

Article

Murder on the VR Express: Studying the Impact of Thought Experiments at a Distance in Virtual Reality

Andrew Kissel ^{1,*} , Krzysztof J. Rechowicz ²  and John B. Shull ²¹ Philosophy and Religious Studies, Old Dominion University, Norfolk, VA 23529, USA² Virginia Modeling, Analysis, and Simulation Center, Old Dominion University, Suffolk, VA 23435, USA; krechowi@odu.edu (K.J.R.); jshull@odu.edu (J.B.S.)

* Correspondence: akissel@odu.edu

Abstract: Hypothetical thought experiments allow researchers to gain insights into widespread moral intuitions and provide opportunities for individuals to explore their moral commitments. Previous thought experiment studies in virtual reality (VR) required participants to come to an on-site laboratory, which possibly restricted the study population, introduced an observer effect, and made internal reflection on the participants' part more difficult. These shortcomings are particularly crucial today, as results from such studies are increasingly impacting the development of artificial intelligence systems, self-driving cars, and other technologies. This paper explores the viability of deploying thought experiments in commercially available in-home VR headsets. We conducted a study that presented the *trolley problem*, a life-and-death moral dilemma, through SideQuestVR, a third-party website and community that facilitates loading applications onto Oculus headsets. Thirty-three individuals were presented with one of two dilemmas: (1) a decision to save five lives at the cost of one life by pulling a switch and (2) a decision to save five lives at the cost of one life by pushing a person onto train tracks. The results were consistent with those of previous VR studies, suggesting that a "VR-at-a-distance" approach to thought experiments has a promising future while indicating lessons for future research.

Keywords: virtual reality; moral dilemma; philosophy; social issues



Citation: Kissel, A.; Rechowicz, K.J.; Shull, J.B. Murder on the VR Express: Studying the Impact of Thought Experiments at a Distance in Virtual Reality. *Societies* **2023**, *13*, 69. <https://doi.org/10.3390/soc13030069>

Academic Editor: Gregor Wolbring

Received: 25 July 2022

Revised: 1 March 2023

Accepted: 5 March 2023

Published: 14 March 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

There is a long history in philosophy of using imagined, hypothetical scenarios called *thought experiments* [1] to aid in philosophical methodology. Despite their pervasiveness, their specific uses have varied. For example, Lucretius used the thought experiment of trying to throw a spear at the boundary of the universe to defend his Epicurean position that the universe is infinite [2]. Here, Lucretius understood the process of reflecting on the thought experiment to reveal something about the nature of reality [3]. By contrast, contemporary experimental philosophers use thought experiments to collect widespread intuitions from non-philosophers in order to supplement philosophical theorizing [4]. Still others use thought experiments as a form of conceptual analysis [5].

Since they are often presented as imagined narratives, thought experiments are also an effective way to illustrate ideas and introduce non-philosophers to complex philosophical ideas [1,6]. For example, a 2017 episode of the TV show *The Good Place* grappled with the implications of the famous "trolley problem" thought experiment [7]. The trolley problem has numerous variations. Its classical presentation involves a simple dilemma: Would you pull a switch to prevent a train from killing five people but, in the process, kill one person [8,9]? By reflecting on this scenario, philosophers and non-philosophers have explored classical ethical principles such as "Do not kill" and "Act so as to maximize the greatest happiness for the greatest number of people".

The trolley problem is not the only thought experiment reaching the public eye. The finale of the popular Marvel television show *WandaVision* included a discussion of

the “ship of Theseus” thought experiment, which deescalated the conflict between the titular Vision and his doppelganger [10]. Even classic movies such as *The Matrix* are based on the thinking of philosophers such as Descartes and Zhuangzi, as well as “brain-in-a-vat” thought experiments, plausibly construing these media as thought experiments themselves [11]. Thought experiments can, thus, be used to make complex philosophical ideas accessible to a broader audience, an audience that has shown a keen interest in embracing them. Reflection on popular interpretations of thought experiments has even turned back around to provide philosophically rich insights [12,13].

The lesson is that broader public engagement with accessible thought experiments can lead to benefits for the audience, as well as for researchers. By presenting the trolley problem and its variations moral psychologists purport to investigate people’s moral judgments [14–17]. Nevertheless, the use of thought experiments as research tools has not been without criticism [11]. In many studies, the trolley problem is presented as a tool for gathering information about people’s actual moral judgments if were to find themselves in the presented dilemma [18–21]. Himmelreich and Cohen (2021) called this the “model view”, as it takes the trolley problem to represent “actual dilemma situations just like models in the sciences represent phenomena of interest” [20].

The model view runs into difficulties on at least two fronts. First, it is unclear whether judgments occurring in “cool” moments in a laboratory will match judgments in an actual moral dilemma. For example, the presentation mode is known to affect moral judgments [22–24]. Second, even if thought experiments reflect moral judgments, they may not reflect moral behavior, as the two acts can come apart [25–28]. It is, thus, difficult to determine whether theoretical reflections on thought experiments in the laboratory reflect how people would *actually* behave were the dilemma to occur in reality [29]. Put simply, sitting in a lab with a researcher is not how moral decisions are usually made in real life. We may be learning more about how people react in laboratory settings than how people make moral decisions.

The ecological validity of such studies is particularly pressing, as the results are already referenced in the development of self-driving cars and other AI machines [18,30]. However, there is debate about the appropriateness of the practice [31]. As one particularly prominent illustration, MIT’s Moral Machine Project uses an online crowd-sourced approach in order to gather moral intuitions from various trolley problem variations [30]. They write, “Even if ethicists were to agree on how autonomous vehicles should solve moral dilemmas, their work would be useless if citizens were to disagree with their solution, and thus opt out of the future that autonomous vehicles promise in lieu of the status quo”. The claim seems to be that, by gathering data about how people make moral judgments, engineers will be better placed to design autonomous vehicles that people would use. Despite the crowd-sourced approach, the Moral Machine Project still relies on abstract reasoning about the situation rather than creating an immersive environment where moral actions can emerge. Typical scenarios are presented as static images. Featureless red figures represent people. Arrows and images of skulls indicate possible options and the adverse effects of those options. As such, it is susceptible to the worries of the model view.

The shortcomings of using thought experiments as models for moral reasoning and behavior are explored in the eponymous episode of “The Good Place”. When presented with the trolley problem, Michael (a central protagonist) is unsure of what he would do were the situation real, complaining, “It’s just so theoretical, you know?”. If Michael is right that thought experiments are *too* hypothetical, then they might not give researchers very good data about people’s moral judgments. Ultimately, Michael “fixes” the problem by using his magic powers to actualize the trolley problem dilemma (to humorous, albeit gory, results).

Of course, that option is off the table for real-world researchers [32,33]. Instead, some investigators have turned to virtual reality (VR) to increase their studies’ ecological validity [27,28,34–38]. The rationale seems to be that the contextual information provided by VR, through its ability to invoke feelings of “presence”, will make it more likely that

people will respond to the trolley problem in the way that they would if the scenario was actual [37]. “Presence” here describes the subjective experience of “being there” in the virtual environment [39]. By making users feel as if they were in the presented environment, the hope is that elicited moral behaviors will more closely track real-world behavior. For example, Niforatos et al. argued that VR studies provide more ecologically valid data on moral behaviors compared to judgment-based pen-and-paper alternatives [38]. Similarly, Patil et al. recognized a discrepancy between user responses to judgment-based pen-and-paper versions of the trolley problem and behaviorally based VR versions of the trolley problem, arguing that the latter better reflects actual moral behavior [27]. As a general trend, the presentational qualities of VR are being used to focus studies on moral behavior, as opposed to moral judgment, with the argument that this will lead to increased ecological validity.

Although there is evidence that historical studies performed in laboratory settings are congruent with the results of similar studies run in VR, previous VR trolley problem studies occurred in laboratory settings, with participants being invited to use hardware in an unfamiliar location in the presence of researchers [40,41]. The presence of researchers in the lab as facilitators may affect participant behavior, although the data in the case of VR are, admittedly, in their infancy [41,42].

The present study is meant as a proof of concept that investigates a new approach to the trolley problem that melds the crowd-sourced approach of MIT’s Moral Machine with the immersive and presentational qualities of recent VR trolley problem studies by creating and disseminating a trolley problem scenario that can be downloaded and experienced on commercially available VR hardware. This allows users to experience the trolley problem from the comfort of their homes, removing the requirement of a laboratory and an in-person researcher on site. As a result, this study at a distance reduces the demand characteristics and increases the potential generalizability of the results to broader populations [43]. The approach has the additional benefit of facilitating research under pandemic conditions in which in-lab research is restricted [44].

As significant as the data collection aspect of thought experiments is their influence on the users that experience them. Recall that thought experiments are also useful tools for bringing philosophical ideas to a broader audience. Philosopher Bertrand Russell writes, “If the study of philosophy has any value at all for others than students of philosophy, it must be only indirectly, through its effects on the lives of those who study it” [45]. Thought experiments allow those who encounter them to reflect upon and modify their attitudes toward the concepts unearthed by thought experiments. In the case of the trolley problem, the dilemma provides those who do not study philosophy an opportunity to reflect on their own beliefs about the morality of maximizing the needs of the many at the cost of the needs of the few and to potentially revise those beliefs in light of new considerations. In turn, users develop a more informed and reflective attitude toward their moral commitments. Such informed reflection on moral commitments is crucial to a well-informed society that must increasingly confront the ethics of emerging technologies, such as self-driving cars and AI [46]. MIT’s Moral Machine experiment recognizes this benefit by allowing users to compare their results to others and even build trolley problem variations. Similarly, the project of which this study is a part aims at exposing a new, non-academic audience to the nuances of the trolley problem through VR. This approach promises a broader public to experience the intricacies of thought experiments in unique new ways.

In what follows, we present the design and distribution of a free and publicly available VR application that is capable of collecting data at a distance and includes two versions of the trolley problem. Much of the utility of the present paper is found in the development of these tools for use by future philosophers and researchers to engage with the general public.

We then present the results of an initial exploratory study. The exploratory study is motivated by three research questions: (1) Will participants in a VR study at a distance overlap with the population of VR users more generally? (2) Are responses to the trolley problem presented at a distance consistent with those in previous VR studies conducted in

laboratory settings? (3) Are there any associations among experiences of presence, study type, outcome, or moral decision making?

The results of the exploratory study are promising. Participants' responses to the two dilemmas are consistent with those found in previous VR studies, suggesting that this sort of VR research at a distance is a viable alternative to laboratory-only studies. This conclusion should not be overstated, however, as each VR study (including the present one) uses a different VR experience, making direct comparisons difficult. Because participants can enjoy the experience from the comforts of their own homes, the study also suggests an exciting venue for engaging the general public in philosophical reflection. The paper closes by discussing some of the strengths and shortcomings of this approach and suggests areas for future development.

2. Materials and Methods

2.1. Participants

Following the protocol approved by Old Dominion University's Institutional Review Board (ODU IRB 20-125), we collected valid data from 33 participants during the approved period. Three failed the Moral Control scenario and were excluded from further analysis. The study followed the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) reporting guidelines [47].

All participants were 18 and above and recruited on a rolling basis through online websites and forums. Ownership of a headset was a requirement for participation in the study. Participants that scored above 19 on the Golding Motion Sickness questionnaire were prevented from completing the experience to avoid simulator sickness [48,49].

Participants were randomly assigned a study condition (push vs. switch) at the beginning of the experiment. We expected close to a 50% split between participants assigned to each condition. However, since some participants were excluded from the study for failing the Moral Control scenario after being assigned the study condition, the distribution of conditions was 6:4 (push vs. switch), with 18 in the push condition and 12 in the switch condition.

2.2. VR Experience

The "VPRL Presents Life and Death Moral Dilemmas" experience (henceforth "the experience") was developed to run on the Oculus (now Meta) Quest 2 and Rift headsets. The development of the experience was designed to be as virtually real as possible within the constraints of the available hardware while also adhering to principles of accessible design and the recommendations of Ramirez et al. for creating virtually real experiences [32,50]. Some of these decisions are outlined below. (1) The onboarding process includes floating text and non-diegetic audio. Once completed, however, the user is moved to an independent virtual space. From this point forward, all audio, including instructions, is presented diegetically. Accompanying text for the hearing impaired is presented via a virtual computer station within the experience. Furthermore, once the experiment has begun, the experience is limited to a single environment. (2) The experience does not involve any kind of heads-up display (HUD). Instead, abstract visual information, such as information concerning which track is "active", is presented through a virtual computer station and through a virtual wristwatch. (3) The virtual computer station's information is presented by using colors that were selected for visual accessibility. (4) One condition tasks users with pushing a worker into the path of an oncoming train in order to bring it to a halt. Early tester feedback suggested that this scenario was implausible. In response, we developed a system (outlined below) whereby pushing the worker to their death triggers an automatic stop system. (5) The experience utilizes a teleportation movement system. While teleportation is less likely to support virtually real experiences, limitations of in-home VR, including lack of space and the increased risk of simulator sickness, supported a teleportation system over other mobility systems.

2.2.1. Onboarding

The experience begins with participants completing an informed consent document, a demographic questionnaire, and a motion sickness questionnaire to screen participants for simulator sickness. The experience then places participants in a small office to acquaint themselves with the experience's teleoperation system and to select accessibility options (Figure 1). From this point, the experience is divided into three successive scenarios: a Training scenario, a Moral Control scenario, and a Moral Dilemma scenario.

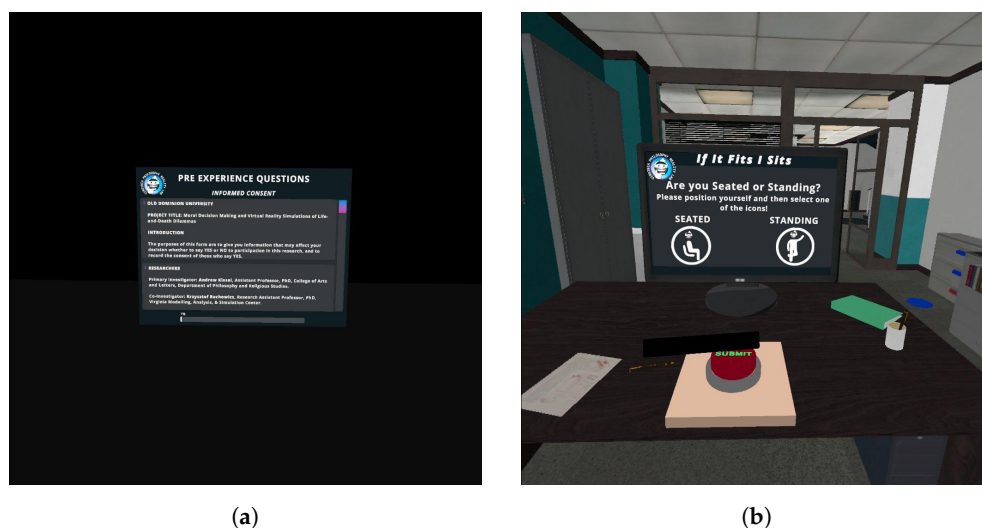


Figure 1. The onboarding process. (a) Informed consent forms and demographic questionnaires are completed before users are placed in a defined location. (b) Users learn navigation inside a small office and make accessibility selections.

2.2.2. Training Scenario

Participants are then moved to a control tower next to a fork between a “main” railroad track and a “spur” track. They are informed through an overhead loudspeaker that they will be trained in the railroad’s safety systems. An in-experience computer screen shows which track is currently “live” (Figure 2).

The Training scenario teaches the participant how to use the railroad’s two safety systems. The first system allows participants to press buttons to switch oncoming trains between the main track and the spur track. To press a button, users must physically reach forward with their hand controllers to bring their virtual finger down onto one of two buttons. The switch system is used to divert trains. The second system allows participants to push a water barrel onto the tracks to create an electrical short that stops the train (Figure 2). Users must physically reach forward to bring their virtual hands into contact with the barrel with sufficient force to push the barrel onto the tracks. Participants are informed that the water barrel system is a backup in case the switch system fails and that anything filled with water would stop the train, including “a water barrel... or even a person”. After demonstrating competency in both systems, participants are shifted to the Moral Control scenario.

2.2.3. Moral Control Scenario

In the Moral Control scenario, five workers begin construction activities on the main track when a train comes down the tracks. The five workers cannot escape because they are working on a narrow bridge. The five workers will be hit by a train if nothing is done. If the switch system is used, the five workers can be saved. In this scenario, there is a clear, morally preferable option: divert the train onto the spur track and save all five workers. This scenario was included as a comprehension check. Participants that failed the Moral Control scenario were not included in the additional analysis for the

study. After completing the Moral Control scenario, participants are shifted to the Moral Dilemma scenario.

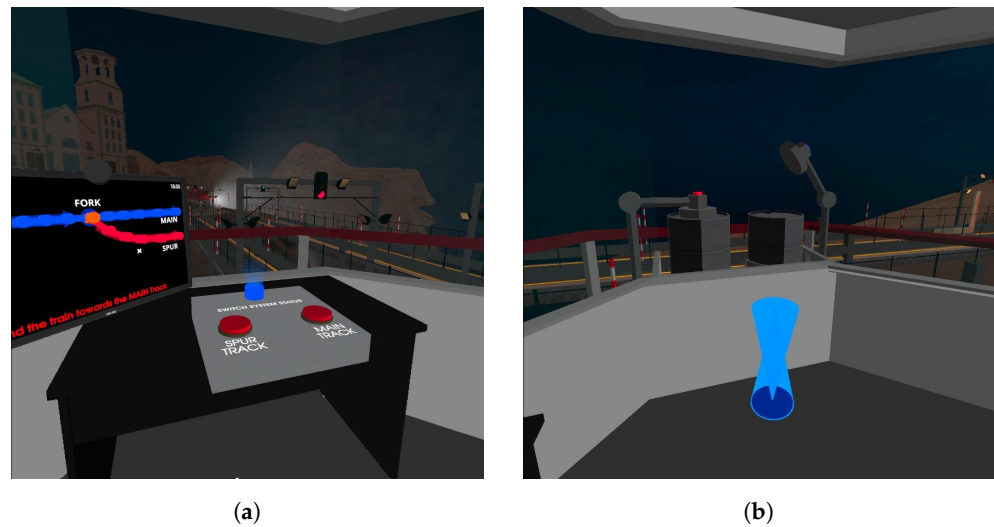


Figure 2. Training with safety systems. (a) A screen shows the main and the spur tracks. Here, the main track is live. Pressing buttons switches between tracks. (b) When the switch system fails, participants can push unlocked barrels to stop a train on the spur track.

2.2.4. Moral Dilemma Scenario

In the Moral Dilemma scenario, participants were randomly assigned to one of two conditions, (1) *The Switch Dilemma* condition or (2) *The Push Dilemma* condition. In the Switch Dilemma, the primary switch system is operational. Five workers begin construction activities on the main track, while one begins construction work on the spur track when a train arrives. If nothing is done, the five workers will be hit, but if the primary switch system is used, the one worker on the spur track will be hit (Figure 3). As a result, choosing to pull the switch saves five lives at the cost of one life. Choosing to do nothing saves one life at the cost of five lives.

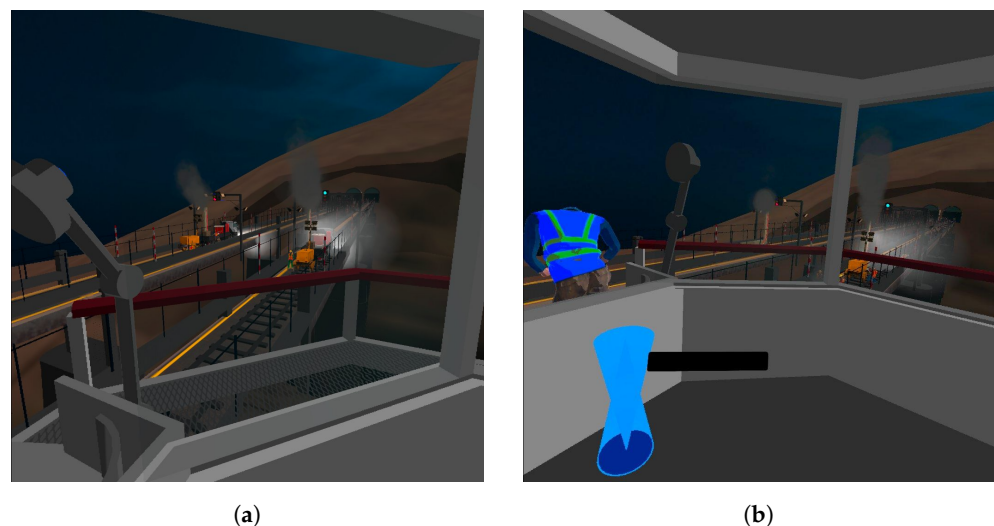


Figure 3. Moral Dilemmas. (a) In the Switch Dilemma, there are five workers on the main track and one worker on the spur track. Doing nothing results in the death of five workers, and using the switch system results in the death of one worker. (b) In the Push Dilemma, there are five workers on the spur track and the switch system is inoperable. Doing nothing results in the death of five workers, and pushing the worker from the control tower results in the death of that worker only.

The primary switch system has failed in the Push Dilemma, and the oncoming train must take the spur track. Five workers begin construction on the spur track. Additionally,

one worker stands in the tower in place of the backup water barrel. The five workers on the spur track will be hit if nothing is done. However, if the worker in the tower is pushed onto the tracks, it will initiate the water barrel backup system, resulting only in the death of the one worker that was pushed. As a result, choosing to push the worker onto the tracks saves five lives at the cost of one life. Choosing to do nothing saves one life at the cost of five lives.

Upon completing the Moral Dilemma, participants are given the opportunity to verbally explain why they chose their particular course of action in the Moral Dilemma scenario.

After completing the recording, participants are presented with a debrief message and destinations for mental health resources.

2.3. Study Design

The experience was hosted on SidequestVR, a third-party website that facilitates loading applications onto Oculus headsets. Participants were recruited through SidequestVR, VR enthusiast forums such as r/virtualreality and r/oculus on Reddit, VR podcasts, and by word of mouth. Ownership of a headset was required for participation in the study.

Upon downloading the experience, participants were first required to complete an informed consent document, a demographic information survey, and a Golding motion sickness questionnaire [48,49].

Upon completing the onboarding questions, participants were randomly assigned to the Push Dilemma or the Switch Dilemma. Each participant completed the Training scenario, the Moral Control scenario, and the Moral Dilemma scenario as described above. Participants that failed the Moral Control scenario by not selecting the option that saved five lives with no casualties were not included in the additional analysis for the study.

After the Moral Dilemma scenario, participants were given the following prompt: “Please explain why you decided to (pull the switch/not pull the switch; push the worker/not push the worker)” according to the dilemma that they faced and the choice that they made. Participants were given one minute to verbally respond to the prompt. Verbal responses were recorded as audio logs, then manually converted into text. We independently scored each verbal response, relying on a broad ethical distinction between *consequentialist* ethical theories and *deontological* ethical theories. Although the specifics of the background theories are not of primary importance, we can provide some brief details to illustrate the distinction. For consequentialists, the normative status of an action depends only on the outcomes, or consequences, of the action [51]. Mill’s utilitarianism is an example of a consequentialist moral theory, where the relative rightness of the action depends on the relative amount of good that that action brings about [52]. For deontologists, by contrast, the normative status of an action does not depend only on the outcomes of the action. Deontological theories tend to hold that the rightness of the action has priority over the good it might bring about [53]. Kantian ethics are a prominent example of a deontological moral theory, where the rightness of an action is determined by its conformity to objectively rational principles, independent of any outcomes [54].

Responses that appealed to the actions’ consequences were thus coded as consequentialist justifications (C). Responses that appealed to prior moral norms, e.g., the wrongness of intentionally taking a life, were coded as deontological justifications (D). Responses that appealed to an unintended outcome were coded as accidental (A) or as providing insufficient information to be categorized (I). While acknowledging philosophical disagreements about what constitutes a “killing”, for the purposes of brevity, here, we refer to any action that results in death as a “kill”.

All responses were automatically logged and stored with a randomly generated six-digit identifier on a secured Microsoft Azure database.

2.4. Data Sources and Measurements

The pre-experience questionnaire provided information about the participants’ demographics: age, gender, education, religion, and prior training in philosophy and psychology.

The experience provided information about the type of headset used, seated or standing pose, dominant hand, completion time, assigned study condition, moral dilemma check result, the decision made by the participant and the justification, and the logic behind it.

The post-experience survey adopted the Virtual Experience Test (VET) developed by Chertoff and Goldiez [55]. The VET is a tool used to evaluate the holistic experience of virtual environments in order to predict increases in users' sense of presence. Questions 1, 2, and 6 are related to the sensory dimension, 3 and 4 are related to the affective dimension, 5 is related to the active dimension, and 7 and 8 are related to the cognitive dimension. Questions 9, 10, and 11 were devised by the authors and were designed to capture the moral dilemmas of the participants.

2.5. Statistical Analysis

We calculated frequency statistics to understand the demographics of our sample and the VR setup they used. We analyzed how many people in each study condition decided to take action compared to those who let the train kill five people. We broke down the study outcomes based on demographic information, such as gender, religion, education, and prior training in philosophy and psychology.

We analyzed the post-experience survey to detect if there were any statistical differences ($\alpha = 0.05$) in the presence (Q1–8) and moral dilemma (Q9–11) when users' groups were considered.

We also analyzed the survey by comparing participants' responses based on their decisions. We investigated statistically significant ($\alpha = 0.05$) correlations between questions for the whole cohort and subset according to the study conditions and decisions made by the participants.

Coding the reflection transcripts allowed us to compare the motives behind the decisions made in the VR experience. Lastly, all analyses were conducted by using R version 4.2.2 with supplementary packages [56].

3. Results

Following the protocol approved by the Old Dominion University's Institutional Review Board (ODU IRB# 20-125), we collected valid data from 33 participants. Three failed the moral dilemma check and were excluded from further analysis.

3.1. Demographics and VR Setup

Most participants (83%) were equipped with Oculus Quest devices, while the rest took part in the experiment by using the Oculus Rift. Two-thirds of the participants engaged in the experience seated, and all were right-handed.

The participants were mostly college-age—18–24 (56%). Almost all participants were younger than 45 years old. More than two times as many participants identified as male (67%) compared to female (27%), while 6% preferred not to say or identified as another gender. Almost 77% of participants declared themselves White, while Asians and African Americans equated to 7% each. A total of 60% of the participants did not associate themselves with any religion, while 30% were Christian.

A significant majority of participants (73.3%) had at least some college education. Almost 17% completed a four-year degree, while another 17% had a graduate degree. Most participants (57%) did not have prior training in philosophy, while the rest had at least one college-level course. A similar distribution could be observed as far as prior experience in psychology was concerned. Table 1 summarizes the demographic information of participants.

Table 1. Demographics: prior experience in psychology and philosophy, VR equipment used, and pose of the participants during the experiment ($n = 30$).

Age	18–24	25–34	35–44	45–54		
	56.7%	20.0%	16.7%	6.7%		
Gender	Female	Male	Other	Not specified		
	27.0%	67.0%	3.0%	3.0%		
Race	American Indian/ Alaskan Native	Asian	Black/ African American	White	Other	Not reported
	3.3%	6.7%	3.3%	76.7%	6.7%	3.3%
Religion	Christian	None	Other			
	30.0%	60.0%	10.0%			
Education	High school and less	Some college	4 year degree	Graduate degree		
	26.7%	40.0%	16.7%	16.7%		
Prior training in Psychology	Very little/None	1–2 college-level courses	3+ college-level courses			
	60.0%	30.0%	10.0%			
Prior training in Philosophy	Very little/None	1–2 college-level courses	3+ college-level courses			
	57.0%	30.0%	13.0%			
VR equipment	Oculus Quest	Oculus Rift				
	83.0%	17.0%				
Pose	Seated	Standing				
	67.0%	33.0%				

3.2. Study Outcomes

All participants who were presented with the *Switch Dilemma* condition decided to change the course of the train to kill one person instead of five (Figure 4). In the case of the *Push Dilemma* condition, almost 40% of participants decided not to take the action that would lead to the death of one worker.

**Figure 4.** Study outcomes by scenario ($n = 30$).

When the participants' reported gender was considered, 62% of women decided to make the decision that resulted in the death of one worker compared to 80% of the participants identifying themselves as men. All participants who identified as *other* and

preferred not to say reported deciding to switch the tracks or push the person. A total of 78% of the participants who identified themselves as Christians and 72% of those who did not belong to any religion decided to kill one person and save five. The participants who had prior training in philosophy were more likely (85% to 71%) to take action to save five workers while sacrificing one. A similar relationship was observed between those who had or did not have prior training in psychology (92% to 67%) and those who had and did not have any college education (82% to 62%).

3.3. Post-Experience Survey

At the end of the experience, the participants, while still in VR, were presented with an eleven-question questionnaire asking them about various aspects of what they went through. The survey was adapted from the Virtual Experience Test (VET), which can be used to predict experiences of presence [55]. Table 2 summarizes the results of that questionnaire for all participants. Below, we indicate places in which there was a divergence between those who chose to act (i.e., pull the switch or push the worker) and those who refrained from acting.

Table 2. Results of the post-experience survey ($n = 30$).

#	Survey Item	% Strongly Agree/ Agree	% Neutral	% Strongly Disagree/ Disagree
1	I feel that the visual display was of high quality.	60.0%	33.0%	7.0%
2	I feel the visual content of the environment to be of high quality.	46.7%	43.3%	10.0%
3	I felt my emotional reaction was appropriate given the events that occurred in the virtual environment.	67.0%	17.0%	17.0%
4	I felt a variety of emotions when working on the task.	60.0%	30.0%	10.0%
5	I felt that I was the character I was controlling.	70.0%	23.0%	7.0%
6	I feel that the audio experience was of high quality.	76.7%	10.0%	13.3%
7	I feel that the virtual environment made it clear what I was and was not allowed to do.	63.3%	20.0%	16.7%
8	I thought that the task I was asked to do in the virtual environment was interesting.	77.0%	20.0%	3.0%
9	I felt like I was in a moral dilemma when making the decision.	70.0%	20.0%	10.0%
10	I thought about my decision before making it.	76.7%	10.0%	13.3%
11	It was difficult for me to make the decision I made in this scenario.	53.0%	13.0%	33.0%

Around 60% of the participants believed that the visual display was of high quality (Q1). However, opinion dropped to 46.7% when the question switched to the quality of the visual content (Q2). For Q2, almost 40% of the participants who took action in the experience—compared to over 70% who did not take action—found the visual content of the experience to be of high quality.

Almost 70% of the participants who took action believed that their emotional reactions were appropriate (Q3), and over 55% of people in that group felt a variety of emotions (Q4). Conversely, of those who refrained from acting, around 57% found their emotional responses to be appropriate and felt a variety of emotions.

Around 70% of all participants felt that they were the character that they were controlling (Q5). For Q6, 76.7% of all participants thought that the audio experience, including the narration and voice prompts, was of high quality. Among those who took action, however, 82% considered the audio experience to be of high quality, while only 57% of those who refrained from acting found it to be of high quality. Slightly more people who took action understood what they were and were not allowed to do in the experience (Q7) (65%) when compared to the group that refrained from acting (57%).

The majority of participants (around 70%), regardless of the decision made, reported experiencing a moral dilemma (Q9). Participants who either switched the tracks or pushed the person were more likely to think about the decision before making it (78%) than people who did not take any action (71%) (Q10). Lastly, more than half of the participants in both groups found it difficult to decide (Q11).

Welch's *t*-test showed no statistical difference ($p > 0.05$) in the reported presence (Q1–8) and moral dilemma (Q9–11) between users who were presented with the *push the person* ($n = 18$) versus *switch the tracks* ($n = 12$) scenario.

Similarly, the same test indicated no statistical difference ($p > 0.05$) in the reported presence (Q1–8) and moral dilemma (Q9–11) between users who took the action that resulted in the “death” of one person ($n = 23$) and who did not prevent the “demise” of five workers.

We also investigated correlations and their statistical significance ($\alpha = 0.05$) between the questions for all participants ($n = 30$) and subgroups based on the study type and outcome.

Considering all participants, the responses related to the visual aspects of the experience (Q1 and Q2) were strongly correlated. In addition, people who felt various emotions (Q4) experienced a moral dilemma (Q9). It was rather difficult for people who experienced a moral dilemma (Q9) to make a decision (Q11) (Figure 5).

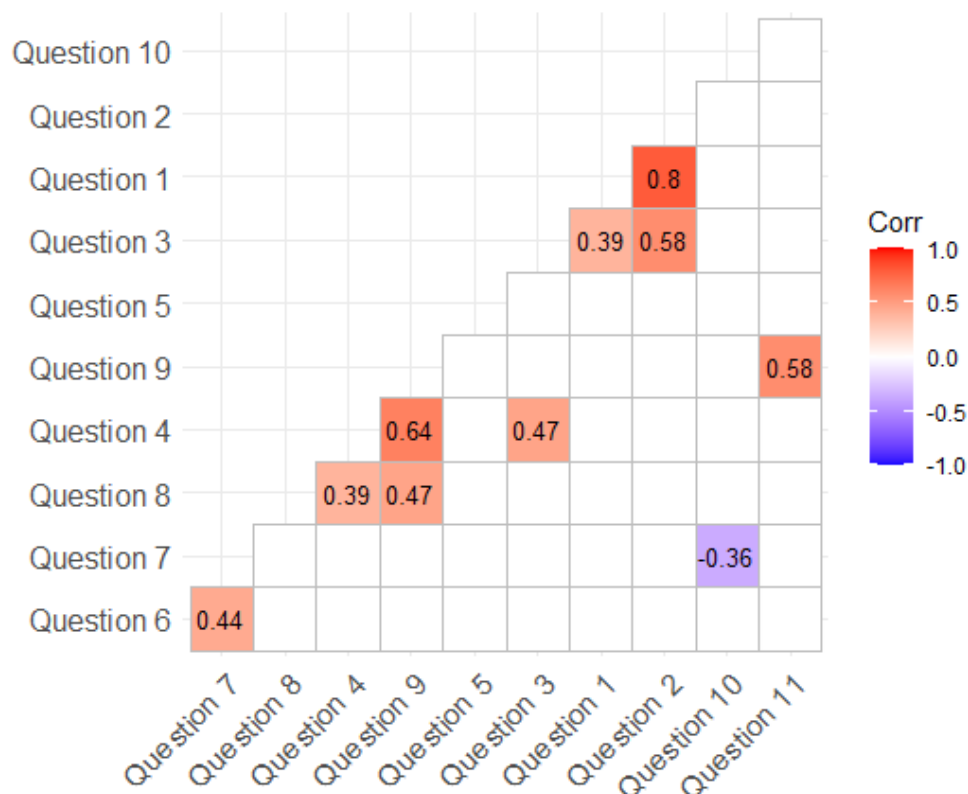


Figure 5. Statistically significant correlations between answers to the post-experience survey questions for all participants ($n = 30$). Statistically insignificant correlations ($p > 0.05$) were omitted.

Considering the participants who decided not to take any action, which resulted in the death of five workers ($n = 7$), it was difficult for people who experienced a moral dilemma (Q9) to make a decision (Q11). The quality of the visual content was positively associated with the appropriateness of participants' emotional reactions given the events that occurred in the virtual environment (Q3). Even though the participants felt various emotions when in VR (Q4), they believed that their emotional reactions were appropriate given the events that occurred in the virtual environment (Q3). It was challenging to decide in the scenario (Q11) if a participant felt that they were in a moral dilemma (Q9). The high-quality visual display

(Q1) helped make the tasks in VR interesting (Q8). Making clear what the participants were and were not allowed to do (Q7) helped them feel in control of the character that they were (Q5).

When considering the participants who decided to push the person or switch the tracks to save five people and sacrifice one ($n = 23$), to a certain extent, having the environment make clear what the participants were and were not allowed to do (Q7) was positively correlated with how interesting the task was to them (Q8). The participants who experienced a moral dilemma (Q9) tended to feel various emotions when working on the task (Q4). When analyzing the correlations between answers of the participants who were presented with the *switch tracks* scenario ($n = 18$), the feeling of being the character that they were controlling (Q5) was strongly correlated with the perception of the audio experience to have been of high quality (Q6). Even though the participants thought about the decision that they had to make (Q10), it was still challenging to make (Q11). Participants who experienced a moral dilemma (Q9) tended to view the task that they were asked to perform as interesting (Q8). Interestingly, the participants who felt this way (Q8) tended to think that the visual content of the environment was not of high quality (Q2). As the perception of the audio experience quality increased (Q6), it was easy for the participants to decide in the *switch tracks* scenario.

Considering the participants who were presented with the *push the person* scenario ($n = 18$), except for the strong correlation between the answers related to the quality of the visual display and content (Q1 and Q2), a group of questions were moderately correlated (0.4–0.6). The participants who found the display and visual content to be of high quality believed that their emotional reaction was appropriate (Q3). It was rather difficult for people who were in a moral dilemma (Q9) to make a decision (Q11).

3.4. Post-Experience Reflections

A total of 19 out of the 30 participants who did not fail the Moral Control scenario provided post-experience reflections regarding their decisions in VR. This included 13 participants who decided to take action to save five workers and six who did not.

We coded their reflections as *accident*, *consequentialist*, *deontological*, and *insufficient* according to the kind of explanation given to justify their action. One reflection could be evaluated with multiple codes.

Five participants reported that the action or inaction logged by our experience was accidental. In those cases, we also coded their intended action if there was a sufficient explanation. For instance, one participant reflected, “*This was by accident. I was trying to pull him back actually, not push him. Then, I realized there are [a] few other people who could have died. I think one life against a number of others is a fair choice*”.

Eleven reflections were classified as *consequentialist*, since they appealed to the outcomes to justify their action, while three were classified as *deontological*, since they appealed to prior principles to justify their actions. Two provided *insufficient* explanations of their decisions. The *consequentialist* character of the decision was more prevalent for those who decided or intended to either push the person or press the switch.

4. Discussion

4.1. Research Question 1: Will Participants in a VR Study at a Distance Overlap with the Population of VR Users More Generally?

One claimed benefit of doing VR studies at a distance is that they can reach a more general audience. In the present study, the majority of participants in the study identified as white, college-aged, and male. This breakdown largely reflects the demographics of VR users more generally [57]. More than half had no prior experience with philosophy or psychology, which decreased the likelihood that they had prior formal experience with the trolley problem.

Although a sample size of thirty participants is consistent with that of some similar studies with human subjects in VR [27], the study’s small size should be kept in mind both

when interpreting the results and when considering the viability of doing VR studies at a distance. The number of participants was capped to those that participated during the prior approved grant period. If surpassing 30 participants requires multiple years of data collection, this speaks against the viability of VR studies at a distance. These numbers were also affected because groups such as Central Washington University's EthicsLab discovered the experience and used it as part of their community outreach projects, skewing to younger and college-educated demographics. Although this study reports on participants during a finite research period, the experience continues to collect data.

Further limitations of the study and the potential for bias should be considered when evaluating these results. The study was limited to owners of the Oculus Quest and Rift. People who suffered before from motion sickness in various contexts had a higher probability of being excluded from studies involving VR experiences. It is also possible that the ads attracted people interested in philosophy and moral dilemmas. The results and limitations suggest that while VR studies at a distance can reach a more general audience, they still fall far short of reflecting broader diversity in world populations.

4.2. Research Question 2: Are Responses to the Trolley Problem Presented at a Distance Consistent with Those of Previous VR Studies Conducted in Laboratory Settings?

If VR studies at a distance are to provide a viable alternative to in-lab studies, they should at least replicate the results of previous studies. Previous VR studies on the trolley problem suggested that an overwhelming majority of participants would pull the switch to kill one and save five, which was a result shared by our study (100% of the users in the Switch Dilemma chose to pull the switch) [27,35,37]. Furthermore, previous VR studies suggested that most participants would push the worker onto the tracks to kill one and save five. Comparing the Switch Dilemma and the Push Dilemma, fewer were willing to push one worker to save five than were willing to pull the switch to kill one to save five (60% of users in the Push Dilemma chose to push the worker). Again, this is consistent with previous studies and was predicted by recent adaptations of Cushman's dual-process model of moral decision making [15,27]. As a result, this provides initial support for the claim that VR can replicate in-lab results without bringing participants into the lab, though more research is needed.

4.3. Research Question 3: Are There Any Associations Among Experiences of Presence, Study Type, Outcome, or Moral Decision Making?

The claim that VR studies (in-lab or at a distance) have greater ecological validity than their text-based counterparts depends crucially on the technology's ability to evoke a sense of presence, the sense of "being there" in a virtual environment. Despite the Quest 2 being less expensive and lower-powered than the state-of-the-art hardware found in laboratories, we see evidence that our experience was still able to evoke a high sense of presence in users. A total of 70% of the respondents strongly agreed/agreed that they felt like they were the character that they were controlling (Q5). In contrast, 67% strongly agreed/agreed that their emotional reaction was appropriate given the events in the virtual environment (Q3). This suggests that participants tended to have experiences of presence while interacting with the virtual environment.

We did not establish statistically significant differences between groups of users who were presented with the push and switch scenarios as far as presence and moral dilemma were concerned. Similarly, no difference was detected when we considered the study outcomes. However, we will reiterate that the group sizes were unequal and relatively small, especially in the case of participants who did not take action in the experience. Currently, the power of the test is 0.27 and 0.22, respectively. To achieve a medium-sized effect at a power of the test level of 0.8, we would need to recruit at least 130 subjects to study the difference between two study conditions and 180 subjects to investigate the difference between participants who were classified based on the study outcome.

There was also a positive correlation between participants' sense that they were the character that they controlled (Q5) and the rated quality of the audio (Q6), particularly

among those that chose to kill one to save five. This suggests that audio played a crucial role in the experience of presence, perhaps more than visual fidelity. Although previous papers have criticized the visual and presentational qualities of VR trolley problem experiences, less attention has been paid to the audio quality [32].

4.4. Future Recommendations

The development process and resultant data suggest numerous lessons for future studies of this type. The first is to identify friction points that reduce the size of the participant pool. In addition to the limitations identified with research question 1, the onboarding process was an obstacle to participation. There was a sharp drop-off between views of the Sidequest page (over 1000) and downloads (around 300). Of the first 100 downloads from SidequestVR, only 12 completed the experience. We believed that part of the drop-off in participation was due to the difficulty of navigating to an independent Qualtrics website where participants completed the onboarding process before returning to the in-headset experience. While this move was meant to make completing the questions more accessible for participants and provide an additional level of anonymity, it also created a friction point that saw a significant drop-off in participants. The experience was revised to include the onboarding questions within the VR experience itself, as presented in Section 2. We recommend that future VR studies at a distance include the onboarding process within the experience to avoid this friction point.

Another lesson that we learned was that the Moral Dilemma scenario should not be assigned until participants have completed the Moral Control scenario. The Moral Dilemma was assigned based on a randomly assigned code upon completion of the onboarding questions, assuming that it would lead to a 50% split between the Push and Switch dilemmas. However, since some participants failed the Moral Control scenario, we could not use their results in the Moral Dilemma scenario, resulting in a 6:4 distribution (Push vs. Switch). We could have achieved a more even split between the two experimental conditions by waiting until after the Moral Control scenario to assign dilemmas.

Based on our results, greater focus should be placed on the audio quality of VR trolley problem experiences going forward. Audio quality contributed substantially to an agreement with questions associated with presence. Furthermore, without a facilitator to respond to questions, all training and instructions must be built entirely into the experience. A high-quality audio experience contributes to an understanding of what interactions are available in the virtual environment and a sense of presence while immersed in the environment.

This final recommendation should be taken with a grain of salt. It was not easy, particularly with experienced VR users, to train for an interaction without incentivizing that interaction. For example, when participants are trained to use the switch system, they may think they are expected to pull the switch when possible. Indeed, some verbal responses suggested that the elicited behavior was testing the VR environment rather than engaging in moral deliberation. For example, one participant responded to the question "Why did you choose to pull the switch?" by saying, "...in the real world I definitely would not have, but in the VR world, I didn't know if he would go down or not." Future study designs should account for this phenomenon.

On the other hand, efforts to avoid unduly incentivizing interaction may have made it unclear what interactions were available. In the case above, the participant reported testing the possible interactions in the VR world rather than making a moral decision. This suggests that the possible affordances were not always clear to participants.

As a final consideration for future research of this kind, there has been some debate about the ethics of this kind of research [57,58]. The primary concern is that, by creating virtually real experiences, participants in these kinds of studies may learn something new about themselves: That they are capable of taking another human life. Our study suggests that in-home technology and experiences are not yet sophisticated enough to create truly virtually real experiences. Nevertheless, these are serious ethical concerns to be taken seriously. For the present study, we took steps to reduce participants' exposure to violence

and gore in the design of the experience by using lighting and occlusion. As such, the kinds of decisions and visuals to which participants were exposed are no worse than those in video games that are currently available on the market. These ethical concerns should be kept in mind for future studies.

5. Conclusions

Our study suggests that researching philosophical thought experiments through VR studies at a distance is in its infancy, but has a promising future. We replicated laboratory results from previous VR studies. Nevertheless, additional efforts must be made to broaden the participant pool beyond the demographics of the VR user base.

We identified several difficulties in running this kind of study at a distance and provided several recommendations to address these difficulties in future studies. Overall, these recommendations concern optimization. As such, the present study provides a template for executing future thought experiments in VR at a distance and tools for building data-collecting experiences. With the increasing popularity of in-home VR technologies comes the promise of spreading philosophical thinking to new users while collecting valuable new data.

Author Contributions: Conceptualization, A.K., K.J.R. and J.B.S.; methodology, A.K. and K.J.R.; software, J.B.S.; formal analysis, A.K. and K.J.R.; investigation, A.K. and K.J.R.; resources, K.J.R.; data curation, A.K., K.J.R. and J.B.S.; writing—original draft preparation, A.K. and K.J.R.; writing—review and editing, A.K. and K.J.R.; visualization, J.B.S. and K.J.R.; supervision, A.K. and K.J.R.; project administration, A.K.; funding acquisition, A.K., K.J.R. and J.B.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Endowment for Humanities (grant HAA-277270-21).

Institutional Review Board Statement: This study was approved by the Institutional Review Board of Old Dominion University (protocol approval reference # 20-125; initial approval: 3 August 2020).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The data obtained during the study are available at <https://osf.io/kzqmr/> (accessed on 22 June 2022).

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

AI	Artificial intelligence
IRB	Institutional review board
VPRL	The Virginia PhilosophyReality Lab
VR	Virtual reality

References

1. Brown, J.R.; Fehige, Y. Thought Experiments. In *The Stanford Encyclopedia of Philosophy (Spring 2022 Edition)*; Zalta, E.N., Ed.; Stanford University: Stanford, CA, USA, 2022. Available online: <https://plato.stanford.edu/archives/spr2022/entries/thought-experiment/> (accessed on 15 March 2022).
2. Bailey, C. *Lucretius on the Nature of Things (Translation of De Rerum Naturae)*; Clarendon Press: Oxford, UK, 1950.
3. Sorensen, R.A. *Thought Experiments*; Oxford University Press: Oxford, UK, 1992.
4. Knobe, J.; Nichols, S. An Experimental Philosophy Manifesto. In *Experimental Philosophy*; Knobe, J., Nichols, S., Eds.; Oxford University Press: Oxford, UK, 2007; pp. 3–14.
5. Cohnitz, D. Modal Skepticism. Philosophical Thought Experiments and Modal Epistemology. In *The Vienna Circle and Logical Empiricism: Re-Evaluation and Future Perspectives*; Springer: Berlin/Heidelberg, Germany, 2003; pp. 281–296.
6. Behmel, A. Was sind Gedankenexperimente. In *Kontrafaktische Annahmen in der Philosophie des Geistes—der Turingtest und das chinesische Zimmer*; Ibidem: Stuttgart, Germany, 2001.
7. Siegel, J.; Morgan, D. writers. The Good Place; Season 2, Episode 6, “The Trolley Problem.” Aired October 19, 2017. NBC, 19 October 2017.

8. Foot, P. The Problem of Abortion and the Doctrine of the Double Effect. *Oxf. Rev.* **1967**, *5*, 5–15.
9. Thomson, J.J. Killing, Letting Die, and the Trolley Problem. *Monist* **1976**, *59*, 204–217. [[CrossRef](#)] [[PubMed](#)]
10. Schaeffer, Z., writer. WandaVision. Season 1, episode 9, “The Series Finale”. Aired March 5, 2021. *Disney+*, 5 March 2021.
11. Stuart, M.T.; Fehige, Y.; Brown, J.R. *The Routledge Companion to thought Experiments*; Routledge: London, UK, 2018.
12. Chalmers, D.J. The Matrix as metaphysics. In *Science Fiction and Philosophy: From Time Travel to Superintelligence*; Wiley Online Library: Hoboken, NJ, USA, 2016; pp. 35–54.
13. Chalmers, D.J. *Reality+: Virtual Worlds and the Problems of Philosophy*; W. W. Norton: New York, NY, USA, 2022.
14. Greene, J.D.; Sommerville, R.B.; Nystrom, L.E.; Darley, J.M.; Cohen, J.D. An fMRI Investigation of Emotional Engagement in Moral Judgment. *Science* **2001**, *293*, 2105–2108. [[CrossRef](#)] [[PubMed](#)]
15. Cushman, F.; Young, L.; Hauser, M. The Role of Conscious Reasoning and Intuition in Moral Judgment: Testing Three Principles of Harm. *Psychol. Sci.* **2006**, *17*, 1082–1089. [[CrossRef](#)] [[PubMed](#)]
16. Mikhail, J. *Elements of Moral Cognition: Rawls’ Linguistic Analogy and the Cognitive Science of Moral and Legal Judgment*; Cambridge University Press: Cambridge, UK, 2011. [[CrossRef](#)]
17. Greene, J.D. Solving the Trolley Problem. In *A Companion to Experimental Philosophy*; Sytsma, J., Buckwalter, W., Eds.; John Wiley & Sons, Ltd.: Hoboken, NJ, USA, 2016; pp. 175–189.
18. Bonnefon, J.F.; Shariff, A.; Rahwan, I. The social dilemma of autonomous vehicles. *Science* **2016**, *352*, 1573–1576.
19. Dickinson, D.L.; Masclet, D. Using ethical dilemmas to predict antisocial choices with real payoff consequences: An experimental study. *J. Econ. Behav. Organ.* **2019**, *166*, 195–215. [[CrossRef](#)]
20. Himmelmreich, J.; Cohen, J. Teaching moral reasoning: Why and how to use the trolley problem. *J. Public Aff. Educ.* **2021**, *27*, 451–471.
21. Jenkins, R.; Cerny, D.; Hribek, T. *Autonomous Vehicle Ethics: The Trolley Problem and Beyond*; Oxford University Press: Oxford, UK, 2022. [[CrossRef](#)]
22. Schwitzgebel, E.; Cushman, F. Expertise in Moral Reasoning? Order Effects on Moral Judgment in Professional Philosophers and Non-Philosophers. *Mind Lang.* **2012**, *27*, 135–153. [[CrossRef](#)]
23. Christensen, J.F.; Flexas, A.; Calabrese, M.; Gut, N.K.; Gomila, A. Moral judgment reloaded: A moral dilemma validation study. *Front. Psychol.* **2014**, *5*, 607. [[CrossRef](#)]
24. Schwitzgebel, E.; Cushman, F. Philosophers? Biased Judgments Persist Despite Training, Expertise and Reflection. *Cognition* **2015**, *141*, 127–137. [[CrossRef](#)]
25. Pan, X.; Slater, M. Confronting a moral dilemma in virtual reality: A pilot study. In *Proceedings of the HCI 2011 the 25th BCS Conference on Human Computer Interaction*, Newcastle-upon-Tyne, UK, 4–8 July 2011; pp. 46–51.
26. Tassy, S.; Oullier, O.; Mancini, J.; Wicker, B. Discrepancies between Judgment and Choice of Action in Moral Dilemmas. *Front. Psychol.* **2013**, *4*, 250. [[CrossRef](#)]
27. Patil, I.; Cogoni, C.; Zangrando, N.; Chittaro, L.; Silani, G. Affective basis of judgment-behavior discrepancy in virtual experiences of moral dilemmas. *Soc. Neurosci.* **2014**, *9*, 94–107. [[CrossRef](#)]
28. Francis, K.B.; Howard, C.; Howard, I.S.; Gummerum, M.; Ganis, G.; Anderson, G.; Terbeck, S. Virtual morality: Transitioning from moral judgment to moral action? *PLoS ONE* **2016**, *11*, e0164374. . [[CrossRef](#)]
29. Bauman, C.W.; McGraw, A.P.; Bartels, D.M.; Warren, C. Revisiting external validity: Concerns about trolley problems and other sacrificial dilemmas in moral psychology. *Soc. Personal. Psychol. Compass* **2014**, *8*, 536–554. [[CrossRef](#)]
30. Awad, E.; Dsouza, S.; Kim, R.; Schulz, J.; Henrich, J.; Shariff, A.; Bonnefon, J.F.; Rahwan, I. The Moral Machine experiment. *Nature* **2018**, *563*, 59–64. [[CrossRef](#)]
31. Keeling, G.; Evans, K.; Thornton, S.M.; Mecacci, G.; Santoni de Sio, F. Four Perspectives on What Matters for the Ethics of Automated Vehicles. In *Proceedings of the Road Vehicle Automation 6*; Meyer, G., Beiker, S., Eds.; Springer International Publishing: Cham, Switzerland, 2019; pp. 49–60.
32. Ramirez, E.J. Ecological and Ethical Issues in Virtual Reality Research: A Call for Increased Scrutiny. *Philos. Psychol.* **2018**, *32*, 211–233. [[CrossRef](#)]
33. Slater, M.; Gonzalez-Liencre, C.; Haggard, P.; Vinkers, C.; Gregory-Clarke, R.; Jelley, S.; Watson, Z.; Breen, G.; Schwarz, R.; Steptoe, W.; et al. The ethics of realism in virtual and augmented reality. *Front. Virtual Real.* **2020**, *1*, 1. [[CrossRef](#)]
34. Parsons, T.D. Virtual reality for enhanced ecological validity and experimental control in the clinical, affective and social neurosciences. *Front. Hum. Neurosci.* **2015**, *9*, 660. [[CrossRef](#)] [[PubMed](#)]
35. Francis, K.B.; Terbeck, S.; Briazu, R.A.; Haines, A.; Gummerum, M.; Ganis, G.; Howard, I.S. Simulating Moral Actions: An Investigation of Personal Force in Virtual Moral Dilemmas. *Sci. Rep.* **2017**, *7*, 13954. [[CrossRef](#)]
36. McDonald, M.M.; Defever, A.M.; Navarrete, C.D. Killing for the greater good: Action aversion and the emotional inhibition of harm in moral dilemmas. *Evol. Hum. Behav.* **2017**, *38*, 770–778. [[CrossRef](#)]
37. Navarrete, C.D.; McDonald, M.M.; Mott, M.L.; Asher, B. Virtual morality: Emotion and action in a simulated three-dimensional “trolley problem”. *Emotion* **2012**, *12*, 364–370. [[CrossRef](#)] [[PubMed](#)]
38. Niforatos, E.; Palma, A.; Gluszny, R.; Vourvopoulos, A.; Liarakis, F. Would you do it?: Enacting moral dilemmas in virtual reality for understanding ethical decision-making. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, Honolulu, HI, USA, 25–30 April 2020; pp. 1–12.
39. Slater, M. Immersion and the illusion of presence in virtual reality. *Br. J. Psychol.* **2018**, *109*, 431–433. [[CrossRef](#)]

40. Blascovich, J.; Loomis, J.; Beall, A.C.; Swinith, K.R.; Hoyt, C.L.; Bailenson, J.N. Immersive virtual environment technology as a methodological tool for social psychology. *Psychol. Inq.* **2002**, *13*, 103–124. [\[CrossRef\]](#)
41. Richesin, M.T.; Baldwin, D.R.; Wicks, L.A.M. The role of contextual information in a virtual trolley problem: A psychophysiological investigation. *Soc. Neurosci.* **2022**, *17*, 246–257.
42. Sterna, R.; Strojny, P.; Rebilas, K. Can virtual observers affect our behavior? *Soc. Psychol. Bull.* **2019**, *14*, 1–18. [\[CrossRef\]](#)
43. Dandurand, F.; Shultz, T.R.; Onishi, K.H. Comparing online and lab methods in a problem-solving experiment. *Behav. Res. Methods* **2008**, *40*, 428–434. [\[CrossRef\]](#)
44. Mottelson, A.; Petersen, G.B.; Lilija, K.; Makransky, G. Conducting unsupervised virtual reality user studies online. *Front. Virtual Real.* **2021**, *2*, 66. [\[CrossRef\]](#)
45. Russell, B. *The Problems of Philosophy*; SMK Books: Radford, VA, USA, 2008.
46. Nyholm, S.; Smids, J. The Ethics of Accident-Algorithms for Self-Driving Cars: An Applied Trolley Problem? *Ethical Theory Moral Pract.* **2016**, *19*, 1275–1289. [\[CrossRef\]](#)
47. Von Elm, E.; Altman, D.G.; Egger, M.; Pocock, S.J.; Gøtzsche, P.C.; Vandenbroucke, J.P. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: Guidelines for reporting observational studies. *Ann. Intern. Med.* **2007**, *147*, 573–577. [\[CrossRef\]](#)
48. Golding, J.F. Motion sickness susceptibility questionnaire revised and its relationship to other forms of sickness. *Brain Res. Bull.* **1998**, *47*, 507–516. [\[CrossRef\]](#)
49. Golding, J.F.; Rafiq, A.; Keshavarz, B. Predicting Individual Susceptibility to Visually Induced Motion Sickness by Questionnaire. *Front. Virtual Real.* **2021**, *2*, 576871. [\[CrossRef\]](#)
50. Zimmermann, G.; Vanderheiden, G. Accessible design and testing in the application development process: Considerations for an integrated approach. *Univers. Access Inf. Soc.* **2008**, *7*, 117–128. [\[CrossRef\]](#)
51. Driver, J. *Consequentialism*; Routledge: London, UK, 2011.
52. Mill, J.S. *Utilitarianism*, ed. Roger Crisp; Oxford University Press: Oxford, UK, 1998; Volume 5, pp. 3–4.
53. Alexander, L.; Moore, M. Deontological Ethics. In *The Stanford Encyclopedia of Philosophy*, Winter 2021 ed.; Zalta, E.N., Ed.; Metaphysics Research Lab, Stanford University: Stanford, CA, USA, 2021.
54. Kant, I. *Groundwork of the Metaphysics of Morals* (HJ Paton, Trans.); Harper & Row: New York, NY, USA, 1964.
55. Chertoff, D.B.; Goldiez, B.; LaViola, J.J. Virtual Experience Test: A virtual environment evaluation questionnaire. In Proceedings of the 2010 IEEE Virtual Reality Conference (VR), Waltham, MA, USA, 20–24 March 2010; pp. 103–110.
56. Ihaka, Ross and Gentleman, Robert. R: A language for data analysis and graphics. *J. Comput. Graph. Stat.* **1996**, *5*, 299–314.
57. Ramirez, E.J.; LaBarge, S. Real moral problems in the use of virtual reality. *Ethics Inf. Technol.* **2018**, *20*, 249–263. [\[CrossRef\]](#)
58. Rueda Etxebarría, J. Hit by the Virtual Trolley: When is Experimental Ethics Unethical? *Teorema Int. J. Philos.* **2022**, *41*, 7–27.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.