), 1–6. https://doi.org/10.1016/j.tics.2008.09.006

Kahneman, D. (2011). *Thinking, fast and slow*. Farrar, Straus and Giroux.

Koenigs, M., Young, L., Adolphs, R., Tranel, D., Cushman, F., Hauser, M., & Damasio, A. (2007). Damage to the prefrontal cortex increases utilitarian moral judgements. *Nature*, *446*(7138), 908–911. https://doi.org/10.1038/nature05631

Krumpal, I. (2013). Determinants of social desirability bias in sensitive surveys: A literature review. *Quality & Quantity*, *47*(4), 2025–2047. https://doi.org/10.1007/s11135-011-9640-9

Lanteri, A., Chelini, C., & Rizzello, S. (2008). An experimental investigation of emotions and reasoning in the trolley problem. *Journal of Business Ethics*, *83*(4), 789–804. https://doi.org/10.1007/s10551-008-9665-8

Lombrozo, T. (2009). The role of moral commitments in moral judgment. *Cognitive Science*, *33*(2), 273–286. https://doi.org/10.1111/j.1551-6709.2009.01013.x

McDonald, M. M., Defever, A. M., & Navarrete, C. D. (2017). Killing for the greater good: Action aversion and the emotional inhibition of harm in moral dilemmas. *Evolution and Human Behavior*, *38*(6), 770–778. https://doi.org/10.1016/j.evolhumbehav.2017.06.001

Mele, A. R. (2022). Weakness of will. In M. Vargas & J. Doris (Eds.), *The oxford handbook of moral psychology* (pp. 349–363). Oxford University Press.

Mikhail, J. (2002). Aspects of the theory of moral cognition: Investigating intuitive knowledge of the prohibition of intentional battery and the principle of double effect. *Georgetown Public Law Research Paper*. 762385.

Moore, A. B., Clarke, B. A., & Kane, M. J. (2008). Who shalt not kill? Individual differences in working memory capacity, Executive control, and moral judgment. *Psychological Science*, *19*(6), 549–557. https://doi.org/10.1111/j.1467-9280.2008.02122.x

Nakamura, K. (2012). The footbridge dilemma reflects more utilitarian thinking than the trolley dilemma: Effect of number of victims in moral dilemmas. *Proceedings of the Thirty-fourth Annual Conference of the Cognitive Science Society*. Paper Series. http://ssrn.com/abstract=762385

Navarrete, C. D., McDonald, M. M., Mott, M. L., & Asher, B. (2012). Virtual morality: Emotion and action in a simulated three-dimensional "trolley problem". *Emotion*, *12*(2), 364. https://doi.org/10.1037/a0025561

Nichols, A., & Maner, J. (2008). The good-subject effect: Investigating participant demand characteristics. *The Journal of General Psychology*, *135*(2), 151–165. https://doi.org/10.3200/GENP.135.2.151-166

O'Neill, P., & Petrinovich, L. (1998). A preliminary cross-cultural study of moral intuitions. *Evolution and Human Behavior*, *19*(6), 349–367. https://doi.org/10.1016/S1090-5138(98)00030-0

Orsi, F. (2012). Moral judgment, sensitivity to reasons, and the multi-system view. *Baltic International Yearbook of Cognition, Logic and Communication*, *7*(1). https://doi.org/10.4148/biyclc.v7i0.1778

Parsons, T. D. (2015). Virtual reality for enhanced ecological validity and experimental control in the clinical, affective and social neurosciences. *Frontiers in Human Neuroscience*, *9*, 660. https://doi.org/10.3389/fnhum.2015.00660

Patil, I., Cogoni, C., Zangrando, N., Chittaro, L., & Silani, G. (2014). Affective basis of judgment-behavior discrepancy in virtual experiences of moral dilemmas. *Social Neuroscience*, *9*(1), 94–107. https://doi.org/10.1080/17470919.2013.870091

Patrick, C. J. (2010). *Operationalizing the triarchic conceptualization of psychopathy: Preliminary description of brief scales for assessment of boldness, meanness, and disinhibition* [Unpublished test manual]. Florida State University.

Petrinovich, L., O'Neill, P., & Jorgensen, M. (1993). An empirical study of moral intuitions: Toward an evolutionary Ethics. *Journal of Personality and Social Psychology*, *64*(3), 467–478. https://doi.org/10.1037/0022-3514.64.3.467

Pilch, I., Sanecka, E., Hyla, M. & Atłas, K. (2015). The Polish adaptation of the TriPM scale measuring psychopathy. *Psychologia Społeczna*, *35*, 435–454. https://doi.org/10.7366/1896180020153506

Ramirez, E. J. (2019). Ecological and ethical issues in virtual reality research: A call for increased scrutiny. *Philosophical Psychology*, *32*(2), 211–233. https://doi.org/10.1080/09515089.2018.1532073

Randall, D. M., & Fernandes, M. F. (1991). The social desirability response bias in ethics research. *Journal of Business Ethics*, *10*(11), 805–817. https://doi.org/10.1007/BF00383696

Rovira, A., Swapp, D., Spanlang, B., & Slater, M. (2009). The use of virtual reality in the study of people's responses to violent incidents. *Frontiers in Behavioral Neuroscience*, *3*, 59. https://doi.org/10.3389/neuro.08.059.2009

Sanchez-Vives, M. V., & Slater, M. (2005). From presence to consciousness through virtual reality. *Nature Reviews Neuroscience*, *6*(4), 332–339. https://doi.org/10.1038/nrn1651

Santucci, F., Frenguelli, F., De Angelis, A., Cuccaro, I., Perri, D., & Simonetti, M. (2020). An Immersive Open Source Environment Using Godot. In O. Gervasir (Ed.), *Computational Science and Its Applications – ICCSA 2020. ICCSA 2020. Lecture Notes in Computer Science* (Vol. 12255). Springer. https://doi.org/10.1007/978-3-030-58820-5_56

Schwitzgebel, E., & Cushman, F. (2012). Expertise in moral reasoning? Order effects on moral judgement in professional philosophers and non-philosophers. *Mind & Language*, *27*(2), 135–153. https://doi.org/10.1111/j.1468-0017.2012.01438.x

Singer, P. (2005). Ethics and intuitions. *The Journal of Ethics*, *9*(3–4), 331–352. https://doi.org/10.1007/s10892-005-3508-y

Skimina, E., Strus, W., Cieciuch, J., Szarota, P., & Izdebski, P. K. (2020). Psychometric properties of the Polish versions of the HEXACO-60 and the HEXACO-100 personality inventories. *Current Issues in Personality Psychology*, *8*(3), 255–278.

Skulmowski, A., Bunge, A., Kaspar, K., & Pipa, G. (2014). Forced-choice decision-making in modified trolley dilemma situations: A virtual reality and eye tracking study. *Frontiers in Behavioral Neuroscience*, *8*, 426. https://doi.org/10.3389/fnbeh.2014.00426

Slater, M. (2009). Place illusion and plausibility can lead to realistic behaviour in immersive virtual environments. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *364*(1535), 3549–3557. https://doi.org/10.1098/rstb.2009.0138

Slater, M., & Sanchez-Vives, M. V. (2016). Enhancing our lives with immersive virtual reality. *Frontiers in Robotics and AI*, *3*, 74. https://doi.org/10.3389/frobt.2016.00074

Slater, M., Spanlang, B., & Corominas, D. (2010). Simulating virtual environments within virtual environments as the basis for a psychophysics of presence. *ACM Transactions on Graphics (TOG)*, *29*(4), 1–9.

Sütfeld, L. R., Ehinger, B. V., König, P., & Pipa, G. (2019). How does the method change what we measure? Comparing virtual reality and text-based surveys for the assessment of moral decisions in traffic dilemmas. *PLoS ONE*, *14*(10), e0223108.

Tassy, S., Oullier, O., Ducols, Y., Coulon, O., Mancini, J., Deruelle, C., Attarian, S., Felician, O., & Wicker, B. (2012). Disrupting the right prefrontal cortex alters moral judgment. *Social Cognitive and Affective Neuroscience*, *7*(3), 282–288. https://doi.org/10.1093/scan/nsr008

Tassy, S., Oullier, O., Mancini, J., & Wicker, B. (2013). Discrepancies between judgment and choice of action in moral dilemmas. *Frontiers in Psychology*, *4*, 250. https://doi.org/10.3389/fpsyg.2013.00250

Tenenbaum, S. (2010). *Akrasia and irrationality in a companion to the Philosophy of action* (T. O'Connor, C. Sandis, Ed.) Blackwell Publishing Ltd.

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

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software*, *36*(3), 1–48. https://doi.org/10.18637/jss.v036.i03

Waldmann, M. R., & 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. https://doi.org/10.1111/j.1467-9280.2007.01884.x

Appendix A – The translations of the original Polish vignettes (the questionnaire condition)

The "pushing" condition

Imagine you are standing on a footbridge under which there are railway tracks. Behind you, on the tracks, there is a group of five railway workers, in front of you, on the footbridge, there is a woman. Due to the noise, the workers cannot hear the approaching railcar. There is a way, however, to save these people's lives. There is a special button under the footbridge. If pressed, the safety lock will activate and the railcar will stop before it hits the group of people on the other side of the footbridge. However, you won't be able to get down from the footbridge and activate the button yourself. The only way to activate the blockade is to throw the woman standing in front of you off the footbridge. However, the footbridge is so high that the fall will certainly cause the death of a woman.

What would you do in this situation?

- I push the woman and throw her off the footbridge.
- I refrain from action

The "gesture" condition

Imagine you are standing on a footbridge under which there are railway tracks. Behind you, on the tracks, there is a group of five railway workers, in front of you, on the footbridge, there is a woman. Due to the noise, the workers cannot hear the approaching railcar. However, there is a way to save these people's lives. There is a special button under the footbridge. If pressed, the safety lock will activate and the railcar will stop before it hits the group of people on the other side of the footbridge. However, you won't be able to get down from the footbridge and activate the button yourself. The only way to activate the blockade is to throw the woman standing in front of you off the footbridge. However, the footbridge is so high that the fall will certainly cause the death of a woman.

The woman is standing behind the glass and you can't touch her. However, you can perform a special gesture that activates the automatic garbage removal system from the footbridge. Activating it will throw the woman off the footbridge. This gesture is to touch the top of your own head with both hands.

What would you do in this situation?

- I touch the top of my head and throw the woman off the footbridge
- I refrain from action