

Confronting a Moral Dilemma in Virtual Reality: A Pilot Study

Xueni Pan
University College London (UCL)
Gower Street, London, UK
s.pan@cs.ucl.ac.uk

Mel Slater
UCL and ICREA-University of Barcelona
Gower Street, London, UK
m.slater@cs.ucl.ac.uk

People tend to respond realistically to situations and events in immersive Virtual Reality (VR). Our research exploits this finding to test the hypothesis that the psychology underlying moral judgement is distinct from the psychology that drives moral action. We have conducted an online survey study with 80 respondents on people's judgments of moral dilemmas. Additionally, we have carried out a pilot study with 36 participants investigating people's responses when confronted with comparable moral dilemmas in two different types of VR: desktop VR and Immersive VR. We recorded participants' behavioural responses and post experimental questionnaire data. The results show that in general, participants' responses in VR were consistent with the patterns obtained from the online survey. However, results also suggest that participants in the Immersive VR condition differed from those in the desktop VR condition in two ways: they 1) experienced more panic and made more mistakes in their immediate action; 2) were more likely to give a utilitarian answer (saving the greatest number of lives) in the post experimental questionnaire. This pilot study provides encouraging evidence for the use of VR in the study of moral psychology, and in particular, for teasing apart the distinction between judgments and actions. They further reveal that although our VR set up only presented abstract human figures, participants had a strong emotional reaction to the dilemma, on both immersive and desktop platforms.

Key Words: Virtual Reality, Immersion, Moral Judgement, Computer Animation

1. INTRODUCTION

It has been frequently demonstrated that in immersive Virtual Reality (VR), people tend to respond to situations and events as if they were real, despite the fact that they are consciously aware of the situation's artificiality (Rovira et al., 2009, Pertaub et al., 2002, Slater et al., 2006). This tendency of acting realistically in such environments distinguishes Immersive VR from any other media, such as films or books, and leads to a wide range of beneficial applications ranging from virtual psychotherapy to training (Sanchez-Vives and Slater, 2005). Moreover, because of its unique properties, VR has great potential to serve as a tool for exploring certain research topics in social science, such as psychology, where hypotheses are traditionally tested on paper based survey studies, video based studies, or studies that portray abstract representations of the situation that have problems regarding their ecological validity (Rovira et al., 2009).

One of the areas that could benefit from using VR is research into the psychology behind people's judgements in moral dilemmas. Two classic examples of such moral dilemmas are the Trolley

case and the Footbridge case (Foot, 1985, Thomson, 1971). In the Trolley case, an empty trolley is running out of control down a track that will run over and kill five people standing on the track. Standing next to the track there is someone who can flip a switch turning the trolley onto another track, and will therefore save the five people but as a consequence kill one other person standing on the other track who would otherwise survive. In the Footbridge case, similarly, a trolley is running out of control and is about to kill five people. But in this case, someone is standing on a bridge over the track and can push a nearby person wearing a heavy backpack, which because of the weight of the backpack will stop the trolley and therefore save the five, but sacrifice the one with the backpack. Here, both scenarios lead to the same consequences (sacrificing one to save five), but nevertheless most participants typically state that they would push the switch in the Trolley case, but not push the man in Footbridge (specifically, Trolley 85%, Footbridge 12% (Hauser et al., 2007)).

This has led to a number of questionnaire-based studies in moral psychology that attempt to understand the cause of the distinction between the

two cases (Hauser et al., 2007, Knobe, 2003, Machery et al., 2004). However, results generated from questionnaire data may not reflect behaviour that people would actually carry out when faced with an actual circumstance but only their judgement at a subjective level (Slater, 2004). It has been frequently shown that there is a difference between what people say and what they would actually “do”. A good illustration of this can be found in the classic experiments by Stanley Milgram on obedience to authority, where most people when asked said that only a tiny percentage of individuals would give fatal electric shocks to a stranger at the behest of an authority figure, whereas in fact in one experimental condition 60% of subjects did so (Milgram, 1963). Therefore, questionnaire data should best be backed up with behavioural observations.

Nevertheless, behavioural observation requires confronting participants with situations where there are moral dilemmas. This creates a problem as some of the dilemmas involve saving or sacrificing other people’s lives – clearly generating great ethical issues even if this were to be pretended in physical reality (as in Milgram’s study), since such pretence would involve unacceptable deception. VR, on the other hand, provides the possibility of putting participants in such situations *in vivo* without actually putting anybody’s life in danger, and without deception (since everyone knows that it is not ‘real’) and therefore alleviates the ethical concerns.

The idea of using VR in social situations where participants have to make difficult choices about their actions is not new and its power has been demonstrated in previous studies. For instance, Slater et al.’s study on the VR equivalent of obedience experiments (Slater et al., 2006) has shown that despite the fact that participants knew that the scenario was not real, many of them were stressed by the requirement to inflict electric shocks on a virtual woman and some of them even withdrew early from the experiment. A more recent study explored people’s responses to violent incidents using VR (Rovira et al., 2009), in which participants in VR witnessed a perpetrator bullying a victim, leading eventually to violence. The results showed that participants became involved in the scenario realistically, and many intervened to try to stop the violence or said that they had wanted to intervene. Both of the above studies were they to be investigated in real life scenarios would be very difficult to implement and would involve ethical problems.

In this project we examine social encounters more directly related to classic moral dilemmas. We attempt to translate the Trolley Case from paper to VR. Our research questions are: 1) would those

who said they would push the switch in the paper version also be likely to actually push a switch when confronted with a similar situation, but in VR? 2) would those who claimed they would not push the switch actually fail to act, allowing 5 people to die?

2. THE EXPERIMENT

2.1 Experimental Design



Figure 1: In an art gallery, an attacker is firing at 5 visitors. Bringing the lift down would save those 5 but put the 1 on the ground floor in danger.

There are a few challenges for translating paper-based moral dilemmas to VR. First, descriptions such as “five people will be killed” when clearly stated on paper, are unmistakably understood by participants. The scenario in VR has to achieve the same level of clarity. Secondly, in order to achieve an unbiased spontaneous response from participants, the scenario has to be “new” to all participants. Therefore it is less plausible to use the exact scene as classic moral dilemmas (such as the trolley and footbridge case as mentioned above), as they might be known to participants. More importantly, some scenarios such as footbridge are highly implausible.

We have designed a scenario that takes place in an art gallery. The participant is trained to use a lift (elevator) that takes visitors up to the first floor. Eventually an attacker on the lift starts shooting at 5 people who happen to be on the first floor. There is a switch next to the participant that controls the lift. Pushing the switch takes the attacker down to the ground floor, where one visitor happens to be standing. Therefore pushing the switch brings this one person on the ground floor into danger but the five on the first floor are saved, as shown in Fig.1. A critical purpose of the pilot experiment is to gather information about whether the scenario ‘works’: do participants understand the situation when the shooting starts? Are they aware that there are 5 visitors upstairs and will they remember the one downstairs? What are their thoughts when the dilemma unfolds? Also, unlike reading a questionnaire, when the event happens in VR it could be shocking for the participants. Will many participants freeze and be unable to act?



(a)



(b)

Figure 2: Virtual Reality Experiment (a) Immersive CAVE (b) Non-immersive desktop

As shown in Table 1, we designed the experiment as a 2x2 factorial between-groups design. The first factor was whether the moral dilemma was an action or omission condition. The Action Condition (AC) involved the scenario described above – the participant has to act to take the lift holding the attacker down to the ground floor, thus saving five but endangering one. In the Omission Condition (OC), everything else was the same except that one visitor was on the first floor and five were on the ground floor. In this case the participant did nothing, the one would be killed and the five saved. The second factor was the VR platform, either an Immersive Cave VR system or a normal desktop screen. Furthermore, to test the validity of the scenario, we also carried out an online survey study to test our scenarios in a questionnaire format. These results were then used as a baseline with which to compare our VR experiments. In the following Sections, we describe first the online survey study and then the VR study in detail.

Table 1: Factorial Design

	AC	OC
Immersive VR	n	n
Non-immersive VR	n	n

2.2 Online survey

In the survey we included 5 moral dilemmas, 3 existing ones (Trolley and Footbridge) and 2 depicting our scenario (Lift), as follows¹:

- Lift Action (LA): our scenario, as described in the previous section. Participants were asked if they would push the switch, which would save 5 people but sacrifice 1.
- Trolley Action (TA): a classic moral dilemma as described in Introduction. Participants were asked if they would push the switch to save 5 but sacrifice 1.
- Footbridge (FA): a classic moral dilemma, as described in Introduction. Participants were asked if they would push and kill the man to save 5.

- Trolley Omission (TO): similar to (b), apart from the consequence of pushing the switch being killing 5 to save 1.
- Lift Omission (LO): similar to (a), except that pushing the switch would kill the 5 but save the 1.

There were two conditions: Survey Action Condition (SAC) where the five dilemmas were presented in the above sequence and Survey Omission Condition (SOC) where LA and LO were exchanged to yield a different order: participants in SAC have LA as their first scenario and for those in SOC have LO. This is because the sequence of those dilemmas could have an impact on the results. To be able to use the results from the survey as a baseline for our VR study, in which both LA and LO would be presented to the participants, we needed to have unbiased results from both scenarios.

2.3 Virtual Reality Study

Here the goal was to observe participants' behavioural responses when confronted with a moral dilemma in Virtual Reality.

Upon arriving, the participants were assigned to one of the four conditions as shown in Table 1 in random order constrained by the need to have equal number of participants per cell. A questionnaire which collected basic information (age, occupation, etc.) was given to the participants. They were told both in writing and verbally that their task was to operate a lift in an art gallery.

They were then introduced into one of the VR Systems, being either an Immersive VR system (Fig 2(a), CAVE-like projection based system (Cruz-Neira et al., 1993)) or a desktop (Fig 2(b), a windows machine with a 17" display). They could see a virtual gallery with two floors (ground floor and first floor) and the only access to the first floor was by a virtual lift. They were taught to operate the virtual lift through a wand (a joystick like device with buttons), which took virtual visitors to the first level gallery.

¹ For the full questionnaire please see <http://web4.cs.ucl.ac.uk/cgi-bin/s.pan>

With the assistance of the experimenter, participants completed some lift-operator training sessions. They were then left on their own to operate the lift. Several virtual visitors walked into the gallery; some stayed at the ground level, others went on the lift to the first level. In AC 5 visitors eventually were at the upper level with one on the ground floor. A seventh visitor entered and asked to go to the upper level. Upon arrival, and while still on the lift, he started firing shots at the 5 visitors; one visitor was immediately injured. Participants were faced with the choice of doing nothing — thereby endangering all 5 — or pushing the switch that controls the lift to send it down again, thereby endangering the 1 visitor at ground level. In OC, all was the same except that there was 1 visitor on the upper level and 5 downstairs, and the participant had to choose between doing nothing — thereby endangering the life of the 1 — or sending the lift down and endangering the lives of the 5.

During the experiment, participants' interactions with the wand were recorded in a log file on the machine. After the experiment, participants were interviewed by the experimenter to discuss their experiences in the VR and also to debrief them about the purposes of the experiment. They were also informed that there was no 'correct' action that they should have taken. Finally, they were given a post-questionnaire which consisted of 3 classic moral dilemmas (same as survey question TA, TO, FA).

3. RESULTS

3.1 Online survey

Table 2: Questionnaire Survey result. The numbers in bold are the first questions in each condition. *P* is the significance level for tests of difference of proportions between the two conditions (SAC and SOC)

	Overall (n=80) %	SAC (n=38) %	SOC (n=42) %	p
LA	84	84	83	.14
TA	86	92	81	.14
FA	13	13	12	.87
LO	31	21	40	.05
TO	6	13	0	.02

We have collected data from 80 participants (38 in SAC, 42 in SOC) who visited our online survey webpage, among them 66% male and the average age is 35 (± 9 S.D.) with no significant difference regarding gender or age between the 2 conditions. The overall scores for Trolley Action (86%) and Footbridge (13%) are very similar to those of previous studies (Hauser et al., 2007) which included a very large sample size ($N > 2000$, Trolley Action: 85%, Footbridge: 12%).

From Table 2 it can be seen that in the case of the two omission conditions there is a significant difference between the proportions, meaning that the order of presentation of the questions had an effect. This is in line with what has been found by (Liao et al.). Since our aim was to collect judgements that were unbiased by answers to previous questions, we used those percentages corresponding to participants' first exposure to these questions as the baseline proportions with which to compare the results of the VR study. These are the bold values in Table 2.

3.2 Virtual Reality

Thirty-six participants attended the study, 9 in each condition. There were 29 males and 7 females with average age 30 (± 6.7 S.D.). Since this was a pilot study participants were only recruited around the Computer Science department at UCL, and the computer literacy level was in general quite high.

3.2.1. Behavioural results

During the experiment participants interacted with the scenario through a wand with two buttons. When they pressed button 1, as shown in Fig. 3, the lift moved from position A to B, or B to A; when button 2 was pressed, the lift moved between B and C. Their button pressing behaviour was recorded in a log file. Participants first went through two training sessions. Here, they learned that when a virtual visitor stepped onto the lift from the lower floor (A), they should press 1 to bring it to B, and then press 2 to bring it to the first floor (C).

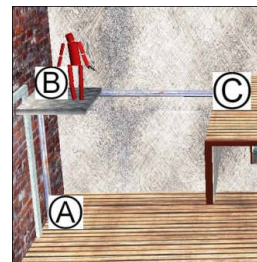


Figure 3. Positions of the Lift

After successfully going through the training sessions, participants were left alone in the scenario, either in the CAVE or in front of the desktop to operate the lift. After 6 virtual characters had entered the gallery and were looking at the paintings, another character entered and stepped onto the lift from the lower floor (position A). This character might have been seen as distinct from the others, by colour, by the way that it was walking; for those who were highly observant, they would see that the individual was carrying a gun. Participants then, as trained, pressed button 1 to bring this character to B. However just before the lift arrived at B, this character started shooting those on the upper floor. When the lift reached B, this put participants in a critical situation where they

had to make a decision according to the unexpected incident. When the lift reached B, they could either press 2 to bring the lift to C (as trained), or press 1 to bring the lift back to A, or do nothing and leave the lift at B. In the experiment, some participants pressed a button and then stopped; others pressed a button by mistake or changed their mind, and pressed another one immediately. Therefore the experimenter waited for 5 seconds after which the scenario was terminated.

A critical point in analysing the button pressing data was to identify the equivalent behaviour to participants' answer to the question: "would you push the switch?" In other words, how do we compare data from participants' behaviour to our survey data?

We answer the above question by analysing the following features that we extracted from our data: the first button they press, the final position of the lift, and how many times they pressed the button.

3.2.1.1 First Button

Right after the shooting, some participants pressed a button immediately, or hesitated for a few seconds before pressing a button, others (only in OC) did not press any button at all. Here for the first button, pressing "1" would bring the lift to the ground floor, and pressing button "2" would leave the lift upstairs which is equivalent to doing nothing. Therefore the percentage of participants who first pressed button "1" is equivalent to answering "Yes" to the question "would you press the switch?"

Table 3: Percentage of participants who first pressed button "1". P value is from Test of Proportions against the proportions in bold from Table 2

	AC (84%)		OC (40%)	
	Proportion %	P	Proportion %	P
CAVE (n=9)	33	.00	22	.20
Desktop (n=9)	44	.02	11	.01
Overall (n=18)	39	.00	17	.01

When compared to the questionnaire proportions from Table 1, there is a significant difference on most variables. This could be because the first button pressed by the participants might be a reaction due to "shock" and panic, and as a result the first button pressed by them during this time might not reflect their real intension. Therefore in the following we examine another feature: the final position of the lift.

3.2.1.2 Final position

As mentioned above, we define a pause between two button-click events longer than 5 seconds as "termination", and the position of the lift at the termination is defined as the final position of the lift. Here we summarise the percentage of participants

who left the lift at position "A", which is equivalent to answering "Yes" to the question "would you push the switch?"

As shown in Table 4, no significant difference was found when compared to the questionnaire proportions in Table 2. In AC, the CAVE condition has a greater proportion than the Desktop, however this is not significant ($p=0.24$).

Table 4: Percentage of participants who pushed the switch to move the lift down to ground level at termination. P is the significance level for the test of against the questionnaire proportion taken from Table 2.

	AC (84%)		OC (40%)	
	Proportion %	P	Proportion %	P
CAVE (n=9)	89	.64	22	.20
Desktop (n=9)	67	.27	22	.20
Overall (n=18)	78	.53	22	.07

3.2.1.3 The number of buttons pressed

As mentioned above, some participants did not press the button, or pressed it once, others pressed it more than once. This could be because they were panicked and pressed the wrong button by mistake. Here we give the percentages of those who pressed the buttons more than once.

Table 5: the Percentage of Participants who pressed buttons more than once

	AC %	OC %	Overall %
CAVE (n=9)	67	56	61
Desktop (n=9)	44	22	33
Overall (n=18)	61	33	47

As shown in Table 5, there was no difference between AC and OC. However, there is a trend that in the CAVE condition, in both AC and OC, a higher percentage of participants pressed more than one key (overall Cave against Desktop, test of proportions, $p = 0.08$). This suggests that it is possible that participants were more likely to panic in the CAVE, as compared to desktop.

4.2.2 Post-questionnaire results

After the experiments, participants were given a questionnaire containing three moral dilemma problems (TA, TO, and FA). These questions were asked in order to reveal whether participants' moral choices were affected by their experience in the experiment. FA asked whether participants should push a fat man with a heavy backpack off a bridge in order to stop a train hurtling towards another 5 people. For this particular question, participants who experienced CAVE were significantly more likely to give a utilitarian answer ('yes': 33%) than those using the Desktop display (0%, test of proportions $p<0.01$). This is also counter to results from our questionnaire survey (13%).

4. DISCUSSION

First, participants' behaviour in VR confirmed that, although their first reaction might have been accidental, caused by shock, their final decision (especially in the CAVE) was consistent with the survey results. Secondly, the behavioural result also pointed out the possibility that participants were more likely to "act" in order to achieve a utilitarian outcome in the CAVE. This tendency also extended to post-questionnaire results (i.e., after experiencing the CAVE participants were more likely to give a utilitarian answer). However, one could argue that it might be that the more utilitarian view of participants in the CAVE condition was something they had before the experiment rather than a result of their VR experience. Taking this view we examined the data and found that, out of 6 participants who chose "yes" for Footbridge (all from CAVE condition), 2 made a non-utilitarian choice in VR. Though our sample size is small, these results do not support the idea that they were utilitarian before the experiment.

5. CONCLUSION

This was a pilot study designed to explore the feasibility of both Immersive VR and desktop VR for exploring the potential differences between hypothetical judgments and more behaviour/action-like responses. The number of participants was low, thus all statistical tests had low power. We were interested in observing, for the first time, how people would respond in practice when faced with a novel moral dilemma, and whether their responses might be different in an immersive Cave system compared to a desktop system. Our preliminary results suggest that VR can be used to effectively study people's responses in this kind of situation. On a more qualitative level, we observed nervous and panicked responses of the participants, and the post-experiment discussions with the participants supported the notion that participants had found themselves responding as if it were real; this will be quantified in future studies by collecting physiological data such as skin conductance and heart rate. Second, there seem to be differences between the responses of people in the Cave compared to the desktop. However, if it turns out that with a larger sample size no such differences emerge then the less expensive and more portable desktop version can be implemented. Third, we found an unexpected effect of the Cave experience on participants' subsequent responses to the dilemmas in the questionnaire: those in the CAVE adopted a more utilitarian response counter to the predominant attitude that people take in such questionnaires. This is interesting, but also raises ethical concerns, since we need to examine whether having the Cave

experience would unduly influence people's beliefs and perhaps decisions in everyday life.

5. ACKNOWLEDGMENT

This research is funded by the Leverhulme Trust project "The exploitation of immersive virtual reality for the study of moral judgements". Special thanks to Prof. Marc Hauser for contributing to the experimental design and for commenting on an earlier draft.

6. REFERENCES

- CRUZ-NEIRA, C., SANDIN, D. J. & DEFANTI, T. A. Year. Surround-screen projection-based virtual reality: the design and implementation of the CAVE. In, 1993. ACM, 135-142.
- FOOT, P. 1985. The problem of abortion and the doctrine of double effect, Blackwell.
- HAUSER, M., CUSHMAN, F., YOUNG, L. & KANG XING, J. 2007. A dissociation between moral judgments and justifications. *Mind & Language*, 22, 1-21.
- KNOBE, J. 2003. Intentional action and side effects in ordinary language. *Analysis*, 63, 190.
- LIAO, S. M., WIEGMANN, A., ALEXANDER, J. & VONG, G. Putting the Trolley in Order: Experimental Philosophy and the Loop Case.
- MACHERY, E., MALLON, R., NICHOLS, S. & STICH, S. P. 2004. Semantics, cross-cultural style. *Cognition*, 92, B1-B12.
- MILGRAM, S. 1963. Behavioral study of obedience. *abnorm. soc. PsychoJ*, 67, 371-378.
- PERTAUB, D. P., SLATER, M. & BARKER, C. 2002. An experiment on public speaking anxiety in response to three different types of virtual audience. *Presence: Teleoperators & Virtual Environments*, 11, 68-78.
- 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.
- SANCHEZ-VIVES, M. V. & SLATER, M. 2005. From presence to consciousness through virtual reality. *Nature Reviews Neuroscience*, 6, 332-339.
- SLATER, M. 2004. How colorful was your day? Why questionnaires cannot assess presence in virtual environments. *Presence: Teleoperators & Virtual Environments*, 13, 484-493.
- SLATER, M., ANTLEY, A., DAVISON, A., SWAPP, D., GUGER, C., BARKER, C., PISTRANG, N. & SANCHEZ-VIVES, M. V. 2006. A virtual reprise of the Stanley Milgram obedience experiments. *PLoS One*, 1, 39.
- THOMSON, J. J. 1971. Individuating actions. *The Journal of Philosophy*, 68, 774-781.