

An Investigation into the Effects of Social Influence on Moral Behaviour using Immersive Virtual Reality

Jacob Thomas Thorn

Thesis submitted to University College London for the degree of
Doctor of Philosophy

Department of Computer Science

University College London

06/01/2020

I, Jacob Thomas Thorn, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the work.

Acknowledgements

First, I would like to thank my supervisor Mel Slater, most of all for his trust over the last three and a half years, but also for his constant guidance and help in all aspects of this thesis and the PhD as a whole.

I would also like to thank David Swapp for his additional support, whether he was piloting my next experiment, helping me with any technical issues in the HMD lab or providing otherwise invaluable advice.

I would also like to thank my company supervisor, Pablo Bermell-Garcia, for supporting this project so wholeheartedly.

I also want to thank my family and friends at UCL. To my wife Zhou for being so supportive for the entire PhD, my parents for providing extra support in the final few months and to Ben, David, Drew, Maria and Sebastian for their friendship over the last three and a half years.

Lastly, I would like to thank all of the voice actors who provided their voices for the characters in my experiments: Alex and Sophie Burr, Ian Giblet, Mark Hall, Olivia Slater-Sanchez, Jared and Hope Somerville, Stephen Thorn and Luke Woodford.

I also have an anti-acknowledgement for my one year old son William, without whom this thesis would have been completed much sooner, but who I am glad is here to see the completion of it.

Abstract

Much of the research surrounding social influence investigates its effects in specifically non-moral situations while almost no research has looked at its effects during moral emergencies. At the same time, studies of moral psychology tend to focus on the intricacies of moral decision-making during the responses of individual participants. This thesis aims to bridge this gap between social influence and moral psychology by having participants respond to moral dilemmas while under the duress of social influence.

In order to investigate the effects of social influence on moral behaviours, immersive virtual reality (IVR) was used, allowing participants to be placed in a life-like virtual simulation of events that they would normally only read about in a text-based vignette, probing their observed moral behaviours instead of just their abstract moral judgments. The benefits of using IVR include the ethical and controllable nature of questionnaires along with the verisimilitude of real-life.

Another focus of this thesis is to compare moral judgments to moral behaviours. In two out of the three studies presented in this thesis, the virtual moral dilemma was replicated in a text-based questionnaire in order to compare the results from the two media. Moral judgments in response to text-based moral dilemma can miss out key contextual information such the motoric feedback of having to physically act out a movement. These factors can lead to a divergence between moral judgments and behaviours.

The thesis starts with a literature review on IVR technology and moral decision-making and social influence research. After this, the three studies conducted as part of this thesis are described. The major findings from these studies include the demonstration of a preference to take action regardless of outcome only when in IVR and the inability for compliance attempts to influence specifically moral behaviour.

Impact Statement

Perhaps the most impactful outcome of this thesis is the finding that participants may have a preference for action in moral dilemma presented in immersive virtual reality (IVR) compared to identical questionnaire descriptions. Previous results found that participants had a preference for utilitarianism in IVR (a response that is intimately coupled with taking an action) compared to questionnaire results and posited that this outcome was related to the saliency of the outcomes of the moral dilemma, i.e., the sensory confirmation that people will die unless an action, no matter how grisly it might be, is taken (Francis et al., 2016; Francis et al., 2017; McDonald, Defever, & Navarrete, 2017). The current result potentially explains these findings by showing that, instead of a preference for utilitarianism, participants may simply have had a preference to act, where the alternative was to do nothing. However, it disagrees with the proposed hypothesis for these results which implicated salient outcomes, where the current results seem to indicate that it is the saliency of actions that contributes to our moral behaviours.

In order to more clearly understand the contribution of salient actions and outcomes to our moral decision-making, as well as any potential preferences for utilitarianism or taking action in IVR, a study should be run that decouples utilitarianism and having to take an action into separate conditions, something previously done by Conway and Gawronski (2013) and Gawronski, Conway, Armstrong, Friesdorf, and Hütter (2016).

Another impactful result showed that participants did not adhere to a compliance request regarding and when responding to a moral dilemma and were similarly less likely to adhere when responding to a moral compared to a non-moral dilemma. This is counter to many previous studies that showed the willingness of people to alter their moral dispositions when faced with a consensus opinion, even when that opinion is very uncommon (Bostyn & Roets, 2017; Kelly, Ngo,

Chituc, Huettel, & Sinnott-Armstrong, 2017; Kundu & Cummins, 2013; Lisciandra, Postma-Nilsenová, & Colombo, 2013). The potential reason for this might be the form of medium used to elicit moral decision-making, where, for example, participants have the option to lie in response to a moral judgment task, e.g., when responding to a questionnaire, but not during a task requiring real actions. Other differences such as the saliency of the scenario might also impact this interaction.

Both the difference between the preference to take an action and the efficacy of social influence in IVR representations of moral dilemmas compared to textual vignettes in this thesis shows the necessity for future moral decision-making studies to make use of current technology, such as IVR, to present participants with more realistic representations of moral dilemmas.

Contents

Abstract.....	4
Impact Statement.....	5
List of Figures.....	12
List of Tables.....	14
Chapter 1 – Introduction.....	15
1.1 Research Problem	16
1.2 Research Questions	18
1.3 Scope	21
1.4 Contributions	23
1.5 Structure	24
Chapter 2 – Background.....	26
2.1 Immersive Virtual Reality	26
2.1.1 IVR in Social Psychology	27
2.1.2 Technology	31
2.1.3 Presence	34
2.2 Morality	40
2.2.1 Evolution of Morality	40

2.2.2 Moral Philosophy	43
2.2.3 The Trolley Problem	47
2.2.4 Moral Psychology	63
2.2.5 Moral Neuroscience	73
2.2.6 Moral Action	81
2.3 Social Influence	87
2.3.1 Models of Social Influence	88
2.3.2 Compliance and Persuasion	92
2.3.3 Social Cognitive Processes	97
2.3.4 Moral Influence	99
2.3.5 Virtual Influence	103
2.4 Chapter Summary	108
Chapter 3 – Methods.....	111
3.1 Moral Dilemmas	111
3.2 Virtual Environments	112
3.2.1 Embodiment	113
3.2.2 Virtual Characters	116
3.2.3 Virtual Conversation	118
3.3 Equipment	118
3.4 Participants	119
3.5 Procedures	120
Chapter 4 – Moral Judgment Versus Moral Behaviour.....	125
4.1 Introduction	126

4.2 Methods	129
4.2.1 Scenario	129
4.2.2 Experimental Design	130
4.2.3 Participants	131
4.2.4. Equipment	131
4.2.5 Procedures	132
4.2.6 Response Variables	133
4.3 Results	134
4.3.1 Presence Data	134
4.3.2 Response Data	136
4.4 Discussion	139
4.5 Conclusion	141
Chapter 5 – Social Influence and Omission Bias on Moral Behaviour.....	143
5.1 Introduction	144
5.2 Study One – Methods	148
5.2.1 Scenario	148
5.2.2 Experimental Design	149
5.2.3 Participants	150
5.2.4 Equipment	150
5.2.5 Procedures	151
5.2.6 Response Variables	153
5.3 Study One – Results	153
5.3.1 Presence Data	153

5.3.2 Behavioural Data	155
5.4 Study Two – Methods	158
5.4.1 Experimental Design	158
5.4.2 Participants	158
5.4.3 Procedures	159
5.4.4 Response Variables	159
5.5 Study Two – Results	159
5.6 General Discussion	161
5.6.1 Social Influence	161
5.6.2 Omission Bias	164
5.6.3 Moral Judgment and Behaviour	166
5.7 Conclusion	168
Chapter 6 – Social Influence During Moral and Non-Moral Dilemmas.....	169
6.1 Introduction	169
6.1.1 Background	169
6.1.2 Present Research	171
6.1.3 Predictions	173
6.2 Methods	175
6.2.1 Experimental Design	175
6.2.2 Participants	176
6.2.3 Equipment	177
6.2.4 Procedures	177
6.2.5 Response Variables	179

6.3 Results	179
6.3.1 Presence Data	180
6.3.2 Behavioural Data	185
6.4 Discussion	189
6.5 Conclusion	191
Chapter 7 – Conclusions.....	192
7.1 Moral Judgment versus Behaviour Experiment	193
7.2 Social Influence and Omission Bias Experiment	194
7.3 Moral versus Non-Moral Experiment	196
7.4 Contributions	196
7.4.1 Methodological Contributions	196
7.4.2 Substantive Contributions	199
7.5 Future Work	200
7.6 Conclusion	201
References.....	203
Appendix A – Moral Dilemmas.....	217
Appendix B – Forms, Sheets and Questionnaires.....	222
Appendix C – Supplementary Material.....	255

List of Figures

Figure 2.1 – Head-mounted display example (Oculus Rift CV1).....	32
Figure 2.2 – The Switch Dilemma.....	48
Figure 2.3 – The Footbridge Dilemma.....	48
Figure 2.4 – Moral emotions.....	66
Figure 2.5 – The Social Intuitionist Model.....	72
Figure 2.6 – Brain regions implicated in moral decision-making.....	74
Figure 2.7 – Threshold of social influence.....	106
Figure 3.1 – Inverse kinematics diagram.....	114
Figure 3.2 – Immersive virtual reality support in Unity.....	115
Figure 3.3 – Embodied character setup.....	115
Figure 3.4 – OptiTrack gloves.....	116
Figure 3.5 – Wrist movements.....	116
Figure 4.1 – The scenario.....	130
Figure 4.2 – Box plots for presence questions.....	134
Figure 4.3 – Box plots for plausibility of the situation.....	135
Figure 4.4 – Box plots for plausibility of the virtual people.....	135
Figure 4.5 - Prior and posterior distributions for the three cases.....	138

Figure 5.1 – The scenario.....	149
Figure 5.2 – Box plots for presence questions.....	154
Figure 5.3 – Box plots for plausibility of the situation.....	154
Figure 5.4 – Box plots for plausibility of the virtual people.....	155
Figure 5.5 – Bar chart showing the proportions who acted by condition.....	156
Figure 5.6 – Bar chart showing the proportions who stopped by condition.....	156
Figure 5.7 – Bar chart showing means and standard errors for the time.....	157
Figure 5.8 – Proportion acted by condition.....	160
Figure 5.9 – Proportion stopped by condition.....	160
Figure 6.1 – The scenario.....	172
Figure 6.2 – Box plots for the Place Illusion questions in Table 3.3.....	180
Figure 6.3 – Scatter diagrams for ypi on each of the scores to the questions in Table 3.3.....	181
Figure 6.4 – Bar chart showing means and standard errors of ypi by the conditions.....	182
Figure 6.5 – Scatter diagram of the factor variable ypsi_situation on each of the original scores to the questions in Table 3.4.....	182
Figure 6.6 – Bar chart showing means and standard errors of ypsi_situation by the conditions	183
Figure 6.7 – Scatter diagram of ypsi_people on the scores to the questions in Table 3.5.....	184
Figure 6.8 – Bar chart showing means and standard errors for ypsi_people by condition.....	185
Figure 6.9 – Proportions and standard errors who saved the child as their first action.....	186
Figure 6.10 – Proportions and standard errors who saved the child as their first action.....	186
Figure 6.11 – Means and standard errors for the number of attempts to save the adults.....	188
Figure 6.12 – Means and standard errors for the number of attempts to save the child.....	188
Figure 6.13 – Number of times the participants tried to save the child and the adults.....	189

List of Tables

Table 3.1 – Unity versions.....	113
Table 3.2 – Animation variables.....	117
Table 3.3 – Place Illusion questions.....	121
Table 3.4 – Plausibility for the situation questions.....	122
Table 3.5 – Plausibility for the virtual people questions.....	122
Table 3.6 – Moral dilemma questionnaire.....	123
Table 4.1 – Proportions who save the 5 in conditions equivalent to the Switch Dilemma.....	127
Table 4.2 – The number who saved the 5 in the questionnaire responses.....	136
Table 4.3 – Results of the Bayesian analysis.....	137
Table 5.1. – Study two conditions.....	150
Table 5.2 – Study two participant breakdown.....	158
Table 5.3 – Frequency of actions and omissions by condition.....	159
Table 5.4 – Frequency of participants who stopped or continued by condition.....	160
Table 6.1 – Study three conditions.....	175
Table 6.2 – Proportion who saved the child (as their last action).....	187

Chapter 1

Introduction

More than 40 years ago, the trolley problem was born from the comparison of a set of moral dilemmas and is most often demonstrated using two specific cases. The first, dubbed the Switch Dilemma, presents the responder with a situation where they can divert an oncoming train away from five people, but by doing so will send it towards one. Although people tend to have a strong aversion to actively participating in the death of a person, acting in this dilemma is generally considered permissible (Hauser, Cushman, Young, Kang-Xing, & Mikhail, 2007). The second, dubbed the Footbridge Dilemma, is identical but instead requires the bystander to push a man off of a footbridge and in front of the train in order to stop it, instead of being able to divert it. Contradictory to the first dilemma, it is generally considered impermissible to act in this dilemma, leading most to let the five die (Hauser et al., 2007). The question the trolley problem proposes then is as such: why do people respond differently to consequentially identical moral dilemmas, where the only differences between them are the actions the person has to take in order to resolve the situation? (Thomson, 1985). At first it might seem obvious, physically assaulting someone is bad in and of itself whereas throwing a switch is not. However, research has shown that there is much more at play here than initially meets the eye. Many studies (Cushman, Young, & Hauser, 2006; Greene et al., 2009; Hauser et al., 2007) have indicated that people unconsciously follow several heuristics embedded in these moral dilemmas such as the doctrine of double effect which makes us apprehensive about using someone as a means rather than an end, or omission bias which hampers our moral decision-making by making us place greater agency over actions than

omissions (a fallacy when the agent has equal intentions and complete information in both scenarios) (Spranca, Minsk, & Baron, 1991).

Beyond any true differences between the two examples given above is the debate about whether or not these differences should even matter. Assault the man or throw the switch, he will die and so the fact that we might have to assault him should be meaningless. In reality however, people do regard the distinction as important and often cite it as justification for not pushing the bystander off of the footbridge while simultaneously agreeing to throw the switch to divert the train (Hauser et al., 2007). Scientists have implicated emotions as responsible for this seemingly contradictory behaviour, claiming that pushing the man off of a height elicits very strong negative emotions such as guilt or shame, which inhibits our ability to comply with the demand (Greene, 2008). Throwing a switch on the other hand is less associated with harm and therefore people are more willing to take action.

The trolley problem and the dilemmas that came with it were adapted at the start of the 21st century for a different purpose. It now provided a useful methodology for probing the constituents of responder's moral judgments. By very slightly altering pairs of moral dilemmas, researchers could isolate the specific qualities of each that contribute to our own personal morality and by using fMRI technology could identify the neural networks underlying our moral judgments (Greene, Sommerville, Nystrom, Darley, & Cohen, 2001). This provided the field of moral psychology with a common procedure, allowing comparisons to be made and models to be formed. The most famous of these, Greene et al's (2008) Dual-Process Theory of Moral Judgment (DPTOMJ), lent enormous credence to the involvement of emotions in our moral decision-making, a claim that was denied for centuries before regaining momentum in the 1980s (Haidt, 2008).

1.1. Research Problem

This methodology however is starting to show its age and an upgrade is desperately needed if further progress is to be made to the field of descriptive morality, i.e., the study of people's real moral dispositions (Bauman, McGraw, Bartels, & Warren, 2014). The issue is not with the dilemmas themselves which are likely to remain timeless, but in the medium in which they are

presented. Text-based vignettes and questionnaires are the method of choice for many research papers over the past two decades due to their simplicity and ethical tolerance. However, these may fail to provide all the necessary contextual cues that impact our decision in a real situation and should only be seen as a coarse estimation of a person's true moral behaviour (Parsons, 2015).

Recently, more ambitious studies have attempted to directly investigate our moral behaviours by using a range of more complex methods in order to invite people to take physical actions as opposed to providing moral judgments to hypothetical questions. The most common method, used currently in eleven published papers, is to use immersive virtual reality to simulate a moral dilemma (Faulhaber et al., 2018; Francis, Gummerum, Ganis, Howard, & Terbeck, 2018; Francis et al., 2016; Francis et al., 2017; Friedman et al., 2014; McDonald et al., 2017; Navarrete, McDonald, Mott, & Asher, 2012; Pan & Slater, 2011; Patil, Cogoni, Zangrando, Chittaro, & Silani, 2014; Skulmowski, Bunge, Kaspar, & Pipa, 2014; Sütfield, Gast, König, & Pipa, 2017). This bypasses many of the ethical problems faced with studying real moral actions such as participant deception and potential psychological harm while retaining realistic responses from participants (Slater, 2009). Aside from this, several other studies have used a variety of techniques to research moral actions such as using mice instead of humans (Bostyn, Sevenhant, & Roets, 2018) or using mild electric shocks instead of severe harm (or death) (FeldmanHall et al., 2012). Many of these studies are taking the first steps, focusing solely on the first question that should be asked in this new field, "Are moral actions distinct from moral judgments?". The answer to this question seems to be yes as many of the newer studies are showing differences between moral actions and judgments although not in the way predicted by the DPToMJ (Francis et al., 2017). This burgeoning new field is still young however and more questions need to be thought of, asked and answered to further the quality of our understanding about moral behaviours.

Another aspect lacking from traditional moral dilemma research regards the lack of investigation into moral judgments or behaviours in an interpersonal setting. To date, only four studies have used the trolley problem methodology in order to investigate how our judgments can be influenced by the presence (real or imaginary) of others, of which all four use written questionnaires in order to elicit moral judgment (Bostyn & Roets, 2017; Kelly et al., 2017; Kundu & Cummins, 2013; Lisciandra et al., 2013). As such, the impact of social influence on moral actions (during a trolley problem style dilemma) is currently unknown. It should be noted that research on the influence of

others on our moral actions does exist, although it is usually targeted at very specific moral norms such as torture or capital punishment (Aramovich, Lytle, & Skitka, 2012; Fiske, Harris, & Cuddy, 2004) or less harmful moral acts involving embarrassment (Warren & Smith-Crowe, 2008).

Furthermore, what research there is looking at socially impacted moral judgments has utilised norm conformity and not compliance from the demands of others. These two differ such that conformity refers to the larger population of a member group asserting an influence on an individual by expressing a shared value, setting it as the norm and thereby encouraging the individual to also hold that value. Compliance attempts instead involve the request of an individual to hold that value without setting it as the norm, potentially asking them to go against something they believe in (Cialdini & Goldstein, 2004). This might have drastically different results in a moral setting as moral norms are generally regarded as much more important and central to our self-concept than social norms (Lisciandra et al., 2013).

Both of these aspects relate to the research of moral dilemmas in a more realistic setting. While the translation of moral judgment into behaviour might relate to this more obviously, it is unlikely that in the few professions that potentially require moral decision-making to take place, e.g., in the military, that this is handled in a solitary fashion and it is more likely that many voices are heard before any final decision is made.

1.2. Research Questions

This thesis focuses mainly on furthering the knowledge around interpersonal moral decision-making, that is, how people respond to moral dilemmas when in a group or otherwise being subjected to compliance-based social influence. This is the main focus of studies two and three. Studies one and two also look at the potential difference between moral judgments and behaviours by constructing moral dilemmas in both questionnaire form and in IVR. Throughout this thesis, I will consider the following questions and attempt to answer each of them.

1. Is there a measurable difference between people's moral judgment and behaviours in response to a moral dilemma involving a child? (Study One)

The inclusion of a child in the moral dilemma was meant to introduce a novelty to the dilemma that has not been studied before. Firstly, moral dilemmas have become more popular recently, with many people at least knowing of problems such as the switch dilemma, the inclusion of a less research aspect such as children ensured that participants did not recognise the moral dilemma before making a decision, preventing any pre-emptive decision-making that could have affected the results.

More importantly however, including a child allowed the comparison of a moral dilemma involving a more concrete character between IVR and questionnaire responses. Past versions of the switch dilemma typically include individual or groups of generic characters that are numerically balanced such that the death of either party is equally contemptable to the other (given what the participant is asked to do to save either party). This is also the case in many of the studies that have looked at moral dilemmas in IVR previously (i.e., Pan & Slater, 2011; Navarrete et al., 2012). However, it is possible that having only generic characters results in participants not truly appreciating the outcomes of the moral dilemma. Five deaths is obviously worse than one, but where each death of a character is equal, the deaths can be factored out of the equation and result in a simple problem of arithmetic, "five is greater than one" (Tassy, Oullier, et al., 2013). Previous studies that have aimed to see if identifying the characters in moral dilemmas as specific people or types of people sways participants judgments found that responses to moral dilemmas indeed changed depending on the identified person (Cikara, Farnsworth, Harris, & Fiske, 2010) and even changed between objective and subjective judgment of the dilemma, where subjective judgment might be thought of as similar to actions in an IVR (Tassy, Oullier, et al., 2013). In order to further investigate how the inclusion of an alternative character impacts moral decision-making between subjective (IVR in our case) and objective (questionnaire) responses, it was decided that a child, a character with an inherently greater value of life compared to an adults (Kawai, Kubo, & Kubo-Kawai, 2014), be included in the moral dilemma.

Beyond this, research into the difference between moral judgment and behaviour is important moving forward as it better informs the wider community on how moral dilemmas should be presented to participants in psychological studies. A moral judgment is the response of a participant or any other person to a text-based moral dilemma, i.e., one presented in a questionnaire. In this sense, no physical action or behaviours are needed and the responder is required to make assumptions about how they would actually act in such a situation. Conversely,

moral behaviours are those that arise from real or simulated moral dilemmas, for example, those presented in IVR. Moral behaviours are different from judgments in that additional contextual information is added to the situation which in turn recruits distinct or further cognitive processes, potentially changing the persons reaction to the dilemma. Finding differences between these two moral expressions would support the use of IVR in future investigations of moral decision-making as it is more likely to reflect actual moral behaviours rather than abstract responses to questionnaires.

2. Is the difference between moral judgment and behaviour modulated by omission bias?

(Study Two)

Building on the first question, this one ponders how omission bias interacts between moral judgment and behaviour given that the difference between actions and omissions is something more likely to be salient in IVR than in a questionnaire. Briefly, omission bias is the preference to not act when the outcomes of that action are negative, even if not acting produces even worse results (Baron & Ritov, 2004). For example, the reader is referred back to the switch and footbridge dilemmas presented at the start of this Chapter where responders are apprehensive about diverting the train or pushing the man even though it would save five lives. This aspect of our morality has been heavily researched over the past 30 years, but not using IVR to measure our moral behaviour instead of judgment.

3. Are moral behaviours influenced by compliance attempts? (Study Two)

a. Is this modulated by omission bias? (Study Two)

Question 3 and 3a begin the focus on interpersonal moral decision-making. Research into social influence has shown the ease at which someone can be influenced, including the studies that investigated the effect during moral decision-making. The current study looking at these questions introduces new aspects by using IVR to study moral behaviour instead of judgment. Where participants in previous studies might have altered their moral outward judgments, but retained their original position in secret, taking moral actions in a way commits the person to that opinion and thus might be harder to influence.

In the same study, we aimed to see whether omission bias (described above) interacts with social influence with the hypothesis being that it would be easier to influence someone to omit rather than to act.

4. Is there a measurable difference between the efficacy of compliance attempts between moral and non-moral dilemmas? (Study Three)

This final question was born from the results of the previous one which showed that compliance attempts were ineffective at influencing people's moral behaviour. To investigate this further, a comparison was needed between moral and non-moral situations to see if a contributing factor towards the failed compliance attempts was due to the moral nature of the situation.

Each of these questions are interesting topics in and of themselves, but also provide the groundwork for next question. Question one provides a start by asking whether morality is altered between judgment and behaviour in a novel scenario involving the life of a child. Question two then takes this a step further and introduces the moral principle of omission bias into the equation. Question three marks the transition into research regarding social influence and four builds on top of the results of the second study by broadening the question to include non-moral situations.

1.3. Scope

Haidt (2007) asserts that there are five foundations of morality: harm, fairness, loyalty, authority and purity and successful theories have been proposed that incorporate all of these aspects, attempting to explain them and their origins in a singular model (Graham et al., 2013). This thesis however will focus primarily on harm-based morality and discuss more in-depth the models of moral judgment, such as the DPToMJ, that deal specifically with our aversion to harm as this is typically the aspect of our morality that is invoked when we deal with dilemmas related to the trolley problem. We worry about harming the man by pushing him to his death, but also about the harm that would come to the five were we to let them die. We are averse to these things and our morals drive this harm-based aversion. Other aspects of our morality can certainly be incorporated into the trolley problem, for example, including family members might elicit a different response due to our loyalty to them or authority might force us to partake in acts we otherwise would not have, such as in the Milgram experiments (Milgram, 1963), but this is not the focus of the current thesis.

Furthermore, this thesis does not aim to decipher how people should respond to certain moral dilemmas, but simply to discover how they do respond to them.

During the three studies conducted, critical variables observed were of participants behaviour, i.e., what option they chose to take during the moral dilemma. Other types of assessments, such as physiological measurements like skin conductance, heart rate and pupil dilation, were not used. These types of measurements have been found to be related to the emotional condition of a person and can thus provide further insight into the state of mind of participants during their behavioural actions. For example, Francis et al. (2016) found that, compared to their respective base levels, moral dilemmas encountered in immersive virtual reality were more arousing than their text-based counterparts. Physiological measurements can also be used to assess the presence that participants feel in a virtual environment, although this is usually only useful in scenarios that include arousing stimuli, such as during a moral dilemma. Previous studies have used physiological responses to great effect to measure participants heart rates while they stand over a virtual pit (Meehan, Insko, Whitton, & Brooks Jr, 2002), showing that participants do (at least sometimes) react realistically to their virtual surroundings. The measures were not used however as the research questions outlined above were addressed specifically by the behavioural responses of participants and not by the amount of stress that the situations caused.

During the second two studies, the source of social influence comes from virtual characters (VCs) present in the environment. These characters were computer controlled and while participants were never explicitly told this, it is possible that they became aware of it very soon into their interactions with them as the VCs followed a very rigid script when talking to the participant. Instead of using computer controlled characters, it might have been possible to use a collaborative environment, where real actors controlled the VCs from other rooms in real-time using their own immersive virtual reality equipment. This technique was used by several other studies not related to moral dilemmas with one showing that people indeed responded more positively towards the human controlled characters over the computer controlled (Bailenson, Blascovich, Beall, & Loomis, 2003). On the other hand, it has been shown that participants will respond realistically to computer controlled VCs as evidenced from other studies. Slater et al. (2006), for example, showed that participants felt concern for a VC when instructed to administer electric shocks to them and Slater et al. (2013) attained realistic responses from participants using a very similar setup to what has been implemented in the current studies, i.e., using a conversation to form a

rapport before presenting a stressful incident. Additionally, a benefit of using immersive virtual reality is that each iteration of a scenario with each participant can be guaranteed to be the same. Computer controlled virtual characters will always say the same lines at the same time with the same inflection on their voices because the lines are pre-recorded. This provides near-perfect internal validity that using real-time actors could not achieve. For example, using actors might result in someone forgetting a line or acting different towards particular participants, possibly changing the way the participants interprets the social influence attempts.

1.4. Contributions

Contributions made over the course of this thesis include both substantive and methodological. The aim of the thesis was to investigate mainly how moral behaviours are affected by an interpersonal environment and secondly to further the discussion on the difference between moral judgments and behaviour. With regards to these aims, two key results were found that provide new information, illuminate the conclusions of previous studies and promote further investigation in the future.

Firstly, with the goal of seeing how social influence affects specifically moral behaviour, it was found that participants responses were not changed by the compliance attempts of others. This stands in contrast to previous studies where moral judgments in response to text-based moral dilemmas could be swayed by the presence (real and imagined) of others (Bostyn & Roets, 2017; Kelly et al., 2017; Kundu & Cummins, 2013; Lisciandra et al., 2013). The main difference here is the presentation of the moral dilemma, where previous studies only required judgments to be made, the studies as part of this thesis demanded physical actions to be taken on the part of the participant. With this result in mind, the third study then aimed to compare participants responses to moral and non-moral dilemmas while under the duress of social influence. These results, while not particularly strong, support the previous ones, showing that social influence is indeed less effective during moral, compared to non-moral, situations. This result specifically is the first to directly compare the effects of social influence during moral and non-moral situations.

Second, the result was found that participants preferred to take an action over an omission (where both options resulted in the same outcome) in response to a moral dilemma in IVR, but not in a questionnaire. This provides a new perspective into previous results showing that participants prefer utilitarianism over deontology in IVR (Francis et al., 2017), where utilitarian responses generally require acts, not omissions. However, previous studies attributed this phenomenon to the saliency of outcomes in IVR, something that this current result does not support, encouraging further research in future.

As for methodological contributions, the second study presented a novel way to present a moral dilemma to participants in IVR, ensuring that participants both understand what options are available to them while also not divulging the events that will happen. The study made use of an autonomous car that could interact with the participant through computerised speech. The participant was not told the details of the moral dilemma, but when the event happened, as the car was already involved in some fashion, it made sense for it to be able to relay the participants options to them. This might be translated into other moral dilemmas by having a bystander shout out participants choices to them as the dilemma is happening.

Finally, the literature review found in this thesis aims to add to the growing literature on immersive virtual reality, moral decision-making and social influence through its own literature review, where each area is discussed and analysed in isolation as well as in conjunction with each other, comparing and contrasting the findings of previously separate research topics.

1.5. Structure

Chapter 2 provides an in-depth review of the literature relevant to the topic at hand. This includes a discussion on immersive virtual reality, an explanation of its technology, the concept of presence inside a virtual environment and how it can be used to study social psychology. Morality is also discussed, from ancient philosophy to the trolley problem and the current research conducted in psychology, neuroscience and moral action research. Finally, social influence is broadly discussed before focusing on its relation to morality and immersive virtual reality.

Chapter 3 covers the methods used for the research. This section explains the decision to focus on trolley problem style moral dilemmas and details many of the shared technical aspects behind the virtual environments developed for the three studies.

Chapter 4 describes the first experiment. The aim of this study was to further investigate moral judgments versus moral behaviour and to do so using a novel moral dilemma in IVR. In this study, participants either encountered a moral dilemma in a virtual environment or read a description of it in a questionnaire. In order to introduce a novel aspect to the scenario, a young child was included where the participant was asked to either save them or five adults.

Chapter 5 describes the second experiment, which consists of two studies. The aim was to see firstly whether social influence was effective at changing moral behaviour, induced using IVR, and also whether this effect interacts with the notion of omission bias, i.e., is it easier to influence someone to act or to omit? The second study replicated the first, but instead presented the moral dilemma in an online questionnaire. Using the data from the first and second studies, we hoped to see whether omission bias operates differently during moral judgment and behaviour.

Chapter 6 describes the final study. The results from the second study indicated that social influence, specifically compliance attempts, was not effective during a moral dilemma. This was unexpected as traditional literature indicates the general success of compliance attempts during non-moral situations, especially towards friendly, in-group confederates (something we attempted to replicate in our virtual scenarios) (Cialdini & Goldstein, 2004). As such, this next study aimed to investigate this further by directly comparing social influence during moral and non-moral contexts thereby confirming whether social influence indeed operates differently during moral and non-moral situations or if compliance attempts failed for some other reason.

Finally, Chapter 7 provides a discussion relating to the three studies and the overall conclusions drawn from them, details the contributions made during the research and what further work could be done in the field.

Chapter 2

Background

2.1. Immersive Virtual Reality

In order to study the moral behaviour of participants in the three studies conducted as part of this thesis, immersive virtual reality (IVR) was used. IVR systems aim to transport people into a virtual world, typically with the use of a head-mounted display (HMD), which provides users with complete visual sensory information, as well as tracking equipment used to transport their body (most importantly their head movements which is used to translate their viewpoint) into the virtual world. Using this technology, participants can be placed into a morally charged scenario, which would be unethical to replicate in real-life, where their behaviours can be monitored and analysed.

Using real constructed events in order to study moral decision-making presents serious ethical issues while at the same time questionnaires are not equipped to evoke the same level of realistic responses, lacking contextual information and forcing responders to mentally simulate the situation. IVR offers a middle ground while retaining some of the benefits of both mediums, e.g., it has the contextual information of real life events, earning it at least some degree of external validation as well as offering the experimental control that questionnaires are known for, giving it internal validity as well.

As the studies conducted as part of this thesis are so reliant on the application of IVR, this section will cover the relevant information regarding it. Previous uses in social psychology are discussed, providing evidence for IVR's correct use in these projects, the technology behind IVR and the vital

concept of presence are then covered, offering more information about how IVR works and why it elicits realistic responses.

2.1.1. IVR in Social Psychology

“An attempt to understand and explain how the thought, feeling, and behaviour of individuals are influenced by the actual, imagined, or implied presence of others” – Allport’s (1985) definition of social psychology.

An issue in social psychology is the desire to study unethical behaviours or behaviours that arise from unethical situations. The issue here is that in order to study these phenomena, researchers would have to put undue stress on and deceive participants by having them take part in a staged (unknown to the participant) ethical dilemma. In the 1960s, Stanley Milgram used this methodology to investigate the effects of authority and persuasion (Milgram, 1963). Regardless of the amazing results from the study itself, the scientific community reacted to the study, along with the Stanford prison experiment (Haney, Banks, & Zimbardo, 1973) conducted some years later, with disapproval and many universities and other research institutions began requiring research experiments to pass through an ethical approval process before being conducted. Most importantly, these check whether participants are being lied to or are at a high risk of being mentally or physical harmed and as such it is no longer possible to run the types of studies as Milgram. While the protection of research participants is correctly a high priority, the unfortunate side effect is the death and no real subsequent replacement of an entire methodology of study.

Since then, researchers studying sensitive behavioural topics have had to choose one of two options: present the eliciting situation as a written vignette (Cushman et al., 2006; Hauser et al., 2007) or present the participant with a similar, but far less harmful, real-life construction of the event (Bostyn et al., 2018; Cushman, Gray, Gaffey, & Mendes, 2012; Eisenberger, Lieberman, & Williams, 2003; FeldmanHall et al., 2012). Written questionnaires have now become a staple of moral judgment research, where almost all evidence is drawn from participant’s responses to a battery of survey questions. In theory, these questionnaires are still a valid methodology given that Allport’s definition above includes the “actual, imagined, or implied presence of others,”

however, some feel that social psychologists have over-extended the inclusion of “imagined” and “implied” and that the stimuli presented in these studies does not do justice to the behaviours that would be found in real life (Blascovich et al., 2002). Parsons (2015) says that “a limitation of these hypothetical moral dilemmas is that while they have been effective in enhancing our understanding of moral decision making, they do little to expand our knowledge of how these vignettes translate into real-world behaviours... Consequently, results from both paper-and-pencil and fMRI investigations may overestimate cognitive processes and underestimate important dynamic situational and affective components of the moral dilemmas,” further arguing that static stimuli such as hypothetical moral dilemmas fail to replicate the multisensory, dynamic and contextually-rich atmosphere of real events.

In fewer words, written vignettes might be said to have low external validity, i.e., the results they produce are unlikely to be replicated in response to the real event. At the same time, (less harmful) real-life reconstructions suffer from low internal validity, i.e., the experimental variables are less controlled and could result in unknown differences between participants, conditions or study replications. For example, confederates and actors taking part in the experiment cannot be expected to act identically for each participant which might unknowingly alter a participants key responses.

The solution is of course IVR (Blascovich et al., 2002; Loomis, Blascovich, & Beall, 1999; Pan & Hamilton, 2018; Parsons, 2015). IVR circumnavigates the issues of harm and deceit that plagued Milgram by presenting virtual dilemmas instead of “real” ones (Slater et al., 2006). Physical harm is not a worry in the virtual world (except for unwanted collisions with real objects) and psychological harm as well as the aspect of deceit is minimised as participants are always aware that the events are virtual, nothing is real and no one is really in any danger. Despite this, participants do respond realistically to situations in IVR provided a strong sense of immersion and presence are felt in relation to the virtual world, meaning that it is potentially externally valid. IVR is also internally valid, giving the experimenter full control over the events that unfold and ensuring that each event, such as a virtual character talking, is performed identically for each participant. Beyond this, virtual worlds can be open-sourced and shared among researchers to allow for near-perfect reproductions.

IVR has been an invaluable research tool to many recent social psychology studies. Slater et al. (2006), as a demonstration of the power of IVR, reprised Milgram's original experiment. Instead of researching obedience however, they were curious about how people would respond to the behaviours of a virtual character during a stressful social situation. The setup was identical to one of the conditions of the original study whereby a Teacher (the participant) was asked to provide shocks to a Learner (the virtual character) if they failed to remember a sequence of words, with the shocks increasing in voltage after each failed attempt. The only difference then was that the 2006 study was conducted entirely in IVR. The results showed that 6 out of 23 participants requested to stop the experiment early and 12 out of 23 (including the first 6) later reported that it at least occurred to them to stop early. Furthermore, the experimenters recorded increased participant stress levels as a result of administering the shocks and observed increased delays between an incorrect answer and the participants giving the shock. Slater et al. (2006) concluded that "our results reinforce this argument [that IVR can be a useful tool for social psychology] and show that virtual environments can provide an alternative methodology for pursuing laboratory-based experimental research even in this type of extreme social situation." Another experiment aimed to investigate the effect of virtual characters' behaviour towards a real participant giving a speech (Pertaub, Slater, & Barker, 2002). Virtual audiences were either depicted as bored, neutral or interested in the participants speech and just as would be expected in real life, the more negative audience induced a greater deal of stress in the speaker by measure of pre- and post-exposure assessments.

A famous phenomena in social psychology is the bystander effect (Darley & Latané, 1968). The effect counter-intuitively states that the more people there are to witness an event such as an crime or medical emergency, the less likely bystanders are to help. The effect is well documented in real-life, such as the death of Kitty Genovese in 1964 and has been used to explain the unbelievable atrocities committed during the Holocaust under the rule of Nazi Germany. However, to study the effect scientifically is much more difficult for the reasons talked about earlier. Participants would theoretically need to be lied to and put in a situations where they think a true emergency is happening, potentially causing undue psychological distress. Furthermore, human actors and confederates would add to the uncontrollability of the scenario. In an attempt to alleviate these issues, Slater et al. (2013) used IVR technology to instead simulate a violent incident. Participants would witness the virtual event where one of the virtual characters would

either look (or not) at the participant for help as things got more violent. Additionally, the impact of social identity was also measured by having half of the participants talk pre-incident with the virtual character about a shared supported football team (participants were specifically chosen as supporters of this team). Results supported both predictions, that participants would be more likely to help when they perceived that the victim was looking at them for help and also if they had a shared social identity.

Despite the overall good that IVR brings to social psychology research, there are still issues that stop it from being immediately adopted by everyone. From a technical perspective, IVR equipment costs more money, takes up more space and requires specialist skills in order to be used effectively (Parsons, 2015). While money may not be so much of an issue for research laboratories, especially with the prices of HMDs these days, acquiring the space needed to accommodate the equipment might be harder to come by. Furthermore, programming a vignette into a three dimensional virtual world requires specialist knowledge not often found in faculties that commonly deal with social psychology research. On top of this, development of a IVR scenario can also take a considerable amount of time, depending on the complexity of the scenario which also brings us onto the fourth problem: how should text-based dilemmas be translated into a virtual world? A questionnaire can (and should) provide only the essential information to the reader, whereas a IVR scenario attempting to recreate the real world as much as possible, is required to provide unessential information. It has to depict the unrelated surroundings of the participant, their location and the weather, it has to depict the virtual characters as real people with different faces and clothes and height and weight, gender, age, ethnicity, social class, etc. Each of these factors must be decided and has the potential to influence the participants responses to the situation. Finally, the designer of the study must ensure that the participants understand what is happening in the situation or even more simply that the participant is looking in the right direction. In a vignette, this is simplified as the writer can literally explain to the reader what is going on and clearly spell out their options. However, telling participants what they are expected to do in the virtual world might affect their responses to the situation, so the experience must be designed such that the participant can piece it together themselves. Two common methods of doing this are: a) giving the participant instructions for a mundane task that they can engage in IVR beforehand that is also how they interact with the eliciting event, for example, getting them to operate an elevator and then using that elevator as

part of a moral dilemma (Pan and Slater, 2011; See also Chapters 4 and 6 in this thesis) and b) having a virtual character or device give the participant instructions at the time of the event (See Chapter 5).

Overall, it is evident that IVR is a positive step forward for the field of social psychology, providing a realistic and controllable methodology for researching ethically charged research questions. However, a prerequisite for this is that the technology behind the methodology is used correctly.

2.1.2. Technology

“As digital computer, tracking, and rendering technologies have advanced and become relatively inexpensive, IVE (immersive virtual reality) systems have begun to proliferate. Sophisticated systems can easily cost less than \$20,000...” – Blascovich et al. (2002)

If you ask the lay person, they might tell you that IVR was invented around 5 years ago and cite the release of the Oculus Rift DK1 as the first IVR headset, or at least the first commercially available headset. Go back about another 30 years and you might get similar answers, but instead they might point to NASA Ames or VPL as the originators. In reality, many of the concepts that are now standard in head-mounted displays (HMDs) were developed in the 1960s by Ivan Sutherland, who designed the “Sword of Damocles” HMD (Sutherland, 1965, 1968). The concepts put forward by Sutherland are now ubiquitous in modern HMDs, such as the generation of two images of a virtual scene (rendered with perspective respect for each eye in the virtual scene) and the presentation of those images on two small displays in front of the user’s corresponding eyes. Furthermore, Sutherland proposed the use of additional hardware to track the user’s head, using this data to update the viewpoint of the user with respect to the virtual scene. With each of these components, a three dimensional virtual world can be presented right in front of a user’s eyes.

Twenty years later, in the 1980s, two more IVR systems were developed independently of each other. NASA Ames developed the VIEW (Virtual Interface Environment Workstation) and VPL the Eyephone HMD, both building in some ways on the work done by Sutherland. Slater and

Sanchez-Vives (2016) attribute the Eyephone HMD as the headset that first publicized IVR to the masses, promising high quality immersive virtual reality at an affordable price. Unfortunately, this was not the case and the optimistic quote from Blascovich et al. (2002) above is evidence that even 20 years later, IVR was still not at a commercially affordable price.

Now a third wave of public interest and excitement in IVR has come, kicked back into motion by the announcement of the Oculus Rift DK1 in 2012 and release in 2013 at a price at \$300. Since then we have seen a resurgence in research into IVR technology and have seen a massive increase in the number of commercially available headsets. As of today the Oculus Go and Rift are £200 and £350 respectively, the HTC Vive and Vive Pro are £500 and £800 respectively, the Sony PlayStation VR is £240, Windows mixed reality headsets cost between £280 to £430 and countless mobile phone headsets that allow users to replace the internal displays with their mobile phone have been released, ranging from £6 (Google Cardboard) to over £100 (Samsung Gear VR). See Figure 2.1 for an example.



Figure 2.1. – Head-mounted display example (Oculus Rift CV1) (Pino, 2018).

Unlike traditional computer displays, such as desktop monitors, the goal of IVR is to substitute all real sensory information with virtually generated sensory data (Slater & Sanchez-Vives, 2016). For example, when a user puts on a HMD, visual information from outside of the headset is lost

and replaced by visual information from inside the headset. This allows visual transportation of the user to arbitrary locations, controlled completely by the designer. In essence, headphones achieve the same thing, but for soundwaves instead of light. By covering our ears (and using the noise cancellation function on newer headphones) we can block out external noises and auditorily transport ourselves to a rock concert. Utilising both of these mediums of sensory replacement, users have been able to watch live concerts as if they were there with the use of IVR technology (www.nextvr.com/live-nation/).

Despite the lulls in publicity, research into IVR continued steadily, not just on the hardware and technology, but on the human side of IVR. Researchers asked questions about how people behave in IVR and about how they respond to simulated events. See Slater and Sanchez-Vives (2016) for a discussion of this research from areas such as psychology and neuroscience, scientific visualization, education and surgical training, sports and exercise, social psychology, cultural heritage, moral behaviour, travel, collaboration and industry and news. All conducted in the last 25 years.

As an example, Meehan et al. (2002) hypothesized that by placing participants on the edge of a virtual room containing a deep pit, they would become discomforted and experience a negative physiological response to the virtual scenario. Measurements of participants' heart rate was congruent with this belief, showing that heart rates increased from being in a room with no pit, to one that did. Additionally, the use of a tactile ledge (in reality only a few inches off of the ground) placed to coincide with the virtual precipice further increased the physiological response to the scenario.

Other studies have looked at the impact of the type of virtual body given to users in a virtual environment. While it is important for experiments studying human behaviour to replicate the real world as much as possible in a virtual one, the power of IVR is that we can change aspects of the world if we want to. Yee and Bailenson (2007) discovered that participants embodied in an avatar with an attractive face would behave more intimately with a virtual character than those given an ugly face, and also showed that taller embodied avatars performed more confidently in a negotiation task. The authors labelled this the Proteus Effect. Research building on this has shown a plethora of phenomena, such as the overestimation of object sizes by participants

embodied as a child (Banakou, Groten, & Slater, 2013) and the reduction of implicit biases by the embodiment of white participants as a black avatar (Banakou, Hanumanthu, & Slater, 2016).

Apart from HMDs, another IVR system called a Cave (Cruz-Neira, Sandin, & DeFanti, 1993) is utilised to some degree in research environments, however, due to the amount of space and money needed to construct and maintain the system, it is unlikely to ever be as well-known or available as HMDs. The Cave works by back-projecting images of a virtual scene onto the walls of a 3m by 3m room with another projector illuminating the floor from above. Two interlaced images are projected at a high frequency and with the use of shutter glasses to block out one of the images to each eye, the brain assimilates visual information as three dimensional. Tracking equipment attached to the glasses can also allow the scene to be positioned and rotated with respect to the user's head. The advantage of this over a HMD is that the user retains their body in the virtual world.

2.1.3. Presence

“Consciousness occurs when we can generate, automatically, the sense that a given stimulus is being perceived in a personal perspective; the sense that the stimulus is ‘owned’ by the organism involved in the perceiving; and, last but not least, the sense that the organism can act on the stimulus” – Damasio (1998)

Given the definition of consciousness above by Damasio (1998), it should be clear that the application of IVR, especially in scenarios where the user is able to touch and manipulate the objects around them, should feel like the user's reality despite their absolute knowledge that what they are seeing and hearing is not real – this concept has been labelled presence (Sheridan, 1996). Sanchez-Vives and Slater (2005) define presence as occurring “when there is successful substitution of real sensory data by computer-generated sensory data, and when people can engage in normal motor actions to carry out tasks and to exercise some degree of control over their environment,” where, “successful [means] that the person responds to the virtual stimuli as if they were real.”

Presence is unique to IVR and is the main factor that makes it different from other forms of computer-generated output. The feeling of 'being there' leads people to respond unconsciously, intuitively and sometimes emotionally to what is happening around them which in turn leads to realistic responses to many situations in IVR. Being in a virtual burning room makes people want to evacuate (Kinateder et al., 2014), witnessing a virtual argument makes people want to intervene (Slater et al., 2013) and being above a virtual precipice forces people to walk around it (Zimmons & Panter, 2003), despite participants knowledge that none of these things truly exist. Social norms are also kept in IVR (Gonzalez-Franco & Lanier, 2017), people will get nervous when giving a speech to a virtual audience (Pertaub et al., 2002) and feel reluctant to electrically shock a virtual learner (Slater et al., 2006). Slater and Sanchez-Vives (2016) argue that "VR effectively relies on this duality" between reflexive and reasoned actions, while participants always know that they are in a virtual world, they struggle to stop themselves reacting automatically to their surrounding and end up walking around a pit or waving back at someone who is not really there.

There are a number of concepts that relate to or are a part of presence, including immersion, place illusion (PI), plausibility illusion (Psi), embodiment and body ownership.

According to Sanchez-Vives and Slater (2005), immersion refers to "the technical capability of the system to deliver a surrounding and convincing environment with which the participants can interact," such that the immersion of a system is determined wholly by the physical properties of the IVR system being used. This includes basic parameters such as the resolution, field of view and frame rate of the display, the complexity of tracking, i.e., just head or entire body, and tracking latency. Additionally, the number of sensory modalities supported and their quality also contributes towards a user's immersion, where, realistically, an IVR system should support at least visual, auditory and haptic feedback (Ellis, 1996; Gonzalez-Franco & Lanier, 2017; Held & Durlach, 1992; Loomis, 1992; Sheridan, 1992, 1996; Slater, 2009). Haptic feedback refers to the feeling of a resistant force when touching an object. This is much harder to accommodate than sight and sound and is typically solved through the placement of real physical objects that correspond to the virtual scene.

Immersion is also based on the valid effectual and sensorimotor actions of a system (Sheridan, 1992, 1996; Slater, 2009). Effectual actions refers to the actions that participants can take in order

to interact with the environment, for example, presence was higher for participants who were given the ability to walk in place to navigate a virtual world than those who had to do so using a button (Usoh et al., 1999). Sensorimotor actions then refer to our perceptual actions, i.e., the actions that we take in order to perceive the world, such as turning our head or moving our eyes. Here, a system that supports more valid sensorimotor actions would obviously be the more immersive system, e.g., a Cave with four walls compared to Cave with three walls, or a HMD with positional and rotational tracking compared with a headset with only rotational tracking.

PI refers to the general concept of presence that we have been talking about thus far, the sense of 'being there', Slater (2009) defines it as "the strong illusion of being in a place in spite of the sure knowledge that you are not there." Like immersion, PI is a function of the physical properties and sensorimotor actions of a system. The relationship between immersion and PI is such that "immersion provides the boundaries within which PI can occur." What is meant by this is that the immersion of one system, as a static property of that system, can be inherently greater than another system based on its valid actions, however, a user's sense of PI would be the same in those two systems if they only performed the valid actions of the lower immersive system. For example, if a user only rotated their heads in a system that supported rotational tracking and another that supported both rotational and positional tracking, their level of PI would be the same, despite the systems having different degrees of immersion.

Interestingly, while graphical realism may be the first thing that comes to mind for many people due to the attention it receives for films, most research evidence actually shows that this has very little to no effect on a user's sense of presence in a virtual scenario. Zimmons and Panter (2003) reported no significant difference in physiological measurements of participants walking over a high or low quality rendered pit. Supplementary to this however is the inclusion of graphically accurate shadows and reflections of the user in a virtual scenario which does increase immersion (Slater, Khanna, Mortensen, & Yu, 2009). This is likely due to relation of those graphical elements to the user themselves, grounding them further as part of the world, instead of just observing it.

While immersion is based on the technical aspects of the IVR system itself and PI the response to this in terms of the illusion of being in the place depicted by the virtual environment, Psi is related to the virtual world presented to the user and is the illusion that the events happening in the virtual world are really happening (Slater, 2009). For example, being able to move your head

in IVR to look past a person standing in front of you increases your sense of PI, but having that person smile at you when you do this would increase your sensation of Psi. The key here is that the event of the person smiling relates directly to the user, reinforcing the notion that they are taking part in the events, not just witnessing them. Similarly, having virtual characters purposefully walk around or look at the user would increase their sensation of Psi. Another aspect of Psi is the believability of the situation, if the events are completely out of the ordinary or the physics of the virtual world do not match our own, then we are less likely to believe that they are really happening. If someone walks into a room through a wall instead of a door or objects float above tables, this is likely to trigger alerts in our brain that what we are seeing is not real. This makes the sensation of Psi much harder to create than PI, where PI is all but determined from the choices of hardware we use, Psi is directly related to the world we create and extra care should be taken to make it as close to reality as possible, especially when researching a participant's behaviour to a specific event. This difficulty is compounded by evidence that when PI breaks, it is likely to return once the user returns to using valid sensorimotor actions, however when Psi breaks, it is unlikely to recover (Slater, 2009).

The hypothesis put forward by Slater (2009) is that a strong sense of PI, the illusion of being in a place, and Psi, the illusion that events are happening in that place, will result in realistic responses.

As described earlier, IVR is the process of removing external visual information and replacing it with one of our own creation. A side-effect of this is the removal of the user's body (unless using a Cave system). This leads to a proprioceptive mismatch where we can still feel our body and know roughly where our arms and legs are, but can no longer confirm this information through sight. At the same time, not having a body can lead a participant to feel as if they are simply viewing the virtual world, instead of taking part in it. The solution to this is embodiment, providing users with a virtual body that coincides with their own. This can either be static or, with the use of tracking equipment, can be moved synchronously with the user's own body. Under the right conditions, the substitution can lead to a sense of body ownership, the illusion that the virtual body is your own (Spanlang et al., 2014). This sensation is not unique to immersive virtual reality and has been studied in real-life with the rubber hand illusion (Botvinick & Cohen, 1998). In the rubber hand illusion, a participant's real hand is hidden and the participant is shown a fake hand which both then receive identical tactile stimulations. The congruence of both visual and tactile sensations leads participants to identify the fake hand as their own. The same process is

happening during embodiment in IVR, the congruence of visual and proprioceptive sensations when users have control over their virtual-self leads to the sensation of ownership of the virtual body. The rubber hand illusion experiment has even been replicated in IVR to a similar effect (Slater, Pérez Marcos, Ehrsson, & Sanchez-Vives, 2008). Slater (2009) asserts that this sensation of body ownership, most prominently when users have synchronous control over their virtual bodies, strengthens the degree of presence felt by the user.

2.1.3.1. The Brain on Presence

Cognitively, presence works due to our brains reliance on existing models of the world, i.e., what it expects to see and happen. This allows us to extrapolate the contents of our surroundings without a complete picture or with a low-detail simulated picture. For example, when looking at a half-concealed apple, we do not believe that we are looking at half an apple, but truly that the other of the apple is behind the concealment, or when looking at a house from far away, we do not believe the house is any smaller than the one next to us, but understand that perspective projection forces objects further away from us to appear smaller. Similarly, when looking at a virtual object, while it may not look real, our brain understands what it represents and fills in the blanks, causing a suspension of disbelief in participants (Sheridan, 1996).

It is these active processes that utilise our brains previous model of the world that makes IVR seem so real (Slater & Sanchez-Vives, 2016). While IVR often gives us a watered-down version of reality, it provides enough visual cues for our brain to extrapolate, based on an internal representation of the world, what we are seeing. Just as we do not need to inspect every nook and cranny of a space to know that it is a room, the limited information of a virtual world can also trigger the same sensation through our brain filling in the gaps.

Gonzalez-Franco and Lanier (2017) asserts that the illusion of IVR is enabled through three classes of neural mechanisms: bottom-up multisensory processing, sensorimotor self-awareness frameworks and top-down prediction manipulations. Multisensory processing is the brain's act of combining the input from multiple senses. If the senses are congruent with each other, then the brain is less likely to question the reality of what is happening. A participant seeing and feeling themselves touch a virtual object will feel more immersed than a participant who can only see an

object, but who's hand goes through it. Incongruent signals would be sent through the brain as their hand passed through the object, warning that something is not right. This is also the process behind simulator sickness, when a person visually sees themselves moving, but does not receive the same vestibular response, the brain struggles to merge the signals and can cause severe headaches. The dominance of the visual sense means that this illusion is easiest when matching other senses with sight, like with touch in the rubber hand illusion (Botvinick & Cohen, 1998).

Sensorimotor self-awareness frameworks and top-down prediction manipulations work similarly to the description above involving an internal representation of the world. Taking actions results in a predictive state in the brain and the more the true results of this actions matches this state, the more likely we are to assume this outcome was real.

The need for a high sensation of presence in participants in virtual moral dilemmas studies is imperative, as in any other study looking at psychological behaviour using IVR technology, due to the previous evidence showing a relationship between the realism of participants responses to virtual events and their degree of presence within the virtual environment (Slater, 2009). The lack of concern and care taken by previous studies in this regard shows that this field is truly in its infancy, where previous studies haven't provided participants with a virtual bodily representation (Navarrete et al., 2012), used unrealistic methods of interaction (Patil et al., 2014), or even neglected to include head tracking (Skulmowski et al., 2014), something that was known to be crucial to presence almost 30 years ago (Sheridan, 1992). As such, it has been a goal of the current studies to demonstrate and advocate for the proper use of immersive virtual reality technology, including using correct visual, auditory and haptic sensory feedback, body tracking, embodiment, and use of realistic scenarios (as far as moral dilemmas can be realistic). Additionally, presence questionnaires (Slater, 2004; Slater, Usoh, & Steed, 1994) are used in each study to measure and ensure that presence standards are adequate, where the main behavioural results would be in jeopardy if presence scores are found to be too low.

2.2. Morality

Morality: Principles concerning the distinction between right and wrong or good and bad behaviour
– Oxford Dictionary

The goal of this thesis is to use immersive virtual reality to further understand the susceptibility of our moral decision-making to the influence of others, as well as to discover how moral judgments differ from actual behaviour. As such, given the centrality of our moral decision-making to these questions, this section aims to illustrate the current research regarding the make-up of our morality. The section describes both evolutionary and biological accounts of how our morals are constructed, illuminating what exactly is being targeted by the social influence attempts, and also covers the philosophical discussions regarding the trolley problem. This is particularly important given the conducted studies reliance on moral dilemmas inspired by the trolley problem as well as the methodological similarities between the conducted studies and previous studies researching the trolley problem (Cushman et al., 2006; Hauser et al., 2007). Finally, an account of the research into moral actions is discussed, with particular emphasis on studies using immersive virtual reality, ending with the current state of the research, hopefully providing a useful background before presenting the studies in Chapters 4 to 6.

2.2.1. Evolution of Morality

“Finally after several unsuccessful tries, I went up to a passenger and choked out the request, ‘Excuse me sir, may I have your seat?’ A moment of stark anomic panic overcame me. But the man got right up and gave me the seat. A second blow was yet to come. Taking the man’s seat, I was overwhelmed by the need to behave in a way that would justify my request. My head sank between my knees, and I could feel my face blanching. I was not role-playing. I actually felt as if I were going to perish.” – Stanley Milgram quoted in Blass (2004)

The neural foundation of human morality is thought to have evolved millions of years ago when hominids still walked the earth before splitting with Pan around 5 to 7 million years ago (Haidt,

2007). Evidence for this comes from the behaviour of chimpanzees (members of the *Pan* genus), who enforce group level norms for mating, interacting with young and the punishment of members who openly transgress against another (De Waal, 1991). Other studies have shown that chimps also share some level of common emotions with humans, exhibiting signs of sympathy when witnessing distress in others (De Waal, 1996).

The most common explanation for the evolution of morals is that it better facilitated the need to live together in groups which in turn increased our ancestors chances of survival (Greene, 2008; Haidt, 2001). While pre-moral hominids might have been able to kill or steal for their own benefit without remorse, the adaptation of morals provided us with the unique and powerful capabilities of sympathy, reciprocity, altruism, emotional inhibition and the ability to judge the actions of others. Through these mechanisms, humans can now thrive in societies by caring for one another through sympathy and altruism and excluding those who pose a danger through judgment and reciprocity. To accommodate this, humans have also evolved the ability to maintain long-term goals (Schaich Borg, Hynes, Van Horn, Grafton, & Sinnott-Armstrong, 2006). By sacrificing something they have now, e.g., sharing food with someone else, they can reap increased benefits in the future through mutual partnerships.

The acts of reciprocity and altruism take the form of three distinct human functions (Haidt, 2007). Kin selection makes us care strongly about our close relatives, reciprocal altruism compels us to care for those who will in turn care for us back, e.g., close friends or acquaintances, and strong (or indirect) reciprocity encompasses a broader tendency to care for others, such as giving a stranger your seat on the bus. Another factor is emotional inhibition which also drives us to act morally by stopping us from harming others (Greene, 2008). While it may have occurred to you in the past to act immorally towards someone else, emotions such as sympathy and guilt cause us to hesitate. Greene (2008) asserts that “these responses evolved as a means of regulating the behaviour of creatures who are capable of intentionally harming one another, but whose survival depends on cooperation and individual restraint.” Recent studies have even shown that deontological responses to moral dilemmas, i.e., those associated with emotional restraint, were perceived as a signal of trustworthiness in the responder (Everett, Pizarro, & Crockett, 2016) (See section 2.2.2.2 for further discussion regarding deontology). Furthermore, the quote at the start of the section detailing one of Stanley Milgram’s experiments on conformity is a good example of

how our emotions can berate us for even the slightest transgression we may take against a fellow person.

It is even postulated that language could have developed as a way to propagate the status of other people's standing in a group, allowing us to keep track of who did what to whom (Haidt, 2001, 2003). This would have helped humans elevate to group-level punishments of individuals, unlike chimpanzees where it is always the offended who are expected to enact punishment (De Waal, 1991). As such, morality can be seen as a strictly social affair. A man or woman living alone would be unable to act immorally until their acts were witnessed by another who would deem those actions immoral. Depending on who was looking however would change the perceived morality of those acts. Different cultures and even people within those cultures hold contrasting ethical values (the divide between the American conservative and liberal political parties is a good example of this), leading to the hypothesis that humans more likely develop the capacity to hold moral values instead of being born with a specific set of rules in mind (Haidt, 2001). This is not unlike many other aspects of the human brain which supports the learning of abstract and complex behaviours at the expense of an extended duration of learning.

This difference in morals can cause divides between cultures, where two groups can see the other as immoral based on their own sacredly held values. And while morals might encourage us to cooperate within groups, they are less effective at enabling cooperation between groups (Greene, 2014). This is likely another evolutionary adaptation. Working together in groups promotes survival, but so does being cautious of the other group over the hill. Today, this is mirrored in the psychological distinction between in-groups and out-groups, with studies showing that participants are much more likely to want to help an in-group member and more likely to dehumanize an out-group member (Cikara, Farnsworth, Harris, & Fiske, 2010), but we can also still see the primitive version of this through our interaction with tribes such as the Sentinelese (www.survivalinternational.org/tribes/sentinelese).

The foundation of human morality is evolutionary which in turn is blindly guided by the goal of reproduction, in other words, human morality is the product of a non-moral mechanism. Knowing this information, should we continue to be sensitive to our own morality? While we may think of it as innate and pure, the truth is that morals can often go awry. For example, Cushman et al.'s (2012) study on their action aversion model revealed that participants attributed, to some degree,

the amount of harm caused by an action to be related to the motoric and perceptual properties of that action. This is likely based on a heuristic, e.g., the damage caused is related to the strength used, which Sunstein (2005) discusses at length, exposing the numerous errors we make at the hands of our makeshift morals. The danger then with Cushman et al. (2012) results, is that with modern technology, a lot of harm can be caused with very little strength. Greene (2015) provides another example of our moral shortcomings. In the trolley problem, it is well known that, in order to save five people, responders find it more acceptable to throw a switch in order to kill someone else than to push them off of a bridge. However, Greene points out that “were a friend to call you from a footbridge seeking moral advice, would you say ‘Well, that depends... Will you be pushing or using a switch?’” The point here being that the detail in reality is superfluous, but our morals make us think that it is an important distinction.

Despite this melancholy outlook on morality, the answer to the question “should we continue to be sensitive to our own morality?” is obviously, yes, we should. While they are not perfect, they have allowed the human species to become the most powerful beings on the planet and that is probably good enough. To quote Spranca et al. (1991), “[Morals] are maintained, even when they cannot be independently justified, by the absence of critical thinking about them... [However,] these principles... could suffice to prevent the most insidious cases of intentional immorality... People might make better moral judgments on the whole by honouring the distinction consistently than by trying to think normatively.”

2.2.2. Moral Philosophy

Although morals have been a part of the human genus for millions of years, only in the last few thousand years have people started to introspectively think about them. The ideas about morality formed millennia ago are still talked about today, a testament to the importance of morality to our society. Here, the three most prevalent normative ethical theories will be discussed. These philosophical theories attempt to define the qualifications of a morally good act by focusing on either the characteristics of the person, the action itself or the consequences it brings about.

2.2.2.1. Socrates, Plato, Aristotle and Virtue Ethics

Some of the earliest work on moral philosophy comes from ancient Greece and the philosophers Socrates, Plato and Aristotle who laid the groundwork for the normative ethical theory, Virtue Ethics. The first of these philosophers, Socrates, was born around 470 BC. The moral consensus of the time was that of “might makes right”, however Socrates thought that good virtues would create a good society. Socrates would eventually be sentenced to death for his non-government-conforming ideas, but his teachings would survive through his students and eventually outlive the government that killed him. Plato and subsequently Aristotle would continue to advocate the importance of proper virtues in society and that those virtuous enough would achieve Eudaimonia, roughly translated as happiness or well-being. With each of these men teaching the last (Socrates taught Plato who taught Aristotle), what was deemed “virtuous” was expanded upon in each generation. Socrates argued that the one true virtue was knowledge, Plato expanded the list to; wisdom, justice, fortitude and temperance and Aristotle eventually identified eighteen virtues which were distinguished as either moral or intellectual virtues.

While virtue ethics diminished in popularity with the advent of deontology and utilitarianism, it is now considered alongside these as one of the three major normative ethical theories. This recent boost in popularity can be traced back to G. E. M. Anscombe and her 1958 essay *Modern Moral Philosophy* (Anscombe, 1958). The more recent version of virtue ethics retains much of ideas set in ancient Greece, focusing on the character of a person rather than their actions or the consequences of their actions like deontology and utilitarianism, respectively. In this sense, it is a person’s intentions that determine their moral status rather than what those intentions bring about (Uhlmann, Pizarro, & Diermeier, 2015). Using the switch dilemma as an example (See appendix A for descriptions of all moral dilemmas discussed in this thesis or section 2.3.3. for a discussion of the switch dilemma), according to virtue ethics, either response to the switch dilemma could be acceptable, so long as the intent was good (“I couldn’t bring myself to kill” or “I wanted to save the five lives”). This leads some philosophers reluctant to regard virtue ethics as an alternative to the other normative ethical theories (Nussbaum, 1999).

2.2.2.2. Kant and Deontology

Immanuel Kant, born in 1724, and is one of the most influential philosophers to discuss morality. Kant argued that the only morally good acts are those that arise from duty and motivation and not from the acts outcomes. He argued that moral qualification from the result of a moral action can lead to mistakes, such as an evil act that accidentally produces a good outcome or a good act that goes awry. As such, the only true good acts are those that follow a person's duty set by a moral law. In 1785, Kant published the book *Groundwork of the Metaphysics of Morals* (Kant, 1785/2013) and with it his theory for a moral law he called the "Categorical Imperative". The categorical imperative is a set of principles that are intrinsically valid, good in and of themselves and good without qualification. For any behaviour to be thought of as morally right (according to Kant) it must follow the rules set forth by the categorical imperative. These principles are:

1. Act only according to that maxim by which you can also will that it would become a universal law.
2. Act in such a way that you always treat humanity, whether in your own person or in the person of any other, never simply as a means, but always at the same time as an end.
3. Every rational being must so act as if he were through his maxim always a legislating member in a universal kingdom of ends.

These principles provide a set of guidelines for acting morally where a moral act is one you commit only if you are happy for other people to be able to commit the same act, in which a person is not used simply as a means and finally, that the action must be acceptable in a world where no one is used solely as a means. As such, the categorical imperative provides a quick and easy check for the morality of an action (at least from a Kantian point of view). Taking once again the switch dilemma as an example, it is obvious that it should not become a universal law that people should be able to kill others, although five are saved, it is not the outcomes of a moral act that qualify it as moral. Also looking at the footbridge dilemma, Kant would view this case as extremely immoral as it involves using another person as means and not solely as an end.

More recently, Kant's work has become one of the more well-known ethical theories to fall under the banner of the normative ethical theory of deontology. Similar to Kant's moral views, deontology bases an actions rightness or wrongness on the action itself, and never the consequences of the action. Additionally, the clarification as right or wrong is based on rules, duties or obligations applicable to the person. For example, we have a duty to not kill each other

in a civilised society as well as duties to respect each other's rights. Once again, using the switch dilemma as an example, deontology would determine that you should not kill the one because that violate his or her rights, specifically the right to live.

2.2.2.3. Bentham, Mill and Utilitarianism

While virtue ethics highlights the importance of the character of the person committing a moral action and deontology focuses on the action itself, consequentialism looks at the final piece of the puzzle, the outcomes of the action. Consequentialism holds that the it is the actions consequences that determine the rightness or wrongness of it and that the morally right action (where a person might have several options) is the one with the best overall consequences. By overall consequences, this means that the ratio of good to bad outcomes is at its highest. Consequentialism provides a useful framework by using the words "good" and "bad" in a general sense and versions of consequentialism have been developed substituting these terms for specific things, the most famous of which is Utilitarianism. Founded by Jeremy Bentham in the late 18th century, utilitarianism is a hedonistic version of consequentialism, focusing on the aggregate happiness of all of those who the moral action affects. Utilitarianism holds that the morally right action is the one that produces the greatest happiness, through the maximization of pleasure (where "pleasure" can be thought to be replacing the keyword "good" in the original definition of consequentialism) and the minimization of pain (where "pain" replaces "bad"). In fewer words, utilitarianism calls for the greatest happiness for the greatest number of people, no matter the cost.

Although consequentialism was alive in different forms before him, Bentham was the first to propose the hedonistic form of it. Bentham believed that the primary motivators for human beings was the pursuit of pleasure and the privation of pain. As such, what is most important in life and subsequently most moral is the generation of the greatest amount of happiness or the least amount of sadness for the greatest number of people. Bentham further thought that the happiness of one man is worth just the same as any other and declared different forms of happiness as equal as well. Where one person might be happy writing poetry, another might enjoy listening to

music. This allows utilitarianism to be objective and resolve difficult decision even in the conflict of interests.

While Bentham defined many of the aspects of utilitarianism, it was John Stuart Mill who gave it such wide spread acclaim, even popularizing the word 'Utilitarianism' (with it previously being called Benthamism). Mill's book *Utilitarianism* (Mill, 1861/2016) is still regarded as one of the best resources of the definition, clarification and defence of the ethical theory. Although Mill's utilitarianism is heavily influenced and remains largely unchanged from Bentham's, one major contribution by Mill was his argument for the separation of pleasures. Where Bentham regarded all forms of happiness as equal, Mill, similar to Socrates, regarded intellectual and moral forms of pleasures as true happiness and physical pleasures as mere contentment. Mill is famously quoted as saying "It is better to be a human being dissatisfied than a pig satisfied; better to be Socrates dissatisfied than a fool satisfied. And if the fool, or the pig, are of a different opinion, it is because they only know their own side of the question" (Mill, 1861/2016).

Applying the switch dilemma one last time, we find that utilitarianism would support the killing of the one in order to save the five, as this sequence of the events would produce the greatest post-dilemma utility, the saving of five lives compared to one, which would undoubtedly produce the greatest happiness or at the very least prevent the most sadness.

2.2.3. The Trolley Problem

"Why is it that the trolley driver may turn his trolley, though the surgeon may not remove the young man's lungs, kidneys, and heart? In both cases, one will die if the agent acts, but five will live who would otherwise die – a net saving of four lives... Since I find it particularly puzzling that the bystander may turn his trolley, I am inclined to call this The Trolley Problem." – Thomson (1985)

Imagine you are standing next to a set of train tracks that diverges a bit further down. There are five men on the main track and one man on the side track. None of the men are able to quickly get off of their respective paths. There is an out of control trolley hurtling down the main track towards the five men. You are also stood next to the control switch which, if thrown, will cause

the trolley to turn when it reaches the fork in the track, aiming it towards the one man instead of the five others. Is it morally permissible to throw the switch? The overwhelming majority say that it is (Hauser et al., 2007).

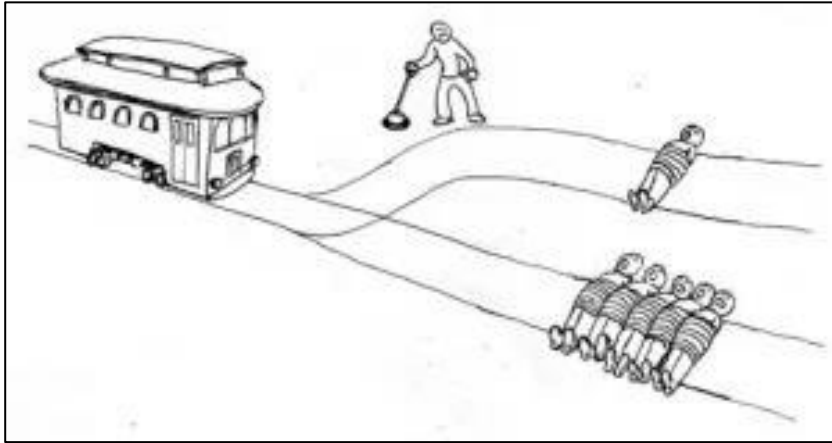


Figure 2.2. – The Switch Dilemma (Valdesolo, 2016).

Now imagine that you are standing on a bridge looking over a set of train tracks next to a large man wearing a heavy backpack. There are five men on the track. None of them are able to quickly get off of the path. Now there is an out of control trolley hurtling down the track towards the five men. The only way to stop the trolley would be to intercept it with a very heavy object. The only object heavy enough around is the man with his backpack next to you. Is it morally permissible to push the man? The overwhelming majority say that it is not (Hauser et al., 2007).

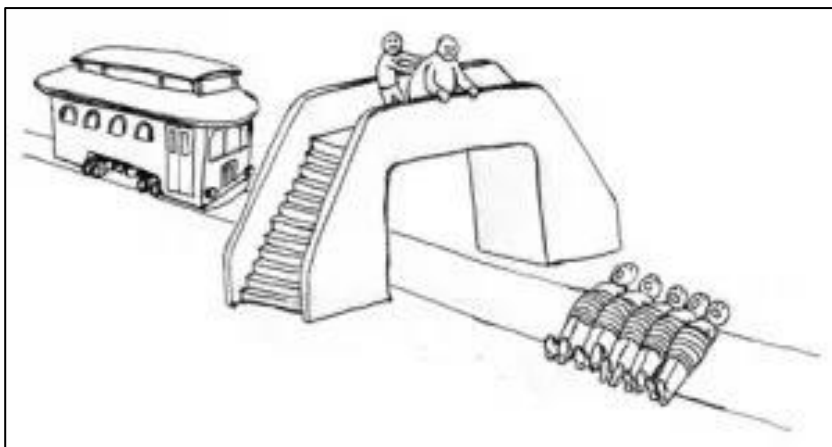


Figure 2.3. – The Footbridge Dilemma (Valdesolo, 2016).

These are the switch and footbridge dilemmas respectively and together they form the trolley problem (Kamm, 2015). This problem refers to the question of why we respond differently to consequentially identical moral dilemmas. In both of these dilemmas we are essentially asked to kill one person in order to save five others with only superfluous details added on top, why then should it matter how the death of the one comes about when the lasting effects are so heavily desired? Shouldn't these take priority? While the switch and footbridge dilemmas are the go to examples, this question refers to all reasonably similar moral dilemmas involving harm that typically invoke different moral judgments, e.g., the surgeon dilemma and the drug dilemma (For a description of these and all other moral dilemmas referred to in this thesis, see appendix A). Furthermore, what are the specific aspects of or differences between the moral dilemmas that trigger such different responses and what psychological and neurological structures recognise these differences?

From our previous discussion, it is obvious that the answer must be related either to virtue ethics or deontology. As consequentialism focuses on the outcomes of a moral act which are typically held constant between the two moral dilemmas when discussing the trolley problem, this is unlikely to play any part in explaining a difference between them. On the other hand, virtue ethics is somewhat ambiguous in examples of moral dilemmas as any option could be viable so long as the responder has virtuous intentions. Therefore, it is left to deontology to help us investigate this phenomenon.

Implicit in the problem is the fact that it is very difficult to pinpoint exactly why we respond differently to the dilemmas (if we could then it would not be a "problem"). You might say that it is impermissible to act in the footbridge dilemma because it is wrong or disrespectful to use the man's body as a means to stop the trolley, but then what about the gas dilemma? (See appendix A). In this situation the person is not being used as a means, but it is still generally considered to be impermissible to kill them, to which you might say that it is his right as a patient of the hospital to not be killed (even for other's benefit) to which I would then refer you to the mayors dilemma where it is arguably permissible for the mayor to kill a man for being in a certain place for which he has promised him he could be, giving him the right to be there and not be killed for it. This strategy of proposing a solution and refuting it by constructing specific counter examples is the tactic taken by early proponents of the trolley problem in order to uncover the mystery of our moral decisions (Foot, 1967; Thomson, 1976, 1985).

The trolley problem (along with many of its supporting hypothetical moral dilemmas) was born out of a discussion started by Philippa Foot on the morality of abortions and the legitimacy of the doctrine of double effect as a moral argument (Foot, 1967). In a continuation of Foot's work looking into the trolley problem, Judith Jarvis Thomson (1976, 1985) emulated Foot's approach to the topic, opting to use written vignettes (often in pairs) in order to prove, refute or simply illustrate a moral problem or hypothesis. Thomson however covers much more ground in her articles, discussing multiple solutions to the trolley problem, including Foot's previous duties solution as well as distributive exemption, rights and physical contact. Thomson's original article also includes the first mention of the term *trolley problem*, saying "Why is it that Edward may turn that trolley to save his five, but David may not cut up his healthy specimen to save his five? I like to call this the trolley problem," referring to the following pair of moral dilemmas:

David has five patients. All five patients need new organs, but have a very rare blood type. A new patient is brought into the hospital with that blood type. David now has the opportunity to kill the new patient in order to save his five patients.

Edward is the driver of a trolley, whose brakes have just failed. On the track ahead of him are five people; the banks are so steep that they will not be able to get off the track in time. The track has a spur leading off to the right, and Edward can turn the trolley onto it. Unfortunately there is one person on the right-hand track. Edward can turn the trolley, killing the one; or he can refrain from turning the trolley, killing the five.

Here Thomson argues that while it is certainly impermissible for David to kill his healthy patient in order to save five others who are dying, it is at the same time permissible for Edward to turn the trolley and subsequently ponders why this is so.

Thomson (1985) takes her fascination out of the armchair and questions others about their responses to the dilemmas she presents in her paper. There is great controversy about mixing an *is* and an *ought*, where a person's opinion (an *is*) about a moral dilemma should not have any impact on what *ought* to be done, where *oughts* include ethical theories such as utilitarianism or deontology (Greene, 2015). Thomson opts to use her respondents' *is*'s as pseudo-*oughts*, giving her moral dilemmas a clean answer, e.g., you should not push the man off of the footbridge, in order to better understand the impact of moral principles. Thus, in the first consensus driven response to the switch dilemma, Thomson found that most people do indeed find it permissible

to throw the switch, a response which has now been replicated many times over (Cushman et al., 2006; Greene et al., 2001; Hauser et al., 2007).

The basic formula of trolley-esque dilemmas is that we have two options: do something bad, or at the very least morally questionable, in order to bring about a good result, or do nothing, but receive a bad outcome. More specifically, they usually represent a scenario whereby the responder has the opportunity to personally kill one person that will ultimately lead to salvation of five others, or do nothing, leading to the indirect death of the five. The dilemmas therefore rely heavily on the distinction between taking actions versus passive omissions, i.e., killing and letting die.

The dilemma itself then arises from the mismatch of good and bad actions and good and bad outcomes. Human nature associates good actions with good outcomes and wants both of these things, on the other hand, we associate bad actions with bad outcomes and typically try to avoid both of these things. Trolley-esque dilemmas force us to compromise with these desires, either offering us a bad action which leads to a good outcome or letting us forego having to take a bad action, but instead giving us a bad outcome. We must then decide what is more important, preventing a bad outcome or abstaining from a bad action?

This again brings us back to normative ethics, mirroring the distinction between deontology and utilitarianism, where deontology advocates for good actions and utilitarianism for good outcomes. Given this, in a trolley-esque situation, we can say that deontology would advocate for the practice of constraint as personally killing the man would be a morally bad action in and of itself and so is not acceptable, whereas Utilitarianism would allow the killing of the man, as the result would be an overall better outcome (five lives is greater than one). That being said, the human brain is neither perfectly deontological or perfectly utilitarian and which of these options is subjectively worse depends on how the scenario is constructed.

2.2.3.1. Moral Principles

In order to explain the trolley problem, philosophers, and more recently psychologists and neuroscientists, have implicated the use of moral principles in our moral judgments. These

principles are aspects of moral dilemmas that contribute, both consciously and unconsciously, towards our evaluation of that moral dilemma. Cushman et al. (2006) defines them as “a single factor that when varied in the context of a moral dilemma consistently produces divergent moral judgments.”

They typically alter our response options and not the outcomes of a moral dilemma. For example, the contact principle asserts that people are less likely to kill someone if they have to physically touch them in order to bring the result about, as opposed to being able to kill them through indirect means. On the other hand, changing aspects of a moral dilemma like increasing the number of people saved is not generally classified as a moral principle although it might change how people respond to the situation.

Two more aspects of moral principles is firstly that they usually only interact with our deontological outlook of a moral dilemma. This is relatively obvious as deontological judgments are those related to the morality of the actions themselves which is what moral principles are altering. And second, the “inclusion” of a moral principle typically acts in favour of a deontological resolution. For example, the switch dilemma acts sort of like a control moral dilemma in that it does not contain any moral principles (bar the action principle), and as such, many people find it permissible to act in that dilemma, i.e., they act in a utilitarian manner. The footbridge dilemma on the other hand contains the intention, contact, action and redirection principles and is thus a famous example of an impermissible moral dilemma.

What this means is that our utilitarian-like cognitive responses to a moral dilemma are identical for each dilemma as most of them rely on the typical outcome of “save five people at the expense of one” paradigm. Moral principles then alter our deontological perspective which determines whether or not we feel it is permissible to kill. Where a low number of principles are present, our deontological response will be weak and utilitarianism will prevail (e.g., in the switch dilemma), but where many principles are present, deontology will overpower us and stop us from acting (e.g., the footbridge dilemma).

The following sections will detail the most common principles researched or otherwise discussed in the literature.

2.2.3.1.1. The Doctrine of the Double Effect

The Doctrine of the Double Effect (DDE), otherwise labelled the intention principle (Cushman et al., 2006), was proposed by the 13th century philosopher Thomas Aquinas in his book *Summa Theologica* (Aquinas, 1485/2012) and is also notably the second law of Kant's three categorical imperatives (Kant, 1785/2013). In general, It states that you should not use people as a means, only as an end. The definition provided by Hauser et al. (2007) says that "it may be permissible to harm an individual for the greater good if the harm is not the necessary means to the greater good but, rather, merely a foreseen side effect." More recently, the doctrine has been used to attack several forms of medical practice such as abortion and euthanasia (Foot, 1967).

What exactly does this mean though? What constitutes using someone as a means? Foot (1967) describes a heuristic to determine whether someone is being used as a means during a moral dilemma which is as follows. If it is the case that we have taken the utilitarian route and have chosen to kill the one in order to save the five, if the one were to miraculously survive their certain death and the five are still saved, then the one was not being used as a means. However, if the five die as a result of the unexpected survival of the one, then the one was being used as a means. The difference here is what is saving the five. In the first instance, it is solely our action that is saving the five (e.g., the turning of the trolley in the switch dilemma) whereas in the latter, it is the death of the one that is saving the five (e.g., the pushing of the man in the footbridge dilemma). For example, if the one were to manage to jump out of the way of the trolley in the switch dilemma, the five would still be saved and thus the one was not being used as a means, however, if the man manages to grab the footbridge as he is falling, not making it all the way to the tracks, the trolley will kill the five despite your best efforts to stop it.

In an example of the use of the DDE in our moral judgments, Foot (1967) posits that most people would advocate to raise the minimum level of education even with the information that this would statistically increase the number of suicide related deaths (scenario A), but would also condemn the act of killing people in order to aid cancer research (scenario B). Taking a step back reveals that the outcome of both scenarios are very similar: a few people will knowingly die (intended or not) so that society in general will benefit. The difference however lies in how the deaths of the few come about. In scenario A, people die as an unintended, but foreseen, side effect whereas in scenario B, the deaths are not only intended but necessary to bring about the benefit. Here it is obvious that the people in scenario B are being used as a means (in order to bring about

advancements in cancer research) whereas those in scenario A are not. If the people in scenario A were to miraculously survive their suicide attempts then we would still retain the benefit of a higher level of education whereas the lack of death in scenario B would result in a lack of cancer research advancement.

However, the DDE is not omnipresent in our moral judgments. For example, Foot further describes the gas dilemma in which five patients can be saved from the manufacture of a certain gas, the process of which will however release toxic fumes into the room of another patient. The one patient in this scenario is not being used as a means (the five patients will still be saved even if he survives the toxic fumes), but the dilemma is still met with apprehension and as such the result is not correctly predicted by the DDE.

Thomson (1985) also provides an example of the shortcomings of the DDE by cleverly constructing a variation of the switch dilemma, labelled the loop variant, in which the only meaningful difference between the two scenarios is the presence of the DDE in the loop variant, where it is not present in the switch dilemma.

“On this case, in which the tracks do not continue to diverge – they circle back... Let us now imagine that the five on the straight track are thin, but thick enough so that although all five will be killed if the trolley goes straight, the bodies of the five will stop it, and it will therefore not reach the one. On the other hand, the one on the right-hand track is fat, so fat that his body will by itself stop the trolley, and the trolley will therefore not reach the five.” – The loop variant, Thomson (1985).

The only difference between this scenario and the switch dilemma is an added piece of railroad connecting the two tracks, such that the body of the one will stop the train circling round to the five and vice versa, thus the DDE is in full effect. This presents an issue for the DDE no matter the public opinion towards these two moral dilemmas. If Thomson’s prediction is true and people find these equally permissible, then the DDE fails and is, at the very least, not a universal moral principle. On the other hand, if people do respond differently to the two dilemmas, then it might signal a problem with our common ethics if the inclusion of an extra piece of track has the impact to kill or save lives.

More recently, psychological studies (Cushman et al., 2006; Greene et al., 2009; Hauser et al., 2007) have confirmed the use of the DDE in people’s judgment of (at least some) hypothetical

moral dilemmas. By comparing judgments across controlled pairs of dilemmas, varied by a chosen principle such as intention, the researchers were able to identify the specific moral principles that contribute towards our moral decision-making. Using a web-based questionnaire, Hauser et al. (2007) conducted the largest study of moral judgment to date, receiving over 5000 respondents. The results showed that the intention principle was generally followed by participants with very little variation between sub-groups, e.g., gender, age or ethnicity. However, a more interesting observation was that a majority of subjects who perceived a difference between the switch and loop dilemmas failed to cite the intention principle as a justification for their responses. Despite participants clearly following it, both Cushman et al. (2006) and Hauser et al. (2007) record that less than a third of participants in both studies were able to correctly identify their pattern of judgments as following the intention principle. These results indicate that the intention principle is employed in our moral judgments, but that the underlying neural algorithm is not available to our conscious mind. It is likely then that those participants who did correctly reference the intention principle were not truly recalling their moral calculations, but simply identified it rationally as the difference between the two dilemmas (Cushman et al., 2006). Theories supporting this style of moral decision-making fall under the emotivist branch of moral psychology (Haidt, 2001) and are discussed later.

2.2.3.1.2. Distributive Exemption

Distributive exemption, or the redirection principle (Hauser et al., 2007), was first mentioned by Thomson (1976). The principle holds that people generally find it morally worse to create a threat than they do to deflect or redistribute an existing one. Since its conception, the principle has been proven to affect people's moral judgments, most commonly between the switch and footbridge dilemmas (Greene et al., 2001; Hauser et al., 2007).

Thomson (1985) defends the morality of the redirection principle by explaining that “[the bystander in the switch dilemma] minimizes the number of deaths which get caused by something that already threatens people... The bystander who proceeds does not make something be a threat to people which would otherwise not be a threat to anyone; he makes be a threat to fewer what is already a threat to more.” In essence, “distributive exemption... permits arranging that something that will do harm anyway shall be better distributed than it otherwise would be”. This

arguably wouldn't apply to the bystander in the footbridge dilemma however, as instead of redirecting an existing threat (the trolley), he introduces an additional one (the danger of falling from a great height).

In another example, in the drug and serum dilemmas, a doctor can either save five patients by giving them a drug that was originally going to be used to save only one patient (who needed five times the dose of the others) or alternatively, he can kill the one patient, creating a serum from his body that can be used to save the five others. The first of these options is commonly thought to be permissible, whereas the other is obviously not. The difference can be explained using the distribution principle which would regard the action of turning a patient's body into a serum as immoral as it is introducing a threat that did not exist before, but the action of better distributing a supply of life preserving drugs as morally good.

Similar to the intention principle, Hauser et al. (2007) found that when participants were asked to justify their decisions regarding moral dilemmas that included the redirection principle, only 30% were able to correctly identify the redirection principle even when clearly following its maxim.

2.2.3.1.3. Rights

Rights, and duties, are the most basic forms of deontology: no matter the outcome, it is your duty to not kill and your right to not be killed. Thomson (1985) thought that an empathic concern for rights, as well as distributive exemption, was the key to understanding the trolley problem, saying that "[the bystander in the switch dilemma] saves his five by making something that threatens them instead threaten the one. And second, he does not do that by means which themselves constitute infringements of any rights of the one's." In other words, in order for an action to be permissible, in Thomson's view, it must first follow the law of distributive exemption while at the same time not infringing on anyone's rights, for example, a right to life.

Thomson (1976) illuminates the importance of rights in another variation of the switch dilemma called the mayor's variant, which goes as follows.

"The five on the track ahead are regular track workmen, repairing the track – they have been warned of the dangers of their job, and are paid specially high salaries to compensate. The right-hand track is a dead end, unused in ten years. The Mayor, representing the City, has set out picnic tables on it, and

invited the convalescents at the nearby City Hospital to have lunch there, guaranteeing them safety from trolleys. The one on the right-hand track is a convalescent having his lunch there; it would never have occurred to him to have his lunch there but for the Mayor's invitation and guarantee of safety. And [the bystander] is the Mayor." – The mayor's variant, Thomson (1976).

Thomson interprets in this situation that the convalescent has a greater right to not be struck by the trolley than any of the five workers and as such it is less permissible for the mayor to turn the trolley than it would be in the original switch dilemma. Other than in the mayors dilemma, people's rights helps us to understand responses to other dilemmas as well, such as the surgeon's dilemma. It is reasonable to think that a patient has the right to expect his or her doctor to care for and protect them instead of butcher them for the sake of others. Not only this, but people in general have a right to what belongs to them and limbs almost certainly belong to the people that they are attached too.

However, the application of rights can be ambiguous. For example, Thomson's reason that the bystander may proceed in the switch dilemma is because he does not violate any rights of the one man he is turning the train towards. Is this truly the case though? If for a moment we imagine a situation where there are not five men on the main track, but the bystander turns the trolley anyway, surely then the bystander infringes on a right of the one man by directing a trolley towards them without reason? If he does, then this raises further questions, such as why the presence of the five men denies the one man that right and also why the five men in the surgeon's dilemma do not strip the healthy patient of his right as they would in the switch dilemma.

Adding further complication to the principle of rights is the notion of the stringency of a right. Thomson (1985) points out that it might be the case that not all rights are equal and that there may be different degrees of stringency in rights, where it would be permissible to infringe some, but not all, in order to save the lives of others. For instance, a person's right to their property is certainly less stringent than their right to life, e.g., a burglar is likely to receive a smaller punishment than a murderer. To put this in the context of a moral dilemma, imagine that in order to throw the switch in the switch dilemma, the bystander must cross the one's private property and steal a tool in order to open a gate. It is obviously more permissible to act in these situations than in the footbridge dilemma because a person's right to private property or their own possessions are much less stringent than their right to life.

Degrees of stringency in rights can also lead to borderline cases, where it is unclear whether an action is permissible or not, depending on the responder's view of that right. Thomson (1985), for example, found it impermissible for the mayor to act in the mayors dilemmas as she believes the right the convalescent has to be on the track a stringent one, however upon asking other philosophers what they would do, she found that she was in the minority and many others thought it permissible for the mayor to throw the switch.

2.2.3.1.4. Contact Principle

According to the contact principle, "using physical contact to cause harm to a victim is morally worse than causing equivalent harm to a victim without using physical contact" (Cushman et al., 2006). For example, in the footbridge dilemma, the bystander has to physically push the heavy man off of the bridge in order to stop the trolley. Acting in this dilemma receives much less support than acting in the switch dilemma where the killing of the one man can be achieved from far away and through indirect, non-threatening means, i.e., throwing the switch.

Thomson (1976) briefly acknowledges this when she says that "what matters in these cases in which a threat is to be distributed is whether the agent distributes it by doing something to it, or whether he distributes it by doing something to a person," offering the analogy, "if there is a pretty shell on the beach and it is unowned, I cannot complain if you pocket it... But I can complain if you shove me aside so as to be able to pocket it."

Similar to the intention principle, Cushman et al. (2006) found that respondents to moral dilemmas do indeed recruit the contact principle in their moral decision-making. However, the researchers found that 60% of participants were able correctly quote the principle, twice as much as the intention principle. Despite this, although most participants were able to cite the principle, they went on to reject the moral relevance of it. In response to this, Cushman et al. (2006) says that "the contact principle guides moral judgment according to the intuitionist model and that a process of post hoc reasoning allows subjects to deduce the principle. However, once deduced, the principle is incorrectly regarded as morally irrelevant."

More recently, Greene et al. (2009) focused on isolating two aspects of the contact principle, personal force and physical contact, with the aim to see whether both or just one of these factors is necessary to affect people's moral judgments. In a between-subjects study, participants

answered one of a possible four moral dilemmas (force and contact, just force, just contact or neither). The results surprisingly indicated that harmful actions involving personal force were judged to be less morally acceptable whereas physical contact had no such effect. In other words, physical proximity to the victim (even close enough to touch) is less morally relevant than having to use personal force in order to kill them. This aligns with later research done by Cushman et al. (2012) who found that people's apprehension to perform a simulation of an assault was mediated by the motoric actions of that assault. Given the evidence, Greene (2015) instead refers to it as the force principle instead of contact principle.

As mentioned earlier, the force principle can cause some problems in the modern world as many very powerful weapons are operated by a simple button. For example, the bombs that were dropped on Hiroshima and Nagasaki were possibly deployed by the press of a button or the flip of a switch. However, if the pilots of the planes had to personally kill all of the people the bomb eventually killed, it is very unlikely that they could have gone through with it.

2.2.3.1.5. Action Principle

Finally, the action principle states that "harm caused by action is morally worse than equivalent harm caused by omission" (Cushman et al., 2006). Sunstein (2005) provides the example of one doctor who refrains from providing life-saving equipment to a patient and another who administers a patient with a lethal injection, both at behest of their patients. The first is seen as morally acceptable by a larger number of people than the latter despite Sunstein's, and many others, claim that they are fundamentally the same.

Although it might seem trivial, the principle is central to Foot's (1967) original dilemma and many others that have been spawned from it. The two choices that the responder is presented with are to take an action to cause one death, or to allow the death of five through omission. The fact that this situation itself is regarded as a dilemma should indicate how significant this principle is to us, given the life and death stakes of it. It should be noted, however, that when responding to a moral dilemma lacking any other principles, e.g., the switch dilemma, the majority of people do opt to act, killing the one in order to save the five (Hauser et al., 2007). Thus, while the principle is strong enough to transform the situation into a dilemma, it can usually be overcome with enough motivation.

In contrast to the intention and contact principle, Cushman et al. (2006) found that the majority of people were able to identify the action principle in their moral judgments and made reference to it when asked to justify their responses to a set of moral dilemmas. As such, while the other principles supported an intuitive view of moral judgments, the action principle lends credence to the older rationalist models of moral judgment (Kohlberg, 1969).

The action principle also extends beyond the trolley problem, into other facets of moral decision-making where it is usually referred to as 'omission bias', defined as "the preference for harm caused by omission over equal or lesser harm caused by acts" (Ritov & Baron, 1990). For example, Spranca et al. (1991) had subjects read a series of scenarios where a harmful outcome was either reached by an actor's action or omission. In the following two scenarios, John is aware that Ivan is allergic to the house dressing.

- A. John recommends that Ivan try the house dressing.
- B. John says nothing while Ivan orders the house dressing.

The outcome of the two scenarios ends in Ivan getting a stomach ache. The results show that participants frequently judged the actions of John in the first scenario to be morally worse than those in the second, citing John's intentions and motives as morally worse in the first scenario. However, this response does not hold up as John's motives were specified and identical in both scenarios (see the original paper for a full description of events). Spranca et al. (1991) explains that "harmful omissions are less blameworthy because the actor is less knowledgeable about the potential consequences... likewise harmful omissions are typically less intentional than commissions, but this difference cannot explain our results." The authors attribute the error in judgment as the misapplication of a moral heuristic, where harm as a result of an omission commonly does signal better intentions than through commissions, however people are too quick to apply this to every scenario without proper justification (Sunstein, 2005).

2.2.3.2. The Solution

Philosophers and scientists alike have proposed solutions for the trolley problem. These do not aim to tell us how to respond to moral dilemmas such as the switch dilemma (there is arguably no solution for this), but at least to explain why the issue of the trolley problem comes about, i.e.,

why we respond differently to consequentially identical moral dilemmas. While single principles might be able to explain the difference in our moral judgment between pairs of dilemmas, e.g., the difference between the switch dilemma and the loop variant can be defined using the intention principle, this only explains some subset of our morality specific to these cases and cannot explain our moral judgments in response to other situations.

2.2.3.2.1. Foot: Negative and Positive Duties

Foot (1967) proposes a system of duties, the terminology of which she borrows from Salmond (1947), where there are both negative and positive duties. Negative duties are those duties that are carried out through omission, for example, refraining from killing or robbing whereas positive duties are carried out through commission, such as looking after children or aged parents. It is also the case, according to Foot, that negative duties are morally superior to positive duties, with Foot saying that “even where the strictest duty of positive aid exists, this still does not weigh as if a negative duty were involved... [For example,] it is not... permissible to commit a murder to bring one’s starving children food.”

Applying Foot’s duty system to the tram dilemma, as in her original argument, presents two negative duties: “do not kill one person” and “do not kill five people”, whereas the judge’s dilemma contains a negative and positive duty: “do not kill the innocent” and “protect the community from the rioters”. Here there is a mismatch of duties (two negatives in the tram scenario and a negative and a positive in the judge scenario) explaining why they invoke different responses and noting importantly that it is the negative duty in the judge’s scenario that is the more morally permissible (in Foot’s opinion). A similar pattern emerges when applied to the drug and serum scenarios as well (two positives in the drug scenario and a negative and a positive in the serum scenario).

Thomson (1976) however refutes Foot’s solution by constructing the switch dilemma where the general response does not support the negative duty (which Foot claims should always triumph over a positive duty). Thomson points out that Foot’s duties system would call for the bystander to let the five die as it is the bystander’s negative duty to refrain from killing the one person and his positive duty to protect the five others, however this does not line up with the general response to the moral dilemma, therefore disputing this solution.

2.2.3.2.2. Thomson: Distributive Exemption and Rights

Thomson's (1985) final conclusion and solution to the trolley problem goes as such, "If the agent must infringe a stringent right of the one's in order to get something that threatens five to threaten the one, then he may not proceed, whereas if the agent need infringe no right of the one's, or only a more or less trivial right of the one's, in order to get something that threatens five to threaten the one, then he may proceed." Here, Thomson is recruiting both the redirection principle and rights in order to explain people's difference between consequentially identical moral dilemmas, specifically saying that people follow the law of distributive exemption, but that they also take into account the stringent rights of others when making their moral judgments. These rules solve many of the moral dilemmas originally proposed by Thomson (1976, 1985) such as the switch versus footbridge, surgeon versus serum and the drug versus gas dilemmas.

2.2.3.2.3. Greene: Intention and Force Principles

The advent of empirical studies targeted towards the trolley problem allowed psychologists to research the properties of human moral decision making with increased vigour and preciseness. As discussed previously, Cushman et al. (2006) identified the intention, contact and action principle all as playing roles in our moral judgment, as well as other studies such as the one conducted by Waldmann and Dieterich (2007) which found a small influence from what they call the 'locus of intervention' which states that people morally prefer to move the threat (as in the switch dilemma) than to move the one (as in the footbridge dilemma).

In Greene et al. (2009), the contact principle was redefined and split into the force and contact principles and it was found that only the force principle contributed significantly to people's moral evaluations. As of 2015, Greene (2015) considers only two key factors to play any large roles in our moral judgments of harm in the trolley problem; the intention and force principles, although Greene admits that "other factors contribute to the switch-footbridge effect but these two factors account for much, if not most, of it."

These solutions attempt to use moral principles as explanations for the trolley problem and purposefully ignore the underlying psychological and neurological mechanisms that respond to them. These will be discussed later in section 2.2.5.

2.2.4. Moral Psychology

Throughout history, the common belief was that moral judgment was the product of conscious reasoning and that emotions either did not factor into the equation or got in the way of it. For example, Plato and Aristotle claimed that reason was rooted in the head which ruled over the passions of the chest and stomach (Haidt, 2003). Jumping forward two thousand years, Kantian ethics declared that there are moral rules which could be followed rationally in order to live a moral life (Kant, 1785/2013).

Even when the study of moral psychology began to take shape towards the end of the 19th century, it was heavily biased towards the study of moral reasoning (Haidt, 2003). One of the first people to study morality in an empirical capacity was Emile Durkheim who strived to separate the burgeoning field of study of sociology from that of philosophy (Haidt, 2008). In Durkheim's view, morality consisted of two parts: moral rules, constructed by each society, that we are obligated, but also desire, to obey as well as rational interpretations of those rules by individuals (www.iep.utm.edu/durkheim). Inspired by the work done by Durkheim, Jean Piaget, proposed what would become the first of many theories of moral cognitive-development (Piaget, 1932/1965). Similar to Durkheim, Piaget thought that our view of morality comes from our peers, however whereas Durkheim thought it came from our respective societies as a whole, Piaget went further and modelled how our morals change from childhood to adulthood, explaining that children focus more on the consequences of actions and adults more on the intentions. A core tenant of Piaget's model (along with many models based on and inspired by it) however is that people come to their moral conclusions rationally and through pure reason alone.

After Piaget came the cognitive revolution of the 1960s when more emphasis was placed not just on how people behaved, but on how and why the brain was making them behave in such a way (Haidt, 2008). During this time Lawrence Kohlberg ruled the realm of moral psychology, expanding Piaget's theory to involve six stages of moral development, the final stage involving justification of actions using Kant's categorical imperative, i.e., that you should not use people as a means (Kohlberg, 1969). Although Kohlberg's model was still based on the assumed fact that people reasoned their way to moral conclusions, Kohlberg accepted that emotions can play a part in moral judgment, but that it was not the driving force and was primarily controlled by reason

(Haidt, 2001). The years following Kohlberg's reign saw little change beyond revisions of Kohlberg's theories from psychologists such as Elliot Turiel, who proposed the moral interactionist model (Turiel, 1983).

Contrary to the overwhelming support for conscious reasoning were a few dissenters, who argued for the role of emotions and intuitions in our moral judgments. The most prominent moral philosopher who advocated for the importance of emotions was David Hume, who famously said that "reason is, and ought only to be the slave of the passions, and can never pretend to any other office than to serve and obey them" (Hume, 1738/2003). Hume thought that reasoning and logic had its place in our morals, but that without sentiment, we would not be able to assign value to life and thus could not reach any sort of moral judgment about it. For example, logic can tell us that five lives is greater than one life, but without sentiment giving us the ability to value those lives saved, it would not matter to us whether we saved the five or the one. It would be like someone asking you whether you want one shiny red button or five, it does not really matter because you do not really want any. You simply do not care. Sentiment also helps us in the footbridge dilemma, sympathy for the man on the footbridge inhibits our urge to push him to his death, although this would save five. A disastrous lack of sentiment can be seen in people with damage to specific parts of the brain (those parts that deal with sentiment and empathy) leading to psychopathic tendencies (Damasio, 2006). In fact, studies have shown an increase in permissibility of the footbridge dilemma among patients with psychopathic tendencies (Koenigs et al., 2007).

Hume's emotivism would return in the 1980s as the "affective revolution" (Haidt, 2003). Triggered by the stagnation of the cognitive revolution, research papers on moral emotions such as anger, shame, guilt, embarrassment and disgust began to skyrocket as papers on moral reasoning began to dwindle. Psychologists realised, like Hume did centuries earlier, that these emotions do indeed contribute greatly to our moral judgments. For instance, we do not need to reason to the conclusion that it is immoral to have sex with a close relative, our disgust would tell us if we thought about it. We do not need to reason to the conclusion that we should not abuse the elderly, our guilt would tell us if we thought about doing it. And we do not need to reason to the conclusion that atrocities happening in less fortunate countries are bad, our anger tells us when we hear about it in the news. Some theorists went even further, claiming that emotional reactions are the driving force of moral judgment and that moral reasoning is often just post hoc rationalization. In

other words, our moral judgments and justifications are not derived from each other, justifications are simply a best guess at why we arrived at a previous moral judgment, with the judgment being reached through unconscious emotional processes. This idea would become the basis for Haidt's renowned Social Intuitionist Theory (Haidt, 2001).

2.2.4.1. Moral Emotions

Emotions provide us with the ability to quickly assess how we should respond to objects and events around us without requiring us to consciously evaluate them. This saves time and effort during situations that can be handled instinctively. We do not consciously choose to feel emotions and we cannot choose what emotion to feel in a specific situation, they are ingrained in us from birth and adapt as we grow. For example, being happy tells us that the situation is good and that we should strive to prolong or replicate it, whereas being sad tells us that the situation is bad and that we should do something to rectify it.

During a morally charged situation, moral emotions help us to gauge how we should react. For example, Valdesolo and DeSteno (2006) say that "considering personal moral violations, such as inflicting direct harm, elicits prepotent negative reactions that appear designed to inhibit amoral acts." The study itself found that altering the emotional state of participants responding to a series of moral dilemmas successfully changed the responses to those dilemmas. Greene et al. (2001) demonstrated this directly by showing an increase in activity in areas of the brain associated with emotions when participants were asked to consider a moral dilemma involving egregious harm.

Moral emotions are not necessarily distinct from normal emotions, but there are certain emotions that are more common during moral situations, such as anger, guilt or disgust. Haidt (2003) defines moral emotions as "those emotions that are linked to the interests or welfare of society as a whole or at least of persons other than the judge or agent." In this definition, the emotions that are triggered by a social event, not necessarily including the agent, qualify as a moral emotion, whereas those triggered by an event relating only to the self would be non-moral, e.g., happiness or sadness can be triggered in non-social settings so these are considered primarily non-moral emotions. Emotions such as guilt, anger or compassion on the other hand are generally provoked

by social situations and relate to the actions of others as well as the self and so might be labelled as moral emotions during in those instances.

Haidt (2003) further provides two features of moral emotions that distinguish them from the non-moral: “the more an emotion tends to be triggered by such disinterested elicitors, the more it can be considered a prototypical moral emotion” as well as the degree “to which there is an increased tendency to engage in certain goal-related actions, such as helping others or upholding the social order.” In other words, emotions more likely to be caused by people unrelated to us are more likely to be moral as well as those that generate a powerful urge to act. These two features make it clearer that there is no neat division between moral and non-moral emotions. An emotion such as anger can be both moral and non-moral depending on the eliciting situation, i.e., anger at the self or anger towards an unrelated someone or something. There are emotions however that are less likely to be elicited by external happenings and thus are less likely to be moral emotions, for an estimation of an emotions likelihood to be elicited as a moral emotion, see Figure 2.4.

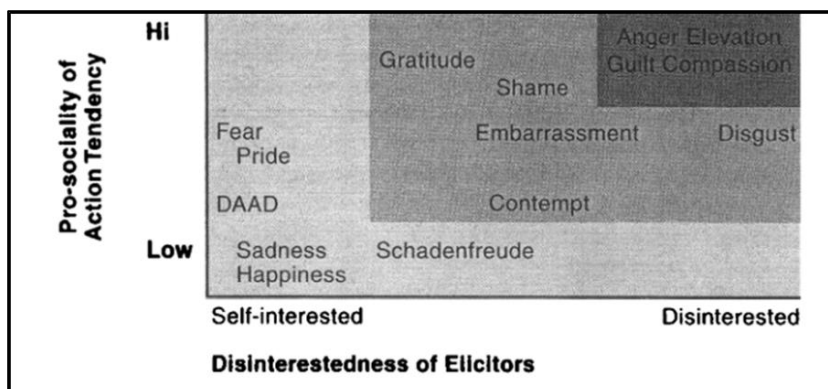


Figure 2.4. – Moral emotions (Haidt, 2003).

Henceforth, the focus will be on the three moral emotions that are most commonly related to the trolley problem: disgust, guilt (or shame) and compassion.

Disgust is a negative visceral response that evolved to stop us doing something or interacting with something unpleasant. It is typically thought of as a response to physical objects such as rotting meat or faeces, but non-physical objects such as obscene actions can also cause a similar feeling of revulsion. In fact, many languages use similar words to represent disgust towards objects and condemnation of actions, for example, a man who behaves chauvinistically is

commonly called a pig, an animal associated with disgusting habits. Disgust can help us during moral dilemmas by making us feel uneasy about partaking in immoral acts, even if there might be a better outcome if we did them. For example, in the surgeon's dilemma, we are offered the chance to cut up a healthy patient, harvesting their organs which can then be used to save five other patients. Our disgust towards killing, not to mention through particularly grisly means, makes us not want to act in this situation and the general consensus to this proposition is that it is morally wrong to do this. Without disgust, cutting up the man might not seem any worse than throwing the switch in the switch dilemma.

Guilt (or shame) works similarly to disgust in that it is an inhibitor, i.e., a negative moral emotion that informs us to stop when we have done or are considering doing something wrong. The difference between this and disgust is that guilt is based on our desire to be highly perceived by others. Humans have an innate desire to be accepted by their member groups and so spend a great deal of effort towards the pursuit of appearing to be a morally good person (Tajfel, Turner, Austin, & Worchel, 1979). Haidt (2003) says that "feelings of shame... indicate to us that we are marked both as a poor partner for future interactions and as an appropriate target for ostracism," which results in an immediate effort to stop or undo the actions that elicited the emotion. In the mayor's dilemma, for example, we have the option to turn a trolley away from five workers towards a patient of a nearby hospital who we (the mayor) have promised can sit where they are, or consider the judge's dilemma in which we (a judge) have the option of sentencing an innocent man to death in order to quell a riot which will result in five deaths. In both situations we would feel shameful knowing that we have sentenced innocent people to die and even more so because of the public nature of the dilemmas. In the end, there may be a case for acting in these situations, but at the very least, our feelings of guilt or shame would make it a more difficult decision.

Finally, Haidt (2003) says that "compassion is elicited by the perception of suffering or sorrow in another person", also saying that it is most strongly felt for one's kin or those in a close relationship, although people can feel compassion for total strangers. In the switch dilemma, compassion is what makes us strive to save the five. While logic and reasoning are often said to be the cognitive defenders of the five (Greene, 2008), these are simply numbers unless we actually care about what happens to them. This reflects the views of David Hume who thought that without sentiment we could comprehend but would not be able to value, the lives of others.

2.2.4.2. Dual-Process Theories

The affective revolution emphasized the importance of emotions in our everyday thinking. This emphasis on non-rational neural processes continued into the 1990s with one of the most important advancements in psychological sciences, the Dual-Process Theory (DPT) (Bargh & Chartrand, 1999). The DPT proposes that we have two cognitive faculties, sometimes labelled system one and system two (Kahneman & Egan, 2011). System one is ancient, evolving millions of years ago in order to handle basic life preserving functions, e.g., eat, sleep and mate. System two on the other hand evolved fairly recently, make us cognizant, sentient or conscious and giving us the ability to reason through complex problems. Mental processes can recruit either or both of these systems when making a decision depending on the decision being made and the deciders mental state.

The two systems differ significantly in how they operate. System one is fast, allowing us to respond to situations quickly and effortlessly, but functioning unconsciously. Therefore we have little say in the conclusion arrived at. On the other hand, system two is conscious, but slow and effortful. We have to work mentally hard when using system two which can be draining. Haidt (2007) defines system one as “the fast, automatic, affect-laden processes in which an evaluative feeling of good-bad or like-dislike appears in consciousness without any awareness of having gone through steps of search, weighing evidence or inferring a conclusion” and system two as “the controlled and ‘cooler’ process; it is a conscious mental activity that consists of transforming information about people and their action in order to reach a moral judgment or decision.”

Furthermore, it is important to note that the DPT does not treat these two processes as equals, stating that system one is the “default” process and that system two is only utilised under the right conditions (Bargh & Chartrand, 1999). This refers first to whether or not system two is needed. As system two requires a lot more effort than system one, it is only activated when necessary, for instance, when solving a complicated math problem. Additionally, system two requires resources to be used, i.e., the time enough to think through the problem or the energy to think effortfully for an extended period of time. Not only this, but even when system two is engaged, it is often biased by the conclusions previously reached using system one. Bargh and Chartrand (1999) say that “it may be, especially for evaluations and judgments of novel people and objects, that what we

think we are doing while consciously deliberating in actuality has no effect on the outcome of the judgment, as it has already been made through relatively immediate, automatic means.”

Greene (2015) adopts the useful metaphor of treating our decision-making like a digital camera. Like our dual-process brains, cameras have both an automatic and manual mode. The automatic mode allows users with less time or ability to take a quick picture that is generally good enough for most situations, but might fail from time to time, requiring the user to enter manual mode. Here, the user can think for themselves and choose the best options available, however, like system two, this takes time, effort and specialist knowledge to be effective at the problem at hand. System one on the other hand matches the cameras automatic mode, allowing us to go about our day without having to effortfully think about every decision, however it can fail and provide us with an incorrect judgment.

How, then, does system one generate decisions so quickly? The most dominant theory is that it uses heuristics based on our motivations and biases. Chen and Chaiken (1999) say that we have three personal motivations: impression, accuracy and defence. The impression motivation drives our interpersonal behaviours and relies on heuristics such as “consensus equals correctness”. Our defence motivation prompts us to protect our privately held views of the world leading to extreme biases in favour of one’s own opinions and against those of the opposition. Finally, our accuracy motivation urges us to have correct information about the world using heuristics such as “length equals strength” – it is generally true that larger things tend to be stronger – and the ‘availability heuristic’ which holds that the easier something is to remember, the greater importance we attribute to it.

A problem can arise with our use of heuristics, however. As they are often applied quickly, automatically and unconsciously, we can often fail to think critically about them and regard them as correct without a second thought. Furthermore, due to the nature of heuristics, while they can be correct some or most of the time, they are certainly not true all of the time and can lead us to make extreme errors (Spranca et al., 1991). Sunstein (2005) provides two examples of this. Consider the following two questions: In the average sentence, how many words end in “ing” and how many words will have “n” as the second-to-last letter? Words ending in “ing” are more easily retrievable than words that have “n” as the second-to-last letter and so due to our availability heuristic, the first question will generally receive a higher answer than the second, even though

this is logically impossible. In another example, given a description of a woman matching a stereotypical feminist and two possible answers, responders will often say that the woman is more likely to be a “bank teller and active in the feminist movement” rather than simple a “bank teller” even though it is logically impossible for it to be more likely for someone to have characteristics A and B than just characteristic A.

Dual-process theories and the characteristics of them described above are designed to explain how our neural mechanisms work during a decision-making process. In recent years, more specific models have been proposed for distinct types of decision-making, such as moral decision-making (Cushman, 2013; Greene, Nystrom, Engell, Darley, & Cohen, 2004; Haidt, 2001). In the next sections, the Social Intuition Model (Haidt, 2001), Dual-Process Theory of Moral Judgment (DPToMJ) (Greene et al., 2004), as well as Cushman (2013) version of the DPToMJ, will be discussed.

2.2.4.3. Social Intuitionist Model

One phenomenon that cannot be explained by rationalism is the inability for people to fully explain their moral judgments. If we look back to the trolley problem, Thomson (1985) asserted that it was permissible to act during the switch dilemma, but not the footbridge dilemma. If we truly used reasoning to come to our moral conclusions, then Thomson should have been able to immediately tell us what the difference was between the two situations that changed her mind. Instead we have trolley-ology, a massive field of moral psychology whose purpose is to identify why people respond differently to almost identical moral dilemmas. Haidt (2001) provides us with another example of this using a short story.

Julie and Mark are brother and sister. They are traveling together in France on summer vacation from college. One night they are staying alone in a cabin near the beach. They decide that it would be interesting and fun if they tried making love. At the very least it would be a new experience for each of them. Julie was already taking birth control pills, but Mark uses a condom too, just to be safe. They both enjoy making love, but they decide not to do it again. They keep that night as a special secret, which makes them feel even closer to each other. What do you think about that? Was it OK for them to make love?

When asked about their opinions on this story, Haidt says that responders will often cite accidental pregnancy or psychological damage even though the story accounts for both of these things; excessive contraception is used and the experience makes them feel closer, not damaged. Even then, responders are recorded as saying something similar to “I don’t know, I can’t explain it, I just know it’s wrong.”

Haidt (2001) attempts to explain this behaviour and others like it using his Social Intuitionist Model (SIM), the primary claim of which is that “moral judgment is caused by quick moral intuitions and is followed (when needed) by slow... moral reasoning.” The “when needed” constitutes the social part of the SIM where demand for justification of a moral judgment tasks the responder with finding post-hoc arguments in favour of their original answer. This stands in contrast to previous rationalist models of the 20th century which assumed a strict causal relationship of moral belief to judgment (Kohlberg, 1969).

When we think about sticking a pin into a child’s hand, most of us have an automatic intuitive reaction that includes a flash of negative affect. Any controlled processes would thus occur after the first automatic process had run and would be influenced by the initial moral intuition, i.e., we would think negatively about it even if were for the greater good. Haidt (2007) then argues that “moral reasoning, when it occurs, is usually a post-hoc process in which we search for evidence to support our initial intuitive reaction” and goes on to provide a plethora of evidence supporting this claim.

The SIM itself is composed of four primary links (labelled 1 to 4 in Figure 2.5) and two secondary ones (labelled 5 and 6 in Figure 2.5) where links are related to moral judgments, justifications and discussions. The first link the model proposes is the “intuitive judgment link”, which holds that moral judgments are likely to be the result of moral intuitions, providing responders with answers automatically and effortlessly, but does not provide them with any of the steps or rationale taken to get to the answer. Link number two proposes that reasoning and justification come after a judgment has already been made, resulting in responders searching for arguments in favour of their stance. It is important to note that these justifications are not retrieved from memory due to the nature of system one heuristics, but instead are generated after the judgment has been made and without any true knowledge of how we actually came to that judgment, therefore they are only best guesses and do not truly reflect our moral intuitions. Link three and four encompass the

social part of the model and detail how we might interact with those around us to discuss our moral judgment. Haidt (2001) has a rather dismal outlook on moral discussions however, saying that “moral discussions... are notorious for the rarity with which persuasion takes place.” Haidt cites the “illusion of objectivity” which refers to how both parties can often regard valid criticisms as irrelevant or immoral and then view the opposition as close minded and insincere for doing the same thing. This can frustratingly be seen time and time again in political debates, where the intention of the discussion is to find the best solution to a problem, but the goal of either side is to champion their own solution and disregard the opponents. When both sides do this, no progress can be made.

The final two subsidiary links mirror the older rationalist models, where Haidt agrees that some people might genuinely reach moral judgments through rationale alone, although he posits that this is very rare.

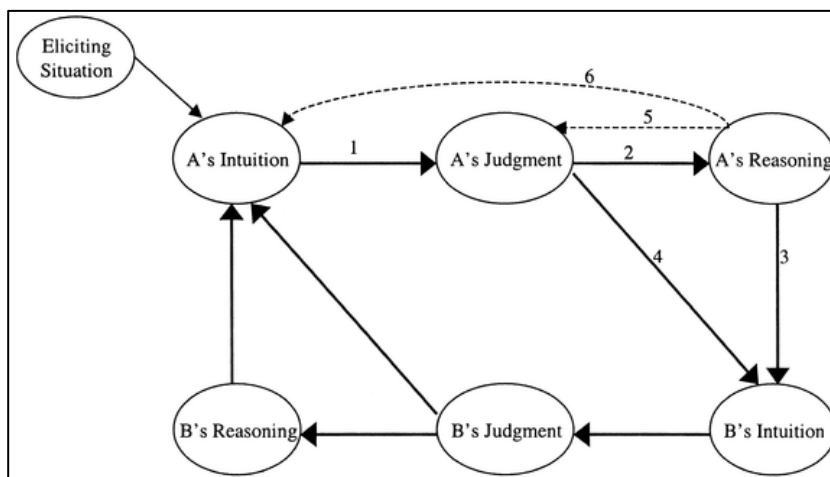


Figure 2.5. – The Social Intuitionist Model (Haidt, 2001).

The SIM relies heavily on the notions of reasoning and intuition. Haidt (2001) describes these in terms of moral decision-making, defining moral reasoning as the “conscious mental activity that consists of transforming given information about people in order to reach a moral judgment.” On the other hand, moral intuition is the “sudden appearance in consciousness of a moral judgment... without any conscious awareness of having gone through the steps of searching, weighing evidence, or inferring a conclusion.” Moral intuition comes equipped with its own moral heuristics which Haidt describes as “a priori moral theories” and regards them as a set of “culturally supplied

norms for evaluating and criticizing others.” For example, we might apply the heuristic “unprovoked harm is bad” when considering pushing the other bystander in the footbridge dilemma.

What is important here is that when we reach judgments through intuition, the true cognitive path to our judgment is not accessible to consciousness meaning we cannot know the actual logic behind our answer. Thus, in order to justify our moral decisions, we must instead use reasoning with consciously available information to produce what we believe to be the real rationale behind our answer. This process, as a conscious one, is effortful and can feel like we are accessing data from working memory, which is why people often regard their justifications as the true reasoning for intuitional judgments, where really they are post-hoc rationalisations (Greene, 2008). This is also not a rare event, as said earlier, many psychologists regard intuition as the default system used to make decisions and that conscious reasoning is only involved when absolutely needed (Bargh & Chartrand, 1999; Kahneman & Egan, 2011).

Haidt (2001) echoes this sentiment, saying that “when asked to explain their behaviours, people engage in an effortful search that may feel like a kind of introspection. However, what people are searching for is not a memory of the actual cognitive processes that caused their behaviours, because these processes are not accessible to consciousness. Rather people are searching for plausible theories about why they might have done what they did.” Here, Haidt is saying that although it might feel as if we are digging into our memory to find out why we hold the moral outlook that we do or took that action that we took, what we are doing instead is extrapolating from known data in order to guess (although it does not feel like guessing) why we hold that view or took that action.

2.2.5. Moral Neuroscience

Similar to how the study of ethics moved from philosophy to psychology around the start of the 20th century, it has recently transitioned into neuroscience as well. Starting primarily around the start of the millennium, kickstarted by the research conducted by Joshua Greene, Jorge Moll and others (Greene & Haidt, 2002; Greene et al., 2004; Greene et al., 2001; Heekeren, Wartenburger,

Schmidt, Schwintowski, & Villringer, 2003; Moll, de Oliveira-Souza, Bramati, & Grafman, 2002; Moll, de Oliveira-Souza, Eslinger, et al., 2002; Moll, Eslinger, & Oliveira-Souza, 2001), the study of ethics in neuroscience has aimed to discover what areas of the brain contribute to our moral judgments and whether there are moral specific neural networks at work.

As was predicted previously by Haidt (2001), many of the neural regions activated during moral contemplation are associated with emotional and social cognitive processing (Young & Dungan, 2012). This recent advancement provides undeniable evidence that the prediction made by Hume hundreds of years ago that our morals are at the mercy of our passions is indeed true and that any model of moral judgment moving forward must account for the human-emotional disposition.

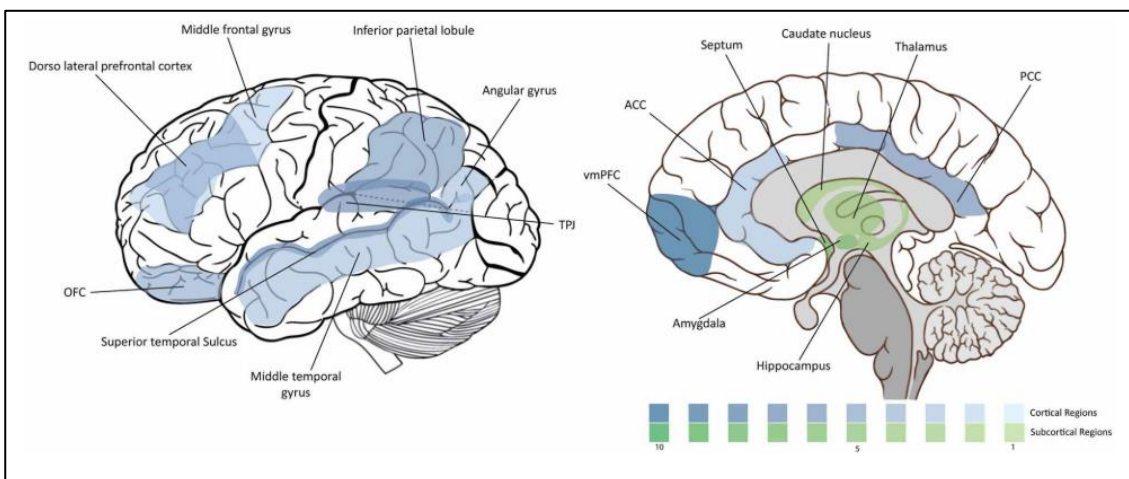


Figure 2.6. – Brain regions implicated in moral decision-making (Pascual, Gallardo-Pujol, & Rodrigues, 2013).

The most consistently engaged brain regions either during presentation of a moral stimulus (e.g., descriptions or indecent images) or contemplation of a moral judgment (e.g., as in response to the switch, footbridge or any other moral dilemma) include the anterior cingulate cortex (ACC, BA 32) (Greene et al., 2004; Pascual et al., 2013), frontopolar cortex (FPC, BA 10) (Moll, De Oliveira-Souza, & Zahn, 2008), inferior parietal lobe (BA 40) (Greene & Haidt, 2002), dorsolateral prefrontal cortex (dlPFC, BA 9 and 46) (Greene et al., 2001), orbitofrontal cortex (OFC, BA 10 and 11) (Moll, de Oliveira-Souza, Eslinger, et al., 2002), medial frontal gyrus (mFG, BA 9) (Young & Dungan, 2012), posterior cingulate cortex (PCC, BA 23 and 31) (Schaich Borg et al., 2006), amygdala (Shenhav & Greene, 2014), superior temporal sulcus (STS, BA 19 and 39) (Moll, de

Oliveira-Souza, Bramati, et al., 2002; Moll et al., 2001) and ventromedial prefrontal cortex (vmPFC, BA 10) (Heekeren et al., 2003; Koenigs et al., 2007; Moll & de Oliveira-Souza, 2007).

Although a complex series of brain regions are shown to become active during moral judgment, the area most commonly associated with morality is the vmPFC (Moll & de Oliveira-Souza, 2007). Several studies have repeatedly shown vmPFC activations in response to moral statements and scenes as well as during moral dilemmas involving evaluative judgments and decision-making (Greene et al., 2004; Heekeren et al., 2003; Moll, de Oliveira-Souza, Eslinger, et al., 2002; Shenhav & Greene, 2014). As well as this, research has shown that damage to the vmPFC generally leads to severe impairments of moral behaviours as a result of diminished social emotions such as compassion, shame and guilt, resulting in dangerous psychopathic behaviour in patients (Koenigs et al., 2007).

The vmPFC, along with a network of other regions including the amygdala, STS and PCC (including the neighbouring precuneus) have been implicated in the processing of emotions and social cognition (Young & Dungan, 2012). The vmPFC specifically plays a critical role in theory of mind tasks such as understanding other people's behaviour and in emotional decision-making (Heekeren et al., 2003; Moll, de Oliveira-Souza, Eslinger, et al., 2002). Synonymous with the vmPFC, the OFC has also been implicated in the guidance of potential moral behaviours through emotionally relevant punishment and reward evaluations (Moll, de Oliveira-Souza, Bramati, et al., 2002). More recently, Shenhav and Greene (2014) found that the vmPFC was more active during moral judgments that utilised both emotional and rational cognitive processes (such as those arising from the dlPFC and inferior parietal lobe) than when either were considered in isolation, resulting in the theory that the vmPFC is involved in the integration of results from emotional cognitive processes with that of rational. This role was previously attributed to the ACC which responds to conflict in the brain, such as during the Stroop task (Greene et al., 2004). However, the conflict hypothesis regarding the ACC is tentative and it is currently unclear how the ACC interacts with moral judgment, being the least researched region with regards to moral decision-making.

Similar to the vmPFC, the posterior STS, located in the temporal lobe, has been implicated in the processing of emotional and social cognition (Greene & Haidt, 2002). The STS, however, is more active when participants are presented with either moral statements or vignettes, rather than

morally indecent pictures (Greene et al., 2004; Moll, de Oliveira-Souza, Eslinger, et al., 2002; Schaich Borg et al., 2006) and is sensitive to stimuli describing the actions or intentions of others (Heekeren et al., 2003). As such, the leading hypothesis regarding the STS is that it is responsible for the perception and representation of socially relevant information, allowing us to understand the beliefs and intentions of others (Moll et al., 2008). This means understanding the difference between seeing someone lamentingly act during the switch dilemma and someone else seemingly enjoying pulling switch. While both actions are the same, biological information such as facial expression and body movement would indicate a difference in motivations.

In contrast to the STS, the PCC/precuneus and amygdala, both situated in the limbic sector of the brain, have been found to be more active during the presentation of emotionally-charged moral imagery, rather than words (Greene & Haidt, 2002). The amygdala specifically is a phylogenetically ancient system that evolved long before our executive faculties and is strongly linked to the emotion of fear, as well as others, something that our distant ancestors would have commonly relied on to survive. Shenhav and Greene (2014) showed a high degree of coupling between the amygdala and vmPFC during purely emotional assessments and a much weaker connection during induced rational assessments. The PCC on the other hand seems to be involved with memory processing and becomes preferentially activated when moral dilemmas involve close relatives (Schaich Borg et al., 2006).

The mFG is also correlated with our moral-emotional processing, evidenced by several studies (Moll, de Oliveira-Souza, Eslinger, et al., 2002; Moll et al., 2001), and is thought to provide us with self-referential capabilities. This is most clearly seen during personal moral dilemmas¹ which are likely to require a greater amount of introspection than other types of moral dilemmas (Greene et al., 2001). This activation of the mFG also lines up with the notion of our morality as a key aspect of our identities (Schaich Borg et al., 2006).

The regions of the brain discussed in-depth thus far have all been observed to correlate with the generation of emotional responses to stimuli. This aligns neatly with more recent theories that emotions such as disgust, guilt or compassion play a large role in our moral judgments. Looking again at Haidt's (2001) SIM, the cognitive processes discussed here would take place during the first main link, i.e., "intuitive judgment". However, rational cognitive processes also seem to take place during moral deliberation as well, although not as prominently as emotional intuitions. This

partially opposes the Haidt's SIM and lends a small amount of support for older models of moral judgment (Kohlberg, 1969).

These rational cognitive processes seem to originate from the dlPFC region and inferior parietal lobe (Schaich Borg et al., 2006). Both of these networks have been recorded activating during non-moral dilemmas (Heekeren et al., 2003), impersonal moral dilemmas² (Greene et al., 2001) and difficult, as compared to easy, personal moral dilemmas³ (Greene et al., 2004). It has been suggested that both of these areas are associated accessing with working memory, abstract reasoning and problem solving. Furthermore, they have been linked with the suppression of negative emotions during moral judgment. This may be what allows responders to overcome their intuitions and take action during moral dilemmas that require indecent acts in order to attain greater utility.

Here it is clear that morality is supported not by a single brain circuitry, but by a multiplicity of circuits that overlap with other general complex regions (Pascual et al., 2013) that have all been implicated in non-moral processes. Greene and Haidt (2002) agree, saying that “there is no specifically moral part of the brain” and that “[we] believe that the ordinary concept of moral judgment refers to a variety of more fine-grained and disparate processes, both ‘affective’ and ‘cognitive’.”

1. A personal moral dilemma is one in which the act that the responder is asked to commit is particularly egregious. As such, people typically answer in favour of deontology when responding to a personal moral dilemma, i.e., they choose not to act. The most well-known personal moral dilemma would be the footbridge dilemma.
2. An impersonal moral dilemma is one in which the act that the responder is asked to commit is relatively innocuous. As such, people typically answer in favour of utilitarianism when responding to an impersonal moral dilemma, i.e., they choose to act. The most well-known impersonal moral dilemma would be the switch dilemma.
3. An easy personal moral dilemma is one in which the act that the responder is asked to commit is particularly egregious and in which the benefits for committing that act are unequivocally not worth the act. An example would be the infanticide dilemma. A difficult personal moral dilemma is one in which the benefits for committing the act are at least debatably worth the act. An example would be the footbridge dilemma.

2.2.5.1. Dual-Process Theory of Moral Judgment

Greene (2008, 2015) asserts that emotional cognitive processes are likely a response to the immoral actions that participants are asked to commit during a moral dilemma in order to save the five. Recall from earlier that acts can be defined by a number of moral principles. For example, in a personal moral dilemmas such as the footbridge dilemma, pushing the man off of the footbridge involves several moral principles and usually invokes a strong negative emotional reaction. However, impersonal moral dilemmas such as the switch dilemma, where one can simply throw a switch in order to save five, involve very few moral principles and typically invoke a weaker emotional response and are also responded too differently. Therefore, the level of emotional reaction in response to a moral dilemma is correlated with the unpleasantness of the action needed to save the five which in turn is related to our willingness to commit that action, i.e., negative emotions lead to deontological responses.

Contrary to emotions, Greene (2008, 2015) asserts that rational processes help responders to calculate the greater good of a scenario. For example, if we have the choice to either kill one person or five people, our executive faculties allow us to calculate that five lives are greater than one and that saving five would result in the most happiness. Thus, in a vacuum, we would rationally choose to kill the one, i.e., behave in a utilitarian fashion. However, moral dilemmas are rarely this easy, and an unpleasant action (triggering a correlated emotional response) is often required to achieve the greatest good.

The move from psychology to neurology was vital to the development of what is now one of the most popular theories of moral judgment: the Dual-Process Theory of Moral Judgment (DPToMJ). The DPToMJ is the culmination of a series of work done by Joshua Greene and colleagues (Greene et al., 2008; Greene et al., 2004; Greene et al., 2001). The DPToMJ is similar to dual-process theories in other areas of psychology and neurology, resting on the basic idea that there are two processes that contribute towards our moral judgment: intuitive (or emotional) and rational. These processes respond to separate aspects of moral dilemmas and guide our moral judgment towards its final answer. Unique to the DPToMJ however is Greene's Central Tension Principle which states that: "Characteristically deontological judgments are preferentially supported by automatic emotional responses, while characteristically consequentialist judgments are

preferentially supported by conscious and allied processes of cognitive control” (Greene, 2015). This neither lines up with Kohlberg’s rational model or Haidt’s emotivist model, suggesting a synthesis of the two viewpoints that acknowledges roles played by both emotion and reasoning (Greene et al., 2008).

This principle is based on findings starting from Greene et al. (2001). In this paper, the researchers used fMRI to scan the brains of participants while they read and considered a series of moral dilemmas, including the switch and footbridge dilemmas. As mentioned previously, people find it permissible to kill the one man in the switch dilemma, but not in the footbridge dilemma and a great deal of effort has gone in to working out why (Thomson, 1985). The theory presented by Greene and his colleagues is that the different responses to the switch and footbridge dilemmas is a product of the distinct levels of emotional engagement that the dilemmas trigger. More specifically, that personal moral dilemmas (i.e., the footbridge dilemma and others similar to it) are much more emotional than impersonal moral dilemmas (i.e., the switch dilemma and others similar to it).

Providing the first neurological evidence as to why people respond differently to the switch and footbridge dilemmas, the results revealed that areas of the brain previously associated with emotional processing, such as the vmPFC and amygdala, were indeed more active during personal moral dilemmas than during impersonal moral dilemmas and non-moral dilemmas, with the latter two having quite similar neural activations such as an increased activation of the dlPFC and inferior parietal lobe, areas associated with working memory. Participants also took longer to choose the inconsistent answer during personal moral dilemmas (i.e., killing the one), possibly a product of the activation of emotional cognitive processes and the subsequent time needed to overcome them.

However, Greene et al. (2001) only compared cognitive patterns between different types of moral dilemmas, i.e., personal and impersonal. Greene et al. (2004) went a step further and looked at and predicted different cognitive patterns based on participants answers to the moral dilemmas and indeed showed that utilitarian responses to personal moral dilemmas positively correlated with increased activity in traditionally cognitive regions of the brain including the dlPFC, inferior parietal lobe and ACC.

Combining these results, we can see that personal moral dilemmas consistently, short of psychopathy (Koenigs et al., 2007), elicit emotional reactions. As said earlier, this is likely a response to the unpleasant acts we are asked to commit in a moral dilemma. Additionally, it is evident that utilitarian responses to personal moral dilemmas correlate with system two processes associated with the dlPFC, inferior parietal lobe and the ACC. The dlPFC is also activated during impersonal moral dilemmas and has thus been implicated in our ability to calculate the greater good, however, the ACC is specific to personal moral dilemmas and is posited to assist in reaching a resolution between both our activated emotional and rational cognitive processes (Greene, 2015). Greene et al. (2004) says that “according to our theory, [the parent’s] dilemma is difficult because the negative social-emotional response associated with the thought of killing one’s own child competes with a more abstract, ‘cognitive’ understanding that, in terms of lives saved/lost, one has nothing to lose and much to gain by carrying out this horrific act. We believe that the ACC responds to this conflict and that control-related processes in the dlPFC tend to favour the aforementioned ‘cognitive’ response.”

In support of the DPToMJ, Valdesolo and DeSteno (2006) found that participants who were happier (induced by showing a comedy clip beforehand) were more likely to choose the utilitarian option in a moral dilemma. Similarly, Koenigs et al. (2007) found that patients suffering from induced psychopathy from damaged to the vmPFC were also more likely to choose the utilitarian option in response to personal moral dilemmas like the footbridge dilemma. Both of these papers altered the emotional state of participants (the first by raising it and the second by effectively eliminating it) and provided causal evidence that the impairment of emotion leads to increased utilitarian answers. To compliment these studies, both Greene et al. (2008) and Paxton, Ungar, and Greene (2012) found that interfering with or aiding cognitive capabilities during moral dilemma judgments impacted utilitarian answers, but not deontological. For many more results in support of the DPToMJ, see Greene (2015).

A question remains however, why do emotional processes support deontology and “cognitive” utilitarian? Greene (2008) suggests that moral emotions are like alarm bells, quickly and loudly telling us, without conscious rhyme or reason, to not do something. This aligns with deontology’s forceful assertion that everybody has certain rights and that it is strictly morally wrong to infringe on those rights. For example, when we think about pushing the man off of the footbridge, we might feel disgust or guilt, stopping us from infringing on the rights of the man. On the other hand,

our “cognitive” capabilities allow us to identify the greater good in a scenario, while knowing nothing of the emotional impact of whatever we are asked to do in order to achieve that greater good.

Greene et al. (2004) says that “a parallel battle can be found within the brains of moral dilemma responders where the social-emotional responses undergird the absolute prohibitions that are central to deontology and utilitarianism is made possible by brain areas that support abstract thinking and high-level cognitive control.” Greene (2008) declares that the terms deontology and utilitarianism actually refer to what he calls psychological natural kinds, explaining that “consequentialist and deontological views of philosophy are not so much philosophical inventions as they are philosophical manifestations of two dissociable psychological patterns.” These patterns being emotional and “cognitive” processes leading to deontology and utilitarianism respectively.

Going full circle and moving back to philosophy, this view of the two normative ethical theories is problematic for deontology, as it is seemingly based on irrational emotions. This does not seem very fitting for such a formal, thought out philosophical theory. Greene (2008) argues in favour of the DPToMJ, praising utilitarianism as genuinely involving moral reasoning, but implicates deontological philosophy as a kind of moral confabulation, saying “What better way to express that feeling of non-negotiable absolute wrongness than via the most central of deontological concepts, the concept of a right: You can’t push him to his death because that be a violation of his rights. Likewise you can’t let that baby drown because you have a duty to save it.”

2.2.6. Moral Action

“Moral cognition needs to be studied in action-relevant environments where the stakes are immediate, emotionally charged, and tangible.” – FeldmanHall et al. (2012)

So far, only moral judgment has been discussed, abstract responses to hypothetical moral dilemmas. And while moral judgment has been extensively researched for the last 100 years and even more so in the last 20, far less work has been done on moral action, especially research

regarding harmful transgressions and the trolley problem. While this makes it difficult to directly compare moral judgments and actions, the increased prevalence of immersive virtual reality (IVR) in the domain of moral psychology has provided an increase in usable data. To date, there are four studies (that have been found) that attempt to research moral action without the use of IVR (Bostyn et al., 2018; Cushman et al., 2012; FeldmanHall et al., 2012; Tassy, Oullier, Mancini, & Wicker, 2013) and eleven in immersive virtual reality (Faulhaber et al., 2018; Francis et al., 2018; Francis et al., 2016; Francis et al., 2017; Friedman et al., 2014; McDonald et al., 2017; Navarrete et al., 2012; Pan & Slater, 2011; Patil et al., 2014; Skulmowski et al., 2014; Sütfield et al., 2017). As well as these papers, research on psychopathy has also contributed to the field of moral action as findings show that the emotional deficits associated with psychopathy affect moral judgments, but less so moral actions (Tassy, Deruelle, Mancini, Leistedt, & Wicker, 2013).

The limitations plaguing moral action research are fairly obvious. While it is fine to ask a participant to read a ghastly passage, it would not be ethical to stage a moral dilemma in real life for a participant to “stumble upon”, not to mention the logistics of constructing an elaborate trolley problem style dilemma in real life. Previous moral action studies have thus focused on the allocation of money in economic games (e.g., am I morally obligated to share this money?), however moral transgressions in these kinds of studies are intrinsically different to those at play in the trolley problem (Haidt, 2007). These factors have left moral action research stunted and the scientific community to use the results of moral judgment studies as our best guess of true moral behaviour.

If research into moral actions is so much more difficult than moral judgment, there needs to be a reason to do it. Is there truly a difference between moral action and judgment? And do moral actions conform to the now robust DPToMJ? These questions have been researched directly in several IVR and general moral action studies and has received mix results. Before continuing however, it is important to qualify the answers to each of these questions. Firstly, a true difference between moral judgment and action should consist of a significant difference in the amount of utilitarian versus deontological responses to moral dilemmas through judgment and action (Patil et al., 2014). Secondly, in order to follow the current DPToMJ, this dissonance should fall in the favour of deontology when responding through moral action (Navarrete et al., 2012; Patil et al., 2014; Skulmowski et al., 2014).

The first study to use IVR to study moral action was conducted by Pan and Slater (2011) whose main goal was to directly test whether the psychology behind moral action was distinct from that of moral judgment. In this study, participants either completed a battery of moral dilemmas in a questionnaire or were presented with a switch-esque dilemma in an immersive virtual reality CAVE system (Cruz-Neira et al., 1993). The authors opted to use a novel but similar moral dilemma to the switch case in order to prevent participants from recognising the situation and pre-empting their responses, something unique to this study and another from the same authors (Friedman et al., 2014). The difference in these conditions then was that the first required only moral judgment whereas the second demanded participants to physically act in order to resolve the situation with the hypothesis being that the transformation from moral judgment to action would recruit additional or distinct neural circuitry and potentially change the results (Navarrete et al., 2012; Pan & Slater, 2011). The results, however, indicated that there is little difference between moral judgment and action, showing that participants chose to kill one person to save five others 89% of the time in the IVR condition and 84% in the questionnaire.

The next study to use immersive virtual reality to research moral action replicated this result (Navarrete et al., 2012) finding that roughly 91% of participants acted in a utilitarian manner in IVR, comparing these results to the 85% found by Hauser et al. (2007). Skulmowski et al. (2014) also found similar results showing no difference between moral action and judgment choices. It is important to note that this is not the results that many of the researchers predicted based on the DPToMJ (Navarrete et al., 2012; Patil et al., 2014; Skulmowski et al., 2014). Specifically, the DPToMJ states that emotional processes contribute towards deontological thinking and it should therefore be that a more emotionally provoking medium such as IVR (Francis et al., 2016) should favour deontological actions compared to mundane questionnaires. This is even supported by evidence from Amit and Greene (2012) who were able to increase the number of deontological responses to moral dilemmas by showing participants simple visual imagery of the dilemmas they were considering.

The lack of increase in deontological responses is possibly explained however by the fact that each study used an impersonal moral dilemma. These types of dilemmas do not tend to activate emotional regions of the brain as opposed to a personal moral dilemma (Greene et al., 2001) and so the increased emotional saliency of IVR might not have had a large enough impact. Navarrete et al. (2012), Patil et al. (2014) and Skulmowski et al. (2014) subsequently all call for the running

of an experiment witnessing a personal moral dilemma in IVR so as to properly test the effects of the medium on moral action. The first study to do this was Francis et al. (2016), followed up by a more comprehensive study a year later (Francis et al., 2017).

Francis et al. (2017) presented participants with a IVR simulation of the footbridge dilemma. The head-mounted display (HMD) used was the Oculus Rift CV1 which also provided head tracking allowing participants to look around the virtual scene. In order to push the man off of the bridge, the participants had to physically push an interactive, life-like sculpture mechanism comprised of foam, silicon and heated wiring, designed to generate haptic feedback, co-located with the man in the virtual scene (Francis et al., 2017). The results revealed that while 10% of the participants endorsed utilitarianism (pushing the man) when answering a questionnaire, a staggering 56% endorsed the same action in IVR. Not only is this a massive difference in attitude between moral action and judgment, but it is in clear contrast of the DPToMJ, showing that instead of the predicted increase in deontological decision making in IVR, utilitarianism seems to prevail. This result is also not isolated and is supported by other IVR studies (Francis et al., 2016; McDonald et al., 2017; Patil et al., 2014) and other general moral action studies (Bostyn et al., 2018; FeldmanHall et al., 2012; Tassy, Oullier, et al., 2013) which all found a similar pattern of results.

Outside of immersive virtual reality, FeldmanHall et al. (2012) studied the reactions of participants in real and hypothetical situations where they could pay up to £20 to prevent harmful shocks being administered to a confederate. In this situation, paying the £20 might be regarded as the deontological option as it prevents harm whereas keeping the money might be regarded as the utilitarian option as it maximizes outcome for momentary pain. The results of the study showed that the participants in the real condition kept seven times as much money as those in the hypothetical condition (£12.52 compared to £1.53), thus showing an increase in utilitarian behaviour in the real condition, supporting Francis et al.'s (2017) conclusion. The authors end by saying that "The findings... illustrate that the proscription to not harm others... has surprisingly little influence when potential significant personal gain is at stake. [In contrast,] imaginary moral scenarios... seem to maximise people's opportunity to adhere to moral duties, such as not harming others" (FeldmanHall et al., 2012).

Another study by (Tassy, Oullier, et al., 2013) presented a series of moral dilemmas to participants, either probing moral judgment by asking "Is it acceptable to do this?" or moral action by asking

“Would you do this?” The authors quantify the difference between these two questions by saying that “[The latter] implies projecting oneself in a direct interaction using egocentric frame of reference with potential self-relevant consequences” (Tassy, Oullier, et al., 2013). The results confirmed the authors hypothesis, showing that the probability of utilitarian response was significantly higher in the action condition than in the judgment condition. A similar result was obtained in a separate study by Schaich Borg et al. (2006).

Finally, (Bostyn et al., 2018) studied a real-life interpretation of the switch dilemma by requiring participants to either passively allow five mice to receive an electric shock or to personally deliver the shock to one mouse. Participants were also presented this scenario as a hypothetical question. The analysis revealed that participants were more than twice as likely to make a utilitarian decision when faced with the real-life dilemma compared to the hypothetical version. However, hypothetical responses were able to predict the amount of remorse felt by participants for shocking the real mice, indicating that there are shared neural processes between moral judgments and actions.

To summarise, three studies (Navarrete et al., 2012; Pan & Slater, 2011; Skulmowski et al., 2014) indicate that there is no difference between moral judgment and action, with the first two directly showing that there was no significant difference between participants moral judgment and actions in response to the switch (or switch-like) dilemma. Counter to this, seven studies (Bostyn et al., 2018; FeldmanHall et al., 2012; Francis et al., 2016; Francis et al., 2017; McDonald et al., 2017; Patil et al., 2014; Tassy, Oullier, et al., 2013) have shown that there is a difference between moral judgment and action, specifically, that moral action tends towards utilitarianism. It should also be noted that the three studies that show no difference all used impersonal moral dilemmas, i.e., those that tend to activate rational parts of the brain such as working memory and not emotions. This may be why no difference was seen, as the emotional saliency of immersive virtual reality did not interact properly with the type of dilemma.

Referring back to our original questions then, “do moral actions differ from moral judgments?” So far, the majority of the data seems to indicate that yes, it does. Moral actions tend to elicit more utilitarian responses than moral judgments. So then, “are moral action responses predicted by Greene’s DPToMJ?” While this is obviously much more debatable, the general consensus seems to be that, no they are not. At the very least, you could say that no difference between moral

judgment and action aligns with Greene's DPToMJ, but even better would have been an increase in deontological responses for moral actions. This is the opposite of what has been recorded and while Greene's DPToMJ has been almost universally predictive of moral judgments, it seems a different model is needed for moral action.

In replacement of this model, Patil et al. (2014) and Francis et al. (2016) propose Cushman's action-outcome model (Cushman, 2013). The action-outcome model is based on two models of reinforcement learning, one aimed at choosing the best action (aligned with deontology) and another at attaining the best outcome (aligned with utilitarianism). Cushman is careful to state however that this coarse division is utilised only to benefit the future development of the model and it is likely that the two systems interact closely. Cushman states that benefits to certain actions (regardless of outcomes) are learned over a life time and start to feel rewarding in and of themselves. After a while, these actions can feel rewarding or punishing, even when the outcome is removed. On the other hand, our "cognitive" processes allow us to anticipate rewards or punishments from novel actions, guiding us towards a maximized outcome.

Support for this model comes from Cushman et al. (2012) which showed that participants are aversive to harmful actions, even when the negative consequences have been removed (e.g., stabbing someone with a rubber knife). This implies that it is not solely the outcomes of these actions that we find morally apprehensible, but also the actions themselves. Furthermore, this provides evidence that outcomes are still considered as it is likely that if participants were given a real knife, they simply would not have taken the action. Additional support comes from the general distinction people tend to have towards harmful actions and omissions where actions are seen as worse than omissions, even if the outcomes are identical (Spranca et al., 1991).

Unlike the DPToMJ, Cushman also argues that emotional and "cognitive" processes do not promote distinct responses and instead work together, contributing towards the valuation of both action and outcome, although they are more prominent when considering actions. For a more detailed account and how the model fits into current accounts of moral judgment and actions, see Cushman (2013).

Patil et al. (2014) and Francis et al. (2016) claim that the outcomes of a moral dilemma are much more salient in IVR and thus, based on Cushman's action-outcome model, explain why people seem to focus more on maximizing outcomes in IVR (i.e., acting in a utilitarian manner) rather

than abstaining from causing harm (i.e., acting in a deontological manner). Patil et al. (2014) says “we hypothesize that in VR participants could have been more sensitive to outcomes because they witnessed distressing consequences of their actions and emotions motivated them to act in order to minimize the distress by choosing the best of two emotionally aversive options in which either one or numerous deaths occur,” going on to say that “with textual descriptions, people need to rely more on mental simulation of the situation.” Agreeing with them, Francis et al. (2016) says that “given the contextual saliency of the virtual footbridge dilemma, outcome-based value representations for not pushing the man and allowing the people on the tracks to be killed, might have had a greater negative value.”

More evidence points to the saliency of outcomes as the defining factor that increases utilitarian choices in reality and immersive virtual reality. The evidence is twofold. First, Cushman et al. (2012) found that performing pretend harmful actions (with no negative consequences) resulted in a correlation with deontological, not utilitarian, choices to moral dilemmas. Therefore, it is unlikely that it is the performance of pretend harmful actions that one performs in immersive virtual reality that leads to increased utilitarianism. Second, in another study conducted in FeldmanHall et al. (2012), a “near-real” condition was added that matched the “real” condition from the first study, but participants were instead asked to imagine the shocks and money. Thus, while pretend, similar to immersive virtual reality, this condition still reduced the need for mental simulation of outcomes when making a moral decision. The “near-real” condition closely matched the results from the “real” condition (i.e., an increase in utilitarian responses over the hypothetical condition).

Therefore, we might conclude that it is this aspect (the salient outcomes) that drive utilitarian behaviour in immersive virtual reality. This is only a supposition however, and future work should be aimed at refuting or supporting both the difference seen between moral judgment and actions and the source of this difference which currently seems to stem from the saliency of outcomes.

2.3. Social Influence

Social influence is a humungous research area, with theories describing the process of how social interactions can function ranging including obedience to authority (Milgram & Gudehus, 1978), diffusion of responsibility (Darley & Latané, 1968), achieving influence over other people’s actions

by way of well-known compliance techniques (Cialdini, 2009) or describing how people in a society can affect each other through passive conformity to the minority or majority consensus (Asch, 1955; Latané, 1981). However, a topic that has been relatively isolated from social influence is that of moral decision-making. As such, the research conducted herein regarding social influence, towards the goal of identifying its relationship with morality, has mainly focused on how it has worked in non-moral circumstances. Where previous studies have been conducted combining the two areas of interest, they focus entirely on moral judgment tasks, leaving an open question as to how our moral actions or behaviours are affected by social influence.

The goal of this review then is to understand how social influence works in both psychological and neurological terms so as to better understand what studies need to be conducted in order to determine its relationship with moral decision-making and what the results of those studies might mean. This will start with a brief review of some of the most well-known research on social influence, move on to describing how social influence works on a neurological level and finish by discussing the small amount of research that has been done on social influence during moral dilemmas as well as how social influence works in virtual environments.

2.3.1. Models of Social Influence

One of the earliest pieces of work on social influence still cited today is the work done by Solomon Asch (1955, 1956) which looked at how people responded to a measuring task when responses were witnessed by a group of others, where those others have already answered previously and gave unanimously incorrect answers to the problem. Asch conducted a series of studies around this idea and found that the greater the consensus amongst the rest of the group, the greater the influence on the participant, such that, when answering on their own, participants gave incorrect answers only 1% of the time, when presented with one incorrect answer from a confederate, participants gave incorrect answers only slightly more often (value not given in paper), two confederates giving incorrect answers increased participant conformity to 13.6% and finally, three confederates increased conformity to 31.8%. More confederates after the initial three (measured up to 15) only increased the conformity from participants marginally, eventually ending at 36.8% at 15 confederates. Asch also demonstrated that by providing participants with a dissenter from

the group, i.e., a confederate who gave the correct answer whilst all other continued to give incorrect answers, conformity to the majority was significantly reduced.

In order to explain these findings and those of many other studies, Bibb Latane proposed the Social Impact Theory (SIT) (Latané, 1981). SIT broadly describes the effects of social influence from other people on an individual and defines social impact as “any of the great variety of changes in physiological states and subjective feelings, motives and emotions, cognitions and beliefs, values and behaviour, that occur in an individual, human or animal, as a results of the real, implied, or imagined presence or actions of other individuals”, going on to say that this can appear under such conditions as “allelomimetic behaviour, behavioural contagion, conformity, compliance, group pressure, imitation, normative social influence, observational learning, social facilitation, suggestion and vicarious conditioning.”

According to the theory, impact from a group of people (the source) on an individual (the target) should be a multiplicative function of the strength, immediacy and number of people in the source group, where the N^{th} person in the source group will exert less influence than the $(N - 1)^{\text{th}}$. The theory goes on to say that when other people stand with the target, social impact should be divided amongst them as a function of the strength, immediacy and number of people in the target group. In SIT, strength is defined as the power, importance or intensity of a message from the source to the target – determined by such things as the source’s status, age, socio-economic status, and prior relationship with, or future power over, the target and immediacy as the closeness in space or time and absence of intervening barriers or filters between the source and target.

The most famous example of the division of social impact when the target is a group instead of an individual is the bystander effect (Latané & Darley, 1970). This effect predicts that in situations where an individual person (the source) might be in need of help, they are less likely to receive aid when there are more onlookers (the target). This happens due to the diffusion of responsibility from the knowledge that other people are also available to respond, the assumed embarrassment from knowing that other people may watch one make a fool of oneself and the influence from seeing that other people are also not responding to the situation.

As noted earlier, SIT is supposed to be an almost universal theory of what the effects of any social interaction might be rather than a theory of the underlying process of any specific type of social

influence (Turner, 1991). As such, although SIT is able to predict what might happen during a social communication, it does not explain why it happens.

Other theories have attempted to answer this question, one of the earliest being Festinger's Cognitive Dissonance Theory (CDT) (Festinger, 1957). CDT holds that cognitive dissonance arises from a psychological inconsistency between linked cognitions such that "if a person held two cognitions that were psychologically inconsistent, he would experience dissonance and would attempt to reduce dissonance much like one would attempt to reduce hunger, thirst or any drive" (Festinger, 1957). The theory has been extended since its conception, most notably by Aronson (2007), who suggested that CDT makes its strongest and clearest predictions when the self-concept of the individual is engaged, more specifically, a cognition about the self and a behaviour that violates that self-concept where self-concepts can include a person's consistency, stability, predictability, competence and moral goodness. For example, a person might attempt to make amends if they have done something immoral, but see themselves as a morally good person.

Other research into CDT has also shown that when there are multiple routes available that lead to a reduction of dissonance, people apparently prefer to reduce dissonance directly by changing attitudes and behaviours, rather than through simple self-affirmations (Steele, 1988). For example, when refusing an initial offer to buy a £20 theatre ticket for charity from a boy scout, a person might be much more likely to subsequently agree to his offer of purchasing a £1 chocolate bar. In the initial exchange, the targets moral goodness has become internally dissonant and in order to rectify the situation through reciprocity, they might actively purchase the £1 chocolate bar (Cialdini, 2007).

More recent theories have merged CDT with other areas of research on normative and informational social influence to form a tripartite model of motivations that lead to attitude change (Wood, 2000). The terms normative and informational social influence were first used by Deutsch and Gerard (1955) in response to previous research on social influence that unknowingly focused on only normative social influences. They defined normative social influence as "an influence to conform with the positive expectations of another" and informational social influence as "an influence to accept information obtained from another as evidence about reality". Together (along with CDT), these form three motivations for attitude change: ensuring the coherence and favourable evaluation of the self (Defence), ensuring satisfactory relations with others given the

rewards/punishments they can provide (Affiliation) and understanding the entity or issue featured in influence appeals (Accuracy) (Wood, 2000). There are two largely detached theories that focus on accuracy, affiliation and defence: the Heuristic-Systematic Model (HSM) by Chen and Chaiken (1999) and the research on persuasion by Cialdini and Goldstein (2004).

Much like the dual-process theories described earlier in section 2.2.4.2, the HSM defines two basic modes of processing which perceivers use to determine their attitudes, systematic processing and heuristic processing (Chaiken, 1980). Systematic processing entails a relatively analytic and comprehensive treatment of judgment-relevant information and given its nature, requires high levels of both cognitive ability and capacity, on the other hand, heuristic processing calls for the activation and application of judgmental rules of thumb, e.g., “experts can be trusted” or “consensus opinions are correct”. Relative to systematic processing, heuristics make minimal cognitive demands and can be applied very quickly, but must be stored in memory, be retrievable and be relevant to the task at hand. It is also possible that the accessibility of a heuristic corresponds to the likelihood of its use and its frequency of use increasing its ease of accessibility, similar to Damasio’s somatic markers theory (Damasio, 2006).

The model also assumes that people are guided by the principle of least effort such that heuristic processing will dominate over systematic whenever possible, with systematic processing only being used when necessary and if cognitive demand and time allows. HSM’s sufficiency principle maintains that people attempt to strike a balance between minimizing cognitive effort on the one hand and satisfying their current motivational concerns on the other. Similar to CDT, the model proposes a continuum of judgmental confidence along which the perceiver’s actual level of confidence and desired level of confidence lie. When low-effort heuristic processing fails to confer sufficient judgmental confidence, dissonance between the levels of confidence will engage systematic processing in an attempt to close the confidence gap.

As mentioned earlier, HSM defines three motivations in which perceivers process information, either through heuristic or information processing. Impression motivation refers to “the desire to hold attitudes and beliefs that will satisfy current social goals” (Chen & Chaiken, 1999). Impression motivated heuristics entails those such as “moderate opinions minimise disagreement” when talking to someone whose opinions are unknown or “go along to get along” when opinions are known.

The accuracy motivation holds that “people are motivated to hold accurate attitudes and beliefs” (Chen & Chaiken, 1999). Given high levels of motivation and sufficient cognitive resources and time, people may engage in systematic forms of processing in order to reach their accuracy sufficiency threshold. When motivation, cognition or time is low however, people may simply base their attitudes on the best suited heuristic-cue information (e.g., “Expert opinions are correct”).

Finally, the defence motivation refers to “the desire to hold attitudes and beliefs that are congruent with one’s perceived material interests or existing, self-definitional attitudes and beliefs” (Chen & Chaiken, 1999). The defence-motivated aims to preserve the self-concept and any associated views, and thus processes information selectively, that is, in a way that is least likely to make us question our internal beliefs. For example, social heuristics are much more likely to be invoked when they represent a congenial group of peers compared to an uncongenial, e.g., “consensus opinions are correct” may exert far less influence when in a crowd of people of whom you intensely disagree. In a study conducted by Giner-Sorolla and Chaiken (1997), participants with a vested interest in a target issue were significantly more likely to criticize a poll when it presented data that opposed the issue rather than support it.

2.3.2. Compliance and Persuasion

Cialdini and Goldstein (2004) define compliance as “a particular kind of response – acquiescence – to a particular kind of communication – a request... The request may be explicit... or it may be implicit, but in all cases, the target recognises that he or she is being urged to respond in a desired way.” Persuasion builds on top of compliance by using some psychological tactic, e.g., the foot-in-the-door technique, to increase the chances that their request is responded to favourably (Cialdini, 2009). The three motivations described earlier in the HSM have also been used in compliance and persuasion literature in order to explain how and why some of these tactics work.

2.3.2.1. Accuracy Motivation

Cialdini and Goldstein (2004) define the accuracy motivation as acting in a way such that “people are motivated to achieve their goals in the most effective and rewarding manner possible.” One way we do this is by using our emotions as cues as to what the correct course of action is to take. For example, Whatley, Webster, Smith, and Rhodes (1999) differentiated between the emotions and related goals associated with public and private compliance in response to a request. They posited that individuals avoid or alleviate feelings of shame and fear via public compliance, and guilt and pity via private compliance. This lines up with other research on moral and social emotions which suggests that shame and embarrassment are social emotions, largely only triggered by the presence of others (you would not feel embarrassed falling over when no one is watching) and that guilt and pity are more moral emotions, not needing the presence others to be felt by an individual (Warren & Smith-Crowe, 2008).

Forgas (1995) argues that the conditions under which affect mediates the responses to a request can be explained by the Affect Infusion Model (AIM), where AIM contends that a target’s mood will permeate the processing of a request to the extent that the processing is more effortful and exhaustive, in other words, the more cognitive resources that are required coming to a decision, the more likely our mood is to interfere with that decision. Forgas (1998) later added to this, suggesting that the processing of a request will be more sensitive to mood if the appeal is made through unconventional means (thus requiring more substantive processing).

One interesting phenomenon that has been researched is the subsequent vulnerability to suggestions, proposals or demands presented in the moment of sudden relief after a previous experience of fear (Fear-Then-Relief) (Dolinski & Nawrat, 1998). A series of experiments conducted by Dolinski and Nawrat which involved inducing fear in participants (e.g., hearing a car horn while crossing the road or seeing a police ticket under your car wiper) and then relief (e.g., realising the car horn and police ticket were fake) found that participants that had experienced the sudden mood change agreed to complete an unrelated 15 minute questionnaire 75% of the time, compared to a control group that only agreed 52.5% of the time, and furthermore that when relief was not induced (e.g., the participants did not learn the horn or police ticket was fake until afterwards), participants only acquiesced 37.5% of the time. Dolinski (2007) posits that this

increased tendency to be influenced is due to the changing of our emotional state, such that when we feel in danger, fear is engaged, but it must almost immediately be replaced again. In this moment of disorientation, before an adequate emotional state has been reached, we act automatically and mindlessly.

As in the HSM, individuals might also rely on social heuristics when they are the target of a compliance request. Individuals are frequently rewarded for behaving in accordance with the opinions, advice and directives of authority figures as we often associate authority figures with accuracy or correctness, at least on issues within their field of expertise. The potentially harmful consequences of this has been famously demonstrated by (Milgram & Gudehus, 1978) in the famous Milgram experiment. In this experiment, an authority figure (the experimenter) was able to successfully convince a significant proportion of participants to apply a dangerous electrical shock to, who they thought, was another participant. In reality, the shocks were fake and the participant was a confederate, but the results hold up and continue to be one of the most disturbing results in a social psychological study. Other research has shown that we are also susceptible to those feigning expertise and that individuals look to social norms to gain an accurate understanding of and effectively respond to social situations, especially during occasions of unfamiliarity, uncertainty or under time constraint. Real-life disastrous effects of this can be seen in events like the Jonestown massacre where cult members were convinced to drink cyanide by the group's leader (www.britannica.com/event/Jonestown-massacre).

2.3.2.2. Affiliation Motivation

In regards to the affiliation motivation, Cialdini and Goldstein (2004) explain that “humans are fundamentally motivated to create and maintain meaningful social relationships with others. Accordingly, we use approval and liking cues to help build, maintain and measure the intimacy of our relationships with others.”

According to work done by Tajfel et al. (1979) on group behaviour, people are more likely to act in accordance with and respond affirmatively to people with which they identify, also known as a reference group or in-group. A reference group is defined as a group that a person belongs to with regards to their attitudes, beliefs and behaviours and not necessarily the group to which they

spatially belong to. Despite the deep connection people can feel towards reference groups, it is often very easy to lead individuals to respond to strangers in ways that belie the absence of a truly meaningful relationship between them (Deutsch & Gerard, 1955). For example, Burger, Soroka, Gonzago, Murphy, and Somervell (2001) found that simply being exposed to a person even for a brief period without any interaction substantially increased compliance with that person's request. In a series of studies, Dolinski, Nawrat, and Rudak (2001) also showed that simply engaging people in a short, trivial dialogue prior to making the target request was sufficient to elevate compliance.

Additionally, the norm of reciprocation – the rule that obliges us to repay others for what we have received from them – can also cause us to comply with requests that we would otherwise have been easily able to decline (Cialdini, 2007). The norm of reciprocity can explain the effectiveness of the door-in-the-face technique where one employs the strategy of preceding a request for a truly desired action with a more extreme request that is likely to get rejected. The target feels a normative obligation to reciprocate the influence agent's concession with a concession of their own by subsequently complying to the second request.

2.3.2.3. Defence Motivation

Finally, the defence motivation dictates that people have “a strong need to enhance their self-concepts by behaving consistently with their actions, statements, commitments, beliefs and self-ascribed traits” (Cialdini & Goldstein, 2004).

This motivation has been demonstrated via the foot-in-the-door technique (Freedman & Fraser, 1966) whereby after securing compliance to a small appeal, the requester makes a second, larger solicitation. After agreeing to the initial request, our goal to appear competent, logical and consistent makes us more likely to acquiesce to further similar requests. This has been demonstrated by Freedman and Fraser whereby homeowners were asked to place a small sign on their front lawns and were later asked if the sign could be replaced with a much bigger one. The compliance rate with the second request compared to those who were not asked to put up the original small sign was significantly greater. Individuals are also driven to be consistent not only with their self-attributions, but with their previous behaviours and commitments as well. The

extent to which one's commitments are made actively and publicly is a powerful determinant of the likelihood of request compliance. For example, Cioffi and Garner (1996) solicited volunteers for an AIDS awareness project by asking participants to indicate their decisions on a form in either an active or passive manner. Irrespective of their choice, participants who made an active rather than passive choice took a more extreme position toward their decision when asked to justify it later on.

Lastly, public displays of commitment have also been utilised by salespeople (Cialdini, 2007). The technique involves gaining a customer's compliance to a sale by offering a discount. Once the commitment to the proposal has been secured the cost of carrying out the deal can increase substantially through admin or some other fees. After accepting the initial offer however, targets are much more likely to accept the further costs "not known previously" by the salesperson.

Counter to prevailing research showing the ease of gaining compliance, some research has been directed at how social influence can fail. One well researched reason why people do not comply is because of attitude strength, such that strongly held attitudes and beliefs can resist persuasion (Sagarin & Wood, 2007). This is linked to our moral behaviour in that some of our strongest attitudes and beliefs are related to our morality, such as opinions on abortion rights, the use of torture or capital punishment (Schaich Borg et al., 2006). Eagly and Chaiken (1993) note that strong attitudes differ from weak ones on both motivational and cognitive dimensions. The motivational dimension is governed by the importance of the attitude to us, the certainty to which the attitude is held and our previous public commitment of that attitude. The cognitive dimension then refers to the embeddedness of the attitude within a web of other attitudes, the degree to which beliefs support or oppose the attitude and the accessibility of the attitude. The end result is that "important, imbedded, accessible attitudes that are held with certainty, expressed publicly, and supported by beliefs are highly resistant to influences" (Sagarin & Wood, 2007). In a study looking at how social influence can alter outward moral judgments, Lisciandra et al. (2013) found that, while moral judgments were swayed towards a group consensus, they were harder to influence than other types of social norms.

This effect of strong attitudes and beliefs has been demonstrated at the neural level. Pincus, LaViers, Prietula, and Berns (2014) demonstrated that individuals with stronger deontological resolve, as measured by greater activity in the ventrolateral prefrontal cortex (vlPFC), displayed

lower levels of conformity. Concluding that unwillingness to conform to others' values is associated with a strong neurobiological representation of social rules.

2.3.3. Social Cognitive Processes

Much like how social impact theory attempts to generally predict the strength of any kind of social influence interaction, Falk, Way, and Jasinska (2012) posited that diverse forms of influence overlap in their underlying neural circuits, such that they can be treated as one, using the umbrella term "normative social influence" to encompass compliance, conformity, susceptibility to peer influence and responsiveness to persuasion. The constellation of brain areas identified by Falk and colleagues include those related to social reward and pain (the same used to respond to primary rewards and punishments, such as food and physical pain). These work as mediators of social influence processes; more specifically, people may respond to social interactions as a joint function of their reward (for conforming to a group norm) and punishment (rejection from the group).

Recent neuroscientific and computational models assume that goal-directed behaviour requires continuous performance monitoring (Klucharev, Hytönen, Rijpkema, Smidts, & Fernández, 2009). Behaviours deemed successful get reinforced while errors in a behaviour's predicted outcome call for adjustments. This model of reward and punishment is known as *reinforcement learning*, which uses prediction error to guide decision making by signalling the need for adjustment of behaviour based on the difference between the expected and obtained outcomes (Izuma, 2013).

Klucharev et al. (2009) successfully demonstrated how this reinforcement learning model works during a social conformity experiment. In the experiment, participants were initially asked to rate the attractiveness of female faces. They were then asked to re-rate them whilst simultaneously being shown the average results from a group of peers (the data was actually fake) and at the same time were subject to an fMRI scan. The results show that group opinion effectively modulated the judgments of individuals even though the group was not physically present and that the conformity effect was especially strong for highly ambiguous faces. The fMRI data revealed that a mismatch with group opinion triggered a neural response in the Rostral Cingulate

Zone (RCZ) and the Nucleus Accumbens (NAc), similar to prediction error signal from previous studies (Schultz, 2006). The activation of the RCZ and deactivation of the NAc region also successfully predicted subsequent conformity to the group. As such, it is likely that group opinion worked as a reinforcer for the individual's behaviour, both being rewarded for being aligned with the group and penalised to being non-aligned.

Later studies reported similar results when attitudes did not align to those of experts (Campbell-Meiklejohn, Bach, Roepstorff, Dolan, & Frith, 2010) and that high trait conformity predicted a larger activation of the areas associated with the RCZ and NAc (Berns, Capra, Moore, & Noussair, 2010).

Other research has also been conducted looking at exactly how the error signal works during social interactions, more specifically, how the brain alarms us that we are being (or that we anticipate being) socially rejected (Lieberman, 2013). Experiences such as social rejection, exclusion or loss are generally considered to be some of the most painful experiences that we can endure and many of us go to great lengths to prevent them. Recent evidence suggests that this is due to our experiences of social pain relying on some of the same neurobiological substrates that underlie experiences of our physical pain. The language of social loss hints at this sharing of neural circuitry when we say things such as "*broken heart*" or "*hurt feelings*", with these phrases being common to many languages (Eisenberger, 2012).

The reason for this shared neural circuitry is likely due to our reliance on each other as a species. From birth, we rely completely on the care of our parents, not being able to feed, care or protect for ourselves years after our birth. Compare this to the maturation of a foal that can run alongside its mother only a few hours after being born. Later into life, we rely on others in our society for many of the same things we relied on our parents for when we were young. Modern life has exploded this onto an enormous scale and it is true now more than ever that we rely on others for common necessities such as food and protection. Due to this profound reliance on others, threats to social connection can be thought of as equally as important, if not more so, than that of physical pain, being just as detrimental to our survival as threats to our physical safety (Eisenberger, 2012).

The specific neural circuitry being shared between physical and social pain has been identified as the Anterior Cingulate Cortex (ACC) and the Anterior Insula (AI) (Eisenberger, 2012). These are the areas of the brain that deal with the affective component of pain, the component associate

with the unpleasant or distressing experience of pain as well as the drive to terminate the stimulus causing the experience. This was demonstrated in an experiment using a computer game called *Cyberball*. In *Cyberball*, a player has to throw a ball to two other online players, after a while however, the two players will start to only throw the ball to each other, excluding the player. When played while under a fMRI scanner, Eisenberger et al. (2003) noted a greater activation in both the ACC and the AI, which subsequently correlated with greater reported feelings of social distress in response to the exclusion.

Another line of research on social conformity investigated whether opinion modulates not only self-reported preference, but also its neural representation (Izuma, 2013). Social psychologists often distinguish between two types of attitude change; public and private compliance, where public compliance refers to our behaviours when in a social situation, with a group of friends or at work and our private compliance as our true beliefs and attitudes about a particular judgment topic. Several studies demonstrated that changes in attitudes (recorded in private) that were induced by others were accompanied by activation in the ventromedial prefrontal cortex (vmPFC), indicating that it has a role in private acceptance of other attitudes or beliefs (Charpentier, Moutsiana, Garrett, & Sharot, 2014; Mason, Dyer, & Norton, 2009; Zaki, Schirmer, & Mitchell, 2011). Charpentier and colleagues also noted that “whether a behavioural modification will occur is determined not necessarily by the brain’s response to the initial social influence, but by how that response is mirrored at a later time when the individual selects their action.”

2.3.4. Moral Influence

The majority of work done in relation to morality has addressed it in terms of individual-level processes while the majority of social influence research has focused on almost exclusively non-moral dilemma (Ellemers, Pagliaro, & Barreto, 2013). This is a shame as the ability to influence others can be annoying when we did not really want to buy that more expensive car, but the salesman was so nice to us, or when we did not really want to help out at a work event, but the boss asked us in front of everyone else so we could not say no. However, it can be catastrophic when a person’s morality can be influenced and altered by others as evidenced by previous real-world disasters.

There is no better demonstration of this than the horrific incidents that occurred at Abu Ghraib. Abu Ghraib was an American run prison facility located in Iraq during the early 2000s, where suspected terrorists would be sent for interrogation. It emerged some time later that prisoners at Abu Ghraib had been tortured without purpose, maimed, killed and sexually assaulted by the guards and soldiers who ran the prison (Hersh & Friedman, 2004).

After several investigations, a number of soldiers were blamed for the mistreatment of the prisoners and were dishonourably discharged from military service, but there were no repercussions for any senior staff who worked at Abu Ghraib. While society holds individuals responsible for their actions, social psychology suggests we should also hold responsible peers and superiors who controlled the social context in which the actions took place. By putting these soldiers in contexts where these acts were not scrutinised to the degree they would have been in society, their morals can be altered to their new environment, such that they would never dream of doing these things in society, but can be suddenly capable when it is the norm. Well-established principles of conformity to peers and obedience to authority may account for the widespread nature of the abuse at Abu Ghraib. In combat, conformity to one's unit means survival, and ostracism means death, thus people likely tried to make sense of a complex, confusing and ambiguous situation by relying on their immediate social group for guidance (Fiske et al., 2004).

Warren and Smith-Crowe (2008) offer a lighter take on moral conformity, positing that internal emotional responses to moral sanctions can trigger a sense of moral deficiency, subsequently shifting the moral judgments of the transgressor. They focus on unintentional moral transgressions, and thus on the emotion of embarrassment or shame that one has done something accidentally immoral.

Embarrassment can be thought of as both a moral emotion, naturally occurring after a moral transgression along with shame or guilt, and a social one as well, in that it is only experienced in the presence of others in reaction to events that threatens a person's social well-being (Berthoz, Armony, Blair, & Dolan, 2002; Finger, Marsh, Kamel, Mitchell, & Blair, 2006). Warren and Smith-Crowe (2008) argue that "[embarrassment] is a form of social control, letting people know when they have crossed a line, especially in instances of ethical ambiguity in which people look to others to know what is right."

A famous real-world example of a lack of embarrassment (as well as other social or moral emotions) comes from Phineas Gage, a railroad worker who suffered an injury to his frontal lobe in 1848 (Damasio, 2006). Despite being described as model citizen and worker before the accident, afterwards he was said to become foul-mouthed and undisciplined, ultimately losing his job and friends (he became what would now be labelled as a psycho- or sociopath). Damasio (2006) argues that the problem appears to have been his diminished capacity for experiencing the moral and social emotions necessary for social interactions.

Leary (2000) postulates that humans have evolved to use these moral and social emotions in order to help us scan the environment for indications that we are being devalued or excluded and to warn against continuing with our current behaviours through eliciting negative emotions as a sign of trouble, working somewhat like an alarm system against ostracism (Lieberman, 2013). This evolution was likely born from the need of humans to be social animals, now dependant on each other within communities for their own physical and emotional well-being. As Moore and Gino (2013) explain, "a solitary organism needs no moral rules, as he has no one but himself to harm, offend, or treat unfairly. The fact that human survival depends on finding ways to live together in peaceful, mutually supportive relations created an evolutionary imperative for fundamental moral behaviours. In other words, we are moral because we are social."

In order to demonstrate exactly how important morals are to us as a society, Leach, Ellemers, and Barreto (2007) ran a series of studies, the results of which indicate that morality is a primary quality of in-groups and suggest that positive evaluations of a group in terms of morality can contribute more to a positive social identity than other dimensions of value, such as competence and sociability. A congruent result was found by Pagliaro, Brambilla, Sacchi, D'Angelo, and Ellemers (2013) who ran a series of studies showing that people are particularly concerned about appearing moral to other group members over other important personal traits.

One study that has looked directly at people's moral judgments during a situation of social influence was conducted by Kundu and Cummins (2013). This study used the Asch paradigm of putting people in a group of confederates and asked each person in turn to give their answers to a problem, with the real participant usually answering last or second to last. In this case, the experimenters asked the group to give their answers to a series of moral dilemmas based on the trolley problem. The results showed a similar trend to the original Asch studies (Asch, 1955, 1956)

with people conforming to the judgment of the group, judging dilemmas that would typically be judged as impermissible as permissible and vice versa based on the responses of the group before them. Lisciandra et al. (2013) also applied the Asch conditioning paradigm to moral dilemmas and found that moral, social and decency norms were all affected by the peer-group, although the change in moral norms was the weakest.

More recently, Bostyn and Roets (2017) conducted a study in which participants responded to a series of moral dilemmas while simultaneously being exposed to fictitious response data regarding the dilemma. The results showed that participants were more likely to give a deontological answer when the data showed a majoritively deontological response, but not when utilitarian favoured data was presented. The authors explain their findings by citing work done by Uhlmann, Zhu, and Tannenbaum (2013) and Everett et al. (2016), both of whom found social preferences for deontologists rather than utilitarians. Finally, Kelly et al. (2017) attempted to replicate the role of social media in our everyday moral attitudes by presenting moral dilemmas along with a social media-like information about how others people might respond to the dilemma. They found that statistical data and comments consisting of high-level reasoning were effective at influencing moral judgments, but that emotionally charged comments were not.

It is important to note however, that each of these studies (Bostyn & Roets, 2017; Kelly et al., 2017; Kundu & Cummins, 2013; Lisciandra et al., 2013) used written questionnaires in order to elicit moral judgment from participants and as such the studies suffer from the same lack of social context and other auxiliary information as previous moral judgment studies. Although it should be noted that the implementation of social influence was real, especially in Kundu and Cummins (2013) and Lisciandra et al. (2013) where real people were used. As such, the most these studies can claim is to have publicly altered the moral judgment of their participant, but it is still unknown whether moral actions or behaviours can be interfered with. For example, one key difference between judgments and actions is the ability to lie during a judgment task, publicly accepting the group norm, but privately resenting it. This is impossible during a behavioural task as the action taken has real consequences. This is in addition to any neurological differences between moral judgment and behaviour that might respond differently to social influence attempts. Given this, it is imperative moving forward to study the effects of social influence on moral behaviour by using IVR to simulate the experience, instead of reading about it in a vignette.

Beyond this, several studies have generally shown that people can adjust their morals or respond differently to moral dilemmas based on their social situation. Waytz and Epley (2012) demonstrated during a study that those led to feel socially connected to other in-groups members were less likely to attribute humanlike mental states to members of various social out-groups and were more likely to recommend harsh treatment for people in those out-groups. Similarly, Lucas and Livingston (2014) found evidence that feeling socially connected to others increased utilitarian choices during the footbridge dilemma. However, it has also been shown that people can be more likely to act deontologically during a moral dilemma in the presence of others and also prefer others who act in a deontological manner as cooperative partners (Everett et al., 2016; Uhlmann et al., 2013). The supposed reason for this is that, as discussed earlier, people are motivated to demonstrate to others (particularly in-group others) that they are morally good. It is possible however that adverse motivational profiles can be mapped onto consequentialist behaviour rendering consequentialist choices less informative of moral values than that of deontological (e.g., did he really want to save the five people the trolley was heading towards, or did just want to kill the other one?). It also the case that carrying out consequentialist actions can require the suppression of empathy in order to follow through with the action, in this case, a truly empathic or morally good person would be less likely to accomplish the consequentialist dictation than the less empathic. As such, people with the goal of appearing as morally good to others would be more likely to adopt a deontological perspective.

2.3.5. Virtual Influence

Allport (1985) defines social psychology as the “attempt to understand and explain how the thought, feeling, and behaviour of individuals are influenced by the actual, imagined, or implied presence of others.” The inclusion of *imagined or implied* is important here as it allows researchers to study the effects of social influence without having to rely on another human to actually bring about that influence. For example, as was discussed earlier, Eisenberger et al. (2003) used a computer game dubbed Cyberball, in which participants initially threw a ball with other computer-controlled players but was later excluded, in order to examine the neural pattern

of social rejection. Although the presence of other people was only implied in the game, fMRI results still showed that same activation in neural regions associated with social pain.

Characters in a virtual environment (VE) are somewhere between *actual* and *imagined or implied*. Factually, they are *imagined or implied*. Virtual characters (VCs) are only there perceptually and reaching out to touch them will almost always fail. They are simply the product of sophisticated hardware and software technologies, moulded by a designer to show what might look like a real person. By programming a VC to be in the environment, the designer is *implying* that there is a person there and any agency that character has over their actions is completely *imagined* by the user. But in truth, we find that participants respond to VCs as if they are *actually there*, altering their own speech, movements, position or decisions in immersive virtual reality in response to completely virtual characters (Bailenson & Yee, 2005; Pertaub et al., 2002; Swinth & Blascovich, 2001).

The likelihood that a person responds in such a way, i.e., realistically, in a VE is related to the amount of presence they felt in that VE (Slater, 2009). This is dependent on a number of factors including the capabilities of the system, the number of sensory modalities supported, the valid actions of the system as well as the physical properties of the VE itself. For a more in-depth discussion on this, see section 2.1.2.

2.3.5.1. Models of Social Influence in Virtual Environments

More specific to social influence is Blascovich et al.'s (2002) model which proposes that social influence will occur in a VE above a certain threshold determined by four factors: the behavioural or communicative realism of the VC, the agency attributed to the VC, the self-relevance of the situation to the user and the user's targeted response system.

Blascovich et al. (2002) define behavioural realism as "the degree to which virtual humans and other objects within IVEs (immersive virtual environments) behave as they would in the physical world." What is most important here is that the eye, head and body movements of the VCs appear realistic. Eye and head movements can be used to engage in mutual gaze between the user and the VC and body movements are essential for expressing non-verbal communications such as

contextual hand gestures. Less integral to the illusion of realistic behaviour is photographic realism. McCall and Blascovich (2009) use the example of a wax model, which can replicate its source material almost perfectly, but, with a speaker in its mouth, would not provide a compelling experience of a real person. A cartoon character, such as Mickey Mouse, on the other hand looks nothing like a photorealistic mouse, but is far more compelling because of the realistic behaviours given to him by his animators.

Related to the behavioural realism of a VC is the perceived agency that a VC has over its own actions. The agency attributed to a VC by a user can range from the understanding that the VC is actually being controlled by another human (high agency) to the knowledge that the VC is completely controlled by a computer algorithm (low agency). De Melo, Gratch, and Carnevale (2015) found that people were more cooperative with and more likely to concede to a human-controlled VC compared to a computer-controlled VC during a negotiation game. Similarly, Bailenson et al's (2003) results showed that those participants who believed that a VC was being controlled by a human (compared to a computer) maintained a greater interpersonal distance between themselves and the VC, indicating a greater amount of consideration for the personal space of a human over the computer. Another result found that agency interacted with the participants gender, such that women were more receptive to female, high agency-VCs and men were more receptive to male, low agency-VCs, offering the explanation that "it may be that men and computers are both expected to be competent relative to women and humans, which are both expected to be warm" (Guadagno, Blascovich, Bailenson, & McCall, 2007). The line between human and computer can become blurry however, if the user is not told this information before entering the VE. If the VC's actions appear realistic enough, the user might assume the VC has agency over its actions even if in truth it is being controlled by a computer. For example, Swinth and Blascovich (2001) found that participants conformed with the actions of a VC during a virtual betting game regardless of whether the VC was human- or computer-controlled.

The model presented by Blascovich et al. (2002) is a continuum modulated by both behavioural realism and attribution of agency such that social influence is likely to occur (where it would in real life) if either of these factors is high or both are moderate (see Figure 2.7).

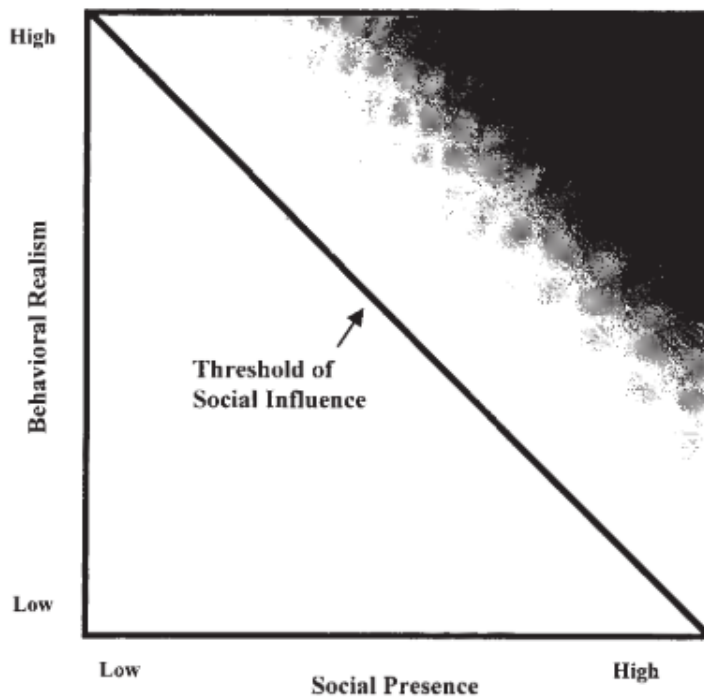


Figure 2.7. – Threshold of social influence (Blascovich et al., 2002).

As such, even if attribution of agency is very low, i.e., the participant is specifically told that they will be interacting with a computer-controlled VC, high behavioural realism can still result in effective social influence. Several studies have shown this to be true. Bailenson and Yee (2005) demonstrated that the chameleon effect, the impact of social influence through mimicry, persisted even after participants were told that the mimicker was being controlled by a computer algorithm and not a real person and another study by Nowak (2004) found no discernible differences in social judgment between participants who were told that they were interacting with a human as compared to a computer agent, offering the explanation that “it is possible that the human brain has not adapted to dealing with entities that look human but are not.”

In addition to the main components of behavioural realism and attribution of agency, Blascovich et al. (2002) describe two more factors that affects the successfulness of social influence in a VE. Self-relevance refers to the importance of the interaction with the virtual character, for example, withdrawing money from a bank teller is of low importance, but performing in a job interview would be of high importance. In the first of these situations, whether the user is interacting with a machine or human would make little difference, in fact in real-life we often choose to take money out at automatic teller machine (ATM) over an actual bank teller, however, a job interview is likely

to be much more intimidating, especially so when conducted with a real person compared to a computer. Finally, the impact of social influence likely depends on the user's activated response system. As was said previously, the illusion of IVR relies on our unconscious, automatic responses. Our gut reactions to events happening in IVR respond as if they are real and it is our rational mind that later reminds us that they are not. Therefore, social events that require visceral gut reactions, such as a VC suddenly coming very close to us in our peripheral vision, are highly likely to exert some amount of influence on us which we might react to through a change in our body language or our position in the VE (Sheridan, 1992, 1996). In contrast to this, events targeting a higher level process such as an in-depth conversation would require a VC with either high behavioural realism or attribution of agency.

McCall and Blascovich (2009) conclude by saying that "we cannot expect normal social influence outcomes in *high level domains* if the user knows she is *interacting with an agent* (a computer-controlled VC), the *communicative realism is low* and the content is *highly self-relevant*. Conversely, we can expect normal influence processes at *lower level domains* if the user is *interacting with an avatar* (a human-controlled VC), the *communicative realism is high* and the domain is of *moderate self-relevance*."

There have been many studies that have successfully studied the effects of social influence using IVR technology. Swinth and Blascovich (2001) conducted a IVR replication of an earlier study by Blascovich and Ginsburg (1974) which analysed the betting patterns of people when exposed to a confederate who would either bet high, low or normal amounts. Results were congruent between the two studies, indicating the usefulness of IVR technology for future social psychology studies. Bailenson and Yee (2005) found that VCs who mimicked the head movements of participants were more persuasive when delivering a message than when being controlled by the head movements of a previous participant, this is more commonly known as the chameleon effect and has been previously studied in real-life situations. Pertaub et al. (2002) were able to induce increased anxiety in participants during a virtual speech by having the audience either react interested or bored. A result that came about even though participants were aware the VCs in the audience were computer-controlled agents. Finally, Pan et al. (2016) reported the acquiescence of medical doctors to a demand for antibiotics by virtual patients and Slater et al. (2006) and Slater et al. (2013) were able to replicate the findings of well-known psychological phenomena such as

the bystander effect, social identity theory and obedience to authority in IVR using computer-controlled VCs.

2.4. Chapter Summary

This review has attempted to highlight the importance of immersive virtual reality (IVR) in the construction of modern moral decision-making studies, how it should be used, how previous studies have used it and why it is integral for the studies conducted as part of this thesis. Furthermore, some of the issues that can occur when translating a moral dilemma from questionnaire to virtual scenario have been highlighted, something that further research should look to address.

The studies and methods described in subsequent chapters in this thesis build on the knowledge gained from this review of IVR and adopt many of the successful methodologies of previous papers. This includes the construction of a virtual conversation presented in Chapters 5 and 6, which took inspiration from Slater et al. (2013) or the moral dilemma used in Chapters 4 and 6 which used techniques similar to Pan and Slater (2011) to ensure participants could respond to a virtual moral dilemma without them knowing the details of what is going to happen.

The studies presented here also took serious notice for the need to ensure participants felt a sense of presence in the virtual environment in order to more reliably attain realistic responses to the virtual events – something that cannot be said for all moral behaviour studies using IVR. For the conducted studies, this included full upper-body tracking which translated into embodiment and agency over a participant's own virtual body as well as ensuring appropriate and modern headset technology was used and scenarios did not break participants' expectations of reality, as much as possible.

In terms of moral decision-making research, the review aimed to fully understand how the morality of people works in evolutionary, psychological and neurological terms, while also recognising the philosophical background behind the trolley problem and how it can be used as a methodological tool to probe the make-up of our morality. This knowledge is critical in order to fully understand the responses from participants to the moral dilemmas presented to them in the conducted studies.

The end of this section focused more on moral actions and behaviours as opposed to the main bulk of research on moral decision-making that has focused on moral judgments. In the conclusion of this section, an aspect of this area that requires further research was identified, namely, why people are acting increasingly utilitarian in IVR compared to what is found in questionnaire studies. The current hypothesis is that the saliency of outcomes incites people to focus on them instead of the physical actions, leading to greater utilitarianist actions being taken. This is considered more closely in the second study in Chapter 5.

Finally, the research behind social influence was analysed. The bulk of the research was aimed at social influence during general decision-making scenarios due to the lack of research in moral situations. It was also identified that, while moral judgment studies have been published utilising social influence during trolley problem style moral dilemmas, no moral behavioural studies have been conducted, leaving a big gap in our current knowledge of how morality and social influence operate together in cases with more tangible circumstances and outcomes. As such, this became the focus of the second and third studies, Chapters 5 and 6.

From all of the areas researched, a lack of research combining both moral decision-making and social influence was noticed which formed the motivation behind the focus of the studies conducted as part of the PhD. Overall, only four studies (Bostyn & Roets, 2017; Kelly et al., 2017; Kundu & Cummins, 2013; Lisciandra et al., 2013) found to date have directly looked at the effects of social pressure on people's responses to moral dilemmas (i.e., the switch dilemma and its family of ethical problems). Furthermore, these studies use somewhat antiquated techniques for probing participants of their moral intuitions, opting to use questionnaires which describe abstract scenarios to participants and elicit intangible judgments from them instead of utilising newer techniques such as IVR technology which require literal actions in order to resolve the situation.

As such, in order to add to this shallow body of knowledge, we decided to use IVR to study whether moral behaviours could be influenced by compliance attempts. The goal of this was to discern whether specifically moral behaviours, as opposed to abstract moral judgment, were sensitive to the opinions of others where the four previous studies found that moral judgment was indeed impacted by the presence and judgments of others.

Following this, we asked the question "Is there a measurable difference between the efficacy of compliance attempts between moral and non-moral dilemmas?" This was born from the results

of the Chapter 5, which revealed that compliance attempts were ineffective at changing participants moral behaviours. A hypothesis was formed that proposed that the reason compliance attempts failed in the second study was due to the fast, intuitive nature of emotions guiding our moral reactions, solidifying a response to a moral dilemma and subsequently inhibiting the power of later social influence attempts. These emotions are not found in non-moral dilemmas however and so the final research question proposed comparing moral and non-moral dilemmas to see if they interacted differently with social influence.

Aside from the research questions involving social influence, we also asked whether the difference between moral judgment and behaviour is modulated by omission bias. Similar to the sparse research on social influence during a moral dilemma, there is almost no research on how the effect of omission bias differs between moral judgment and behaviour. This is surprising considering it is one of key differences between the mediums of questionnaire and IVR, the fact that IVR scenarios require participants to literally carry out the action (or not). The omission bias, as well as it's related psychological effects, normality and action bias, have been studied extensively in non-moral behavioural studies and moral questionnaire-based studies, the knowledge of which contributed greatly to understanding the results to the second study.

Finally, the we also tackled the question of whether there is a measurable difference between people's moral judgment and behaviours in response to a moral dilemma involving a more concrete character (a child in our case). Previous studies (e.g., Tassy, Oullier, et al., 2013) have shown that naming or otherwise identifying characteristics of characters involved in moral dilemmas can sway participants reactions to those dilemmas, but this has only been tested in questionnaire-based studies. One motivation for the first study then was to conduct something similar in IVR using a moral dilemma with a child, a person with a typically greater value of life compared to an adult, and five adults where the participant had a much more visual sense of the characteristics of who they were saving or letting die.

Chapter 3

Methods

3.1. Moral Dilemmas

The three studies conducted and presented in this thesis were partly inspired by the trolley problem, which is discussed at length in the literature review above (Section 2.2.3). The moral dilemmas associated with the trolley problem revolve around our aversion to harm and all follow a basic pattern whereby good and bad outcomes are mismatched with good and bad actions, forcing responders to decide what is the least ethically taxing: committing a bad action or allowing a bad outcome to come about. An example of this type of moral dilemma might be whether or not it is morally permissible to kill one person in order to save five others (Foot, 1967). Chapters 4 and 6 present studies which offer participants the option to either save a young child or five adults, the option to save both is not available. In Chapter 5, the scenario instead revolves around the decision to either help a man get to hospital or to acquiesce to the demands of a policeman, upon which doing would put the health of the man in jeopardy.

There are several benefits to using artificial moral dilemmas. Firstly, scenarios can be altered along a single dimension in order to distinguish specific aspects of our moral judgment. Previous studies have used this methodology to test our responses to actions versus omissions (Cushman et al., 2006), the doctrine of double effect (Hauser et al., 2007) and the use of personal force (Greene et al., 2009). The presentation of a moral dilemma can also be changed, such as the wording of text-based vignettes (Tassy, Oullier, et al., 2013) or by using immersive virtual reality

to present a life-like simulation of the events (Pan & Slater, 2011). In the studies presented later, dimensions were altered in a between-subjects design in order to investigate the effects that this has on moral judgment or behaviour. Specifically, study one focused on the difference between moral judgment and behaviour by presenting a moral dilemma in either a questionnaire or immersive virtual reality experience, study two investigated the effects of social influence and omission bias on moral behaviour in immersive virtual reality as well as the effect of omission bias between moral judgment and behaviour and finally study three focused solely on the effects of social influence during moral and non-moral dilemmas.

Additional benefits of using artificial moral dilemmas include reducing the possibility that participants have any sort of personal connection with the events of the scenario (Hauser et al., 2007). Using real-world events might evoke different responses from participants depending on their experiences, whereas manufactured dilemmas are less likely to resonate personally with any one subject. Furthermore, using a common methodology such as switch-esque dilemmas (Greene, 2015) allows the comparison of results against those of previous studies and the use of robust models based on similar moral dilemmas, such as Greene's dual-process theory of moral judgment (Greene et al., 2008) in order to more accurately predict results. Cushman (2013) regards these types of harm-based dilemmas as the *lingua franca* of moral judgment studies and also as "a useful proving ground" for theories regarding aversion to harm.

3.2. Virtual Environment Development

Each study utilised immersive virtual reality (IVR) technology in order to study moral behaviour in response to simulated moral dilemmas. The Unity content creation engine was used to develop all IVR scenarios, although different versions were used for each study (see Table 3.1).

Scenarios were developed individually, but some code and other asserts were shared between projects. Assets unique to projects are discussed in their own methods section.

Study	Unity version
Study One	5.4.1f1
Study Two	5.6.3f1
Study Three	2017.3.1f1

Table 3.1. – Unity versions.

3.2.1. Embodiment

In each scenario, participants were given a gender matched virtual body that they could control from a first person perspective. This was accomplished by tracking the position of participants head and hands and extrapolating from these the position of the rest of the body. To simplify the implementation, participants were always asked to keep their feet and legs still so only the upper body needed to be tracked. In the first two studies, participants were sat down, on a stool in the first study and in a *Playseat DiRT 3 Edition* stationary gaming chair in the second in order to more accurately replicate the feeling of being seated in a car. In the final study, participants were stood up.

Once the tracking data was sent to Unity, inverse kinematics (IK) was used to correctly place other parts of the body such as chest and elbows, giving the illusion of full upper body tracking. In the first study, inverse kinematics was programmed from scratch, whereas the Final IK Unity package was used in the second and third studies in order to achieve a more realistic effect.

Briefly, IK is the method of calculating the position in space of a node based on a child lower in the transformational hierarchy. This is typically done the other way around where children are transformed by the position and rotation of their parent. For example, a hand is lower than an elbow in its hierarchy because the position of an elbow will determine the position of its hand. However, using IK, we can track a person's hand and, assuming the shoulder is fairly static, can estimate the position of the elbow, allowing multiple joints to be calculate with minimal tracking. With only three points this can be calculated using simple mathematics, where the position of the shoulder (again assuming its relatively static) and hand are known and the length of the of upper arm, lower arm and distance between the hand and shoulder are known. See Figure 3.1 for a diagram of this.

The tracking data for the head was always taken from the position and rotation of the headset. This is relatively easy with both the Oculus DK2 and HTC Vive given their integration with Unity. In order to have the scenario's camera match the movements of the headset, a simple checkbox can be selected and the appropriate SDKs for the headset chosen, see Figure 3.2 where OpenVR maps to the HTC Vive. This was then translated into actual head movements of the embodied character by having the head bone situated identically with the camera and also as a child of it, meaning any transformations applied to the camera are also applied to the head, see Figure 3.3 for how this looks in Unity.

The illusion of virtual hands was somewhat more complicated to achieve as the position of the object being tracked (either an OptiTrack rigid body or the Vive controllers) was not identical to the virtual object that it was motivating, i.e., the virtual character's wrist. For example, the OptiTrack rigid body sits on the back of the hand (gloves shown in Figure 3.4), so rotating the wrist in the motion seen in Figure 3.5 would cause the rigid body to not just rotate, but also translate (change position), whereas the person's wrist would only be rotating. To accommodate for this, an offset was subtracted from the position of the tracked object from its original position towards the virtual characters wrist, cancelling out the positional difference between rigid body (or Vive Controller) and wrist. This sounds simple in theory, but required the knowledge of vector mathematics as well as a lot of manual testing required to get the offset correct. For more information on tracking equipment used, see section 3.3.

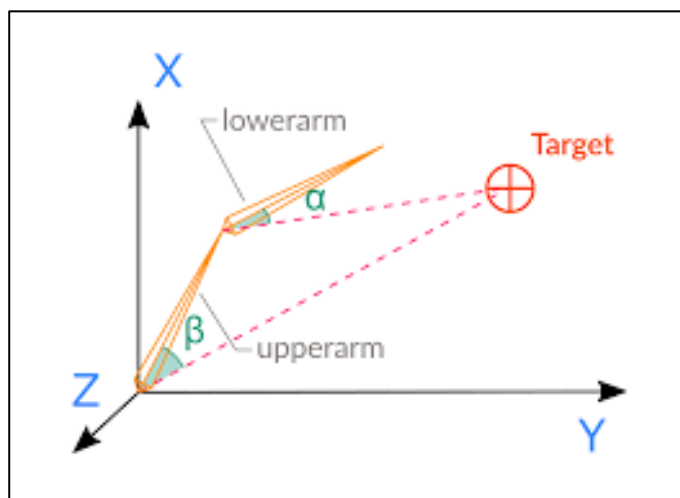


Figure 3.1. – Inverse kinematics diagram

(docs.godotengine.org/en/3.0/tutorials/3d/inverse_kinematics.html).

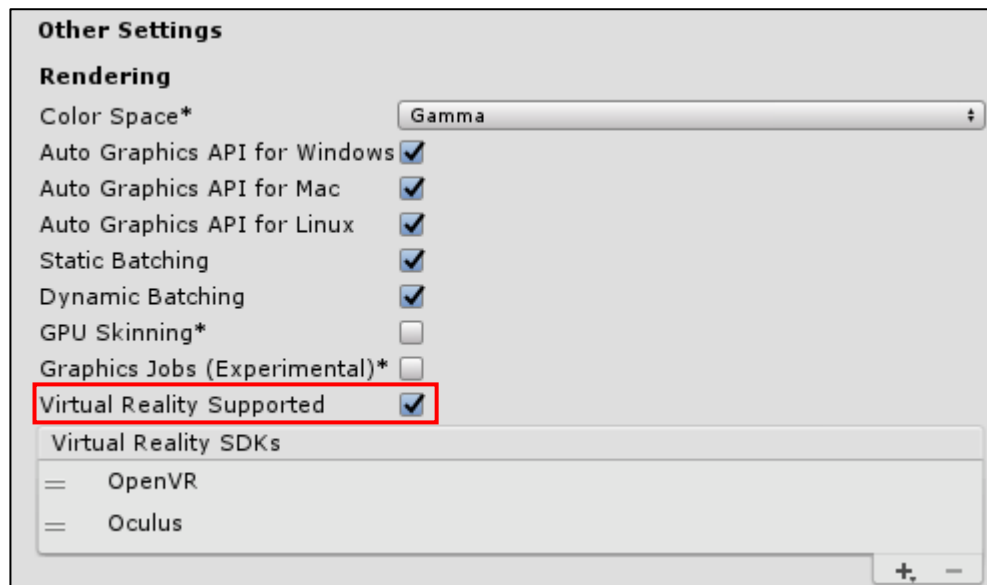


Figure 3.2. – Immersive virtual Reality support in Unity.

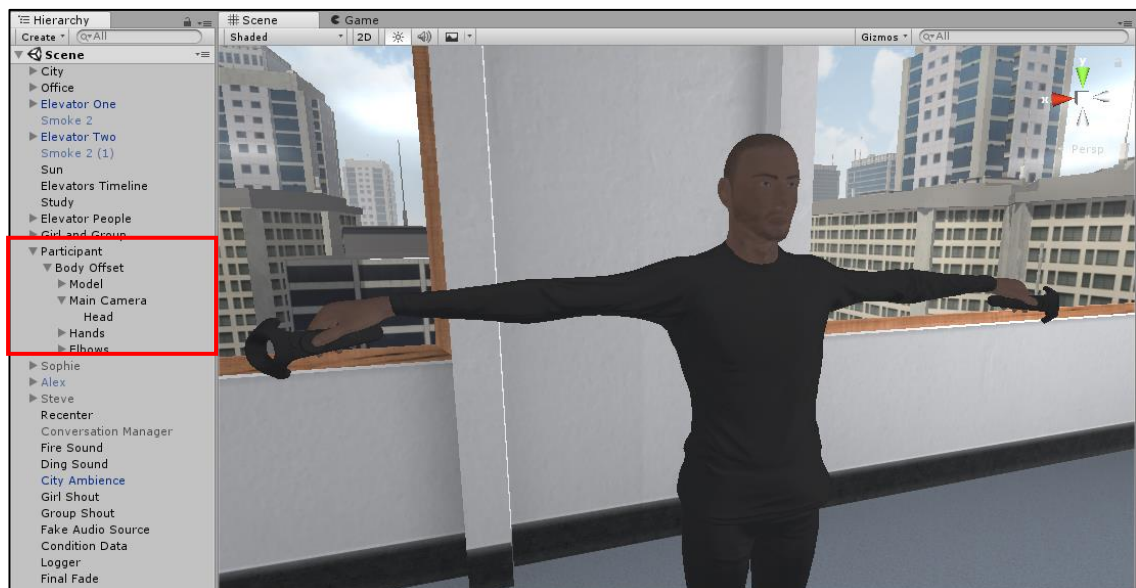


Figure 3.3. – Embodied character setup.



Figure 3.4. – OptiTrack gloves.

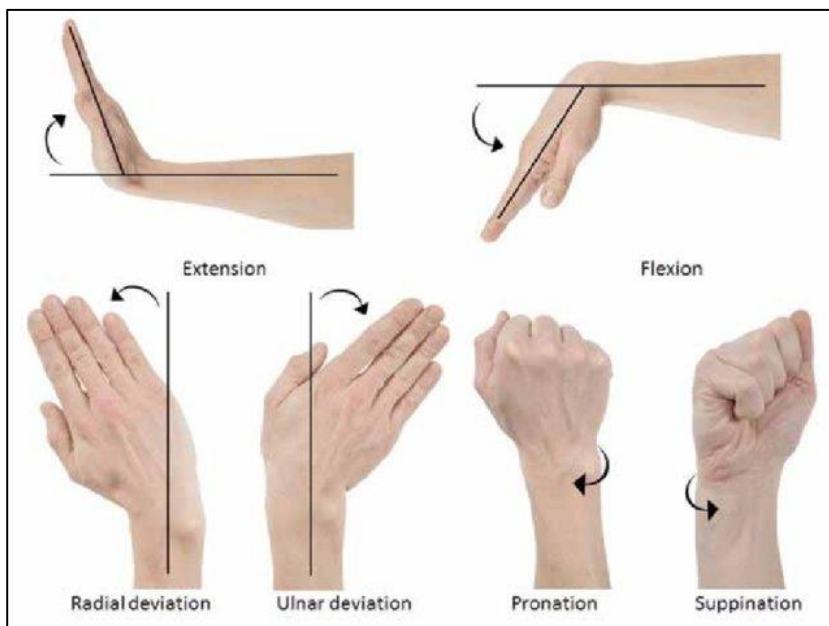


Figure 3.5. – Wrist movements (Raven, 2017).

3.2.2. Virtual Characters

All virtual characters (VCs), as well as the virtual body given to the participants, were modelled and rigged using Adobe Fuse (Beta). The only exception to this are the pedestrians in the second study which were purchased on the Unity asset store. This was because many unique looking

characters were needed to simulate a city environment, but quality could be lower due to their distance from the participant.

In the second and third studies, participants engaged in a conversation with several VCs in order to build up a rapport with them before the characters eventually attempted to influence the participants during a moral dilemma. To accomplish this, dialogue was written and recorded with age and gender appropriate voice actors and body movements were recorded to match the conversation.

To facilitate the recording of body movements, a custom animation system was developed for Unity in which the transformation matrix of entities could be saved every fixed frame (60Hz) and then written as a Unity readable mark-up file. These files could also store other useful information such as transformation offsets (see Table 3.2). Using this system, an actors head and hand movements were recorded while acting out body movements appropriate to the conversation for each character. This animation was then applied to the respective character during the scenario at the necessary time. Lines were recorded individually and idle animations were recorded to allow VCs to retain some movement while waiting for a reply from the participant. Similar to the participant's embodied avatar, Final IK was used to give the VCs realistic body movements. Additionally, the Unity package LipSync Pro was used to give VCs mouth and facial animations when speaking to the participant.

Name	Description
Loop	Continually loop the animation.
Mirror	Play the animation in reverse.
Playback speed	Playback speed of the animation.
Position offset	Initial position offset of the animation.
Rotation offset	Initial rotation offset of the animation.
Line	The audio file of the conversation line.
Line offset	When to play the audio file.
Frames	The transformations.

Table 3.2. – Animation variables.

3.2.3. Virtual Conversation

The code controlling the conversation between participants and the virtual characters was shared between studies two and three. In order to greatly simplify development, a linear conversation system was implemented as opposed to branching system. This means that no matter what the participant said to the VCs, they would always respond with the same line of dialogue. As such, the only control the experimenter had over the conversation was to initiate the next line of the conversation when the VCs were waiting and had received a response from the participant or to repeat the previous line of dialogue if the participant had not heard it clearly. Other than this, the conversation played out automatically. In order for the conversation to seem real using this template, the VCs would ask the participant a question and then wait for a response. Once a response had been given by the participant, the experimenter would trigger the next line which would usually be a generic response to the participants answer, for example:

VC: What course are you doing at UCL?
Participant: Computer science...
VC: Cool. I'm doing Psychology. Are you enjoying it?
Participant: Sure...
VC: I'm struggling with my course a little bit. What are you planning to do after university?
Participant: Find a job...
VC: I'm not sure what I'm going to do. I'll be going to a talk at the students union later. You should come.

For a full transcript of the conversations used in studies two and three, see Appendix C.

3.3. Equipment

Each study utilised IVR technology to bring their respective moral dilemmas to life and head-mounted displays (HMDs) were used as opposed to other options such as a Cave (Cruz-Neira et al., 1993). In the first experiment, the Oculus Rift DK2 was used. This headset has a resolution of 960x1080 per eye, a refresh rate of 75Hz and a 100-degree field of view. The final two studies

instead used the newer HTC Vive which has a resolution of 1080x1200 per eye, a refresh rate of 90Hz and a 110-degree field of view. The computer used in all three studies was a custom built desktop computer running Windows 8.1 with a quad-core, 4GHz i7-4790K processor, GeForce GTX 1080 graphics card and 8Gb of RAM. The graphics card was upgraded just before the second study, with a GeForce GTX 980 being used in the first study.

Aside from this, the only other equipment made use of was tracking equipment, used to follow participants body movements as well as to record the animations for some virtual characters. This was achieved in the first two studies with a combination of the respective HMDs head tracking technology and OptiTrack to track rigid body markers placed on the back of participants hands. Twelve OptiTrack Flex 3 cameras were connected to a computer running the Motive software. This computer was a Dell Optiplex 990, running Windows 7 with a 3.4GHz i7-2600 processor, 8Gb of RAM and an ATI Radeon HD5450 graphics card. Data from this computer was then sent over a local area network to the computer running Unity using the NatNet 2.9 SDK. In the final study, HTC Vive controllers were used to track participant hand movements instead of OptiTrack.

3.4. Participants

Participants were recruited from a variety of Departments and Faculties throughout UCL with all levels of student being permitted to take part. This distinction was more important in the second and third studies as the virtual characters would assume that the participant was a student at UCL during the conversation. Experiments one and three were counter-balanced, consisting of an equal number of male and female participants whereas experiment two admitted only male participants to reduce the impact of gender on decision-making.

Each experiment was approved by the UCL Research Ethics Committee, and was carried out with written informed consent from each participant. Participants had to pass a brief trauma questionnaire before being allowed to take part in each study due to the potentially stressful nature of the scenarios and were also verbally reminded at several points throughout the experimental training procedure that they would be free to stop at any time without having to give reasons or lose any other rights to which they were entitled. They were paid 7 pounds (UK).

Participants were not permitted to take part in more than one study as this might have impacted their responses.

3.5. Procedures

There are some common procedures that were shared among the three studies. Firstly, as just alluded to, participants in all studies were required to complete a brief trauma questionnaire which can be found in Appendix B. In the first study this was done in the lab before the study started, but was sent to prospective participants by email in the second and third studies. The survey consists of 10 questions about stressful or traumatic past experiences. If participants indicated that they had any past traumatic experiences, they were declined from taking part in the study. The only exception to this was during the first study, where they were permitted to take part in the questionnaire condition. The reason we opted to use this questionnaire was to reduce the likelihood of any psychological trauma to participants that might have been caused by the salient nature of the virtual dilemmas.

Assuming participants passed the trauma questionnaire, they were assigned a participant ID and given a time to come to the lab. Each participant was assigned a 1 hour block usually between the times of 10am and 7pm. Although each participant usually took less than 30 minutes to run, the hour was used to simplify booking times and allowed participants to be late or account for any other type of disruption.

At the lab, participants were greeted and asked to read an information sheet, ask any questions they wanted (although not all could be answered, such as questions regarding the purpose of the study) and then sign a consent form (see Appendix B for information sheets and consent forms for each study). Once the consent form was signed, participants were asked to complete a demographic questionnaire and were relayed instructions about the virtual scenario. It was never mentioned that a moral dilemma would occur, but participants were warned that the scenario might be stressful. Instructions usually included information about how to open and close the elevators (study 1 and 3) or about the autonomous car (study 2). Once it was clear that the

participant understood what they had to do, participants were seated or otherwise placed in the correct position and IVR equipment was put on.

Each scenario made use of a “training” phase which allowed participants to get used to IVR in case they had not used it previously and become accustomed to any tasks they being asked to carry out. During the studies, the training phase was usually referred to as “phase 1” so as to not make participants feel like they are simply part of an experiment. After the training phase, participants were asked a final time if they are happy to continue and the main scenarios were started if so. These usually lasted for 2 to 3 minutes, ultimately ending at the conclusion of the moral dilemma at which point the display would fade to black.

Once out of the virtual environment, participants were helped to remove the equipment and were immediately asked to fill in a presence questionnaire regarding their experience of IVR. This includes questions relating to the virtual world itself, the virtual characters and their general feeling of being in the scenario rather than in the lab. We adopted the categorisation presented in Slater (2009) which distinguishes between Place Illusion (PI), and Plausibility Illusion (Psi). PI is a perceptual illusion that the person in IVR is somewhere other than their true location that is likely to arise when the VR system affords natural sensorimotor contingencies for perception – the use of the body for perception in a way similar to perception in reality (head turns, bending, reaching, etc). Psi is a cognitive illusion that the events being perceived are occurring. The questionnaire, along with the encoded values for each question, are shown in Tables 3.3 to 3.5.

there	Please rate your sense of being in the virtual environment where 7 represents your normal experience of being in a place.
real	To what extent were there times during the experience when the virtual environment was the reality for you? 1 = At no time... 7 = Almost all the time.
visited	When you think back about your experience, do you think of the virtual environment more as images that you saw, or more as somewhere that you visited? 1 = Images that I saw... 7 = Somewhere I visited.
lab	During the time of the experience, which was strongest on the whole, your sense of being in the virtual environment, or of being in the real world of the laboratory? 1 = Being in the virtual environment... 7 = Being in the lab.
overwhelm	During the time of the experience, did you often think to yourself that you were just sitting in a laboratory or did the virtual environment overwhelm you? 1 = Most of the time... 7 = Rarely.

Table 3.3. – Place Illusion questions.

situationreal	How much did you behave within the virtual environment as if the situation were real? 1 = Not at all... 7 = Very much so.
emotionreal	How much was your emotional response in the virtual environment the same as if it had been real? 1 = Never... 7 = Almost all the time.
thoughtsreal	How much were the thoughts you had within the virtual environment the same as if it had been a real situation? 1 = Never... 7 = Almost all the time.
behavingreal	How much were you thinking things like 'I know this isn't real' but then surprisingly finding yourself behaving as if it was real? 1 = Never... 7 = Almost all the time.
physreal	To what extent were your physical responses within the virtual environment (e.g., heart rate, blushing, sweating, etc.) the same as if it had been a real situation? 1 = Never... 7 = Almost all the time.
envreal	Overall how much did you treat the virtual environment as if it were real? 1 = Not at all... 7 = Very much so.

Table 3.4. – Plausibility for the situation questions.

peoplereal	How much did you behave as if the virtual people were real? 1 = Not at all... 7 = Very much so.
emotionpeople	How much was your emotional response to the virtual people as if they were real? 1 = Not at all... 7 = Very much so.
thoughtspeople	How much were your thoughts in relation to the virtual people as if they were real? 1 = Not at all... 7 = Very much so.
physpeople	How much did you have physical responses (such as change in heart rate, blushing, sweating, etc.) to the virtual people as if they were real? 1 = Not at all... 7 = Very much so.
behavepeople	How much were you thinking things like 'I know these people are not real' but then surprisingly finding yourself behaving as if they were? 1 = Not at all... 7 = Very much so.

Table 3.5. – Plausibility for the virtual people questions.

The first and third studies also required participants to fill in moral dilemma questionnaires which consisted of four moral dilemmas: the elevator, switch, inverted-switch and footbridge dilemmas (Table 3.6). Finally, participants were debriefed about the nature of the study and encouraged to ask questions.

The only exception to these procedures is the second part of the second study which instead involved an online questionnaire which participants around the university were encouraged to respond to through use of the universities social media. This study consisted of two conditions and to ensure a somewhat equal number of participants in each condition, responders were asked to enter the day of the month they were born on. Participants born before the 16th were put into one condition and those born on or after were put into another. In the end, the split between the conditions came to 138 in one and 146 in the other.

Elevator	You are controlling the doors to two elevators when they arrive at your floor. At some moment, a fire has spread between two elevators which cannot be opened from the inside. From one, you can hear the screams of five adults and from the other the screams of one child. You only have time to open one elevator. Who should be saved?
Switch	An empty boxcar is running out of control down a track. In its path are five people standing on the track; these people are not aware of the oncoming danger. If the boxcar continues, it will kill all five people. You are standing next to a switch. If you flip the switch, it will cause the boxcar to turn off the main track and onto a side track. On the side track, there is one person who is also unaware of the boxcar. If the boxcar goes down this side track, the one person will die but the five people on the main track will survive. Would you push the switch?
Omission	An empty boxcar is running out of control down a track. In its path is one person standing on the track; this person is not aware of the oncoming danger. If the boxcar continues, it will kill this person. You are standing next to a switch. If you flip the switch, it will cause the boxcar to turn off the main track and onto a side track. On the side track, there are five people who are also unaware of the boxcar. If the boxcar goes down this side track, the five people will die but the one person on the main track will survive. Would you push the switch?
Footbridge	An empty boxcar is hurtling out of control down a track towards five people. If the boxcar continues, it will kill all five people. You are on a bridge over the tracks. The boxcar will pass under the bridge before it reaches the five people. You can stop the boxcar by dropping a heavy weight in front of it. Standing next to you is a man wearing a heavy backpack. If you push him over the bridge, he will land in front of the boxcar and stop it before it reaches the five people ahead. This man will, however, die. Would you push the man?

Table 3.6. – Moral dilemma questionnaire.

Chapter 4

Moral Judgment Versus Moral Behaviour

In this study, two groups of participants either responded to a moral dilemma through moral action, by responding to an immersive virtual reality representation of it, or through moral judgment, by writing down their answers after reading a text-based version of the dilemma. The moral dilemma involved the choice to either save one child or five adults from a pair of burning elevators, where the participant only has time to open one elevator. The goal was to investigate the difference between moral judgment and actual moral behaviour by comparing the results between the two mediums and add to the current body of knowledge on moral behaviour by introducing a novel aspect to the moral dilemma, notably, the inclusion of a child. As the life of a child is inherently more valuable than that of an adult's (Kawai, Kubo, & Kubo-Kawai, 2014), the inclusion of a child in the moral dilemma forces participants to think more carefully about the value of the lives they are saving rather than simply being able to compare the lives of five versus the life of one, such as in the switch dilemma. The question then is whether this assessment of lives changes between making a judgment or taking an action in immersive virtual reality (Tassy, Oullier, et al., 2013).

Somewhat congruent with earlier studies looking at impersonal moral dilemmas in immersive virtual reality, there was no significant difference between moral judgments and moral actions. A subsidiary result showed that those participants in the moral judgment task (who still experienced

IVR, but did not witness a moral dilemma) consistently rated higher presence scores than those in the moral action task.

4.1. Introduction

As discussed in Chapter 2, the moral dilemmas born from the trolley problem, such as the switch or footbridge dilemma, have allowed us to probe the constituents of peoples moral decision-making for decades. Originally starting in philosophical discussions, the trolley problem has moved to psychology and neurology and is now starting to transition into the study of moral actions, aiming to understand people's real moral behaviour as opposed to abstract moral judgments.

In order to more fully understand the results of newer moral actions studies, it is important to recognise the theories and conclusions drawn from previous research. In the domain of psychology and neurology, moral judgments are often considered in terms of emotional and rational cognitive processes, working in tandem whenever a response to a moral dilemma is required. It has also been speculated that while rational processes are biased towards utilitarianism, emotional processes support deontological thinking and that these emotions likely evolved to inhibit our immoral behaviour in support of a cohesive society (Greene et al., 2008; Haidt, 2001).

Moral actions on the other hand are concerned with what people actually do in response to a moral dilemma instead of their opinion about what should be done. Previous results have supported a difference between these two moral expressions. FeldmanHall et al. (2012) showed that participants gave significantly more money to prevent an electric shock during a real compared to a pretend event and Bostyn et al. (2018) showed that those in a real-life condition were twice as likely to shock a mouse in order to prevent the same thing happening to five other mice. While the results of these two studies reveal differences between moral responses to real and pretend events, they do not denounce the conclusions from purely hypothetical studies, but hint that they should be used more as a "moral barometer" (FeldmanHall et al., 2012) than as a true reflection of our moral behaviours.

A drawback of these studies – FeldmanHall et al. (2012) and Bostyn et al. (2018) – is their reliance on non-lethal, in fact relatively harmless, actions. This places their conclusions about morality in a more precarious position than if more dire consequences, such as death, were involved. As explained in Chapter 2, due to ethical considerations, it is impossible to put people in a real-life situation where they would need to take some action to resolve a deadly moral dilemma, but that a solution is to use immersive virtual reality (IVR). It was observed by Blascovich et al. (2002) and more recently by Pan and Hamilton (2018), that IVR provides an ideal methodology for social psychology, and this has been demonstrated in previous studies, including partially replicating the Stanley Milgram Obedience to Authority studies (Milgram, 1963) in IVR (Gonzalez-Franco et al., 2018; Slater et al., 2006) and another investigating bystander responses to a violent incident (Slater et al., 2013).

Recent moral dilemma issues have also been studied using IVR. For example, Patil et al. (2014), Francis et al. (2016) and Francis et al. (2017) all found that, in IVR, participants acted in a more utilitarian manner in response to a moral dilemma compared to a non-IVR control group answering a traditional questionnaire. The results, shown in Table 4.1, show a striking difference between simulated and hypothetical responses to the situation.

Paper	IVR		Questionnaire	
	Switch	Footbridge	Switch	Footbridge
Hauser et al. (2007)	-	-	89%	11%
Pan and Slater (2011)	89%	-	84%	13%
Navarrete et al. (2012)	91%	-	-	-
Patil et al. (2014)	95%	-	76%	-
Skulmowski et al. (2014)	96%	-	-	-
Francis et al. (2016)	-	70%	-	20%
Francis et al. (2017)	-	56%	-	10%
McDonald et al. (2017)	93%	52%	-	-

Table 4.1. – Proportions who save the 5 in conditions equivalent to the Switch Dilemma (to throw the switch to save 5) and the Footbridge Dilemma (to throw a heavy man off the footbridge to save the 5).

The proportions who act in a utilitarian manner (i.e., those who save the 5 compared to the 1) across various studies are shown in Table 4.1. This distinguishes between situations equivalent to the Switch Dilemma (throw a switch to avert the trolley killing the 5 but therefore kill the 1) and the Footbridge Dilemma (throw a heavy man off the footbridge to save the 5, but at the cost of killing the heavy man).

The table shows that on the whole the switch case leads to much greater utilitarian behaviour than the footbridge, but also that studies using IVR show greater utilitarian behaviour than those based on questionnaires.

The aim of the present study is to shed further light on the difference between moral judgment and action in response to a moral dilemma by presenting the dilemma in either a questionnaire or an IVR experience. This study also utilises a different moral dilemma from the standard switch dilemma used in previous similar studies (i.e., Navarrete et al., 2012), introducing a child into the moral dilemma rather than only the (presumed) adult potential victims of previous research. This has only been looked at in one previous study which predictably found that sacrificing adults or the elderly was less disapproved than sacrificing children (Kawai et al., 2014). Additionally, we were interested to see whether the IVR experience would lead people to respond to the moral dilemma questionnaire differently compared to those who did not have the IVR experience. Although not utilising IVR, (Tassy, Oullier, et al., 2013) found that responses to moral dilemmas including non-generic characters changed between objective and subjective perceptions of the dilemmas. Conceptually, the inclusion of a child in the current moral dilemma should force participants to think more carefully about the lives of those at stake instead of relying on the innate knowledge that five lives is greater than one like in the switch dilemma which has previously been researched in IVR studies (Navarrete et al., 2012).

Previous results from studies looking at the difference between moral judgment and action have shown little difference between the two responses when targeting impersonal moral dilemmas. As the current moral dilemma is more akin to an impersonal moral dilemma rather than a personal one, these previous results form the basis of what could be expected to happen in the current study. However, the inclusion of a child in the scenario is possible to make the situation more emotionally-charged (Kawai, Kubo, & Kubo-Kawai, 2014) and thus could impact participant's judgment and actions differently given the increased emotional saliency of IVR. In this case, we

might expect participants to save the child in IVR more than in the questionnaire, as their emotions are likely to be more in favour of saving the child while rationally, it could be argued that it is better to save the five.

4.2. Methods

4.2.1. Scenario

The IVR scene depicted a 12.5m by 7.5m room. Participants were seated on a physical chair registered with a corresponding virtual chair in the centre of this room. They had a sex-matched virtual body that they would see if they looked down towards themselves. They were seated by a virtual table registered with a real table that had two red buttons on top. In front of where they were seated was a virtual wall which showed the (usually closed) doors to two elevators. Every so often one of the buttons on the table would become green signalling to the participant that they needed to press it in order open its respective elevator (the button to the left corresponding to the left elevator and the one on the right to the right). When the elevator door opened a virtual human character would walk out past the participant and leave the room through one of hallways behind. Participants were instructed that if both buttons showed green then pressing both simultaneously would result in neither door opening and that after pressing one button, the other would become disabled. This was demonstrated to the participant during a training phase of the study.

After four virtual characters had exited from the elevators, a young girl (aged about 10) and a group of five adults walked into the room from one of the hallways behind the participant, appearing from the same side to ensure the participant noticed both parties. The side that the virtual characters appeared from was counterbalanced across the participants. The group would stand in front of the elevator on one side while the young girl made her way over to the other elevator, walking in front of the participant. After a few seconds the elevator doors opened and both parties, the group of 5 adults and the young girl, would enter their respective elevator. As soon as the doors closed, smoke came out of both elevators and the doors appeared to turn red hot, indicating that there was a fire inside. Screams from the young girl and group would also be

heard from the respective elevators. After several seconds, both buttons became green and the participants were able to choose whether to save the young girl or the group. Figure 4.1 illustrates the scenario.



Figure 4.1. – The scenario (A) A participant seated wearing the HMD and tracking equipment showing the table with the buttons (B) The virtual representation of the table, buttons and body. (C) A character walking out of the elevators (D) The child and 5 adults walk towards the elevators (D) The child walks alone into the elevator. (E) the fire.

4.2.2. Experimental Design

There were two conditions referred to as 'Dilemma' and 'No Dilemma'. In the Dilemma condition, participants experienced the moral dilemma at the end of the IVR experience after the girl and group had walked into their respective elevators, when the doors closed and the fire started. The participant then had the option of opening the door for either the group or the girl. In the No Dilemma condition, the IVR experience stopped as soon as the girl and group entered the elevator

and the doors closed. No fire was seen and there were no other events. Both conditions then exited the virtual scenario and completed the moral dilemma questionnaire detailed later.

4.2.3. Participants

Fifty participants took part in the study with 25 in each condition (No Dilemma and Dilemma) in a between-groups design. Participants were recruited from a variety of Departments and Faculties throughout UCL with all levels of student and staff being permitted to take part. Departments such as Computer Science, Psychology and Philosophy were not informed of the study since students from these schools might have known too much about immersive virtual reality or moral psychology/ philosophy, possibly affecting the results. The mean age of the participants was 26 ± 6.8 (SD) years. The number of men in the Dilemma condition was 13 and there were 12 in the No Dilemma condition. In the No Dilemma 21/25 were students (8 doctoral) and in the Dilemma condition 22/25 were students (10 doctoral). The remainder were Research Assistants or Administrative staff. None of the participants had any children.

The experiment was approved by the UCL Research Ethics Committee, and was carried out with written informed consent from each participant. Participants were also verbally reminded at several points throughout the experimental training procedure that they would be free to stop at any time without having to give reasons or lose any other rights to which they were entitled. They were paid 7 pounds (UK).

4.2.4. Equipment

The scenario was created using the Unity 5 game engine. All 3D models in the scenario were made in Unity, with appropriate textures being downloaded online, except for the virtual characters which were modelled and rigged using Adobe Fuse (Beta) and then animated using Unity's built-in animation system, Mecanim. The head-mounted display used was the Oculus Rift DK2 which has a resolution of 960x1080 per eye, a refresh rate of 75Hz and a 100-degree field of view. Body tracking was achieved using 12 Flex 3 cameras manufactured by OptiTrack in

conjunction with the software Motive and NatNet 2.9 networking SDK to send hand position and rotation data to Unity.

For this experiment, the participants were asked to always remain seated and to keep their feet together. This was to simplify the implementation of motion capture. The Oculus Rift position tracking camera was used to track the position and rotation of the participant's head. The virtual body of the participant was interpolated at runtime to adjust to the height and movement of the participant achieving a sense that it was the participant's virtual body. The only other part of the participant's body that was tracked were the hands. This was accomplished by placing OptiTrack rigid bodies on the backs of participants' hands. Elbow movements were calculated using inverse kinematics based on the length of the arm and the current position and rotation of the hand.

A real table and buttons were registered with their virtual counterparts in the lab so that participants had haptic feedback when touching objects in the virtual environment. The stool they sat on was also equal in size to the one in the virtual scenario.

4.2.5. Procedures

Participants arrived at the lab individually at a pre-booked time slot. They were not told before entering the scenario that the study was about moral dilemmas, however they were informed that they may (or may not) find the situation depicted in immersive virtual reality stressful or upsetting and of course that they were free to continue or leave as they preferred.

Participants first completed a brief trauma questionnaire to determine whether it was appropriate to let them experience the moral dilemma in immersive virtual reality. If the participant scored even one 'yes' to any question then they were assigned to the No Dilemma condition, otherwise they were assigned to the Dilemma condition. The number of participants assigned to the No Dilemma condition for this reason was 20 out of 25. This is because, based on the results of the questionnaire, the Dilemma condition filled up faster and thus the last 5 participants to take part all were all assigned to the No Dilemma condition regardless of the outcome of the questionnaire due to safer nature of the No Dilemma condition.

Participants were then given an information sheet detailing the experiment, prompted by the experimenter whether they had any questions and then given the consent form to sign if they wanted to continue. Finally, a demographic questionnaire was given to them before entering the virtual environment.

A series of instructions were read to each participant about what they needed to do in the virtual environment. The instructions were read twice and participants were asked to relay the instructions back to ensure they understood what they needed to do. Participants then attached the OptiTrack markers to their hands, sat down on the stool and put on the Oculus DK2. They first went through a training period (called stage 1 during the experiment) where people would come and go from the elevators requiring them to press the buttons to open the elevator doors. At the end of the training period both buttons turned green. This was a chance for participants to try and open both at the same time, learning that it was impossible.

After this stage 2 began, this being the scenario that included the moral dilemma detailed earlier. Participants in the Dilemma condition saw the fire and were given the option to save either the child or the group of adults whereas the No Dilemma participants stopped after the child and adults entered the elevator without seeing the fire. Participants in the Dilemma condition were also asked afterwards who was in both elevators and who they chose to save to ensure that they understood what had happened.

Participants then filled out a presence questionnaire, and finally a moral dilemma questionnaire, before being debriefed on the nature of the experiment.

4.2.6. Response Variables

The primary response variables for this study was the number of participants who saved the group of 5 adults compared to the child in the IVR scenario, as well as the answers to the moral dilemma questionnaire shown in Table 3.6. Additional response variables include the answers to the presence questionnaire shown in Tables 3.3 to 3.5.

4.3. Results

4.3.1. Presence Data

The presence questionnaire was adapted from Slater, Usoh, and Steed (1995) and captures the self-reported degree of participants feelings of place illusion (PI) and plausibility illusion (Psi) within the virtual environment. PI refers to the experience of being in a place other than the one in which the participant is actually situated and Psi to the experience that the events around the participant in the virtual world are really happening (Slater, 2009).

Figure 4.2 shows the responses of the presence questions found in Table 3.3. It can be seen that the scores are moderate, except for the control question *lab*, which is suitably low, and there are no systematic differences between the two conditions.

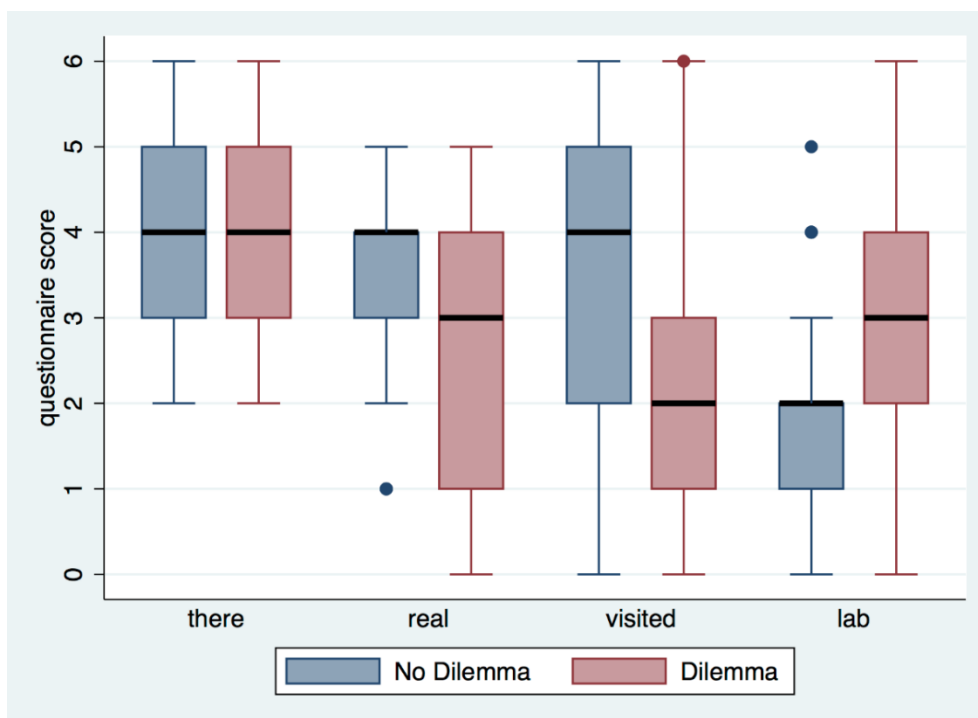


Figure 4.2. – Box plots for presence questions.

Figure 4.3 and 4.4 show the box plots for questions referring to the plausibility of the situation and virtual people found in Tables 3.4 and 3.5 respectively. A trend can be seen between the two

conditions whereby the responses from the participants in the Dilemma condition are consistently lower in almost all gauges bar *thoughtspeople*.

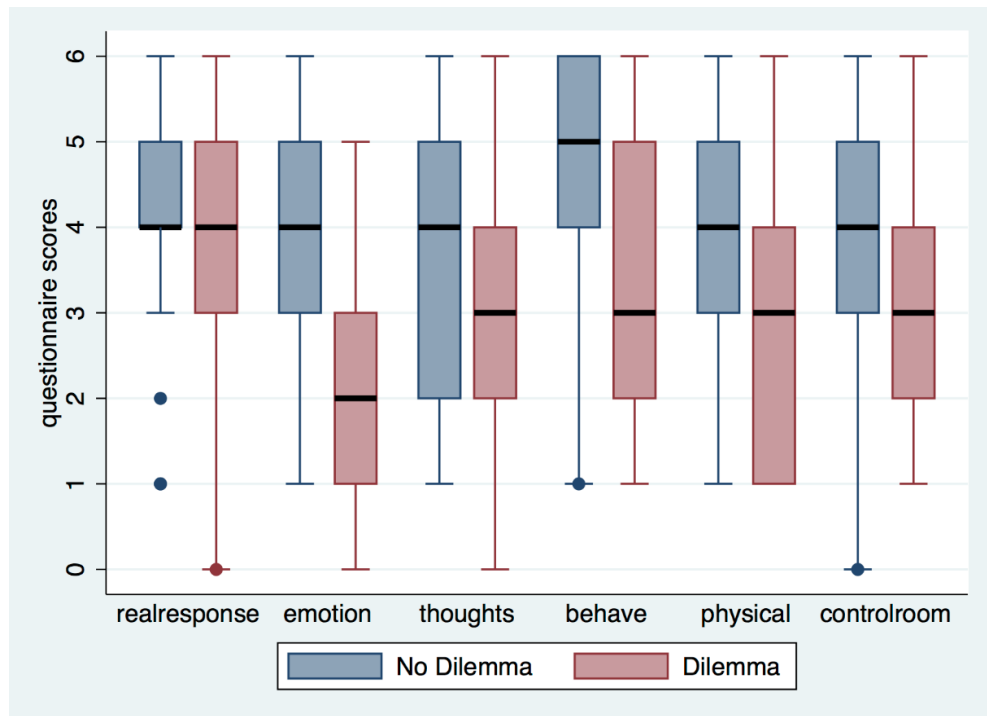


Figure 4.3. – Box plots for plausibility of the situation.

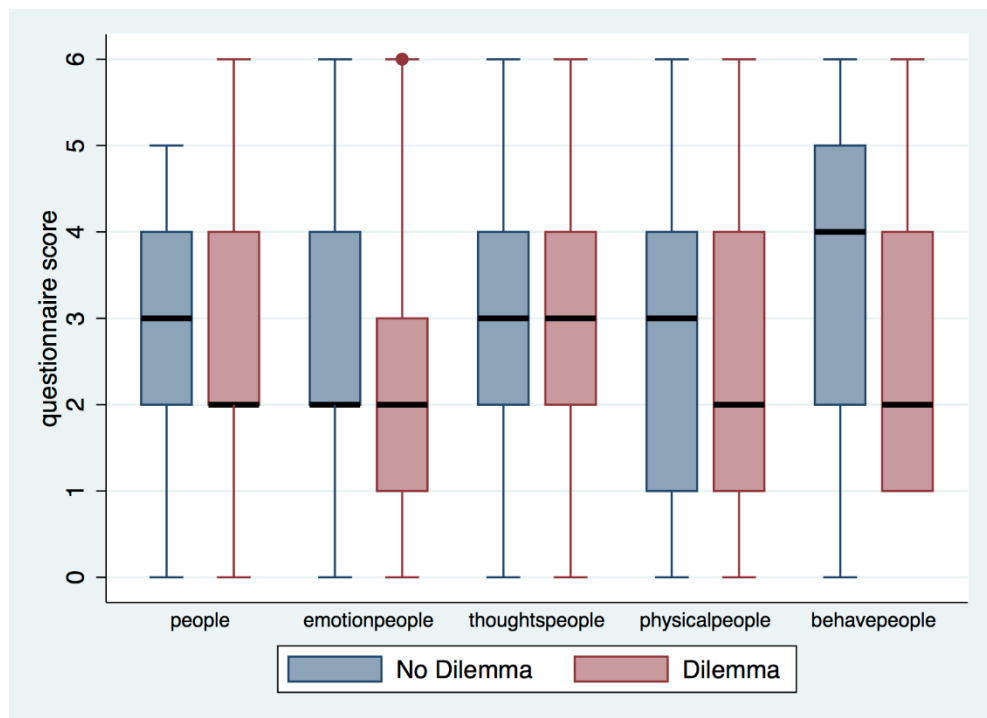


Figure 4.4. – Box plots for plausibility of the virtual people.

4.3.2. Response Data

Table 4.2 gives the questionnaire results corresponding to Table 3.6 showing that there was no influence of whether or not participants had experienced the No Dilemma or Dilemma conditions. However, regarding the question ‘Elevator’ both groups had seen 5 adults in one elevator and one child in the other, so that although those in the ‘No Dilemma’ condition had not actually been faced with the fire, they nevertheless may be able to better imagine this than people who had never experienced this scenario in IVR. Note that the 80% response to this question is less than all entries in Table 4.1 except for one, as are the percentages for Switch. Even though these are questionnaire responses it is possible that they were influenced by the scenario.

Condition	Elevator /25	Switch /25	Omission /25	Footbridge /25
No Dilemma	20 (80%)	18 (72%)	25 (100%)	4 (16%)
Dilemma	20 (80%)	17 (68%)	25 (100%)	2 (8%)

Table 4.2. – The number who saved the 5 in the questionnaire responses.

Participants in the Dilemma condition had the choice of opening one elevator door to save the child or the other to save the five adults. In fact, $x = 18$ out $n = 25$ saved the five (72%). We use a Bayesian analysis to understand this finding taking into account previous work in this area. The random variable x has a binomial distribution with parameter θ as the probability of saving the five, and $n = 25$. We let the prior distribution of θ be $\text{Beta}(\alpha, \beta)$. Then the posterior distribution of θ is known to be $\text{Beta}(x+\alpha, n-x+\beta)$, from elementary distribution theory.

From Table 4.1 we would expect that θ , for the Switch case should be between approximately 0.85 and 0.95. For the IVR studies the mean is 93% and for the Questionnaire studies it is 83%, and overall it is 89%. We examine three alternative methods for choosing the prior. The first (uniform case) is to choose a uniform prior for θ , i.e., ignoring the fact that we do have prior information from previous studies. This prior has $\alpha = \beta = 1$. (This is essentially equivalent to a classical statistics approach, giving equal prior probability for all possibilities between 0 and 1). The second (middle case) is to choose α and β to give a 0.5 probability for $0.85 < \theta < 0.95$. This

results in $\alpha = 14.22$ and $\beta = 1.58$. The third (strong case), is to choose the prior so as to give a 95% credible interval for the range $0.85 < \theta < 0.95$, thus taking into account results from previous experiments. This leads to $\alpha = 119.61$ and $\beta = 13.29$. This prior is heavily weighted towards results from previous experiments – for example, Hauser et al. (2007) found a (classical) 95% confidence interval concentrated narrowly around 0.89. Note that in the second two cases the mean of the prior distribution is 0.9.

Prior for θ	Mean \pm SD of prior	Mean \pm SD of posterior	Prior 95% credible interval (equal tails)	Posterior 95% credible interval (equal tails)	$P(\theta < 0.9 \mid x=18)$
uniform	0.5 ± 0.289	0.70 ± 0.016	0.025 - 0.975	0.522 - 0.857	0.997 (prior 0.9)
middle	0.9 ± 0.073	0.79 ± 0.063	0.717 - 0.991	0.654 - 0.899	0.977 (prior 0.410)
strong	0.9 ± 0.026	0.87 ± 0.026	0.844 - 0.945	0.815 - 0.919	0.862 (prior 0.469)

Table 4.3. – Results of the Bayesian analysis for the likelihood $x \sim \text{binomial}(\theta, n)$, where $x = 18$ is the observed number out of $n = 25$ who save the 5, and θ is the unknown probability of saving the 5. The prior distribution for θ is $\text{Beta}(\alpha, \beta)$ and posterior is $\text{Beta}(x+\alpha, n-x+\beta)$.

Table 4.3 sets out the results with the corresponding priors and posteriors shown in Figure 4.5. These all show that the posterior distribution for θ has the probability shifted to the left (in the case of the middle and strong priors) or more narrowly concentrated in the case of the uniform prior. For the middle case the posterior probability of being less than 0.9 is more than double that of the prior, and even in the strong case there is a notable (almost double) increase in this posterior probability compared to the prior. Figure 4.5 suggests that IVR implementations typically result in greater utilitarian behaviour (more likely to save the 5) than questionnaire responses, and the smallest proportion saving the 5 is 0.89. Our results suggest (but do not prove of course) that 0.89 is at the right hand boundary of the posterior distributions for θ in all three cases.

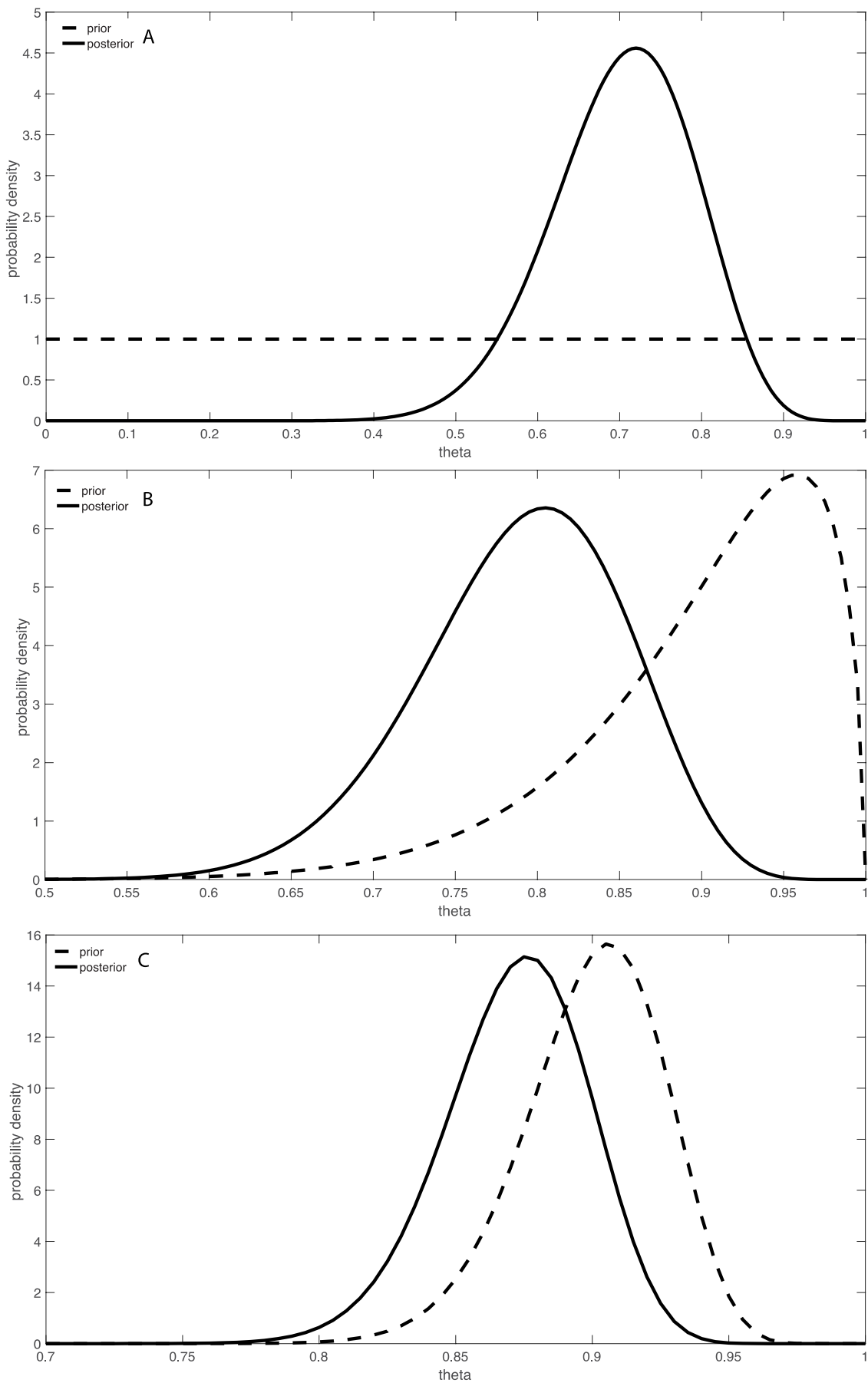


Figure 4.5. – Prior and posterior distributions for the three cases: (A) Uniform case, (B) Middle case and (C) Strong case.

4.4. Discussion

The model for the child was purchased on the Unity Asset Store as part of the “Sporty Girl” 3D model package. The model is supposed to represent a young girl (around 10), but could be seen as somewhat short teenager, undermining the point of the study (comparing the life of a vulnerable child to that of an adults). Participants were, however, given two opportunities after the virtual experience to provide comments back to the experimenter and during the first occasion were specifically asked “Did you understand what was happening in the scenario?” and later to confirm whether they saved the 5 adults or the 1 child. Based on the answers given by the participants, there was no confusion about the moral dilemma and no participants mentioned that they thought the dilemma was comparing the lives of 5 adults against that of 1 adult or 1 teenager.

Although previous studies looking at impersonal moral dilemmas in IVR (Navarrete et al., 2012; Pan & Slater, 2011) have shown little difference between moral judgment and action, the difference they do show is generally in the favour of utilitarianism, although one study (Patil et al., 2014) did actually find a significant difference in the favour utilitarianism. This matches up with other studies studying personal moral dilemmas that find an increased tendency to act in a utilitarian manner in IVR (Francis et al., 2016; Francis et al., 2017; McDonald et al., 2017).

Similar to Pan and Slater (2011) and Navarrete et al. (2012), our results show little difference between judgment and action, however, the difference present does not favour utilitarianism (72% saved the five adults in IVR compared to 80% in the questionnaire). A possible explanation for the difference between our results and those found in the other studies might simply be because the state of our moral dilemma is different to those used in the previous studies, most of which based their dilemmas on the Switch Dilemma. In particular our moral dilemma differs from the Switch Dilemma in its use of actions and omissions. In the earlier studies, the participants only needed to act to achieve the utilitarian outcome, whereas in ours, the participants must act to achieve either option. Therefore, if it was the case that in the earlier experiments, being in immersive virtual reality increased people’s tendency to act, thus increasing utilitarian responses in those studies, this could differently effect participant’s judgments in the present study.

However, there is uncertainty about attributing this to the differences in responses between this and the earlier studies. There is a drawback with the Switch Dilemma itself in that it puts

utilitarianism, deontology and action tendencies at two ends of single scale, with utilitarianism and action tendency at one end and deontology at the other (Gawronski et al., 2016). When someone chooses to take an action that would kill the one to save the five, they are at the same time conforming to utilitarianism, opposing deontologically and choosing to act instead of omit. What this means is that if there is a change in moral judgment towards the Switch Dilemma between two conditions (e.g., between immersive virtual reality and a questionnaire), it is difficult to know whether this was because of a change in attitude towards utilitarianism, deontology or their tendency to act. The previous studies observed that people acted in an increasingly utilitarian manner in immersive virtual reality compared to a questionnaire, but is this because immersive virtual reality increased people's following of utilitarianism, decreased their following of deontology or increased their tendency to act? If it is the case that immersive virtual reality decreased their attitude towards deontology or increased their tendency to act, then this might explain the change in results between our experiments as these are the qualities that have changed between our dilemmas. However, if it was a change in utilitarian attitude, then we would need to look for alternative explanations as the utilitarian state of our moral dilemmas was the same (acting to save five by killing one) thus any change between conditions in utilitarianism would have affected our results equally as the previous ones. Unfortunately, it is impossible to know what has been affected due to the construction of the Switch Dilemma.

Another result, as shown in Figures 4.2 to 4.4, was that those in the Dilemma condition tended to record a lower or equal level of presence compared to participants in the No Dilemma condition. According to the theory (Slater, 2009), since the sensorimotor contingencies were equal in both conditions, the level of "being there" (PI) should also have been the same. The difference in level of plausibility illusion (Psi) is less surprising since this could be because the situation itself is not a plausible one, either conceptually or from the way that it is implemented. Having a fire start the elevators would be a very unlikely occurrence which could cause people to lose the illusion that the event was really happening, or the graphical quality of the fire and smoke or the quality of the voice acting might also have affected the presence scores of the people in the dilemma condition, but not in the no dilemma condition.

Another explanation, supported by participant feedback, might be that due to the several warnings received before the virtual scenario regarding the stressful or upsetting nature of the experiment, those in the dilemma condition who subsequently saw the moral dilemma, were expecting it to

happen and were reminded that they were part of the experiment when they saw it whereas the participants who did not see the moral dilemma were never reminded of this and were possibly paying more attention to the virtual room in the anticipation of a stressful event that never came.

Before participants experienced the virtual scenario, they were presented with several different sources informing them that what they will experience may be stressful. This came in the form of a warning in the initial recruitment email and in the information sheet and the completion a brief trauma questionnaire. They were also asked several questions about children in the consent form and demographic questionnaire and were also reminded several times that they were free to stop at any time without consequence. These messages were given to participants in both conditions in an attempt to keep them as similar as possible.

To reiterate, the difference between the conditions happens at the end of the virtual scenario when the child and group of people walk into the elevators. For the Dilemma condition, the participants saw and had to respond to the moral dilemma whereas the scenario in the No Dilemma condition finished as soon as the elevators closed and no dilemma was witnessed. The difference then might be that because the participants in the Dilemma condition actually saw the event that they were warned about, although they were never told exactly what was going to happen, it reminded them that this was simply all part of an experiment. The No Dilemma condition participants on the other hand never witnessed a moral dilemma, despite still receiving the same warnings. Due to this, it is possible that they grounded themselves much more in the virtual room in anticipation of the stressful event.

It would be important in the future to see what another study can show us by adding a third condition where there are no warnings and no stressful event. Does having warnings and no stressful event raise feelings of presence or does having warnings and then experiencing the subsequent stressful event lower feelings of presence?

4.5. Conclusion

The fundamental result of this study is that the evidence suggests that participants were similarly likely to save the child or five adults in both questionnaire and immersive virtual reality versions

of the moral dilemma. This is congruent with previous studies investigating impersonal moral dilemmas in immersive virtual reality. We also found the surprising result that participants who saw a moral dilemma in immersive virtual reality compared to those who saw a mundane scenario reported lower presence scores. This is suspected to be due to ethical warnings presented before the experiment suggesting what was going to happen, thus reminding the Dilemma participants that they were in an experiment, and also the unlikely nature of the event reducing plausibility. Further studies should be aimed at this area to see if this is truly the case.

Chapter 5

Social Influence and Omission Bias on Moral Behaviour

The main focus of this study was to investigate moral behaviour in an interpersonal setting. The goal was to determine whether social influence, specifically compliance attempts, are effective during moral decision-making by having participants respond to a virtual moral dilemma either while hearing the opinions of other virtual characters or responding to the situation on their own. Another condition was included which alternated the action/omission state of the moral dilemma, allowing us to see whether it is easier to influence someone to act or omit during a moral dilemma.

The conclusion of section 2.2.6, as well as the results from the first study, both call for more information regarding the difference between moral judgment and behaviour, specifically, how the saliency of actions and outcomes affects people's moral decision-making. As such, this study also includes a questionnaire replication of the virtual moral dilemma, minus the social influence conditions, i.e., keeping the action/omission conditions. By comparing the results from people's moral behaviour in immersive virtual reality (IVR) and moral judgment from the questionnaire in situations in which participants may achieve outcomes through either action or omission, we hope to see whether this – the option to act or omit – is perceived differently between judgment and behaviour exercises. For context, previous hypotheses have placed greater emphasis on the outcomes of moral dilemmas in IVR and less so on actions in the difference between peoples moral judgments and behaviours.

IVR was used again to overcome ethical issues present in placing people in a moral dilemma, while still retaining life-like responses, simulating the moral dilemma as well as the virtual

characters attempting to persuade the participant and the actions the participant must take. From the results, it was found that people were very much unaffected by the virtual group, acting similarly whether they were or were not under the duress of social influence. However, we did find the surprising effect that people preferred to take an action to resolve the situation, rather than to do so through omission. Although this is unexpected based on numerous previous studies showing a preference for inaction during moral dilemmas (Cushman et al., 2006), this is congruent with the possibility that people have a greater preference to act during a moral dilemma in IVR than when responding to the same dilemma on a questionnaire (possibly accounting for the increased utilitarian responses seen in previous studies) (Francis et al., 2017), but is inconclusive without moral judgment data to compare it to. In order to investigate this further then, we ran a duplicate of the study in an online survey and found that this preference for action did not exist, unlike in the IVR version of the study.

5.1. Introduction

Through the effective methodology of the trolley problem, many fascinating aspects of our moral judgments have become illuminated, such as the first-hand witness of the use of specific motivations, biases and heuristics that guide our moral judgments (Cushman et al., 2006; Hauser et al., 2007), the neural mechanisms that underlie them (Greene et al., 2004; Greene et al., 2001) and how these have been formed into normative ethical theories like utilitarianism and deontology (Greene, 2008, 2015). However, there are other ongoing debates regarding moral decision-making. One of these is the interpersonal nature of moral dilemmas and another is the translation of moral judgment into moral behaviour.

The vast majority of papers referring to the trolley problem examine people's responses to it in isolation (Ellemers et al., 2013). While this is an obvious and necessary first step, real-life moral judgments are likely seldom made by a single person (Aramovich et al., 2012; Fiske et al., 2004; Hersh & Friedman, 2004; Warren & Smith-Crowe, 2008) and thus the effects of social influences such as conformity or compliance during situations involving harm are key to understanding exactly how they would be tackled in more tangible circumstances. One recent example of this is

the resistance against vaccinations which, without the influence of other people, especially on social media platforms, is unlikely to have become such an issue.

As described in Chapter 2, some research has already started to scratch the surface on this area of moral judgment. Kundu and Cummins (2013) and Lisciandra et al. (2013) both examined the effects on group conformity on a participants moral judgment and found that participants judgments indeed tended towards the group's norm, although Lisciandra and colleagues note that moral norms were harder to influence than both social or decency norms. Similarly, Bostyn and Roets (2017) and Kelly et al. (2017) presented solitary participants with a series of moral dilemmas each accompanied with aggregate information about how other people had responded to that specific dilemma and again, were successful in manipulating the moral judgments of their participants, even without the actual presence of another human-being.

Even more directly related to our real-life moral decisions is the translation of moral judgment into moral behaviour. Judgments are confined to the mind and are forced to make assumptions about how we might actually act in a specific scenario. The lack of auxiliary contextual information in written moral dilemmas compared to real-life ones can make these assumptions very inaccurate, leading to differences between what we say would do and what we actually do. These contextual cues come in the form of sensory stimulus, such as visual (Amit & Greene, 2012) and auditory, which can provide important social context to the scene (Parsons, 2015), subjective reasoning (Tassy, Oullier, et al., 2013), motor neuron activations (Cushman et al., 2012) and self-control (FeldmanHall et al., 2012).

Previous studies have used IVR to study the relationship between moral judgments and behaviours, with the most striking result showing that people opted to push the man in the footbridge dilemma only 10% in a written vignette, comparable to results found by Hauser et al. (2007), but 56% of the time in a IVR replication (Francis et al., 2017). This result is also not an anomaly, being supported by several other similar studies (Francis et al., 2016; McDonald et al., 2017; Patil et al., 2014).

This study also considers how both the areas of social morality and moral judgments versus behaviours interacts with omission bias, a previously well explored aspect of moral decision-making (Baron & Ritov, 2004; Cushman et al., 2006; Ritov & Baron, 1990; Spranca et al., 1991). Omission bias is the "tendency to favour omissions over otherwise equivalent commissions"

(Ritov & Baron, 1990). This preference can be seen in real-life where there is still a moral distinction between active and passive euthanasia in some parts of the world, in the reluctance to vaccinate and in our legal system which is far more punishing to those who harm versus those who do not help. The distinction continues to our responses to hypothetical moral dilemmas where it is seen as morally worse to cause harm through purposeful action even when the overall gains are positive (Thomson, 1985). For example, Cushman et al. (2006) found evidence for the use of the omission bias in people's evaluation of the switch and footbridge dilemmas, Ritov and Baron (1990) found that participants were reluctant to promote the vaccination of a child knowing the vaccine had a chance to be fatal, despite being told that the child had a higher chance of dying without the vaccine and Spranca et al. (1991) found that, in a series of studies where the intentions and knowledge of actors as well as the outcome of events were known and held constant, participants still regarded omission as the more morally sound option.

Omission bias arises from an overgeneralization of a moral heuristic in causing us to perceive actions that cause harm as morally worse than omissions (Sunstein, 2005). This is because omissions are more likely to be the results of ignorance and less likely to be the results of malicious intentions, where the opposite can be said for actions. However, this distinction breaks down when knowledge and intentions are held constant for both actions and omissions, yet the bias apparently remains. In the case of euthanasia, the intentions of the doctor are always at the will of the patient, either from asking the doctor to administer a lethal drug or from orders not to resuscitate, but the administering of a lethal drug is usually met with more severe moral discomfort. In another example, participants responding to a moral dilemma will often justify their choice of inaction by stating that, had they been absent or ignorant, the negative outcome still would have happened, despite the fact they were not absent and not ignorant to the events going on (Ritov & Baron, 1990).

During this study, participants took part in an IVR scenario where they were required to respond to a moral dilemma. The dilemma involved the choice to either immediately drive a man who has just suffered a stroke to hospital or to stop the car at the behest of a police officer, potentially delaying medical assistance, but upon not doing so would break the law.

The scenario presented was modified depending on which condition the participant was appointed, where there was a total of four conditions, made up of two two-part factors. The first

factor altered whether or not the participant would receive pressure from a group of virtual characters regarding how they should respond to the moral dilemma and the second factor altered the way in which participants could respond to the moral dilemma, either allowing them to take purposeful, physical action to continue to the hospital, otherwise stopping through omission or vice versa, i.e., taking action to stop or omitting to continue.

In a second study, the same moral dilemma was instead described in an online questionnaire which students at the university were encouraged to respond to. Only two conditions were employed in this study being equal to those from the action/omission factor from the first study.

Combined, the goal of both studies is first to investigate the effects of social influence on moral behaviour by using IVR technology, where previous studies only required participants to make abstract moral judgments. Typically, there has been evidence that moral behaviour is at least partially distinct from moral judgment, however it is unclear how this difference relates to social influence and whether moral behaviour is more or less likely to be effected by the opinions of others. One supposition is that it might be the case that people are more likely to lie during moral judgment, compared to moral behaviour, as their opinion does not have any tangible repercussions, allowing them to publicly adhere to social influence, while still privately ignoring it. This is less likely to occur in a person's moral behaviour however as they would be more sensitive to the outcomes of that decision, making them prefer their own opinion over others and thus make social influence less effective. Support for this hypothesis comes from previous studies where it is proposed that the cause of the difference between moral judgments and behaviours is due to the saliency of outcomes in IVR compared to in written vignettes (Patil et al., 2014). As such, our first prediction is that social influence will be less effective in this study compared to earlier studies looking at this effect during moral dilemmas, although we doubt social influence will fail completely.

Second, we aim to study how omission bias interacts with social influence, for example, whether it is easier to influence someone to take action or to omit and predict that, due to the more passive nature of omissions, influence will be greater in these cases.

Finally, by the comparison of results from the two conducted studies, we hope to illuminate whether omission bias plays a larger role in either moral judgment or behaviour. Previous studies have shown that responders are sensitive to omission bias when responding to written versions of moral dilemmas (Ritov & Baron, 1990) so we would expect to see participants in the

questionnaire condition choose to omit more than act. However, moral behavioural studies have placed greater emphasis on the outcomes of a moral dilemma in determining how a person responds to that dilemma (Francis et al., 2017), so the effect of omission bias in IVR might be weaker.

5.2. Study One – Methods

5.2.1. Scenario

The IVR scene depicted the participant sat in an S7 Audi autonomous car, parked on the road in an unnamed city. They had a sex-matched virtual body that they would see if they looked down towards themselves as well as in the rear-view mirror. Next to the participant, in the front passenger seat, was a virtual character. After a few seconds, the virtual character would introduce himself and the car would start to drive around the city autonomously. As the car drove around the city, the virtual character would continue to talk to the participant, engaging in trivial conversations such as what the participant does for a living or what they thought about the autonomous car.

After about two minutes of conversation, the virtual character starts to show symptoms of a stroke (exclaiming he has a headache, acting confused, having half his face has fallen), before losing consciousness. A few seconds later the car indicates that it will start driving towards the nearest hospital. As the car drives to the hospital, a police car starts driving behind, indicating that it wants the autonomous car to pull over. The car then gives the participant the option to either: stop the car for the police, potentially increasing the time it takes to get to the hospital or continuing to the hospital, breaking the law in the process.

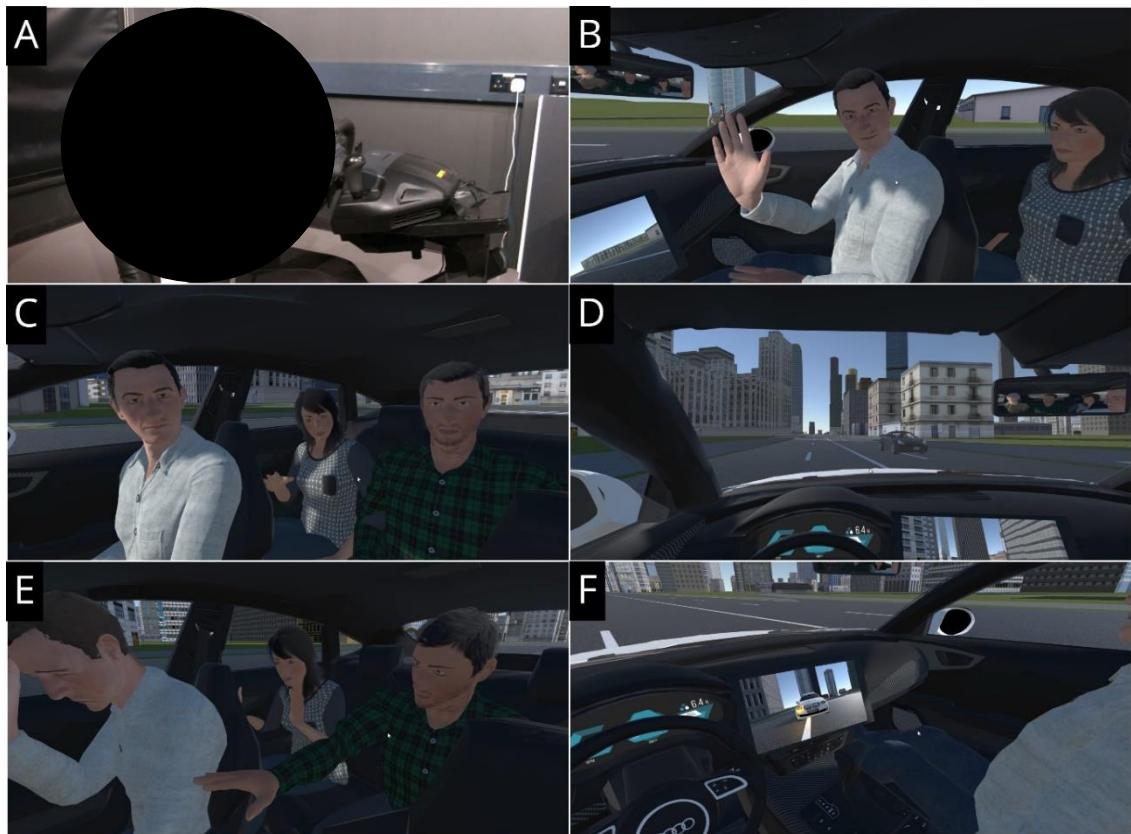


Figure 5.1. – The scenario, A) Participant with equipment, B) Characters introducing themselves to the participant, C) Characters engaging with the participant in conversation, D) Autonomous car driving participants and other characters around the unnamed city, E) Front seat passenger suffers a stroke, other passengers show concern, F) The autonomous car has been alerted and is automatically going to the nearest hospital, police attempt to stop the car.

5.2.2. Experimental Design

There were four conditions in a 2 by 2 factorial pattern. The first factor (Influence) altered whether or not there were three additional virtual characters sat in the back of the car. These virtual characters would join the conversation as the car drove around the city and would encourage the participant to ignore the police and continue to the hospital during the dilemma. The decision to have the virtual characters encourage participants to keep going instead of to stop was based on a series of pilot studies showing that the majority of people stopped for police when left alone. The second factor (Action) altered whether the participants had to take an action to stop the car (otherwise it would keep moving) or to keep the car moving (otherwise it would stop).

	<u>Influence</u>	
<u>Action</u>	No_group	Group
Action_stop	Condition 1	Condition 2
Action_continue	Condition 3	Condition 4

Table 5.1. – Study two conditions.

5.2.3. Participants

Sixty participants took part in the study with 15 in each condition in a between-groups design. Participants were recruited from a variety of Departments and Faculties throughout UCL with only students being permitted to take part in order to simplify the conversation with the virtual characters and increase the chance of reference group status. The mean age of the participants was 24 ± 4.8 (SD) years. Only male participants were used in the study to reduce the chance of any confounding variables on the results.

The experiment was approved by the UCL Research Ethics Committee, and was carried out with written informed consent from each participant. Participants were also verbally reminded at several points throughout the experimental training procedure that they would be free to stop at any time without having to give reasons or lose any other rights to which they were entitled. They were paid 7 pounds (UK).

5.2.4. Equipment

The scenario was created using the Unity 5 game engine. The 3D models in the scenario were re-used from a previous Masters students final year project. The virtual characters were modelled and rigged using Adobe Fuse (Beta) and animated during the scenario using a custom animation system for body movements and hand gestures. The LipSync Pro Unity package was also used to enable mouth movements and facial expressions.

For this experiment, participants were sat on a *Playseat DiRT 3 Edition* stationary gaming chair, registered with the corresponding virtual car seat. A gaming steering wheel was also attached to

the chair so participants could reach out and feel the virtual car's steering wheel. The head-mounted display used was the HTC Vive which has a resolution of 1080x1200 per eye, a refresh rate of 90Hz and a 110-degree field of view. The participants were asked to always remain seated and to keep their feet together. This was to simplify the implementation of motion capture. The HTC Vive lighthouse tracking system was used to track the position and rotation of the participant's head. The virtual body of the participant was interpolated at runtime using the Final IK Unity package to adjust to the height and movement of the participant, achieving a sense that it was the participants own body. The only other part of the participant's body that was tracked were the hands. This was accomplished by placing OptiTrack rigid bodies on the backs of participants' hands and tracking these using 12 OptiTrack Flex 3 cameras that were connected to a computer running OptiTracks Motive software. Information from this computer was then sent over a local area network to the computer running Unity 5 using the NatNet 2.9 SDK. Elbow movements were also calculated using the Final IK Unity package.

5.2.5. Procedures

Two weeks before arriving at the lab, participants were emailed a pack of information and a set of surveys to complete (see appendix B). The only important information in the pack was about the danger of strokes and about the law regarding ignoring a police officer while in a vehicle. We included other information about healthcare, artificial intelligences, highway codes and hate crime statistics so as not to give away the exact nature of the study. Doing this ensured that all participants were equally aware of the urgency of bringing medical attention to someone who is suffering a stroke and the seriousness of not complying with a request from a police officer while driving. The questionnaires included the Mehrabian Conformity Scale (Mehrabian & Stefl, 1995), a survey which gauges how likely someone is to conform to others and a brief trauma questionnaire used to confirm if the participant was psychologically able to take part in the experiment.

Participants then arrived at the lab individually at a pre-booked time slot. They were not told beforehand that the study was about moral dilemmas, however they were informed that they may (or may not) find the situation depicted in IVR stressful or upsetting and of course that they were

free to continue or leave as they preferred. They were given an information sheet detailing the experiment, prompted by the experimenter whether they had any questions and then given the consent form to sign if they wanted to continue. Finally, a demographic questionnaire was given to them before entering the virtual environment (VE).

No actions were required in the VE until the moral dilemma, at which time the autonomous car would relay to the participants their options. As such, no instructions were given to the participant before entering the VE, except that they would be in an autonomous car driving around a city. After this, participants sat in the chair, put on the IVR equipment and experienced a short car journey through a different part of the virtual city to accustom them to the experience of IVR. Once this had finished, participants were asked one more whether they were happy to continue. If they were, the true scenario was ran.

When in the VE, participants would engage in a 2 minute conversation with either the virtual character sat next to them (the one who would eventually have a stroke) or this character as well as three more in the back of the car (conditions 2 and 4). The conversation was constructed such that the front seat character would say the same lines at the same time in all conditions while still maintaining a realistic conversation. The conversation included the other passengers introducing themselves to the participant and proceeding to discuss various things such as the university, the weather and the autonomous car. The virtual characters would always start the topic at hand and then wait for a response from the participant before giving a generic reply or follow up question, this pattern made conversations seem real without having to program in a large amount of conversational combinations.

After the conversation, the moral dilemma described earlier would ensue. This would be identical for all participants, however, those in conditions 2 and 4 would also hear the three backseat passengers encourage them to ignore the police and continue to the hospital.

Once the participant had made a decision the VE would fade to black. Once the IVR equipment had been removed, participants filled in a presence questionnaire, were paid and debriefed on the nature of the experiment.

5.2.6. Response variables

The response variable for each condition was the frequency with which people stopped for the police. Specifically, we were looking at whether the pressure from a group on a public decision would have any effect in a moral setting. Additionally, we looked at the frequency at which people took an action compared to an omission. In conditions 1 and 2, this lined up with stopping for the police, but had the opposite effect in conditions 3 and 4. Thus, while these response variables are related, they are not identical and were analysed separately. This response variable is important in order to determine if people perceived a difference between a response based on whether they must take an action to follow through with it or can accomplish it through omission.

5.3. Study One – Results

Using R, a power analysis was performed to confirm if the minimum number of participants per condition resulted in a sufficient statistical power. The analysis yielded a power of 0.68 from an effect size of 0.15 and a significance level of 0.05.

5.3.1. Presence Data

The following results show the responses of participants to the presence questionnaire. This questionnaire is based on the one developed by Slater et al. (1995) and targets participants feelings of place illusion (PI) – the feeling of a being in a place other than the one in which the participant is actually situation – and plausibility illusion (Psi) – the feeling that events happening around the participant in the virtual world are really happening. The full questionnaire along with the encoded values (e.g., *there*, *real*, *visited*, *lab* and *overwhelm*) shown in Figures 5.2 through 5.4 can be found in Tables 3.3 to 3.5.

Figure 5.2 shows the responses to the presence questions in Table 3.3. It can be seen that the scores are overall high, except for the control question *lab*, which is suitably low, and there are no systematic differences between the conditions.

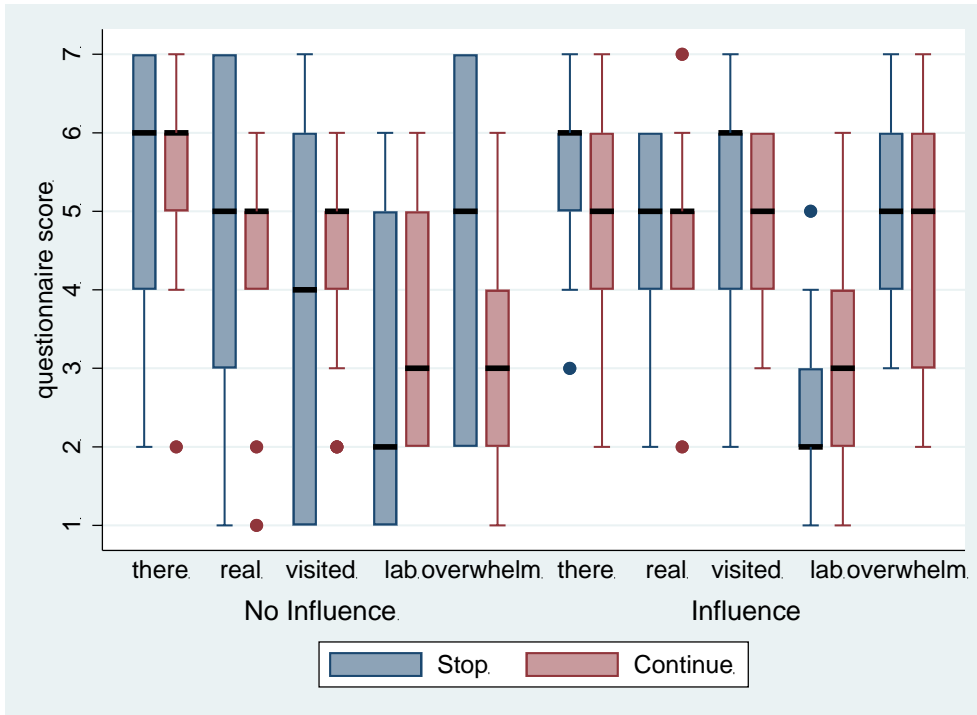


Figure 5.2. – Box plots for presence questions.

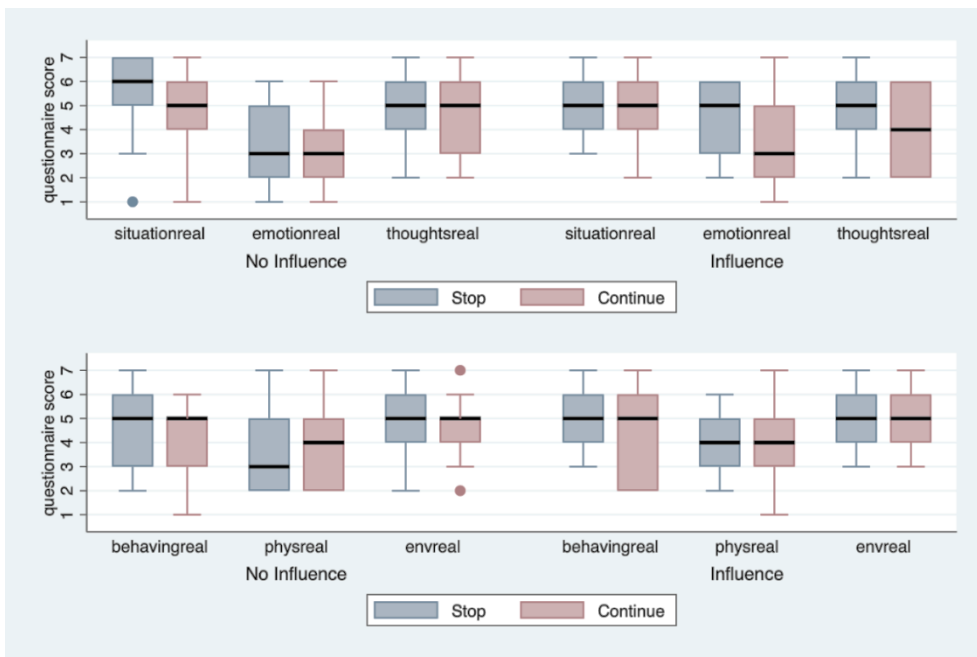


Figure 5.3. – Box plots for plausibility of the situation.

Figure 5.3 shows the box plots for plausibility of the situation questions in Table 3.4. Generally the scores are high except for *emotionreal* which is much lower for the No Influence condition, and lower also for the Influence*Continue condition.

Finally, Figure 5.4 shows the box plots for the questions regarding the plausibility of the virtual people in Table 3.5. On the whole the scores are low, with the exception of *behavepeople* in the Stop*No Influence and most in the Stop*Influence condition. In other words the people and their behaviour were considered as most plausible in these conditions.

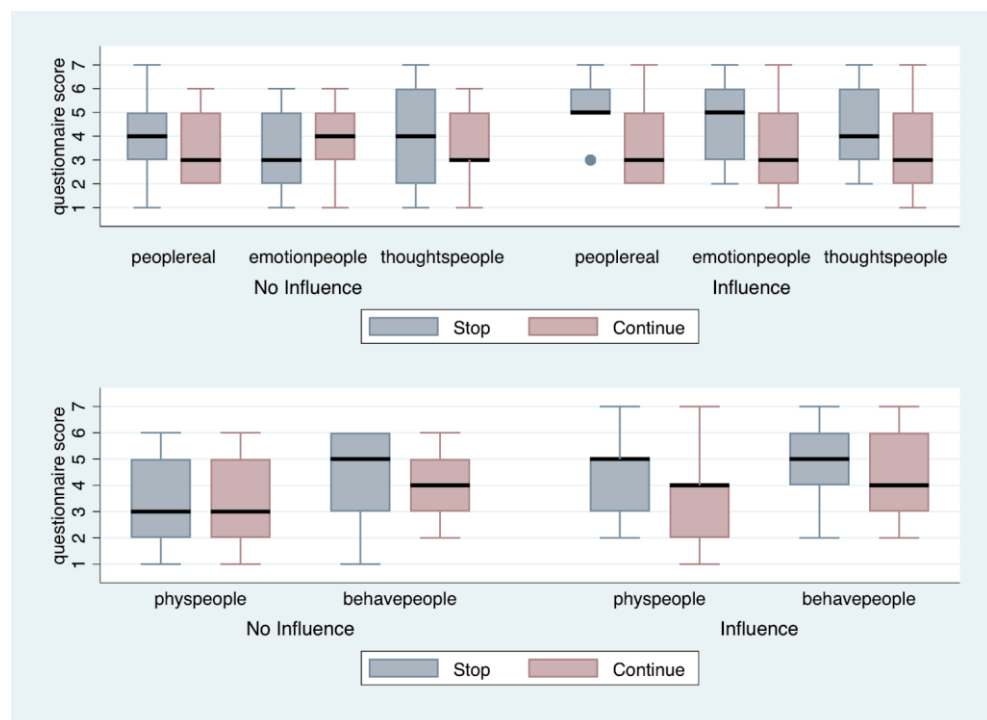


Figure 5.4. – Box plots for plausibility of the virtual people.

5.3.2. Behavioural Data

Figure 5.5 shows the proportions who acted under the conditions. It is clear that Influence had no effect, but there is a difference between Stop and Continue. Binary logistic regression of acted on Influence + Action + Influence×Action shows that the interaction term Influence×Action is not significant ($P = 0.379$) but that the Action is significant ($P = 0.03$). The main effect of Influence is

not significant ($P = 0.304$) – results in line with Figure 5.4. Eliminating the interaction term Action is significant ($P = 0.007$), but Influence is not ($P = 0.546$).

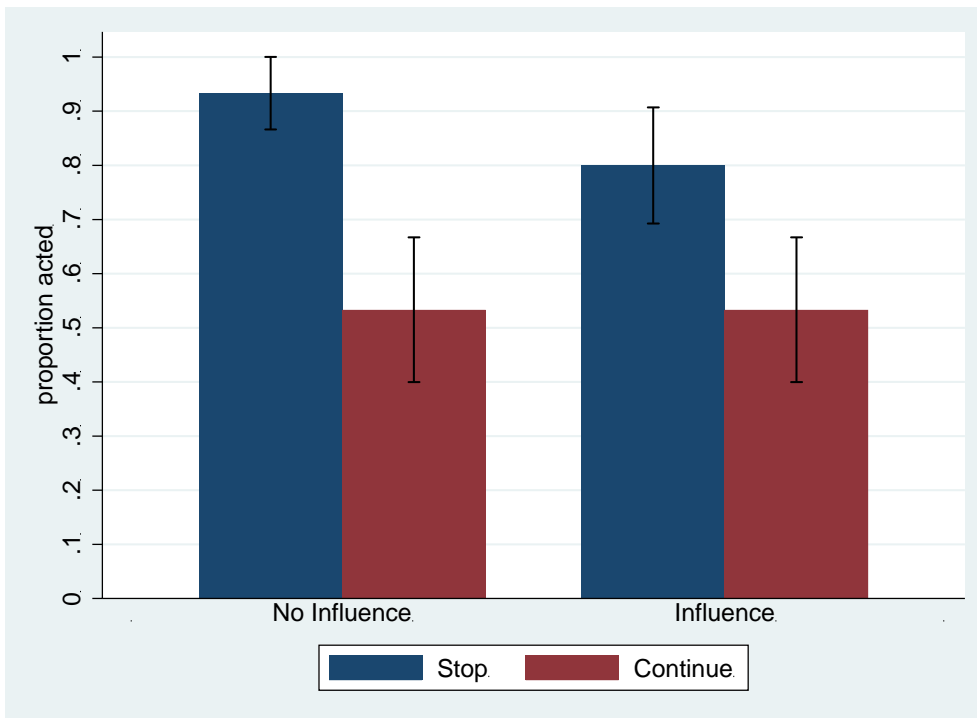


Figure 5.5. – Bar chart showing the proportions who acted by condition.

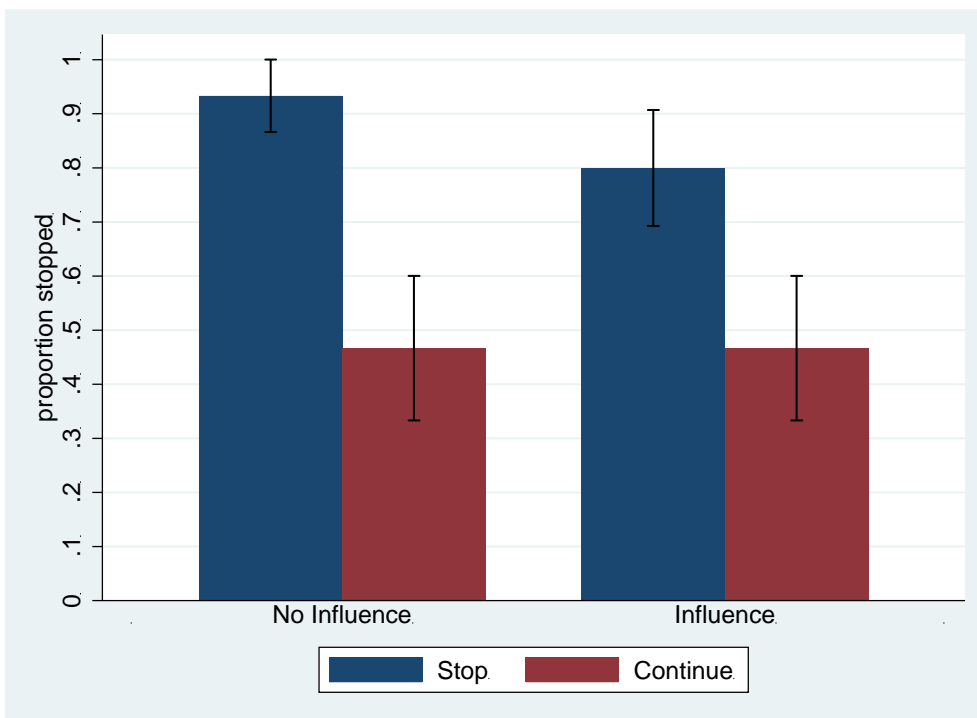


Figure 5.6. – Bar chart showing the proportions who stopped by condition.

The results for stopped (Figure 5.6) are almost identical to acted. The logistic regression shows that in the full model the interaction term is not significant ($P = 0.379$) but the main effect of Action is significant ($P = 0.017$). The main effect of Influence is not significant ($P = 0.304$). Eliminating the interaction term the main effect for Action is $P = 0.002$, and for Influence $P = 0.546$. There is no influence of confidence based on the Mehrabian Conformity Scale (Mehrabian & Stefl, 1995) on either of these two response variables.

The response variable time is only relevant when participants acted, and there are 42 such cases. Figure 5.7 shows the means and standard errors for time for these 42 cases. Waiting times do not typically follow a normal distribution, but rather an exponential or Gamma distribution. Using a generalised linear model with Gamma distribution and log link on the model Influence + Action + Influence \times Action we find no significant interaction effect ($P = 0.22$) but the main effect for Action has $P = 0.086$. However, on eliminating the interaction term nothing remains significant.

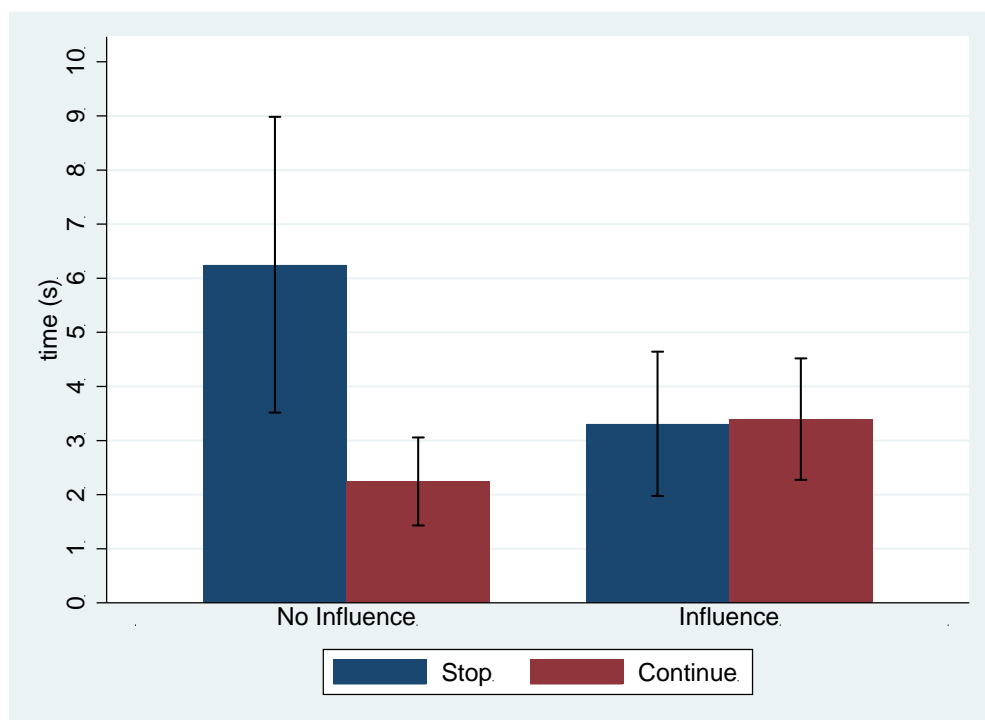


Figure 5.7. – Bar chart showing means and standard errors for the time ($n=42$ cases where participants acted).

5.4. Study Two – Methods

Study 2 was a replication of study 1, but was presented in the form of a survey instead of a IVR experience. The Influence factor was also removed, focusing solely on the Action factor. In other words, there were only two conditions, Action_stop in which participants were given the option to stop the car by using an action or continue through omission and Action_continue in which participants were given the option to stop the car through omission or continue by using an action. By contrasting these results with those regarding omission bias obtained in study 1, we hope to be able to see if preference for action or omission is altered between mediums.

5.4.1 Experimental Design

There were only two conditions in study 2 referred to as Action_stop and Action_continue, mirroring the Action factor of study 1.

5.4.2. Participants

This research project was approved by the designated ethics officer in the Computer Science Department at UCL. 283 participants took part in the study with 146 in the Action_stop condition and 138 in the Action_continue condition in a between-groups design (see Table 5.2). Participants were recruited from a variety of Departments and Faculties through their respective University social media accounts.

Condition	Frequency	Percentage
Action_stop	138	49%
Action_continue	146	51%
Total	284	

Table 5.2. – Study two participant breakdown.

5.4.3. Procedures

For this study, participants were asked to fill out an online survey. In order to ensure an equal number of participants in both conditions, participants were first asked to enter the day of the month they were born on. If they were born on the 1st to 15th, they were assigned to the Action_stop condition, otherwise if they were born after the 16th, they were assigned to the Action_continue condition. After this they were shown the webpage with the moral dilemma question specific to their condition.

5.4.4. Response variables

For this study, we were looking at the frequency with which people say they would take an action compared to omission between the two conditions, similar to the Action condition of the first study.

5.5. Study Two – Results

Figure 5.8 suggests that there is little difference between the conditions with respect to the proportion who acted. Binary logistic regression confirms this ($P = 0.564$). The same is true for the proportion stopped (Figure 5.9). Significance level = 0.564. These results can also be seen in the following frequency tables.

Condition	Omitted	Acted	Total
Action_stop	69	69	138
Action_continue	78	68	146

Table 5.3. – Frequency of actions and omissions by condition.

Condition	Continued	Stopped	Total
Action_stop	69	69	138
Action_continue	68	78	146

Table 5.4. – Frequency of participants who stopped or continued by condition.

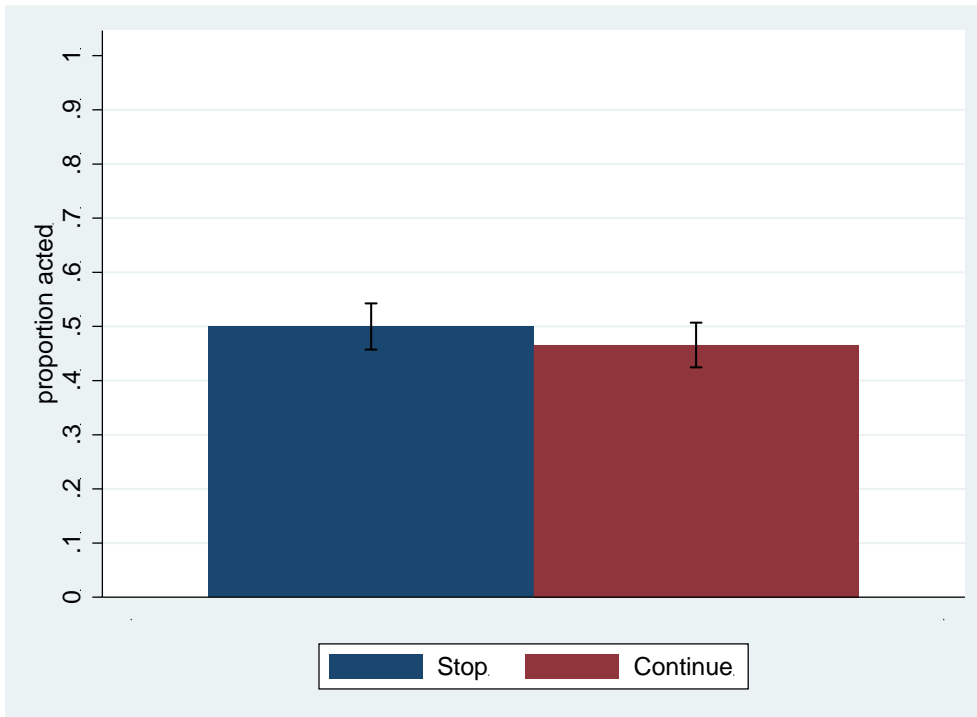


Figure 5.8. – Proportion acted by condition.

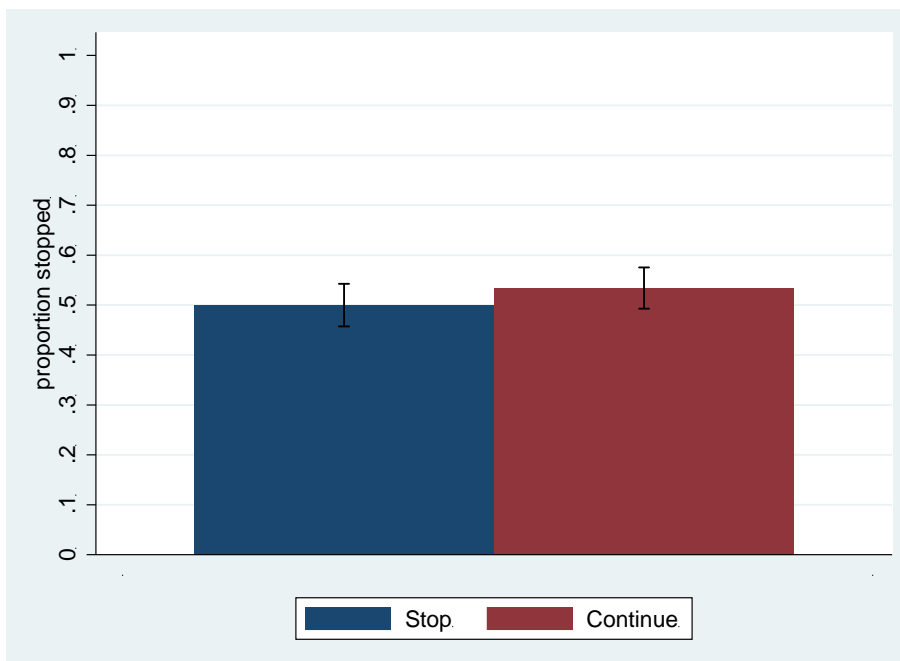


Figure 5.9. – Proportion stopped by condition.

5.6. General discussion

5.6.1. Social Influence

Interestingly, the results showed that there was no effect from the presence of the group with regards to whether or not participants stopped or continued to the hospital. This was not predicted and it is curious why so many studies in the past have shown how easily people can be influenced by small, trivial and artificial groups (Burger et al., 2001; Deutsch & Gerard, 1955; Dolinski et al., 2001), as well during previous moral influence studies (Bostyn & Roets, 2017; Kelly et al., 2017; Kundu & Cummins, 2013; Lisciandra et al., 2013), but this one has failed.

We did predict that social influence would be less, although at least somewhat, effective in this study than in previous ones due to people's ability to lie during moral judgments which is not available during moral behaviour. Specifically, it might be that moral behaviour is harder to influence than moral judgment as behaviour solidifies a person's opinion by outwardly expressing it whereas judgments can be publicly lied about. However, firstly we do not feel that this would completely negate the effects of social influence as it's unlikely that all previous responders were lying and second that this does not explain why social influence does not work during situations eliciting moral behaviour, but more so provides evidence as to why social influence is effective during situations requiring moral judgment.

With that being said, one explanation might be that the virtual characters (VCs) were not convincing enough. Blascovich et al. (2002) asserts that in order for social influence to work in immersive virtual reality as it would in real-life, at least one of two prerequisites must be met. VCs must either have high behavioural realism, high agency or a moderate amount of both. Behavioural realism refers to a VCs realistic attributes, i.e., the qualities of it that make it look like a human. This goes beyond, as well as includes, graphical realism, but also requires realistic animations, body movements, facial expressions and more. Agency on the other hand refers to the knowledge of the participants that they are either talking to a human- or computer-controlled VC, where human VCs are more likely to successfully elicit social influence, although this is not always the case (Swinth & Blascovich, 2001). Looking at Figure 5.4 which shows the responses

of participants when asked about the realism of the VCs, the results are somewhat low compared to other aspects of the virtual environment, staying around the 3 to 4 and sometimes 5 mark. While the scores are far from perfect, Blascovich et al.'s (2002) model presents a continuum and not a hard threshold for the effects of social influence and as such would predict the social influence attempts to at least have some effect where no effect has been found.

There are potentially other factors at play here. One theory we have is that the emotional response to the moral dilemma worked quicker and provided the participant with a sufficient enough answer to the moral dilemma before the group had a chance to influence the participant's decision. With an answer to the moral dilemma at hand, people would be much less likely to look to outside sources of information, such as what the group in the back of the car are saying, and would simply act on their own internal answer. This may work similarly to the way that people who have strong pre-existing beliefs about a certain topic are much harder to influence with regards to that topic than those who only have weak beliefs (Hornsey, Majkut, Terry, & McKimmie, 2003; Sagarin & Wood, 2007). Similarly, while Lisciandra et al. (2013) found that people could in fact be influenced by others during a moral judgment task, moral norms were much less malleable than other types of social norms, perhaps due to their importance to our self-concept.

This phenomenon has not been observed in the typical social influence research in the past as most studies in that domain focus on non-moral dilemmas (Ellemers et al., 2013). When responding to a non-moral dilemma, people will either think cognitively about the problem or use heuristics to obtain an answer without needing to exert too much cognitive effort, depending on their cognitive state (Chaiken, 1980). Heuristics are useful when we do not have much time to decide or when we lack the cognitive abilities to come to the correct conclusion, but a lot of these heuristics can be social-based, such as "experts' statements can be trusted" or "consensus opinions are correct". For example, when buying a new household item, we very often look at reviews from experts or check customer feedback to see what others are saying. Therefore, during a non-moral dilemma, people are likely to conform or comply with others about how they should act, at least in a situation that necessitates the use of heuristics over cognitive thinking.

In a moral setting on the other hand, people can have strong internal intuitions about how to act. These intuitions are our moral emotions, e.g., disgust, guilt or shame (Haidt, 2003), which work quickly and autonomously to provide us with answers to problems we have never faced before.

How these emotions work is still debated, but one prominent theory suggests that our emotional cognitive processes assign a value, e.g., a disgust value, to a set of moral choices in response to a moral dilemma. We then feel different levels of repulsion considering each one and choose the one which disgusts us the least, for example, killing one person rather than letting five die (Moll et al., 2008).

Merging these two theories together reveals that people are likely to already have an answer to a moral dilemma before any social influence can take place. As such, people's reliance on others will be greatly reduced and will be subsequently much less likely to use social heuristics to come to an answer. At the very least, this will reduce the efficacy of social influence by the simple inclusion of an opposing choice of response. This has been observed in other areas of social influence looking at how prior attitudes affect people's susceptibility to being influenced (Eagly & Chaiken, 1993). For example, Zuwerink and Devine (1996) found that when people held a strong belief about a certain topic, they were much less likely to be influenced by other people's opinions. They also found that increased emotions correlated with the pattern of influence resistance providing further evidence that emotions can interfere with social influence mechanisms.

As to why this has not been the case in previous moral influence studies (Bostyn & Roets, 2017; Kelly et al., 2017; Kundu & Cummins, 2013; Lisciandra et al., 2013), there are two differences between the current study on the ones that came prior. The first is the change from written questionnaire to life-like immersive virtual reality simulation. This has already been seen to dramatically change people's moral decision-making by showing an increase in utilitarian responses (Francis et al., 2017) and so other changes cannot be ruled out. Second is the type of social influence. The four previous studies focused on conformity, the urge to follow a groups norm, whereas the current study focused on compliance. The participants were told what to do by the VCs in the back of the car, although they could (and apparently did) ignore them. Previous research on persuasion, a type of social influence specific to compliance requests, shows a reduced effectiveness of compliance techniques when targeting a strongly held value of the responder (Sagarin & Wood, 2007), a description that can certainly be applied to a moral judgment.

5.6.2. Omission Bias

Looking at the Action_stop condition in Figure 5.6, a large majority of people opted to stop. This shows either (or both) a preference to act or to stop. By further looking at the Action_continue condition in the same figure, we can see that this preference to stop disappears, therefore, we cannot conclude that this (the preference to stop) can account for participants full reasoning, but also that they had a preference to act, causing them to continue more often in the Action_continue condition than in the Action_stop condition.

Looking beyond moral decision-making into the general domain there are several instances of action bias, the opposite to omission bias, which might explain our results. Lucas Jr and Taic (2015) say that “rather than focusing on the expected costs and benefits of their behaviour, people sometimes succumb to an irrational penchant for action... [because] in some instances, taking action simply seems normal.” They later say that “action bias also manifests itself in situations in which taking action facilitates claiming credit for a good outcome,” citing Patt and Zeckhauser (2000) who showed that subjects would rather spend money to clean up a resource than spend the same amount to prevent more pollution in the future. Baron and Ritov (2004) add to this, positing that action bias is likely the over-application of a heuristic biased against passivity and irresponsibility, eluding to the homunculus in our heads shouting “Don’t just sit there. Do something!” even when the situation calls for patience.

Zooming out on omission and action bias reveals a grander scheme, something Kahneman and Miller (1986) dubbed norm theory, the preference to do what seems normal whether through action or omission, or the aversion to harm caused by abnormal events compared to harm from normal (Baron & Ritov, 2004). Bar-Eli, Azar, Ritov, Keidar-Levin, and Schein (2007) note that “almost all the literature so far has studied cases in which people are biased in favour of inaction... because inaction is often the norm.” This is certainly the case in previous harm-based moral dilemmas where the action option is often to commit murder, where murder is certainly an abnormal event. In a non-moral context, Bar-Eli and colleagues studied the action effect in the behaviour of goal-keepers and noticed that, although the optimal strategy would be to stay in the centre of the goal, goal-keepers pre-emptively jumped to the left or right 93.7% of the time, most likely because this feels like the normal thing to do. Similarly, Landman (1987), Gleicher et al.

(1990) and Baron and Ritov (1994) all found that participants predicted being happier when winning an amount of money than by finding it by chance, showing a preference for action when results are positive, but also being more bitter when losing money from a bet than by misplacing it, showing a preference for omission when results are negative.

With this being said, perhaps participants perceived the normal course of events for our moral dilemma was to act, instead of omit, leading to a penchant for action instead of omission. While this seems counter to previous moral dilemmas, it can be explained by the use of proscriptive and prescriptive descriptions of morality (Janoff-Bulman, Sheikh, & Hepp, 2009). Proscriptive morality refers to the suppression of harmful behaviours that could lead to negative outcomes whereas prescriptive morality refers to the activation of helpful behaviours which aim to produce positive outcomes. While research into moral decision-making generally focuses on proscriptive morality where the outcome is harmful, and thus the norm is not to act, our moral dilemma involves the helping of a stroke victim and would then recruit prescriptive morality where the norm would be to take action, explaining our diverging results. This was not planned or predicted, but may in fact contribute somewhat to a small body of scientific knowledge, where Janoff-Bulman et al. (2009) notes a serious lack of research on prescriptive morality.

This perspective might also explain similar results obtained by Friedman et al. (2014). In this study, participants were presented with a moral dilemma in IVR similar to the switch dilemma, but were also given the option to press an alarm button. This was a big red button that, if immediately pressed, essentially did the same as taking no action at all. Despite this, the researchers found that 28 out of 30 participants chose to press the alarm button over the other two available options: killing one person in order to save five others or letting five people die. The probable reason that participants were more likely to act, over omit, in order to reach the same conclusion was likely due to the prescriptive, helpful appearance and thus normality of pressing the alarm button. If put in a dangerous situation and given a button called an "alarm button", the normal thing would certainly be to press it, especially if the other options involve doing nothing to help or actively killing someone. As said before, this result is likely the incorrect application of a neural heuristic urging participants to help, used necessarily due to the sudden nature of the situation. Further results showed that, when responding the same moral dilemma a second and third time, thus having more time to prepare and think, participants made better, more thought out decisions.

Another reason the norm might be to act in our moral dilemma is because of the prior events that happened before the moral dilemma took place. Zeelenberg, Van den Bos, Van Dijk, and Pieters (2002) posit that “when prior outcomes are negative, people may feel inclined to take action to improve future outcomes... [making] action more normal than inaction.” In our scenario then, the negative event of the passenger having a stroke would have made participants more eager to actively rectify the situation than to passively sit by and watch. In contrast, previous moral dilemmas such as the switch dilemma place participants in the situation without any prior events happening and so no precedent for action has been set. In another study, Haidt and Baron (1996) found that participants were more likely to take action when in a role of responsibility for others, something that might also be said for the participants of this study.

These claims go some ways towards explaining the behaviour of our participants in IVR. In a situation where someone is in need of help, it would certainly be the norm to help instead of idly sitting by and watching and nobody telling you otherwise is likely to change that opinion. At the same time, the chance to look like the good Samaritan might have also increased the chances that people wanted to take an action.

This sort of pattern of behaviour has been shown before (Uhlmann et al., 2013) where people are more inclined to do the more moral-looking option rather than the simply more moral option. Uhlmann and colleagues explain this through their concept of “act-person dissociation” where acts can be judged morally different to the people who commit them. Knowing this, participants that want to appear as morally good choose the option with the less ambiguous motive, which is in this case, always taking an action to help the other passenger.

5.6.3. Moral Judgment and Behaviour

Finally, we found that the preference for action (as well as the preference to stop) that was observed in the first study was not present when the same scenario was given as a questionnaire, providing evidence for the dissociation between people’s judgments and behaviours. As we showed in the results, people were more likely to choose the option that required an action when experiencing the moral dilemma in IVR regardless of whether this action kept the car moving or

stopped it for the police. However, during the survey we found that roughly 50% of people chose to take an action and as well as to omit and so the preference was not present. Given the previous discussion, this might not simply be a preference to act, but instead a preference to do what is normal when in IVR compared to a questionnaire. However, this conclusion does not line up with the results of previous studies which found a preference to push the man off of the bridge in the footbridge dilemma in IVR, which is almost certainly not a normal act (Francis et al., 2017). Given this, it seems more likely that it is indeed a preference for action.

These results provide some insight on data from previous experiments looking at how people respond to moral dilemmas in IVR compared to questionnaires (Francis et al., 2016; Francis et al., 2017; Patil et al., 2014). Each of these studies found that participants were significantly more likely to act in a utilitarian manner in IVR than those who responded to the same dilemmas in a questionnaire. As demonstrated in this study, people have at least some greater preference to choose an action over an omission in IVR compared to when answering the same dilemma in a questionnaire. This reveals a potentially contributing factor for the reason why people were more likely to act in a utilitarian manner in IVR as the utilitarian option during each of the moral dilemmas observed in the previous studies requires the user to take an action, whereas the deontological options could be achieved through omission. Some support for this theory comes from the previous chapter, which showed that this trend of increased utilitarianism does not appear when the user is required to act to achieve both outcomes.

This contributing factor however does not agree with hypotheses made by Patil et al. (2014) and Francis et al. (2016) as well as conclusions made in section 2.2.6 where it was said that “evidence points to the saliency of outcomes as the defining factor that increases utilitarian choices in reality and virtual reality... therefore, it is unlikely that it is the performance of pretend harmful actions that one performs in virtual reality that leads to increased utilitarianism.” The results from this study, that actions are preferentially taken in IVR, puts these assumptions in jeopardy. The type of response the participant had to take was clearly an important factor in their decision making process given that actions were clearly preferred and is supported by the fact that previous utilitarian options are confounded with taking an action which could explain their increased popularity. Finally, while the outcome was also clearly important to participants, as they also had a preference to stop, it was apparently just one part of their motivation when choosing how to respond to the situation.

5.7. Conclusion

In conclusion, we have shown several important findings during these two studies. Namely that social influence, specifically compliance attempts, might be much less effective during moral situations than non-moral. This is currently unclear however due to the lack of non-moral condition in this study so this will be the focus of the thesis moving forward.

We also found that people have a preference to act when responding to this particular moral dilemma in IVR and that the preference disappears when the scenario is presented on a questionnaire. Moving forward, these results and discussions could be further enlightened by utilising Conway and Gawronski's (2013) design for a moral dilemma where the utilitarian outcome and requirement to take an action are separated. This would allow the separation of a preference to act from a preference for utilitarianism and would make clear whether either or both of these inclinations are responsible for the diverging responses between moral judgment and behaviour tasks. Unfortunately due to time constraints, this task will have to be left to others and will not be present in this thesis.

Chapter 6

Social Influence During Moral and Non-Moral Dilemmas

The focus of this paper will be to compare the efficacy of social influence, specifically compliance attempts, during moral and non-moral dilemmas. The justification for this investigation is based on the results of the previous chapter which demonstrated the inability for compliance attempts to affect participant's moral behaviour. In order to determine if this inability stems from the setting as a morally charged scenario, the current study will compare this type of setting with one which is more akin to a scenario that might be employed in a traditional compliance study, i.e., an non-emotional, non-moral situation. The results found a weak trend in the predicted direction whereby participants were more likely to acquiesce to compliance attempts when in a non-moral, but not a moral, setting.

6.1. Introduction

6.1.1. Background

In order to understand the difference between moral and non-moral thinking, researchers have identified the neural mechanisms that support specifically moral thinking by using functional magnetic resonance imaging (fMRI) to scan participants brains while they respond to statements with both moral and non-moral connotations.

Moll et al. (2001) for example, had subjects silently judge a number of moral statements such as “They hung an innocent” and “Every human has the right to live” and non-moral statements like “Stones are made of water” and “Walking is good for your health.” Subjects would then either rate these sentences as ‘right’ or ‘wrong’ which are applicable in both moral and factual contexts. The results showed that certain regions activated specifically during moral judgment, with these including the frontopolar cortex (FPC) and medial frontal gyrus, both of which are located in the prefrontal cortex, as well as the right anterior temporal cortex. Another study by Schaich Borg et al. (2006) also found that the FPC and medial frontal gyrus were more active in moral scenarios than non-moral. The FPC has been implicated in the maintenance of long-term goals. This is especially important for moral behaviours as most are rewarded in the future at the detriment to a present self, e.g., donating an organ to save a sick family member. On the other hand, the medial frontal gyrus is specifically activated during self-referential thinking, which is also important when being put into a morally-charged life-changing situation.

In a follow-up study, Moll, de Oliveira-Souza, Bramati, et al. (2002) compared moral statements with emotionally aversive non-moral statements, e.g., “He licked the dirty toilet.” Moral statements tend to inherently elicit emotional cognitive processes and so any neural differences between reactions to moral and non-moral statements might possibly be due to the emotional reaction to the moral scenario. By also having the non-moral statements set in an emotional context, these cognitive processes can be identified and removed from the results. The researchers found that the medial orbitofrontal cortex (OFC) and superior temporal sulcus (STS) were preferentially activated only by moral statements whereas the amygdala and lateral OFC were activated only by the affect-laden non-moral statements. The medial OFC is a part of the brain responsible for understanding social rules and emotions during moral processing and is assumed to partially take over the role of an emotion processor during moral judgments whereas the amygdala does a similar job during non-moral contexts. Schaich Borg et al. (2006) and Greene et al. (2004) also found increased activation of the STS during moral contemplation, offering the explanation that the posterior STS is likely active during first-time thought-provoking moral judgments.

Finally, Schaich Borg et al. (2006) and Greene et al. (2001) found preferential activation of the right dorsolateral prefrontal cortex (dlPFC) during non-moral processing compared to moral. This is an area of the brain associated with high-level conscious deliberation and working memory.

Additionally, in the field of moral psychology, Cushman et al. (2012) found that participants scored higher on a TPR (Total Peripheral Resistance) measure when carrying out pretend immoral acts, such as stabbing someone with a knife, compared to pretend non-moral such as cutting a piece of bread. This measurement is associated with negative stress responses in participants, indicating a significant emotional response from those who were pretending to stab someone else.

All of these results suggests that the main (although not the sole) difference between moral and non-moral decision-making is related to our experience of emotion during moral situations.

As said several times previously, research surrounding social influence has almost exclusively targeted the process during non-moral settings (Ellemers et al., 2013), with only a few studies examining social conformity during moral judgment situations (Bostyn & Roets, 2017; Kelly et al., 2017; Kundu & Cummins, 2013; Lisciandra et al., 2013) and none directly comparing the two in a single study.

6.1.2. Present Research

With this being said, the current study then seeks to further understand the relationship between social influence, specifically compliance requests, and the receptivity of people to these attempts during both moral and non-moral social situations. In order to do this, similar to the two previous studies, immersive virtual reality (IVR) technology was used to immerse participants in a realistic scenario where they were presented with the following moral dilemma: You are standing in a building corridor in front of two elevators. The elevators are malfunctioning and you have been tasked with opening the doors for people who are getting in and out. You do this with the use a remote controller which you can use to point-and-click on the elevator door that you want to open. The doors are slow to open and close and only one can be opened at a time (all due to the current malfunction). After some time, a child enters the left elevator and a group of five adults the right. The doors close behind them. After a few seconds, a fire erupts in the elevator shaft and sets both elevators on fire. Due to the malfunction, you only have time to open one elevator before the other is consumed by the flames. Do you save the child or the five adults?

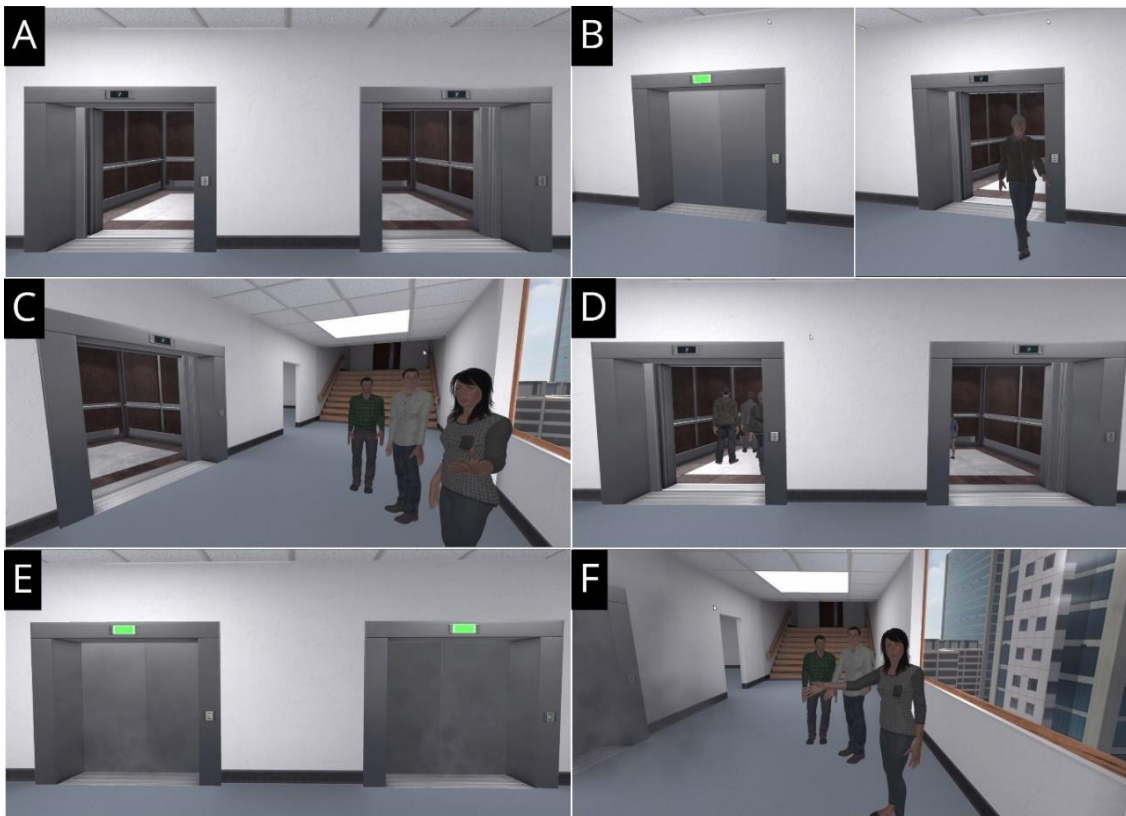


Figure 6.1. – The scenario, A) Scenario start, B) Participant had to open elevator doors for virtual characters signalled by the green light, C) Virtual characters would engage the participant in conversation, D) a group of five adults and a young child would enter the left and right elevators, E) The elevators would spew out smoke and fire could be heard from behind the doors, F) The virtual characters would encourage the participant to either save the child or the adults.

Six conditions (formed from a two by three factor design) were constructed that each altered the above scenario by a particular aspect. In the first factor, participants either dealt with the moral dilemma described above or with a similar non-moral dilemma where the child and adults requested to exit the elevators, but not because of any significant danger. At the same time, participants either responded to this dilemma on their own, with the inclusion of a group urging them to save the girl or with a group urging them to save the adults. The goal of this design was to be able look at the difference in the responses of participants between the social influence conditions in both moral and non-moral settings. If a difference exists between the effectiveness of social influence during moral and non-moral dilemmas, we would imagine to see responses of participants align more with the social influence attempts in one, but not both or neither of these situations.

6.1.3. Predictions

Specifically, our predication regarding these conditions is that social influence will be less effective during moral, compared to the non-moral, conditions.

Our hypothesis regarding this predictions builds both on models of moral judgment and of social influence. There is a plethora of evidence showing that social influence attempts are most effective when targeting our system one processes (refer back to section 2.2.4.2 for more information on cognitive processes), either because of the sudden nature of the request or because our mental faculties are already strained (Dolinski & Nawrat, 1998). For example, Blascovich et al. (2002) mentions the targeting of lower-level responses in their model of social influence in virtual environments as one of the four factors that social influence depends on in order to function properly in an immersive virtual reality scenario.

This happens because of our system one's reliance on the use of heuristics based on our internal motivations (Cialdini & Goldstein, 2004). One of these is the impression motivation which is defined as the "desire to hold attitudes and beliefs that will satisfy our current social goals." This causes us to adopt the opinions of those in our social in-groups and reject those from out-groups. However, in-groups can be formed in trivial ways (Burger et al., 2001; Dolinski et al., 2001), allowing social influence to be effective in arbitrary situations. Another motivation is our desire to have an accurate view of the world. This is hard to achieve at the rate at which our system one needs to provide us with answers, so shortcuts are used such as "length equals strength;" bigger things do tend be sturdier. More relevant to this conversation though is the heuristic "consensus equals correctness" which makes us believe that, if something is a majority opinion, it is more likely to be correct than a minority opinion.

Moral dilemmas also tend to activate intuitive cognitive processes, however, they primarily utilise the emotional parts of our brain, e.g., the ventromedial prefrontal cortex (vmPFC), in order to provide us with quick answers to demanding social situations, for example, compassion drives us to help others and anticipated guilt stops us from hurting others. This has been shown to be the case many times over in fMRI studies that scan that brains of participants while they respond to a battery of moral dilemmas (Greene et al., 2004; Greene et al., 2001; Koenigs et al., 2007) and

has been incorporated into several dual-process theories of moral judgment (Cushman, 2013; Greene et al., 2008).

Contrary to this, non-moral dilemmas have been consistently recorded to activate higher-level cognitive processes such as those associated with working memory and data manipulation like the dlPFC or inferior parietal lobe. Schaich Borg et al. (2006) also reported that subjects answered more quickly at the end of moral scenarios than non-moral scenarios. Later adding that “when lives are not at stake, considerations such as required effort or time are likely to carry more weight and to ultimately contribute to cognitive load.” In other words, it is easier to weigh the lives and deaths of two groups of people than something less consequential, leading to quicker decisions being made during moral situations compared to non-moral. Emotions provide us with gut reactions to moral dilemmas as a result of their extreme consequences, while non-moral dilemmas must rely somewhat on conscious reasoning in order to reach a judgment.

Two similarities can be identified between the cognitive processes elicited by moral dilemmas and those targeted by compliance requests. First is that they both involve, at least in some capacity, system one processes. However, just because this is so does not guarantee that they are fighting for the same neural pathways; as previously mentioned, social influence generally prompts the use of heuristics while moral dilemmas elicits the activation of emotions.

What is more certain, however, is the temporal relationship between the two processes. Both targeted neural processes are designed to provide answers extremely quickly, without any conscious input and as such could interfere with each other if competing resolutions are reached. This can happen anyway due to the nature of moral dilemmas, where emotional cognitive processes have been implicated in fighting for both restraint and action (Cushman, 2013). In these instances, either the vmPFC (Shenhav & Greene, 2014) or the anterior cingulate cortex (ACC) (Greene et al., 2004) are recruited to resolve the cognitive conflict in favour of one side.

If a similar sequence of a events happens when a compliance request is introduced then it is going to be inherently less effective than during a non-moral situation due to its competition against the emotions elicited by the moral dilemma. Furthermore, we often consider our morals to be crucially defining parts of who we are and thus people might interpret their strong moral-emotional reaction to the situation as an indication of its resolutions correctness, similar to the accuracy motivation’s availability heuristic (Sunstein, 2005), putting the victory of the compliance

request in further jeopardy. Related to this, Lisciandra et al. (2013) predicted and found that moral norms were the most resistant to peer-group judgments, although still somewhat malleable, concluding that “behaviour and opinions that involve violations of moral norms are more insulated from conformity effects.” This is supported by evidence that the medial frontal gyrus, a part of the brain related to self-referential processing, is more active during moral, rather than non-moral, judgments (Schaich Borg et al., 2006).

To summarise, it is hypothesized that the presence of intuitive emotional responses to moral dilemmas, which are not present during non-moral situations and which provide us with strong answers, are likely to reduce people’s reliance for answers from external sources and thus reduce the effectiveness of social influence attempts.

6.2. Methods

6.2.1. Experimental Design

The study took the form of a two by three factorial design consisting of the conditions presented in Table 6.1 which each alter the above scenario by a single variable.

	<u>Social</u>		
<u>Situation</u>	No Influence	Influence Child	Influence Adults
Moral Dilemma	Condition 1A	Condition 1B	Condition 1C
	25 participants	15 participants	15 participants
Nonmoral Dilemma	Condition 2A	Condition 2B	Condition 2C
	15 participants	15 participants	30 participants

Table 6.1. – Study three conditions.

The scenario described earlier matches the scenario presented to participants in condition 1A (it was a moral dilemma and there was no influence from bystanders directed towards the participant). The same events happened in condition 1B, except that there was a group of three

bystanders standing near the participant who urged him or her to save the child instead of the adult. Condition 1C is similar again, except that the bystanders were instead in favour of saving the five adults, not the child.

Conditions 2A, 2B and 2C are identical to their factor 1 counterparts except that the participants were instead presented with a non-moral dilemma instead of a moral one. This consisted of one change to the scenario which was to remove the fire. Thus the participants simply heard both the child and the five adults ask to be let out from the elevators, perhaps assuming that the malfunction has gotten even worse and the elevators were not moving between floors anymore. As the only change was the removal of the fire, it is important to note that the audio for the child and adults in the elevators was the same for both the moral and non-moral conditions. To accommodate both situations voice actors were instructed to sound urgent in their tone of voice, but not panicked or scared. This was done in order to reduce the variation between conditions that might have occurred from using different voice recordings.

In the conditions involving social influence, the participants also engaged in a two minute conversation with the other bystanders in order to build a rapport with them before any social influence attempts were made. This included introductions, talk about their classes at the university (all participants were UCL students), the weather and a career event that the virtual characters were planning to attend.

6.2.2. Participants

One hundred and fifteen participants took part in the study with a minimum of 15 in each condition in a between-groups design. Due to an error, 30 participants were ran for condition 2C instead of the usual 15 and 25 participants were ran for condition 1A so as to compare the results to the first study which had 25 participants per condition. Participants were recruited from a variety of Departments and Faculties throughout UCL with all levels of student being permitted to take part. The experiment consisted of 62 men and 53 women.

The experiment was approved by the UCL Research Ethics Committee, and was carried out with written informed consent from each participant. During recruitment, participants were asked to

complete a brief trauma questionnaire to determine if it was safe for the to take part in the study. Participants were also verbally reminded at several points throughout the experimental training procedure that they would be free to stop at any time without having to give reasons or lose any others rights to which they were entitled. They were paid 7 pounds (UK).

6.2.3. Equipment

The scenario was created using the Unity content creation engine (version 2017.3.1f1). The 3D models in the scenario, as well as the fire and smoke seen during the moral dilemma, were all purchased from the Unity Asset Store, except for the virtual characters (VCs) which were modelled and rigged using Adobe Fuse (Beta). VCs were animated during the scenario using a combination of a custom animation system and Final IK for the body movements and hand gestures and the LipSync Pro Unity package for mouth movements and facial expressions.

The head-mounted display (HMD) used was the HTC Vive which has a resolution of 1080x1200 per eye, a refresh rate of 90Hz and a 110-degree field of view. Body tracking was achieved using the HTC Vive's built-in tracking technology in conjunction with Unity, with two Lighthouses being secured in opposite corners of the tracking area, roughly 4 meters apart and 2 meters off of the ground. Tracked components included the HMD to provide position and rotational head-tracking and two Vive Controllers (one in each hand) to provide upper-body tracking with the additional use of the Final IK Unity package to calculate chest and elbow positions. The participants' legs were not tracked and so participants were asked to stand still for the duration of the scenario. In order to open the elevator doors, participant were instructed that they could do so by point-and-clicking towards them using the Vive Controllers provided.

6.2.4. Procedures

Participants arrived at the lab individually at a pre-booked time slot. They were not told beforehand that the study was about moral dilemmas, however they were informed that they may (or may not)

find the situation depicted in IVR stressful or upsetting and of course that they were free to continue or leave as they preferred.

Participants were given an information sheet detailing the experiment, prompted by the experimenter whether they had any questions and then given the consent form to sign if they wanted to continue. Finally, a demographic questionnaire was given to them before entering the VE.

A series of instructions were read to each participant about what they needed to do in the VE.

Most importantly, that:

1. They could use the Vive controllers to open an elevator.
2. The elevator doors were slow to open and close.
3. They could only open one elevator at a time.

The instructions were read twice and participants were asked to relay the instructions back to ensure they understood what they needed to do. Participants then stood somewhere in the tracking area and put on the HMD. They first went through a training period where people would come and go from the elevators requiring them to open the elevator doors for them. At the end of the training period, a situation occurred where the participant was required to open both elevators, this reinforced to the participants that they could only open one at a time and that waiting to open a second elevator after a first one opens and then closes again would take a long time.

After this, participants were asked a final time if they were happy to continue, if yes, they were reintroduced into the same VE where they again had to open elevator doors for VCs. During this time, if a participant was in conditions 1B, 1C, 2B or 2C they would also engage in a conversation with three bystanders who were standing to the right of them. The bystanders introduced themselves and would ask the participant their name. They would then talk about various other things, always including the participant in the conversation. Participants in conditions 1A or 2A were alone in the virtual corridor and did not have a conversation with any VCs.

After about 2 minutes the dilemma started. The participant heard the voices of the child and five adults come from their respective elevators and in the moral dilemma cases (conditions 1A, 1B and 1C), smoke was seen coming from the elevator doors. Participants could attempt to open one of the elevators at this time although nothing would happen. At the same time, the bystanders in the social influence conditions (1B, 1C, 2B and 2C) would start to encourage the participant to

either save the child (1B and 2B) or the five adults (1C or 2C). After a few seconds, the doors would become operational (signalled by a green light above them) and participants were able to make their choice. All data was recorded during this time, such as the elevator participants first tried to open during the dilemma, the door they ultimately chose and the number of times they tried to open either of the elevators.

Once out of the VE, participants filled in a presence questionnaire and finally a moral dilemma questionnaire before being debriefed on the nature of the experiment.

6.2.5. Response Variables

The main response variable in this study is whether the participant chose to save the child or the five adults during the dilemma. It is predicted that there will be little change in this variable in the moral dilemma conditions (1A, 1B and 1C) and greater variation in the non-moral dilemma conditions (2A, 2B and 2C), specifically that people will be more likely to save the child in condition 2B and more likely to save the adults in condition 2C, with condition 2A being somewhere in-between. It is also possible to look at which elevator participants attempted to open first and which they attempted to open the most. These should hopefully be relatively congruent with each other along with participants final choices.

6.3. Results

Using R, a power analysis was performed to confirm if the minimum number of participants per condition resulted in a sufficient statistical power. The analysis yielded a power of 0.79 from an effect size of 0.15 and a significance level of 0.05.

6.3.1. Presence Data

Questionnaire data was obtained on presence (Place Illusion, the plausibility of the scenario, and the plausibility of the virtual characters). The questionnaire used was a modified version of the one proposed by Slater et al. (1995) and can be found in Tables 3.3 to 3.5 along with each questions encoded value used in the following figures. The questionnaire is designed to gauge participants sense of presence, such as their illusion of being in the virtual world as well as the illusion that the people and events taking place around them are really happening.

Figure 6.2 shows the box plots for the Place Illusion questions in Table 3.3. It can be seen that mostly the median score is 5 out of a maximum of 7. Appropriately the scores on the control question 'lab' are clearly lower than for the rest of the questions. Mostly the scores are otherwise very similar across conditions.

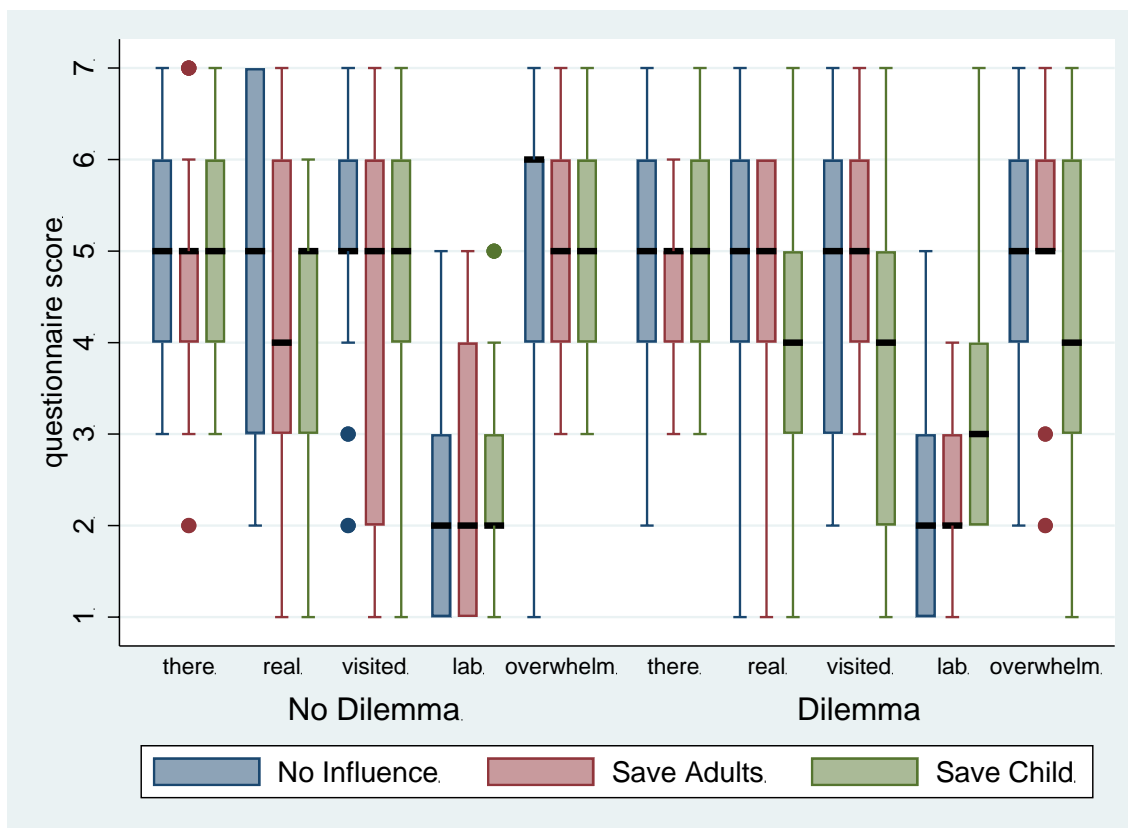


Figure 6.2. – Box plots for the Place Illusion questions in Table 3.3.

We can combine the scores into one overall score using a Principle Components Factor analysis. Using the stata command ‘factor’ the first principle component accounts for 55% of the total variance, and only this factor is retained. The loadings on this factor are almost all of equal magnitude except that the loading for ‘lab’ is negative whereas all the others are positive. The factor is essentially the sum of all the scores minus the score for lab. We can compute the corresponding factor score, which we refer to as ypi.

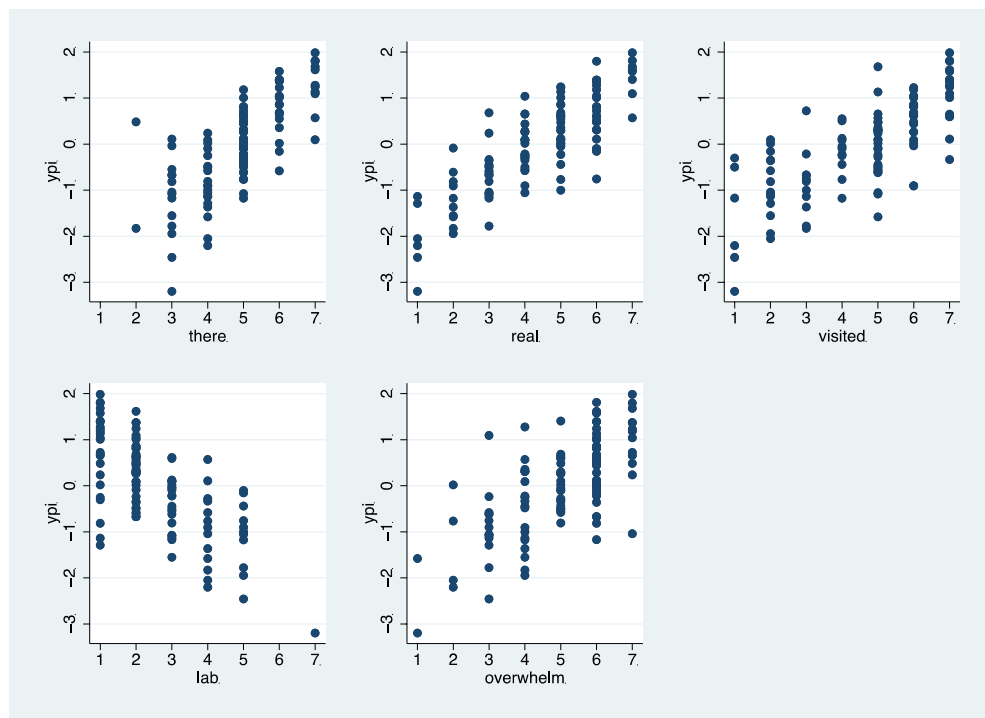


Figure 6.3. – Scatter diagrams for ypi on each of the scores to the questions in Table 3.3.

Figure 6.3 shows the scatter diagrams of the new factor scores variable ypi against each of the original scores to the place illusion questions in Table 3.3. It can be seen that there are strong positive slopes in each case except for lab, where there is a strong negative slope.

Figure 6.4 shows the bar charts for ypi by the conditions. It seems that mean ypi is lower in the condition where there is the Dilemma and the bystanders are influencing to save the child. However, ANOVA (or regression) of ypi on the factors shows no interaction effect ($P = 0.20$) and if the interaction term is removed then there are still no main effects ($P = 0.72$ for situation, $P = 0.18$ for Social).

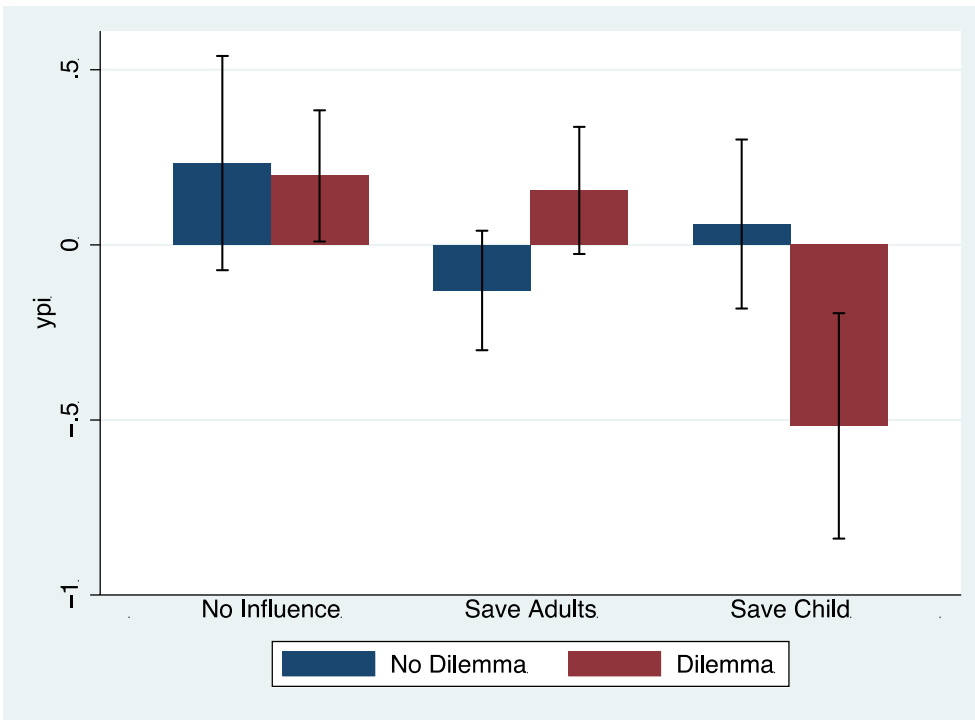


Figure 6.4. – Bar chart showing means and standard errors of ypi by the conditions.

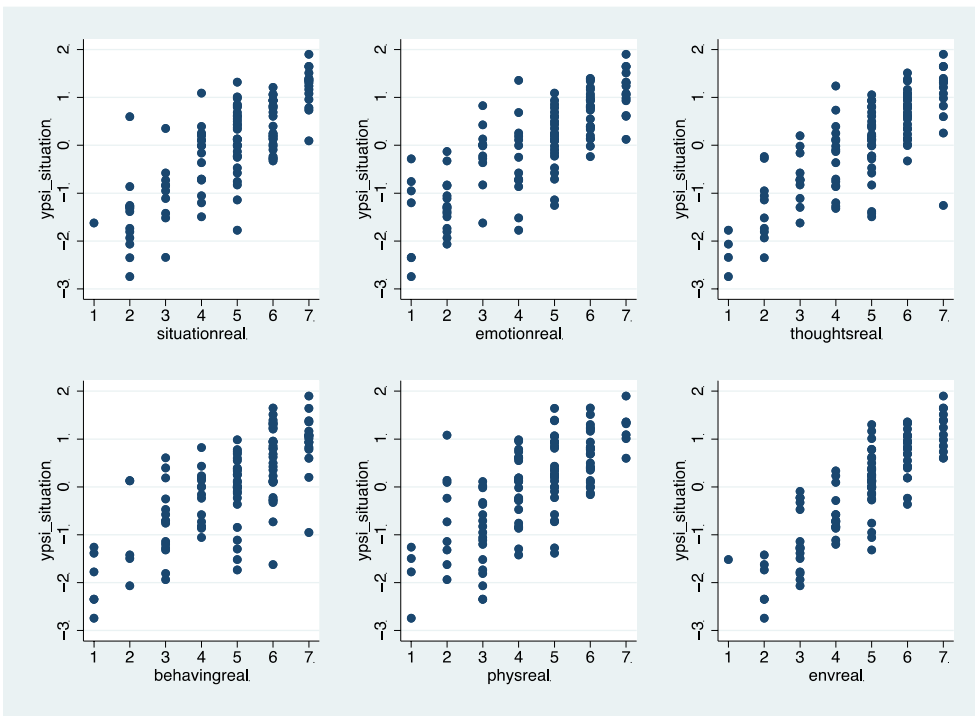


Figure 6.5. – Scatter diagram of the factor variable ypsi_situation on each of the original scores to the questions in Table 3.4.

We carry out the same analysis for the plausibility of the situation leading to the new factor variable `ypsi_situation`. One factor is retained which explains 59% of the total variance. The factor loadings are approximately equal, so that the derived factor `ypsi_situation` is like the sum of all the scores across these six variables. Figure 6.5 shows the relationship between the factor variable `ypsi_situation` and the original scores to the plausibility illusion questions in Table 3.4, showing strong positive correlations in each case.

Figure 6.6 shows the means and standard errors of the new variable `ypsi_situation` by the conditions. Again this shows that in the save the child and Dilemma situation that the scores seem to be less. However, as before ANOVA (or regression) shows no significant differences at all. Nevertheless, it is interesting that in both cases that the scores are lower for this configuration.

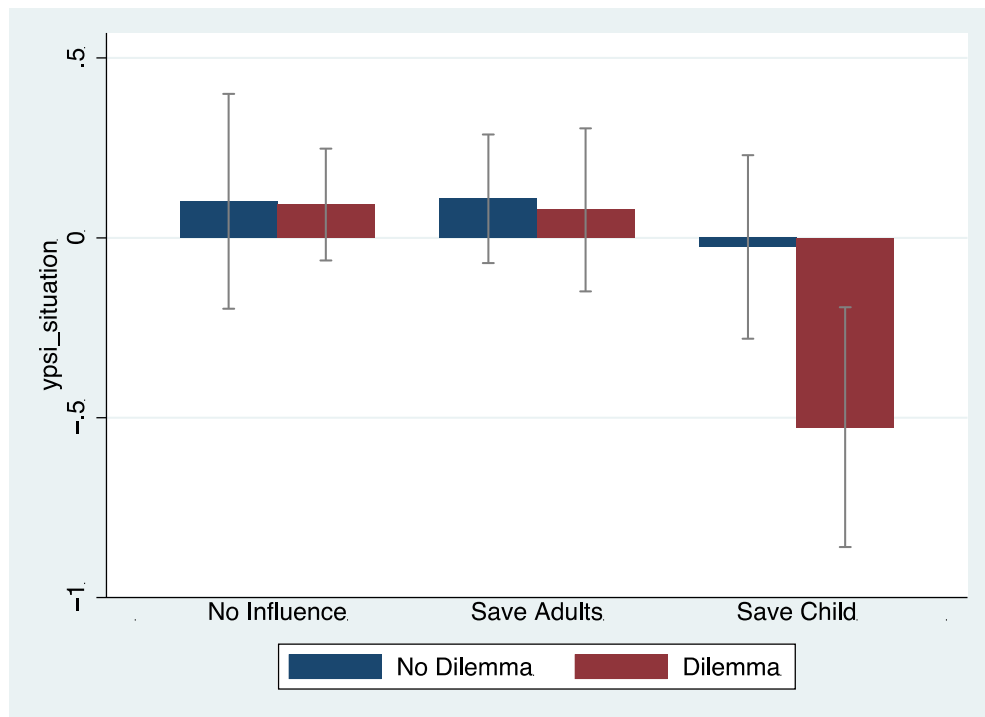


Figure 6.6. – Bar chart showing means and standard errors of `ypsi_situation` by the conditions.

Finally, we carry out the analysis on the plausibility of the virtual people leading to the factor variable `ypsi_people`. One factor is retained explaining 69% of the total variance. The factor loadings are equivalent to the new variable being proportional to the sum of the scores.

Figure 6.7 shows the scatter diagrams of the factor variable `ypsi_people` on the original scores to the plausibility illusion questions in Table 3.5 show a strong positive correlation in each case.

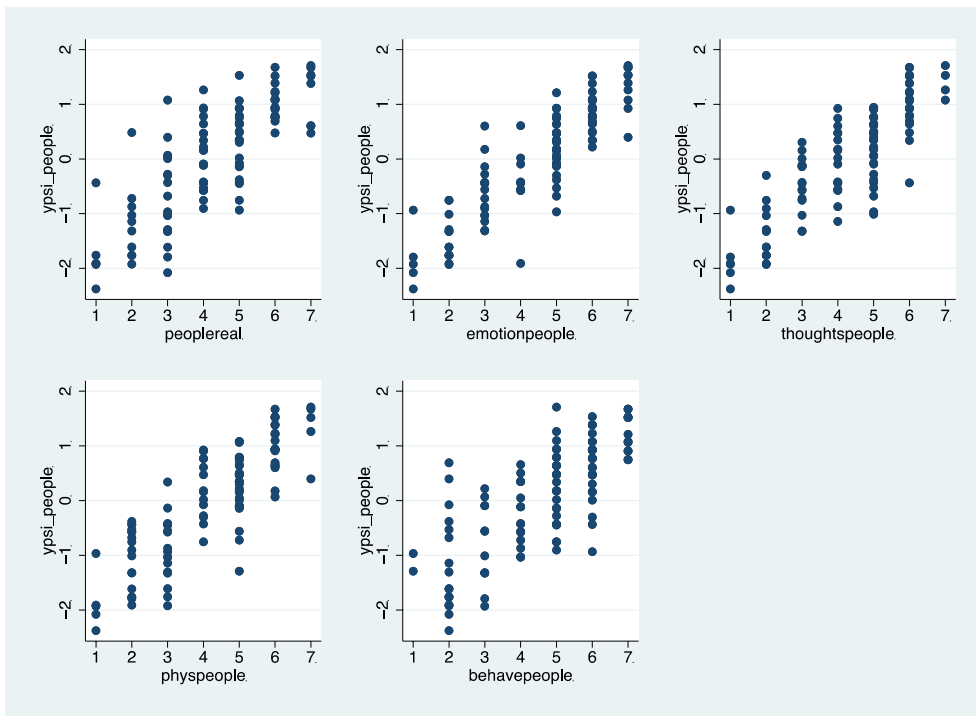


Figure 6.7. – Scatter diagram of `ypsi_people` on the scores to the questions in Table 3.5.

Figure 6.8 shows the means and standard errors of `ypsi_people` by condition. ANOVA (or regression) shows no significant difference at all between these (interaction term has $P = 0.32$, eliminating interaction the smallest $P = 0.48$).

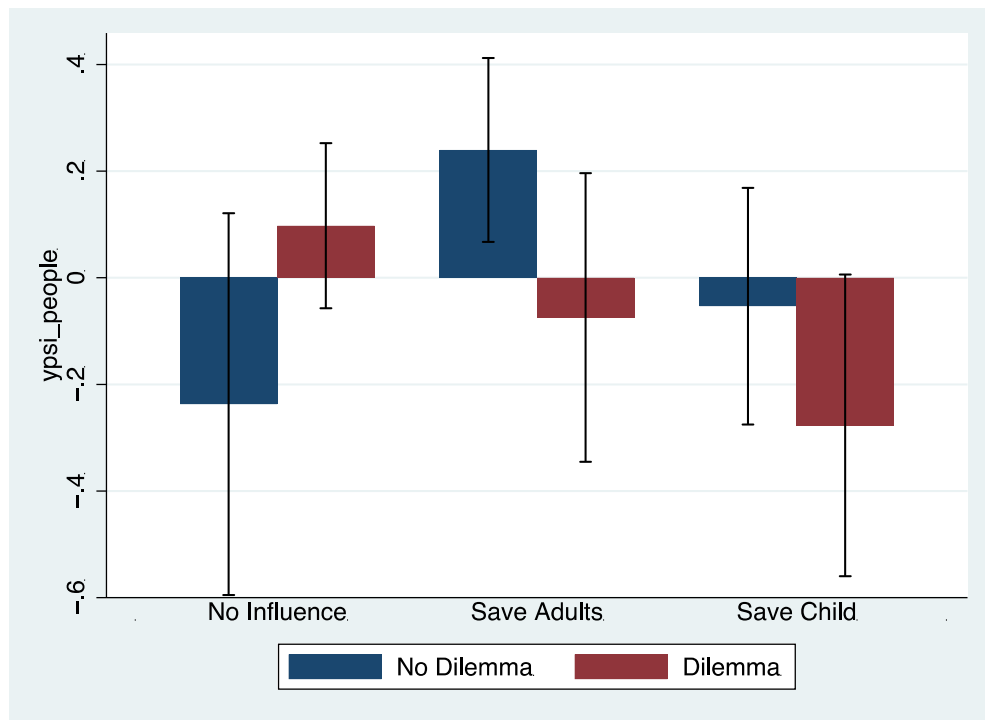


Figure 6.8. – Bar chart showing means and standard errors for ypsi_people by condition.

6.3.2. Behavioural Data

Here we consider what the participants actually did in response to the situation. We consider their first actions, their final actions, and finally the total numbers of actions.

Figure 6.9 shows the proportions of first actions to save the child. Although the proportion is highest for those in the (No Dilemma, Save Child) condition there do not seem to be any significant effects. Since the response variable is binary, we carry out a logistic regression on the factors. The interaction terms are not significant (Dilemma \times Save Adults, $P = 0.91$; Dilemma \times Save Child, $P = 0.96$). Eliminating the interaction term, Dilemma shows a significant reduction in the action of saving the child ($z = -2.17$, $P = 0.030$). This is evident in Figure 6.9, where in each case the Dilemma proportion is less than the No Dilemma proportion. If we consider all pairwise marginal comparisons, at an overall significance level of 5% using Scheffe's method, then we find that the only confidence interval that does not include zero is the comparison between Dilemma and No Dilemma, showing that the proportion for Dilemma is less than for No Dilemma.

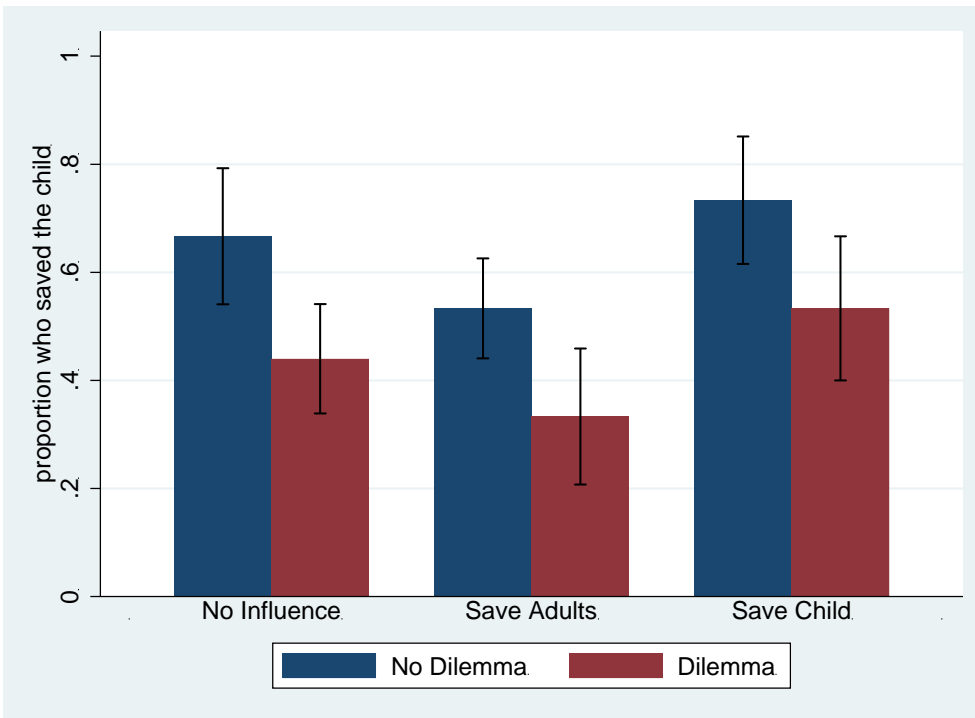


Figure 6.9. – Proportions and standard errors who saved the child as their first action.

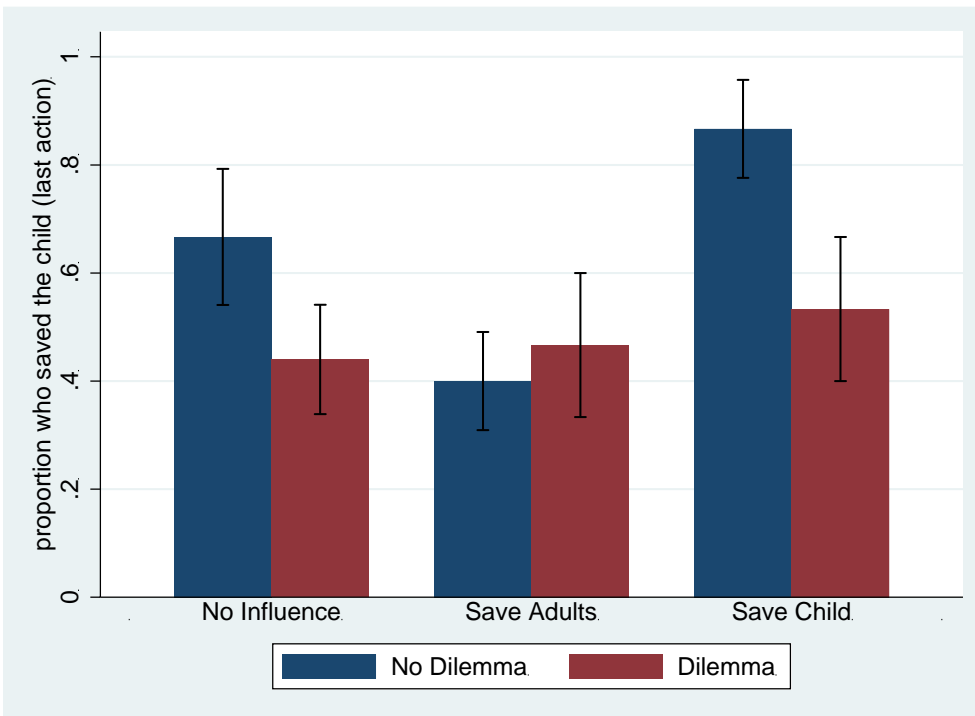


Figure 6.10. – Proportions and standard errors who saved the child as their first action.

Next we consider the final action of the participants, this is what would ultimately occur, i.e., the adults or the child being saved. Figure 6.10 shows the proportions saving the child corresponding to the last actions. Again we can see that the greatest proportion for saving the child was in the (No Dilemma, Save Child) condition. The logistic regression, however, shows no interaction effect ($P = 0.20$, $P = 0.48$). Eliminating the interaction term, there are no remaining significant effects. The proportion of final actions can also be seen in Table 6.2.

Proportion who saved the child (as their last action)			
	No Influence	Save Adults	Save Child
No Dilemma	0.667	0.4	0.867
Dilemma	0.44	0.467	0.533

Table 6.2. – Proportion who saved the child (as their last action).

Next we consider the total numbers of attempts to save the adults or the child during the period that the dilemma had started but before the doors were active. These are count variables that we model with Poisson log-linear regression. Figures 6.11 and 6.12 show that more attempts were made to save both the child and adults in the Dilemma condition compared to the No Dilemma condition. The Poisson regression shows that the only significant effect is the main effect for Dilemma in both sets of data, with $P < 0.0005$, confirming what can be seen in Figures 6.11 and 6.12. Figure 6.13 alleviates some worry that participants were simply pressing random buttons in the Dilemma condition by showing that the more participants tried to save the adults, the less they were focused on the child and vice versa, clearly showing some sort of preference instead of taking manic actions.

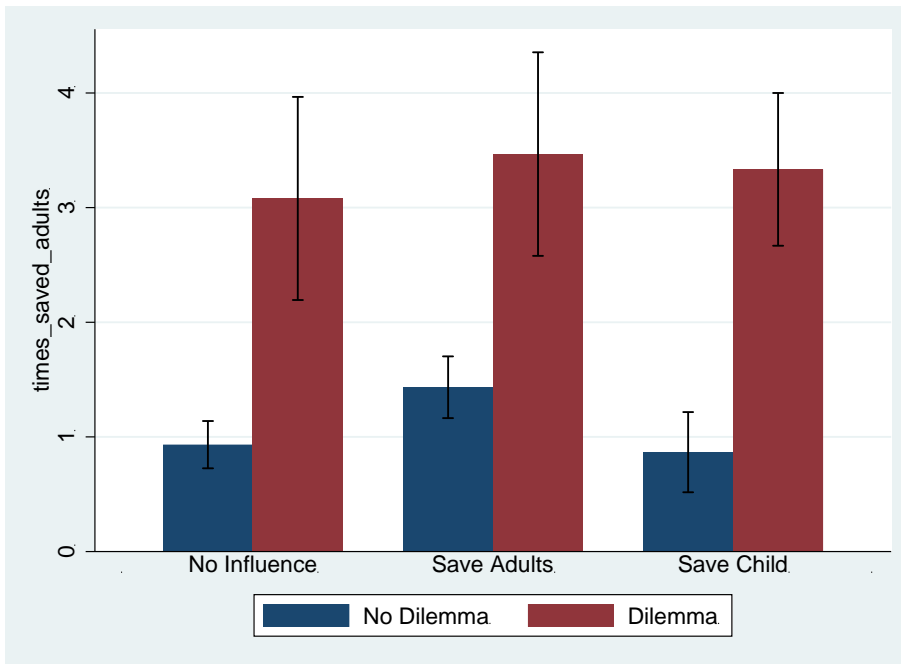


Figure 6.11. – Bar chart showing means and standard errors for the number of attempts to save the adults.

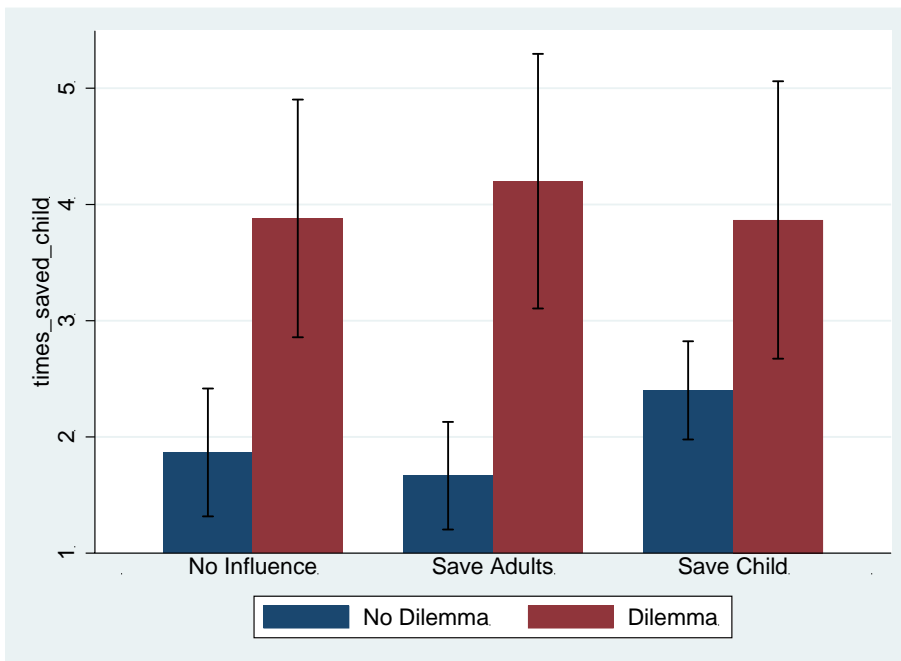


Figure 6.12. – Bar chart showing means and standard errors for the number of attempts to save the child.

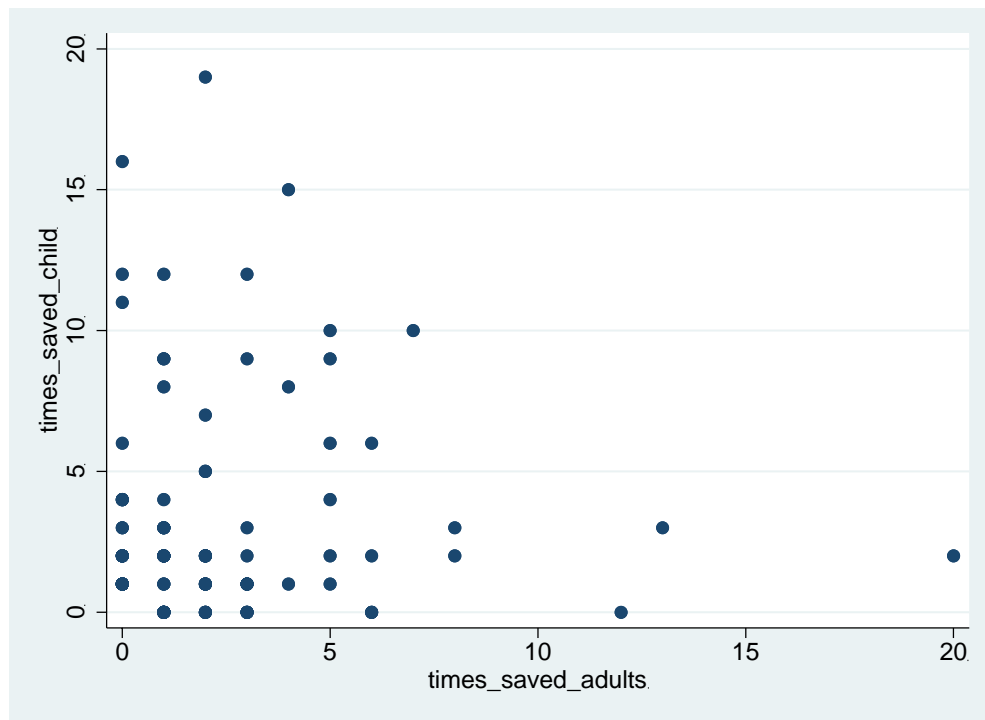


Figure 6.13. – Scatter diagram of the number of times the participant tried to save the child and the adults.

6.4. Discussion

The aim of the study was to identify if there is a difference in the influence of compliance requests between moral and non-moral situations, where the compliance requests relate to the resolution of the dilemma, with our specific hypothesis being that this type of social influence would be less effective during a moral dilemma compared to a non-moral due to emotional difference between the two situations. Figure 6.10 and Table 6.2 shows the proportion of participant’s final actions in response to both the moral and non-moral dilemmas. The results seem to point in the direction previously predicted, but were not strong enough say for certain that random variation was not the responsible factor. Nevertheless, it can be see that in general, responses to the moral dilemma fluctuate less than their non-moral counterparts under the duress of social influence, where participants saved the child 44% of the time under no influence, 47% of the time when influenced to save the adults and 53% of the time when influenced to save the child. This is compared to the non-moral dilemma conditions where participants saved the child 67% of the time under no influence, 40% of the time under influence to save the adults and 87% of the time

under influence to save the child. Here there is greater variation and also in a sensible direction (influence to save the child produced a greater chance to save the child and the same for the adults). Thus, while the results seem to point in a positive direction, they are still too ambiguous to draw any real conclusions from, leaving our original theory still in contention.

The shortcoming of our data might be because the moral dilemma was not emotionally inductive enough to interact meaningfully with the social influence. Previous literature has shown that moral dilemmas can produce overwhelming emotional responses which guide our choices during a moral decision-making. It was our theory that this immediate emotional response would in some ways negate the effects of a compliance request by providing responders with a prepotent answer, reducing our reliance on outside sources of information. However, studies have also shown that it is typically only the most appalling moral dilemmas (also known as personal moral dilemmas) that tend to trigger these powerful emotions (Greene et al., 2001). Examples of these include the footbridge dilemma or parent's dilemma which ask the responder to commit a terrible act (pushing a man to his death or smothering your own baby). These are then met with a significant negative emotional response. Shame and disgust at the thought of smothering our own baby takes over and urges the responder to stop. On the other hand, impersonal moral dilemma, such as the switch dilemma, have previously been found to be more similar to non-moral dilemmas in terms of brain activations. The issue then is that the moral dilemma presented to participants in this study is in many ways more similar to the switch dilemma than a personal moral dilemma and thus the emotional reaction might not have been any different between the moral and non-moral conditions. Although this seems damning, a direct comparison cannot be made between the moral dilemma presented in this study and the impersonal ones from Greene et al. (2001) due to the difference in medium. While Greene and colleagues used text-based questionnaires, the current study opted to use a life-like immersive virtual reality simulation which is a far more emotionally salient medium (Francis et al., 2016) and thus the emotional levels of participants responding to the moral dilemma were likely much higher than if they were responding to the same moral dilemma in a questionnaire. Furthermore, Navarrete et al. (2012) found that those participants who took action in a moral dilemma in immersive virtual reality had greater autonomic arousal, measured through skin conductance. This is positively related to this study as participants were required to take action in order to resolve the moral dilemma. Behavioural data presented in Figures 6.11 to 6.13 also supports the hope that emotional levels were different between the two

conditions by showing that participants were much more eager to save either the child or the adults in the Dilemma conditions compared to the No Dilemma conditions, implying a greater amount of urgency and stress in their behaviour. Unfortunately, no physiological readings were taken during this study however, meaning, while it is likely emotional level was high due to presentation of moral dilemma, the number of attempts made to open the elevators and the level of emotional arousal seen in other virtual dilemma studies, it is impossible to say for sure whether emotional levels were sufficiently different between moral and non-moral conditions to produce a difference in response to social influence.

6.5. Conclusion

Overall, while the results were not strong enough to draw any real conclusions, they were heading in the correct direction predicted by our hypothesis and perhaps could have been stronger with the inclusion of more participants or some slight differences to the procedures that were taken. Had the results been significant, this would reveal an interesting and novel interaction between two previously separate psychological processes, notably, that compliance attempts are less effective during moral situations, potentially due to theory described earlier that prepotent emotional reactions to moral situations reduce the likelihood of compliance attempts succeeding.

Chapter 7

Discussion

One of the goals of this thesis was to identify whether and how social influence works during moral situations. This was in response to the lack of research regarding this in the fields of both social influence and moral psychology. Where these topics have overlapped in the past, investigations have been limited to social conformity instead of compliance and text-based questionnaires instead of realistic immersive virtual reality (IVR). Another aim was to investigate the difference between moral judgments and behaviours, adding to the small number of research articles that have moved from text-based moral judgment experiments to more realistic moral behaviour research with the use of IVR technology and other innovative techniques.

IVR has been instrumental in bringing moral decision-making research into the modern era, where studies using IVR are now showing different responses to some types of moral dilemma compared to older text-based studies (Francis et al., 2017). Whereas before, the data of participants in a moral dilemma study was simply a representation of their best guess as to what they might do in the given situation, IVR more accurately represents real-life situations and produces realistic responses as long as certain presence-based conditions are met. The distinction between what people say they would do and what they actually do has been shown to be different in the past, for example, in the original Milgram experiments (Milgram, 1963) the number of participants who agreed to administer a dangerous electrical shock to another person was much higher than predicted and more recently FeldmanHall et al. (2012) found that participants kept over seven times as much money to themselves in a condition containing real

morally charged actions and outcomes compared to a “pretend” condition where participants were asked to use their imagination.

Perhaps most worrying of all for moral judgment research is the divergence of results of moral behaviour studies from Greene's Dual-Process Theory of Moral Judgment (DPToMJ). This theory posits that emotional cognitive processes support deontological actions, whereas rational processes support utilitarian outcomes and has been extremely robustly supported in moral judgment research. It has struggled however to get a foothold in moral behavioural research, with the more emotional medium, IVR, actually producing greater utilitarian responses in many studies. The results from the second study went somewhat towards explaining these diverging results, positing that an increase in tendency to act in IVR might be responsible. Obviously it is still very much early days for moral behaviour research and increased research in the future will be met with much enthusiasm as it reveals more about our true moral dispositions.

7.1. Moral Judgment versus Behaviour Experiment

Previous studies looking at moral judgment versus behaviour found diverging results based on the type of dilemma that was being utilised. In the cases that involved an impersonal moral dilemma, they generally found that participants acted in IVR much the same way that they did in a companion questionnaire whereas cases that involved a personal moral dilemma consistently found that participants were more likely to act in a utilitarian manner in IVR than in traditional text-based questionnaires. One exception to this rule comes from Patil et al. (2014) who found the same pattern, but in an impersonal moral dilemma.

Our study largely conforms to these results, utilising an impersonal moral dilemma where the responder has the choice to press one of two buttons which will save either a young child or a group of five adults, our results showed no significant difference between responding to this scenario in a realistic virtual environment or writing down the answer in a questionnaire.

A supplementary result showed that presence scores were slightly but consistently lower for those participants who witnessed the moral dilemma in IVR compared to those who did not. This could be due to the unlikely scenario (i.e., a fire in a pair of elevators) breaking the plausibility of the

scenario, but might also be linked to the warnings participants received before entering IVR. In order to comply with strict ethical standards, several warnings were given to participants before they entered the virtual environment due to the nature of the experiment. These warnings were given to both participants who did and did not see the moral dilemma in IVR in order to reduce variance and thus all participants were expecting to see something unsettling (they were never told it was specifically a moral dilemma). As such, those participants who never saw the moral dilemma might have been waiting and grounding themselves in the experience in the anticipation of an exciting event that was never going to come. The other participants however who did see the moral dilemma and might have been reminded at the sight of it that this was all part of the experiment, lowering their immersion.

7.2. Social Influence and Omission Bias Experiment

The research on compliance attempts and persuasion techniques is vast, but a main take away from the field is that people can be influenced to acquiesce to a request through very simple means ranging from knowing a person's name (Burger et al., 2001) to one of the several "salesman" techniques used by the stereotypical car salesman in order to squeeze every last penny from their clients (Freedman & Fraser, 1966).

Similarly, Asch (1955, 1956) found that social conformity was extremely effective in altering people's responses to a public task even when the consensus formed is clearly wrong. This phenomenon has recently been translated into moral contexts as well (Kundu & Cummins, 2013; Lisciandra et al., 2013) and achieved similar, but not identical results, with Lisciandra and colleagues noting that moral norms were the hardest to manipulate compared to other social norms, although they could be manipulated.

The current study, again using IVR, aimed to build on both of these areas by having participants be subject to the social influence of compliance attempts during a moral dilemma or having them respond to the moral dilemma individually. Given the past effectiveness of social influence, it came as surprise to find the results of the study showed that the compliance attempts had had no effect on the behaviour of the participants. There were two main differences between this study

and those that previously investigated social influence during a moral decision-making task. The first is the change from social conformity to compliance attempts and the second was the change from questionnaire to virtual environment. Given these changes, while it was not foreseen, it is well within the imagination that our results could correctly diverge so much.

Causes for the lack of efficacy of social influence might be due to the realism of the virtual characters (VCs) where the plausibility scores relating to the VCs were only moderate. Otherwise, given that this is the first study to look at the influence of compliance on moral behaviour in IVR, it is possible that compliance attempts simply do not work in this type of situation, i.e., a moral dilemma. The conclusion offered was that moral intuitions, not present in non-moral problems, provide responders with powerful prepotent answers to the situation, negating the effects of social influence. This hypothesis came from a plethora of previous results showing or suggesting that moral dilemmas induce intuitive emotional responses (Greene et al., 2004; Greene et al., 2001) and that our morality is considered to be a critical part of our identity, not easily swayed by others (Schaich Borg et al., 2006). The goal moving forward then was to investigate the effect in a situation more likely to be effective, i.e., a non-moral dilemma.

In another result, it was found that participants had a preference to take action in the moral dilemma in IVR, even when the action produced opposite results. This was posited to be due to an action bias, the lesser known cousin of omission bias, in which taking an action feels like the normal thing to do and so it is preferentially chosen. This has been shown before in the behaviour of goal-keepers (Bar-Eli et al., 2007), in response to negative outcomes (Zeelenberg et al., 2002) and in the behaviour of people in a position of responsibility (Haidt & Baron, 1996). Given our studies use of prescriptive morality instead of the traditional proscriptive morality, this explanation makes sense.

In order to investigate this further a questionnaire replication of the IVR study was conducted and found that the preference for action had disappeared. Contributions this results makes to the wider community are discussed in section 7.4.1.

7.3. Moral versus Non-Moral Experiment

In the previous experiment it was found that participants were not influenced during a moral dilemma. A potential reason for this is that moral decision-making, at least in IVR and in response to compliance attempts, is somewhat resistant to social influence possibly due to its reliance on intuitive emotional cognition. Given this, the focus of the current experiment was to investigate compliance attempts during both moral and non-moral dilemmas where emotions are less likely to arise.

The results found a weak trend towards our predicted results where responses to non-moral dilemmas were more varied under different social influence conditions compared to moral dilemmas, but were not divergent enough to overcome the potential to be just random variation. This might be due to some of the choices made in the construction of the IVR scenarios or due to small sample size, where each condition only had 15 participants.

7.4. Contributions

7.4.1. Methodological Contributions

In all three studies, IVR was used to create immersive simulations which prompted participants to take (or refuse to take) a moral action. While this has been done before, the research is still in infancy and the addition of these studies to the growing pool of research papers further proves IVR's usefulness in the domain of moral decision-making.

One contribution made was to demonstrate a novel way to inform the participant how to act during a moral dilemma in IVR. Unlike questionnaires, it is impossible to directly tell participants how to respond to a moral dilemma in IVR without explaining the situation to them beforehand. This is obviously not desired as it could alter participants "in the moment" responses to the moral dilemma by being able to pre-evaluate their options. Previous studies have opted to teach the participant an action, such as moving a platform up and down and then presenting them with a

moral dilemma which can be resolved through the use of this action (Pan & Slater, 2011). This is the method used in Chapters 4 and 6. Chapter 5 however, used a novel approach to the problem. In this dilemma, the participant was not given any instructions before entering IVR and did not have any tasks to do. Instead, when the moral dilemma occurred, a known entity within the virtual scenario relayed the instructions to the participant. In our study, this came from the autonomous car that the participant was driving which had previously made itself known by introducing itself to the participant at the start of the simulation. In future studies, this could easily come from a virtual character present in the scenario. For example, a recreation of the switch dilemma in IVR could include a character – too far away from the switch to pull it themselves, but close enough to be within shouting distance of the participant – that informs the participant that pulling the switch would move the train. While this technique requires additional development time of the virtual scenario, it should be preferable to the alternative of revealing the moral dilemma to the participant before experiencing the events for fear of increased contemplation time which would obviously not happen in real-life.

Two out of three of the studies presented in this thesis also contribute to research regarding the effects of social influence in IVR. To facilitate this, these studies utilised the method of engaging in a conversation before the moral dilemma in order to build up a rapport with the characters in the hopes that this further enables social influence. This has been shown to be true in the past outside of IVR scenarios even without the conversation, where a study demonstrated that social influence was more effective when simply knowing a person's name or even having sat in the same room as them (Burger et al., 2001). Previous studies that used this technique in IVR include Slater et al. (2013) who had participants engage in a conversation about football with a virtual character before seeing that character get into a fight. In the current studies, similar to Slater et al. (2013), the virtual conversation was implemented in a linear fashion – meaning that virtual characters would always speak the same response, no matter how the participant answered. They were constructed in such a way that, assuming that the participant gave a sensible response, would always make sense by having the virtual characters either asking questions or giving generic responses. This linear approach has two main benefits. The first is a more streamlined conversation which allowed for a single animation to be recorded, making the virtual characters movements look more realistic than having them swap between animations for different lines of dialogue. The second benefit is the reduced variability between participants. It is possible that a

branching conversation could result in longer or shorter, or more or less pleasant conversations which may impact how the virtual characters were viewed by the participant. By having every conversation identical, virtual characters were more likely to be viewed similarly by each participant.

The studies presented in this thesis are the first to simultaneously investigate both moral decision-making and social influence using IVR. This produced a unique challenge of making sure that participants were aware of the moral dilemma, what their response options to the moral dilemma were and what the virtual characters were telling them what to do. In study two, this involved ensuring the participant saw the passenger next to them have a stroke, hearing the autonomous car say that it is going to the hospital, seeing the police car behind them (which is the reason a rear-view camera was shown on a screen inside the car), hearing the car relay their options (stop or continue) and then hearing the shouts of social influence from the rear-seat passengers. While this task was accomplished successfully (in the writers opinion), it required extensive testing and the timing of each event carefully such that no two things were vying for the attention of the participant at the same time and could serve as influence for researchers who intend to create complex IVR scenes in the future.

Finally, the current studies demonstrate good methodological practices with regards to awareness of presence by utilising technology to track participants head and body movements, up-to-date IVR headset technology, a virtual sex-matched body given to participants and realistic human-computer interactions (such pressing physical buttons also present in the virtual environment). Where previous studies neglected some (or all) of these aspects (e.g., Patil et al. (2014) and Skulmowski et al. (2014)), the current studies advocate for the best use of IVR possible and sets a good example of how it should be used by those in the future.

In conclusion, novel methodologies were implemented over the three studies and those used continue to solidify the use of IVR as an effective and necessary methodological tool for social psychology studies. Where previous researchers had the excuse of ethical and technological limitations that forced them to use questionnaires to elicit moral judgment from participants, the popularity, availability, ease of use and effectiveness of IVR should convince them otherwise.

7.4.2. Substantive Contributions

The results presented in this thesis both reinforce current scientific information by repeating results previously found and add to this knowledge through novel findings, revealing more information about the results of previous studies and spawning questions for future experiments.

Results from the first study are congruent with previous papers (Navarrete et al., 2012; Pan & Slater, 2011) and thus adds support for the results from these earlier studies with the main finding showing that impersonal moral dilemmas receive similar responses between IVR and questionnaire setups. While not producing significant results, the study did add to the body of work putting pressure on Greene's DPToMJ by not conforming to its predictions. Namely, early researchers of moral behaviour using IVR predicted that there should be an increase in deontological responses in IVR based on its greater propensity to elicit emotional reactions compared to traditional questionnaires, where deontological responses are driven largely by emotional cognitive processes (Navarrete et al., 2012). This prediction has not come true however and in fact seems to be going in the opposite direction, i.e., towards utilitarianism in IVR.

Results from the second study add more information and a new dimension to the results from Francis et al. (2017) and others by showing that people might simply prefer to take action in a virtual dilemma, compared to in response to a questionnaire, rather than to take an action in order to receive a utilitarian outcome. These two responses are confounded by the fact that utilitarian responses require an action to be achieved and so the result from this study, showing a preference for action that did not lead to a utilitarian outcome means this could be a contributing factor to the results seen in the past. This result also puts doubt on the hypothesis that this increase in utilitarian nature is due to the salient outcomes of moral dilemmas in IVR. Where the outcome was indeed seen as important to responders, it could not account for their complete motivation which also seems to be driven by action-based tendencies.

Further results from the second study revealed an inability for moral behaviours in IVR to be influenced by compliance attempts. This was the first study to utilise all of these parts and so would have produced a novel result no matter the outcome. However, based on the results from previous studies, the prediction was that social influence would prevail which ultimately did not happen. This contributes to the scientific community by potentially showing that, predicted by

Lisciandra et al. (2013), morals can be difficult to influence, especially through means such as compliance. While adding to common knowledge about the interaction between social influence and moral decision-making, this also illuminated a lot of questions, the most obvious being whether there is a difference between compliance attempts in moral and non-moral situations. This question would inspire the final study in this thesis.

Results from the third study are tentative, but point in the right direction, showing some increased variance as a result of social influence in a non-moral compared to a moral dilemma. If we take for a moment that these results are correct, it provides evidence that the theory postulated in Chapter 6 is correct and that moral-emotional cognitive processes interfere with the social influence behind compliance attempts. This would be a novel and ground-breaking finding, showing a new interaction between two previously separate areas of psychological research. The goal of the study was to simply add knowledge for the scientific community and so it is left to others to predict the impacts, if any, this has on the wider community. However, before this happens, further work, perhaps a follow-up study, needs to be performed in order to further verify the current results.

7.5. Future Work

To summarise, two key results have been found. From the results of the second study, it was found that people might have a penchant for action in IVR over an omission. An obvious follow up would be to try understand whether the preference to act in a utilitarian manner in IVR is due instead to the preference to take action. In order to do this, the study would have to dissociate the utilitarian option with the requirement to achieve it through action. One way to do this would be to follow the steps taken by Conway and Gawronski (2013) where they describe a moral dilemma in which utilitarianism and action are separated. This type of moral dilemma would then have to be reconstructed in a virtual environment. If the results from the second study are to be trusted, we would predict to see a preference for whatever outcome is achieved through action, but the important result would be whether preference for the utilitarian option is still present. The results from this study would be incredibly important for the future of moral behaviour studies, illuminating

a real inclination, whether this is for action or utilitarianism, in our moral psychology which would likely impact how further moral dilemma studies are designed.

Additionally, from the third study, the possibility was found that social influence is less effective during moral dilemmas compared to non-moral. As said previously, in order for the results for the final study to be supported, a follow-up study is likely needed due to our weak results. The additional study would have to improve several aspects of the scenario and should also strive to include more than 15 participants per condition to ensure robust results.

7.6. Conclusion

In this thesis, two of the newer discussions regarding moral decision-making have been furthered: how moral judgments differ from moral behaviour and how people respond to a moral dilemma in an interpersonal setting. The point of both of these goals is to research moral decision-making in a more realistic environment where physical actions need to be taken and interpersonal influences must be dealt with. This, I feel, is the way forward for the field of moral psychology. Older studies had to rely on text-based questionnaires to elicit moral judgment due to ethical reasons, however, with the advent of affordable and available immersive virtual reality (IVR) equipment, this issue has been removed. IVR allows participants to witness and physically respond to a moral dilemma, instead of imagine and mindfully come up with an answer. In IVR, participants are provided with salient imagery, actions, outcomes and social and behavioural cues all of which contribute towards their cognitive evaluation of the situation and none of which can be replicated by using a questionnaire.

In previous interpersonal moral dilemma studies, questionnaires were used to elicit moral judgment and all major results found that moral judgments could successfully be influenced by others. However, studies conducted as part of this thesis found that moral behaviours could not be influenced so easily. Due to other differences between the current and previous studies, direct comparisons cannot be made and so the difference in medium is not certainly the cause of this difference, although it is easy to imagine that it is at least part of it. In other research, results have shown that people are more likely to act in a utilitarian manner in IVR compared to a questionnaire

and within this thesis it has also been shown that participants have a penchant for action in IVR compared to questionnaire.

These results and others lead me to believe an overhaul of the current methodologies used to research moral decision-making is needed. Due to recent results within and outside of this thesis indicating a definite difference between what people say they would do and what they actually do in response to a moral dilemma, I feel it is imperative that IVR is used prominently in future moral decision-making research.

References

- Allport, G. W. (1985). The historical background of social psychology. In G. Lindzey & E. Aronson (Eds.), *Handbook of social psychology* (3rd ed., Vol. 1, pp. 1-46). New York: Random House.
- Amit, E., & Greene, J. D. (2012). You see, the ends don't justify the means: Visual imagery and moral judgment. *Psychological science*, *23*(8), 861-868.
- Anscombe, G. E. M. (1958). Modern moral philosophy. *Philosophy*, *33*(124), 1-19.
- Aquinas, T. (1485/2012). *Summa theologica*: Authentic Media Inc.
- Aramovich, N. P., Lytle, B. L., & Skitka, L. J. (2012). Opposing torture: Moral conviction and resistance to majority influence. *Social Influence*, *7*(1), 21-34.
- Aronson, E. (2007). The evolution of cognitive dissonance theory: A personal appraisal. In *The science of social influence: Advances and future progress*. (pp. 115-135). New York, NY, US: Psychology Press.
- Asch, S. E. (1955). Opinions and social pressure. *Scientific American*, *193*(5), 31-35.
- Asch, S. E. (1956). Studies of independence and conformity: I. A minority of one against a unanimous majority. *Psychological monographs: General and applied*, *70*(9), 1.
- Bailenson, J. N., Blascovich, J., Beall, A. C., & Loomis, J. M. (2003). Interpersonal distance in immersive virtual environments. *Personality and social psychology bulletin*, *29*(7), 819-833.
- Bailenson, J. N., & Yee, N. (2005). Digital chameleons: Automatic assimilation of nonverbal gestures in immersive virtual environments. *Psychological science*, *16*(10), 814-819.
- Banakou, D., Groten, R., & Slater, M. (2013). Illusory ownership of a virtual child body causes overestimation of object sizes and implicit attitude changes. *Proceedings of the National Academy of Sciences*, *110*(31), 12846-12851.
- Banakou, D., Hanumanthu, P. D., & Slater, M. (2016). Virtual embodiment of white people in a black virtual body leads to a sustained reduction in their implicit racial bias. *Frontiers in human neuroscience*, *10*, 601.

- Bar-Eli, M., Azar, O. H., Ritov, I., Keidar-Levin, Y., & Schein, G. (2007). Action bias among elite soccer goalkeepers: The case of penalty kicks. *Journal of economic psychology*, *28*(5), 606-621.
- Bargh, J. A., & Chartrand, T. L. (1999). The unbearable automaticity of being. *American psychologist*, *54*(7), 462.
- Baron, J., & Ritov, I. (1994). Reference points and omission bias. *Organizational Behavior and Human Decision Processes*, *59*(3), 475-498.
- Baron, J., & Ritov, I. (2004). Omission bias, individual differences, and normality. *Organizational Behavior and Human Decision Processes*, *94*(2), 74-85.
- Bauman, C. W., McGraw, A. P., Bartels, D. M., & Warren, C. (2014). Revisiting external validity: Concerns about trolley problems and other sacrificial dilemmas in moral psychology. *Social and Personality Psychology Compass*, *8*(9), 536-554.
- Berns, G. S., Capra, C. M., Moore, S., & Noussair, C. (2010). Neural mechanisms of the influence of popularity on adolescent ratings of music. *Neuroimage*, *49*(3), 2687-2696.
- Berthoz, S., Armony, J., Blair, R., & Dolan, R. J. (2002). An fMRI study of intentional and unintentional (embarrassing) violations of social norms. *Brain*, *125*(8), 1696-1708.
- Blascovich, J., & Ginsburg, G. P. (1974). Emergent norms and choice shifts involving risk. *Sociometry*, 205-218.
- Blascovich, J., Loomis, J., Beall, A. C., Swinth, K. R., Hoyt, C. L., & Bailenson, J. N. (2002). Immersive virtual environment technology as a methodological tool for social psychology. *Psychological inquiry*, *13*(2), 103-124. doi:10.1207/S15327965PLI1302_01
- Blass, T. (2004). *The man who shocked the world: The life and legacy of Stanley Milgram*: Basic Books New York.
- Bostyn, D. H., & Roets, A. (2017). An asymmetric moral conformity effect: Subjects conform to deontological but not consequentialist majorities. *Social Psychological and Personality Science*, *8*(3), 323-330.
- 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. doi:10.1177/0956797617752640
- Botvinick, M., & Cohen, J. (1998). Rubber hands 'feel' touch that eyes see. *Nature*, *391*(6669), 756.
- Burger, J. M., Soroka, S., Gonzago, K., Murphy, E., & Somervell, E. (2001). The effect of fleeting attraction on compliance to requests. *Personality and social psychology bulletin*, *27*(12), 1578-1586.

- Campbell-Meiklejohn, D. K., Bach, D. R., Roepstorff, A., Dolan, R. J., & Frith, C. D. (2010). How the opinion of others affects our valuation of objects. *Current Biology, 20*(13), 1165-1170.
- Chaiken, S. (1980). Heuristic versus systematic information processing and the use of source versus message cues in persuasion. *Journal of personality and social psychology, 39*(5), 752.
- Charpentier, C. J., Moutsiana, C., Garrett, N., & Sharot, T. (2014). The brain's temporal dynamics from a collective decision to individual action. *Journal of Neuroscience, 34*(17), 5816-5823.
- Chen, S., & Chaiken, S. (1999). The heuristic-systematic model in its broader context. *Dual-process theories in social psychology, 15*, 73-96.
- Cialdini, R. B. (2007). *Influence: The psychology of persuasion*: Collins New York.
- Cialdini, R. B. (2009). *Influence: Science and practice* (Vol. 4): Pearson education Boston.
- Cialdini, R. B., & Goldstein, N. J. (2004). Social influence: Compliance and conformity. *Annu. Rev. Psychol., 55*, 591-621.
- Cikara, M., Farnsworth, R. A., Harris, L. T., & Fiske, S. T. (2010). On the wrong side of the trolley track: Neural correlates of relative social valuation. *Social cognitive and affective neuroscience, 5*(4), 404-413.
- Cioffi, D., & Garner, R. (1996). On doing the decision: Effects of active versus passive choice on commitment and self-perception. *Personality and social psychology bulletin, 22*(2), 133-147.
- Conway, P., & Gawronski, B. (2013). Deontological and utilitarian inclinations in moral decision making: a process dissociation approach. *Journal of personality and social psychology, 104*(2), 216.
- Cruz-Neira, C., Sandin, D. J., & DeFanti, T. A. (1993). *Surround-screen projection-based virtual reality: the design and implementation of the CAVE*. Paper presented at the Proceedings of the 20th annual conference on Computer graphics and interactive techniques.
- Cushman, F. (2013). Action, outcome, and value: A dual-system framework for morality. *Personality and social psychology review, 17*(3), 273-292.
- Cushman, F., Gray, K., Gaffey, A., & Mendes, W. B. (2012). Simulating murder: The aversion to harmful action. *Emotion, 12*(1), 2.
- 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), 1082-1089. doi:10.1111/j.1467-9280.2006.01834.x

- Damasio, A. R. (1998). Investigating the biology of consciousness. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, 353(1377), 1879-1882.
- Damasio, A. R. (2006). *Descartes' error*: Random House.
- Darley, J. M., & Latané, B. (1968). Bystander intervention in emergencies: Diffusion of responsibility. *Journal of personality and social psychology*, 8(4p1), 377.
- De Melo, C. M., Gratch, J., & Carnevale, P. J. (2015). Humans versus computers: Impact of emotion expressions on people's decision making. *IEEE Transactions on Affective Computing*, 6(2), 127-136.
- De Waal, F. B. (1991). The chimpanzee's sense of social regularity and its relation to the human sense of justice. *American Behavioral Scientist*, 34(3), 335-349.
- De Waal, F. B. (1996). *Good natured: The origins of right and wrong in humans and other animals*: Harvard University Press Cambridge, MA.
- Deutsch, M., & Gerard, H. B. (1955). A study of normative and informational social influences upon individual judgment. *The journal of abnormal and social psychology*, 51(3), 629.
- Dolinski, D. (2007). Emotional see-saw. *The science of social influence. Advances and future progress*, 137-153.
- Dolinski, D., Nawrat, M., & Rudak, I. (2001). Dialogue involvement as a social influence technique. *Personality and social psychology bulletin*, 27(11), 1395-1406.
- Dolinski, D., & Nawrat, R. (1998). "Fear-then-relief" procedure for producing compliance: Beware when the danger is over. *Journal of Experimental Social Psychology*, 34(1), 27-50.
- Eagly, A. H., & Chaiken, S. (1993). *The psychology of attitudes*: Harcourt Brace Jovanovich College Publishers.
- Eisenberger, N. I. (2012). The pain of social disconnection: examining the shared neural underpinnings of physical and social pain. *Nature reviews neuroscience*, 13(6), 421.
- Eisenberger, N. I., Lieberman, M. D., & Williams, K. D. (2003). Does rejection hurt? An fMRI study of social exclusion. *Science*, 302(5643), 290-292.
- Ellemers, N., Pagliaro, S., & Barreto, M. (2013). Morality and behavioural regulation in groups: A social identity approach. *European Review of Social Psychology*, 24(1), 160-193.
- Ellis, S. R. (1996). Presence of mind: A reaction to thomas sheridan's "further musings on the psychophysics of presence". *Presence: Teleoperators & Virtual Environments*, 5(2), 247-259.
- Everett, J. A., Pizarro, D. A., & Crockett, M. (2016). Inference of trustworthiness from intuitive moral judgments. *Journal of Experimental Psychology: General*, 145(6), 772.
- Falk, E., Way, B., & Jasinska, A. (2012). An imaging genetics approach to understanding social influence. *Frontiers in human neuroscience*, 6, 168.

- Faulhaber, A. K., Dittmer, A., Blind, F., Wächter, M. A., Timm, S., Sütfeld, L. R., . . . König, P. (2018). Human decisions in moral dilemmas are largely described by utilitarianism: Virtual car driving study provides guidelines for autonomous driving vehicles. *Science and engineering ethics*, 1-20.
- FeldmanHall, O., Mobbs, D., Evans, D., Hiscox, L., Navrady, L., & Dalgleish, T. (2012). What we say and what we do: the relationship between real and hypothetical moral choices. *Cognition*, 123(3), 434-441. doi:10.1016/j.cognition.2012.02.001
- Festinger, L. (1957). *A theory of cognitive dissonance* (Vol. 2): Stanford university press.
- Finger, E. C., Marsh, A. A., Kamel, N., Mitchell, D. G., & Blair, J. R. (2006). Caught in the act: The impact of audience on the neural response to morally and socially inappropriate behavior. *Neuroimage*, 33(1), 414-421.
- Fiske, S. T., Harris, L. T., & Cuddy, A. J. (2004). Why Ordinary People Torture Enemy Prisoners. *Science*, 306(5701), 1482-1483. doi:10.1126/science.1103788
- Foot, P. (1967). The problem of abortion and the doctrine of double effect. *Oxford Review*, 5, 5-15.
- Forgas, J. P. (1995). Mood and judgment: the affect infusion model (AIM). *Psychological bulletin*, 117(1), 39.
- Forgas, J. P. (1998). Asking nicely? The effects of mood on responding to more or less polite requests. *Personality and social psychology bulletin*, 24(2), 173-185.
- 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.
- Francis, K. B., Howard, C., Howard, I. S., Gummerum, M., Ganis, G., Anderson, G., & Terbeck, S. (2016). Virtual morality: Transitioning from moral judgment to moral action? *PLoS One*, 11(10), e0164374. doi: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), 13954.
- Freedman, J. L., & Fraser, S. C. (1966). Compliance without pressure: the foot-in-the-door technique. *Journal of personality and social psychology*, 4(2), 195.
- Friedman, D., Pizarro, R., Or-Berkers, K., Neyret, S., Pan, X., & Slater, M. (2014). A method for generating an illusion of backwards time travel using immersive virtual reality—an exploratory study. *Frontiers in psychology*, 5, 943.
- Gawronski, B., Conway, P., Armstrong, J. B., Friesdorf, R., & Hütter, M. (2016). Understanding responses to moral dilemmas: Deontological inclinations, utilitarian inclinations, and

- general action tendencies. In *The social psychology of morality* (pp. 91-110):
Routledge.
- Giner-Sorolila, R., & Chaiken, S. (1997). Selective use of heuristic and systematic processing under defense motivation. *Personality and social psychology bulletin*, *23*(1), 84-97.
- Gleicher, F., Kost, K. A., Baker, S. M., Strathman, A. J., Richman, S. A., & Sherman, S. J. (1990). The role of counterfactual thinking in judgments of affect. *Personality and social psychology bulletin*, *16*(2), 284-295.
- Gonzalez-Franco, M., & Lanier, J. (2017). Model of illusions and virtual reality. *Frontiers in psychology*, *8*, 1125.
- Gonzalez-Franco, M., Slater, M., Birney, M. E., Swapp, D., Haslam, S. A., & Reicher, S. D. (2018). Participant concerns for the Learner in a Virtual Reality replication of the Milgram obedience study. *PLoS One*, *13*(12), e0209704. doi:10.1371/journal.pone.0209704
- Graham, J., Haidt, J., Koleva, S., Motyl, M., Iyer, R., Wojcik, S. P., & Ditto, P. H. (2013). Moral foundations theory: The pragmatic validity of moral pluralism. In *Advances in experimental social psychology* (Vol. 47, pp. 55-130): Elsevier.
- Greene, J. D. (2008). The secret joke of Kant's soul. *Moral psychology*, *3*, 35-79.
- Greene, J. D. (2014). *Moral tribes: Emotion, reason, and the gap between us and them*: Penguin.
- Greene, J. D. (2015). Beyond point-and-shoot morality: Why cognitive (neuro) science matters for ethics. *The Law & Ethics of Human Rights*, *9*(2), 141-172.
- Greene, J. D., Cushman, F. A., Stewart, L. E., Lowenberg, K., Nystrom, L. E., & Cohen, J. D. (2009). Pushing moral buttons: The interaction between personal force and intention in moral judgment. *Cognition*, *111*(3), 364-371.
- Greene, J. D., & Haidt, J. (2002). How (and where) does moral judgment work? *Trends in cognitive sciences*, *6*(12), 517-523.
- 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. doi: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. doi: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. doi:10.1126/science.1062872

- Guadagno, R. E., Blascovich, J., Bailenson, J. N., & Mccall, C. (2007). Virtual humans and persuasion: The effects of agency and behavioral realism. *Media Psychology, 10*(1), 1-22.
- Haidt, J. (2001). The emotional dog and its rational tail: a social intuitionist approach to moral judgment. *Psychological review, 108*(4), 814. doi:10.1037/0033-295X.108.4.814
- Haidt, J. (2003). The moral emotions. *Handbook of affective sciences, 11*(2003), 852-870.
- Haidt, J. (2007). The new synthesis in moral psychology. *Science, 316*(5827), 998-1002.
- Haidt, J. (2008). Morality. *Perspectives on Psychological Science, 3*(1), 65-72.
- Haidt, J., & Baron, J. (1996). Social roles and the moral judgement of acts and omissions. *European Journal of Social Psychology, 26*(2), 201-218.
- Haney, C., Banks, C., & Zimbardo, P. G. (1973). A Study of Prisoners and Guards in a Simulated Prison. *Naval Research Review, 30*, 4-17.
- Hauser, M., Cushman, F., Young, L., Kang-Xing, J., & Mikhail, J. (2007). A dissociation between moral judgments and justifications. *Mind & language, 22*(1), 1-21. doi:10.1111/j.1468-0017.2006.00297.x
- Heekeren, H. R., Wartenburger, I., Schmidt, H., Schwintowski, H.-P., & Villringer, A. (2003). An fMRI study of simple ethical decision-making. *Neuroreport, 14*(9), 1215-1219.
- Held, R. M., & Durlach, N. I. (1992). Telepresence. *Presence: Teleoperators & Virtual Environments, 1*(1), 109-112.
- Hersh, S. M., & Friedman, P. (2004). *Chain of command*: Penguin London.
- Hornsey, M. J., Majkut, L., Terry, D. J., & McKimmie, B. M. (2003). On being loud and proud: Non-conformity and counter-conformity to group norms. *British Journal of Social Psychology, 42*(3), 319-335.
- Hume, D. (1738/2003). *A treatise of human nature*: Courier Corporation.
- Izuma, K. (2013). The neural basis of social influence and attitude change. *Current opinion in neurobiology, 23*(3), 456-462.
- Janoff-Bulman, R., Sheikh, S., & Hepp, S. (2009). Proscriptive versus prescriptive morality: Two faces of moral regulation. *Journal of personality and social psychology, 96*(3), 521.
- Kahneman, D., & Egan, P. (2011). *Thinking, fast and slow* (Vol. 1): Farrar, Straus and Giroux New York.
- Kahneman, D., & Miller, D. T. (1986). Norm theory: Comparing reality to its alternatives. *Psychological review, 93*(2), 136.
- Kamm, F. M. (2015). *The trolley problem mysteries*: Oxford University Press.
- Kant, I. (1785/2013). *Moral Law: Groundwork of the Metaphysics of Morals*: Routledge.

- Kawai, N., Kubo, K., & Kubo-Kawai, N. (2014). "Granny dumping": Acceptability of sacrificing the elderly in a simulated moral dilemma. *Japanese Psychological Research*, 56(3), 254-262. doi:10.1111/jpr.12049
- Kelly, M., Ngo, L., Chituc, V., Huettel, S., & Sinnott-Armstrong, W. (2017). Moral conformity in online interactions: rational justifications increase influence of peer opinions on moral judgments. *Social Influence*, 12(2-3), 57-68.
- Kinateder, M., Ronchi, E., Gromer, D., Müller, M., Jost, M., Nehfischer, M., . . . Pauli, P. (2014). Social influence on route choice in a virtual reality tunnel fire. *Transportation research part F: traffic psychology and behaviour*, 26, 116-125.
- Klucharev, V., Hytönen, K., Rijpkema, M., Smidts, A., & Fernández, G. (2009). Reinforcement learning signal predicts social conformity. *Neuron*, 61(1), 140-151.
- 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.
- Kohlberg, L. (1969). *Stage and sequence: The cognitive-developmental approach to socialization*: Rand McNally.
- Kundu, P., & Cummins, D. D. (2013). Morality and conformity: The Asch paradigm applied to moral decisions. *Social Influence*, 8(4), 268-279.
- Landman, J. (1987). Regret and elation following action and inaction: Affective responses to positive versus negative outcomes. *Personality and social psychology bulletin*, 13(4), 524-536.
- Latané, B. (1981). The psychology of social impact. *American psychologist*, 36(4), 343.
- Latané, B., & Darley, J. M. (1970). *The unresponsive bystander: Why doesn't he help?*: Appleton-Century-Crofts.
- Leach, C. W., Ellemers, N., & Barreto, M. (2007). Group virtue: the importance of morality (vs. competence and sociability) in the positive evaluation of in-groups. *Journal of personality and social psychology*, 93(2), 234.
- Leary, M. R. (2000). Affect, cognition, and the social emotions.
- Lieberman, M. D. (2013). *Social: Why our brains are wired to connect*: OUP Oxford.
- Lisciandra, C., Postma-Nilsenová, M., & Colombo, M. (2013). Conformability. A study on group conditioning of normative judgment. *Review of Philosophy and Psychology*, 4(4), 751-764.
- Loomis, J. M. (1992). *Presence and distal attribution: Phenomenology, determinants, and assessment*. Paper presented at the Human Vision, Visual Processing, and Digital Display III.

- Loomis, J. M., Blascovich, J. J., & Beall, A. C. (1999). Immersive virtual environment technology as a basic research tool in psychology. *Behavior research methods, instruments, & computers*, 31(4), 557-564.
- Lucas, B. J., & Livingston, R. W. (2014). Feeling socially connected increases utilitarian choices in moral dilemmas. *Journal of Experimental Social Psychology*, 53, 1-4.
- Lucas Jr, G. M., & Taic, S. (2015). Behavioral public choice and the law. *W. Va. L. Rev.*, 118, 199.
- Mason, M. F., Dyer, R., & Norton, M. I. (2009). Neural mechanisms of social influence. *Organizational Behavior and Human Decision Processes*, 110(2), 152-159.
- McCall, C., & Blascovich, J. (2009). How, when, and why to use digital experimental virtual environments to study social behavior. *Social and Personality Psychology Compass*, 3(5), 744-758.
- 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. doi:10.1016/j.evolhumbehav.2017.06.001
- Meehan, M., Insko, B., Whitton, M., & Brooks Jr, F. P. (2002). *Physiological measures of presence in stressful virtual environments*. Paper presented at the Acm transactions on graphics (tog).
- Mehrabian, A., & Stefl, C. A. (1995). Basic temperament components of loneliness, shyness, and conformity. *Social Behavior and Personality: an international journal*, 23(3), 253-263.
- Milgram, S. (1963). Behavioral study of obedience. *The journal of abnormal and social psychology*, 67(4), 371. doi:10.1037/h0040525
- Milgram, S., & Gudehus, C. (1978). Obedience to authority. In: Ziff-Davis Publishing Company.
- Mill, J. S. (1861/2016). Utilitarianism. In *Seven masterpieces of philosophy* (pp. 337-383): Routledge.
- Moll, J., & de Oliveira-Souza, R. (2007). Moral judgments, emotions and the utilitarian brain. *Trends in cognitive sciences*, 11(8), 319-321.
- Moll, J., de Oliveira-Souza, R., Bramati, I. E., & Grafman, J. (2002). Functional networks in emotional moral and nonmoral social judgments. *Neuroimage*, 16(3), 696-703.
- Moll, J., de Oliveira-Souza, R., Eslinger, P. J., Bramati, I. E., Mourão-Miranda, J. n., Andreiuolo, P. A., & Pessoa, L. (2002). The neural correlates of moral sensitivity: a functional magnetic resonance imaging investigation of basic and moral emotions. *Journal of Neuroscience*, 22(7), 2730-2736.
- Moll, J., De Oliveira-Souza, R., & Zahn, R. (2008). The neural basis of moral cognition. *Annals of the New York Academy of Sciences*, 1124(1), 161-180. doi:10.1196/annals.1440.005

- Moll, J., Eslinger, P. J., & Oliveira-Souza, R. d. (2001). Frontopolar and anterior temporal cortex activation in a moral judgment task: preliminary functional MRI results in normal subjects. *Arquivos de neuro-psiquiatria*, *59*(3B), 657-664.
- Moore, C., & Gino, F. (2013). Ethically adrift: How others pull our moral compass from true North, and how we can fix it. *Research in organizational behavior*, *33*, 53-77.
- 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. doi:10.1037/a0025561
- Nowak, K. L. (2004). The influence of anthropomorphism and agency on social judgment in virtual environments. *Journal of Computer-Mediated Communication*, *9*(2), JCMC925.
- Nussbaum, M. C. (1999). Virtue ethics: a misleading category? *The Journal of Ethics*, *3*(3), 163-201.
- Pagliaro, S., Brambilla, M., Sacchi, S., D'Angelo, M., & Ellemers, N. (2013). Initial impressions determine behaviours: Morality predicts the willingness to help newcomers. *Journal of Business Ethics*, *117*(1), 37-44.
- Pan, X., & Hamilton, A. F. d. C. (2018). Why and how to use virtual reality to study human social interaction: The challenges of exploring a new research landscape. *British Journal of Psychology*, *109*(3), 395-417. doi:10.1111/bjop.12290
- Pan, X., & Slater, M. (2011). *Confronting a moral dilemma in virtual reality: a pilot study*. Paper presented at the Proceedings of the 25th BCS Conference on Human-Computer Interaction.
- Pan, X., Slater, M., Beacco, A., Navarro, X., Rivas, A. I. B., Swapp, D., . . . Hamilton, A. F. d. C. (2016). The Responses of Medical General Practitioners to Unreasonable Patient Demand for Antibiotics-A study of medical ethics using immersive virtual reality. *PLoS One*, *11*(2), e0146837.
- 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.
- Pascual, L., Gallardo-Pujol, D., & Rodrigues, P. (2013). How does morality work in the brain? A functional and structural perspective of moral behavior. *Frontiers in Integrative Neuroscience*, *7*, 65.
- 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. doi:10.1080/17470919.2013.870091
- Patt, A., & Zeckhauser, R. (2000). Action bias and environmental decisions. *Journal of Risk and Uncertainty*, *21*(1), 45-72.

- Paxton, J. M., Ungar, L., & Greene, J. D. (2012). Reflection and reasoning in moral judgment. *Cognitive Science*, 36(1), 163-177.
- 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(1), 68-78.
- Piaget, J. (1932/1965). *The moral judgment of the child*. New York, NY, US: Free Press.
- Pincus, M., LaViers, L., Prietula, M. J., & Berns, G. (2014). The conforming brain and deontological resolve. *PLoS One*, 9(8), e106061.
- Pino, N. (2018). Oculus Rift Review. *Gaming accessories*. Retrieved from <https://www.techradar.com/reviews/gaming/gaming-accessories/oculus-rift-1123963/review>
- Raven, E. (2017). Osteoarthritis and rheumatoid arthritis of the carpus.
- Ritov, I., & Baron, J. (1990). Reluctance to vaccinate: Omission bias and ambiguity. *Journal of Behavioral Decision Making*, 3(4), 263-277.
- Sagarin, B. J., & Wood, S. E. (2007). Resistance to influence. *The science of social influence: Advances and future progress*, 1, 321-340.
- Salmond, J. W. (1947). Jurisprudence.
- Sanchez-Vives, M. V., & Slater, M. (2005). From presence to consciousness through virtual reality. *Nature reviews neuroscience*, 6(4), 332.
- Schaich Borg, J., Hynes, C., Van Horn, J., Grafton, S., & Sinnott-Armstrong, W. (2006). Consequences, action, and intention as factors in moral judgments: An fMRI investigation. *Journal of cognitive neuroscience*, 18(5), 803-817.
- Schultz, W. (2006). Behavioral theories and the neurophysiology of reward. *Annu. Rev. Psychol.*, 57, 87-115.
- Shenhav, A., & Greene, J. D. (2014). Integrative moral judgment: dissociating the roles of the amygdala and ventromedial prefrontal cortex. *Journal of Neuroscience*, 34(13), 4741-4749.
- Sheridan, T. B. (1992). Musings on telepresence and virtual presence. *Presence: Teleoperators & Virtual Environments*, 1(1), 120-126.
- Sheridan, T. B. (1996). Further musings on the psychophysics of presence. *Presence: Teleoperators & Virtual Environments*, 5(2), 241-246.
- 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.
- Slater, M. (2004). How colorful was your day? Why questionnaires cannot assess presence in virtual environments. *Presence: Teleoperators & Virtual Environments*, 13(4), 484-493.

- Slater, M. (2009). Place illusion and plausibility can lead to realistic behaviour in immersive virtual environments. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 364(1535), 3549-3557. doi:10.1098/rstb.2009.0138
- Slater, M., Antley, A., Davison, A., Swapp, D., Guger, C., Barker, C., . . . Sanchez-Vives, M. V. (2006). A virtual reprise of the Stanley Milgram obedience experiments. *PLoS One*, 1(1), e39. doi:10.1371/journal.pone.0000039
- Slater, M., Khanna, P., Mortensen, J., & Yu, I. (2009). Visual realism enhances realistic response in an immersive virtual environment. *IEEE computer graphics and applications*, 29(3), 76-84.
- Slater, M., Pérez Marcos, D., Ehrsson, H., & Sanchez-Vives, M. V. (2008). Towards a digital body: the virtual arm illusion. *Frontiers in human neuroscience*, 2, 6.
- Slater, M., Rovira, A., Southern, R., Swapp, D., Zhang, J. J., Campbell, C., & Levine, M. (2013). Bystander responses to a violent incident in an immersive virtual environment. *PLoS One*, 8(1), e52766. doi:10.1371/journal.pone.0052766
- Slater, M., & Sanchez-Vives, M. V. (2016). Enhancing our lives with immersive virtual reality. *Frontiers in Robotics and AI*, 3, 74.
- Slater, M., Usoh, M., & Steed, A. (1994). Depth of presence in virtual environments. *Presence: Teleoperators & Virtual Environments*, 3(2), 130-144.
- Slater, M., Usoh, M., & Steed, A. (1995). Taking steps: the influence of a walking technique on presence in virtual reality. *ACM Trans. Comput.-Hum. Interact.*, 2(3), 201-219. doi:10.1145/210079.210084
- Spanlang, B., Normand, J.-M., Borland, D., Kilteni, K., Giannopoulos, E., Pomés, A., . . . Muncunill, X. N. (2014). How to build an embodiment lab: achieving body representation illusions in virtual reality. *Frontiers in Robotics and AI*, 1, 9.
- Spranca, M., Minsk, E., & Baron, J. (1991). Omission and commission in judgment and choice. *Journal of Experimental Social Psychology*, 27(1), 76-105.
- Steele, C. M. (1988). The psychology of self-affirmation: Sustaining the integrity of the self. In *Advances in experimental social psychology* (Vol. 21, pp. 261-302): Elsevier.
- Sunstein, C. R. (2005). Moral heuristics. *Behavioral and brain sciences*, 28(4), 531-541.
- Sütfeld, L. R., Gast, R., König, P., & Pipa, G. (2017). Using virtual reality to assess ethical decisions in road traffic scenarios: applicability of value-of-life-based models and influences of time pressure. *Frontiers in behavioral neuroscience*, 11, 122.
- Sutherland, I. E. (1965). The ultimate display. *Multimedia: From Wagner to virtual reality*, 506-508.
- Sutherland, I. E. (1968). *A head-mounted three dimensional display*. Paper presented at the Proceedings of the December 9-11, 1968, fall joint computer conference, part I.

- Swinth, K., & Blascovich, J. (2001). *Conformity to group norms in an immersive virtual environment*. Paper presented at the annual meeting of the American Psychological Society, Toronto, Ontario.
- Tajfel, H., Turner, J. C., Austin, W. G., & Worchel, S. (1979). An integrative theory of intergroup conflict. *Organizational identity: A reader*, 56-65.
- Tassy, S., Deruelle, C., Mancini, J., Leistedt, S., & Wicker, B. (2013). High levels of psychopathic traits alters moral choice but not moral judgment. *Frontiers in human neuroscience*, 7, 229.
- Tassy, S., Oullier, O., Mancini, J., & Wicker, B. (2013). Discrepancies between judgment and choice of action in moral dilemmas. *Frontiers in psychology*, 4, 250.
doi:10.3389/fpsyg.2013.00250
- Thomson, J. J. (1976). Killing, letting die, and the trolley problem. *The Monist*, 59(2), 204-217.
- Thomson, J. J. (1985). The trolley problem. *The Yale Law Journal*, 94(6), 1395-1415.
doi:10.2307/796133
- Turiel, E. (1983). *The development of social knowledge: Morality and convention*: Cambridge University Press.
- Turner, J. C. (1991). *Social influence*: Thomson Brooks/Cole Publishing Co.
- Uhlmann, E. L., Pizarro, D. A., & Diermeier, D. (2015). A person-centered approach to moral judgment. *Perspectives on Psychological Science*, 10(1), 72-81.
- Uhlmann, E. L., Zhu, L. L., & Tannenbaum, D. (2013). When it takes a bad person to do the right thing. *Cognition*, 126(2), 326-334.
- Usoh, M., Arthur, K., Whitton, M. C., Bastos, R., Steed, A., Slater, M., & Brooks Jr, F. P. (1999). *Walking > walking-in-place > flying, in virtual environments*. Paper presented at the Proceedings of the 26th annual conference on Computer graphics and interactive techniques.
- Valdesolo, P. (2016). On Emotions and Their Role in Morality. Retrieved from <http://emotionresearcher.com/on-emotions-and-their-role-in-morality/>
- Valdesolo, P., & DeSteno, D. (2006). Manipulations of emotional context shape moral judgment. *PSYCHOLOGICAL SCIENCE-CAMBRIDGE-*, 17(6), 476. doi:10.1111/j.1467-9280.2006.01731.x
- 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.
- Warren, D. E., & Smith-Crowe, K. (2008). Deciding what's right: The role of external sanctions and embarrassment in shaping moral judgments in the workplace. *Research in organizational behavior*, 28, 81-105.

- Waytz, A., & Epley, N. (2012). Social connection enables dehumanization. *Journal of Experimental Social Psychology, 48*(1), 70-76.
- Whatley, M. A., Webster, J. M., Smith, R. H., & Rhodes, A. (1999). The effect of a favor on public and private compliance: How internalized is the norm of reciprocity? *Basic and Applied Social Psychology, 21*(3), 251-259.
- Wood, W. (2000). Attitude change: Persuasion and social influence. *Annual review of psychology, 51*(1), 539-570.
- Yee, N., & Bailenson, J. (2007). The Proteus effect: The effect of transformed self-representation on behavior. *Human communication research, 33*(3), 271-290.
- Young, L., & Dungan, J. (2012). Where in the brain is morality? Everywhere and maybe nowhere. *Social neuroscience, 7*(1), 1-10.
- Zaki, J., Schirmer, J., & Mitchell, J. P. (2011). Social influence modulates the neural computation of value. *Psychological science, 22*(7), 894-900.
- Zeelenberg, M., Van den Bos, K., Van Dijk, E., & Pieters, R. (2002). The inaction effect in the psychology of regret. *Journal of personality and social psychology, 82*(3), 314.
- Zimmons, P., & Panter, A. (2003). *The influence of rendering quality on presence and task performance in a virtual environment*. Paper presented at the IEEE Virtual Reality, 2003. Proceedings.
- Zuwerink, J. R., & Devine, P. G. (1996). Attitude importance and resistance to persuasion: It's not just the thought that counts. *Journal of personality and social psychology, 70*(5), 931.

Appendix A

Moral Dilemmas

A.1. The switch dilemma

You are a passenger on a train whose driver has fainted. On the main track ahead are five people. The main track has a side track leading off to the left, and you can turn the train on to it. There is one person on the left hand track. You can turn the train, killing the one; or you can refrain from turning the train, letting the five die (Hauser et al., 2007).

A.2. The footbridge dilemma

You are on a footbridge over the train tracks. You see a train approaching the bridge out of control. There are five people on the track. You know that the only way to stop the train is to drop a heavy weight into its path. But the only available, sufficiently heavy weight is one large man, also watching the train from the footbridge. You can shove the one man onto the track in the path of the train, killing him; or you can refrain from doing this, letting the five die (Hauser et al., 2007).

A.3. The surgeon's dilemma

You have five patients, each of whom is about to die due to a failing organ of some kind. You have another patient who is healthy. The only way that you can save the lives of the first five patients is to transplant five of this young man's organs (against his will) into the bodies of the other five patients. If you do this, the young man will die, but the other five patients will live. Is it appropriate for you to perform this transplant in order to save five of your patients? (Greene et al., 2001).

A.4. The drug dilemma

We are about to give a patient who needs it to save his life a massive dose of a certain drug in short supply. There arrive, however, five other patients each of whom could be saved by one-fifth of that dose (Foot, 1967).

A.5. The gas dilemma

There are five patients in a hospital whose lives could be saved by the manufacture of a certain gas, but that this will inevitably release lethal fumes into the room of another patient whom for some reason we are unable to move (Thomson, 1985).

A.6. The mayor's dilemma

The five on the straight track are regular track workmen. The righthand track is a dead end, unused in ten years. The Mayor, representing the City, has set out picnic tables on it, and invited the convalescents at the nearby City Hospital to have their meals there, guaranteeing them that

no trolleys will ever, for any reason, be turned onto that track. The one on the right-hand track is a convalescent having his lunch there; it would never have occurred to him to do so if the Mayor had not issued his invitation and guarantee. The Mayor was out for a walk; he now stands by the switch (Thomson, 1985).

A.7. The loop dilemma

You are walking near the train tracks when you notice a train approaching out of control. Up ahead on the track are five people. You are standing next to a switch, which you can throw to turn the train on to a side track. There is a heavy object on the side track. If the train hits the object, the object will slow the train down, giving the men time to escape. The heavy object is one man, standing on the side track. You can throw the switch, preventing the train from killing the five people, but killing the one man. Or you can refrain from doing this, letting the five die (Hauser et al., 2007).

A.8. The serum dilemma

Several dangerously ill people can be saved only if we kill a certain individual and make a serum from another person's body (Foot, 1967).

A.9. The tram dilemma

You are the driver of a runaway tram which you can only steer from one narrow track on to another; five men are working on one track and one man on the other; anyone on the track he enters is bound to be killed (Foot, 1967).

A.10. The parent's dilemma

Enemy soldiers have taken over your village. They have orders to kill all remaining civilians. You and some of your townspeople have sought refuge in the cellar of a large house. Outside you hear the voices of soldiers who have come to search the house for valuables. Your baby begins to cry loudly. You cover his mouth to block the sound. If you remove your hand from his mouth his crying will summon the attention of the soldiers who will kill you, your child, and the others hiding out in the cellar. To save yourself and the others you must smother your child to death. Is it appropriate for you to smother your child in order to save yourself and the other townspeople? (Greene et al., 2001).

A.11. The judge's dilemma

Suppose that a judge or magistrate is faced with rioters demanding that a culprit be found for a certain crime and threatening otherwise to take their own bloody revenge on a particular section of the community. The real culprit being unknown, the judge sees himself as able to prevent the bloodshed only by framing some innocent person and having him executed (Foot, 1967).

A.12. The infanticide dilemma

You are a fifteen-year-old girl who has become pregnant. By wearing loose clothing and deliberately putting on weight you have managed to keep your pregnancy a secret. One day, while at school, your water breaks. You run to the girls locker room and hide for several hours while you deliver the baby. You are sure that you are not prepared to care for this baby. You think to yourself that it would be such a relief to simply clean up the mess you have made in the locker room, wrap the baby in some towels, throw the baby in the dumpster behind the school, and act

as if nothing had ever happened. Is it appropriate for you to throw your baby in the dumpster in order to move on with your life? (Greene et al., 2001).

Appendix B

Forms, Sheets and Questionnaires

Consent form – Chapter Four

UCL DEPARTMENT OF COMPUTER SCIENCE



UCL Ethics Project ID Number: 8941/001

Investigators: Mel Slater, Jacob Thorn

Virtual Reality Study Consent Form

Please read and answer the following questions carefully:

- | | |
|---|--------|
| Have you read the information sheet about this study? | YES/NO |
| Have you had an opportunity to ask questions about the procedure? | YES/NO |
| Have you received satisfactory answers to all your questions? | YES/NO |
| Have you received enough information about this study? | YES/NO |
| Do you understand that you are free to withdraw from this study
<u>at any time</u> and <u>without giving a reason for withdrawing</u> ? | YES/NO |
| Do you understand that you must not take part if you have a child of
younger than 13 or have scored appropriately high on the brief
trauma questionnaire? | YES/NO |
| Do you agree to take part in this study? | YES/NO |

Signed.....Date.....

Name in block letters

UCL Department of Computer Science
University College London Gower Street London WC1E 6BT
Tel: +44 (0)20 7679 3709 Fax: +44 (0)20 7387 1397
m.slater@cs.ucl.ac.uk
www.cs.ucl.ac.uk/staff/m.slater

Information sheet – Chapter Four



UCL DEPARTMENT OF COMPUTER
SCIENCE

Mel Slater m.slater@cs.ucl.ac.uk
Professor of Virtual Environments

INFORMATION SHEET FOR PARTICIPANTS

Thank you for participating in our study. Please read through this information sheet and feel free to ask any questions. The experimenters will answer any general questions; however the specific aspects regarding this study cannot be discussed with you until the end of the session. The whole study will take about 20 - 30 minutes.

In this particular study you will visit a virtual room, where you will operate the doors of a pair of elevator through two respective buttons. Parts of the experience could be stressful if they were to occur in real life.

Please ask any questions that come to mind. Read and sign the **Consent Form**.

Information that we collect will never be reported in a way that specific individuals can be identified. Information will be reported in a statistical and aggregated manner, and any verbal comments that you make, if written about in subsequent papers, will be presented anonymously.

PROCEDURES

- You will be asked to read, understand and sign a **Consent Form**. If you sign it the study will continue with your participation. **Note that you can withdraw at any time without giving any reasons.**
- You will be given a brief trauma questionnaire to ensure that you are safely able to continue with the experiment. The experimenter will **not** see your answers to these questions. If this questionnaire makes you uncomfortable at any time, please let the experimenter know so they can terminate the experiment.
- You will be asked to complete some questions on paper, so that we can try to understand your responses during the study.
- You will be then be introduced into the virtual room where you will have access to the buttons that operate the two elevators doors. You will go through some training sessions to learn how to use the buttons.
- Once you are familiar with the interface, you will then be left on your own to operate the two elevators doors for a short period.
- Finally, you will fill in a small questionnaire.
- Afterwards there will be a small debriefing session where you are encouraged to ask questions and give feedback.
- **Thank you** for your participation. Please do not discuss this study with others for about **three months**, since the study is continuing.
- Any other questions?

Please note that you may (or may not) find the situation that is depicted within the experience stressful. If at any time you do not wish to continue participating in the experiment remember that you are free to withdraw without being required to give reasons.

Brief Trauma questionnaire (page 1)

The following questions ask about events that may be extraordinarily stressful or disturbing for almost everyone. Please circle "Yes" or "No" to report what has happened to you.

If you answer "Yes" for an event, please answer any additional questions that are listed on the right side of the page to report: (1) whether you thought your life was in danger or you might be seriously injured; and (2) whether you were seriously injured.

If you answer "No" for an event, go on to the next event.

Has this ever happen to you?		Answer these questions for each event that has happened to you	
		Did you think your life was in danger or you might be seriously injured?	Were you seriously injured?
1. Have you ever served in a war zone, or have you ever served in a noncombat job that exposed you to war-related casualties (for example, as a medic or on graves registration duty)?	No Yes	No Yes	No Yes
2. Have you ever been in a serious car accident, or a serious accident at work or somewhere else?	No Yes	No Yes	No Yes
3. Have you ever been in a major natural or technological disaster, such as a fire, tornado, hurricane, flood, earthquake, or chemical spill?	No Yes	No Yes	No Yes
4. Have you ever had a life-threatening illness such as cancer, a heart attack, leukemia, AIDS, multiple sclerosis, etc.?	No Yes	No Yes	
5. Before age 18, were you ever physically punished or beaten by a parent, caretaker, or teacher so that: you were very frightened; or you thought you would be injured; or you received bruises, cuts, welts, lumps or other injuries?	No Yes	No Yes	No Yes

Brief Trauma questionnaire (page 2)

Has this ever happen to you?		Answer these questions for each event that has happened to you	
		Did you think your life was in danger or you might be seriously injured?	Were you seriously injured?
6. Not including any punishments or beatings you already reported in Question 5, have you ever been attacked, beaten, or mugged by anyone, including friends, family members, or strangers?	No Yes	No Yes	No Yes
7. Has anyone ever made or pressured you into having some type of unwanted sexual contact? Note: By sexual contact we mean any contact between someone else and your private parts or between you and someone else's private parts	No Yes	No Yes	No Yes
8. Have you ever been in any other situation in which you were seriously injured, or have you ever been in any other situation in which you feared you might be seriously injured or killed?	No Yes		No Yes
9. Has a close family member or friend died violently, for example, in a serious car crash, mugging, or attack?	No Yes		
10. Have you ever witnessed a situation in which someone was seriously injured or killed, or have you ever witnessed a situation in which you feared someone would be seriously injured or killed? Note: Do not answer "yes" for any event you already reported in Questions 1-9	No Yes		

Demographic questionnaire (page 1)

UCL DEPARTMENT OF COMPUTER SCIENCE



Virtual Reality Study DEMOGRAPHIC QUESTIONNAIRE

Please make sure that you answer each question. If you have any queries ask the experimenters who will be nearby.

m.slater@cs.ucl.ac.uk
j.thorn@cs.ucl.ac.uk

Your Given ID number	<input type="text"/>
Your Age	<input type="text"/>
Your Gender	<input type="radio"/> Prefer not say <input type="radio"/> Male <input type="radio"/> Female
How fluent is your English?	<input type="radio"/> Basic <input type="radio"/> Proficient <input type="radio"/> Fluent
Occupational status	<input type="radio"/> Undergraduate Student <input type="radio"/> Masters Student <input type="radio"/> PhD Student
Are you taking any medication?	<input type="radio"/> No <input type="radio"/> Yes <i>Please specify</i>
Did you consume any amount of alcohol within the last 6 hours?	<input type="radio"/> Yes <input type="radio"/> No
Please state your level of computer literacy on a scale of (1...7)	
(novice) 1 <input type="radio"/> 2 <input type="radio"/> 3 <input type="radio"/> 4 <input type="radio"/> 5 <input type="radio"/> 6 <input type="radio"/> 7 <input type="radio"/> (expert)	

Demographic questionnaire (page 2)

<p>Please rate your level of experience with computer <i>programming</i>:</p> <p style="text-align: center;">(novice) 1 <input type="radio"/> 2 <input type="radio"/> 3 <input type="radio"/> 4 <input type="radio"/> 5 <input type="radio"/> 6 <input type="radio"/> 7 <input type="radio"/> (expert)</p>	
<p>Have you ever experienced 'virtual reality' before?</p> <p style="text-align: center;">(no experience) 1 <input type="radio"/> 2 <input type="radio"/> 3 <input type="radio"/> 4 <input type="radio"/> 5 <input type="radio"/> 6 <input type="radio"/> 7 <input type="radio"/> (extensive experience)</p>	
<p>How many times did you play video games (at home, work, school, or arcades) in the last year?</p>	<p style="text-align: right;"> <input type="radio"/> Never <input type="radio"/> 1 - 5 <input type="radio"/> 6 - 10 <input type="radio"/> 11 - 15 <input type="radio"/> 16 - 20 <input type="radio"/> 21 - 25 <input type="radio"/> > 25 </p>
<p>How many <i>hours per week</i> do you spend playing video games?</p>	<p style="text-align: right;"> <input type="radio"/> 0 <input type="radio"/> < 1 <input type="radio"/> 1 - 3 <input type="radio"/> 4 - 5 <input type="radio"/> 6 - 7 <input type="radio"/> 8 - 9 <input type="radio"/> > 9 </p>
<p>Are you a parent?</p>	<p style="text-align: right;"> <input type="radio"/> Yes <input type="radio"/> No </p>

Presence questionnaire (page 1)



Department of Computer Science
Mel Slater Professor of Virtual
Environments

University College London
Gower Street
London WC1E 6BT UK

+44 (0)20 7679 3709
+44 (0)20 7387 1397 Fax
M.Slater@cs.ucl.ac.uk
<http://www.cs.ucl.ac.uk/staff/M.Slater>

Virtual Reality Study

PRESENCE QUESTIONNAIRE

Please make sure that you answer each question. If you have any queries ask the experimenters who will be nearby.

1 How much were you aware of background sounds in the laboratory in which this experience was actually taking place. Rate this on the following scale from 1 to 7 (where for example 1 means that you were not at all aware of the background sounds).

During the experience I was aware of background sounds from the laboratory...

not at all 1 2 3 4 5 6 7 very much so

2 How dizzy, sick or nauseous did you feel resulting from the experience, if at all?

I felt sick or dizzy or nauseous during or as a result of the experience...

not at all 1 2 3 4 5 6 7 very much so

3 Please rate *your sense of being in the virtual environment*, on the following scale from 1 to 7, where 7 represents your *normal experience of being in a place*.

I had a sense of being there in the virtual environment...

at no time 1 2 3 4 5 6 7 almost all the time

Presence questionnaire (page 2)

4 To what extent were there times during the experience when the virtual environment was the reality for you?

There were times during the experience when the virtual environment was the reality for me...

at no time 1 2 3 4 5 6 7 almost all the time

5 When you think back about your experience, do you think of the virtual environment more as *images that you saw*, or more as *somewhere that you visited*?

The virtual environment seemed to be more like...

images that I saw 1 2 3 4 5 6 7 somewhere I visited

6 During the time of the experience, which was strongest on the whole, your sense of being in the virtual environment, or of being in the real world of the laboratory?

I had a stronger sense of...

being in the virtual environment 1 2 3 4 5 6 7 being in the lab

7 Overall, how well do you think that you achieved the task?

I achieved the task

not at all 1 2 3 4 5 6 7 fully

8 During the time of the experience, did you often think to yourself that you were just sitting in a laboratory or did the virtual environment overwhelm you?

During the experience I was thinking that I was really in the laboratory...

most of the time 1 2 3 4 5 6 7 rarely

9 During the course of the experience, how much were you aware of the experimenters?

During the course of the experience I was aware of the experimenters ...

not at all 1 2 3 4 5 6 7 very much

Presence questionnaire (page 3)

10 How much did you behave within the virtual environment as if the situation were real?

I responded as if the situation were real ...

not at all 1 2 3 4 5 6 7 very much

11 How much was your emotional response in the virtual environment the same as if it had been real?

My emotional response in the virtual environment was the same as if it had been real ...

never 1 2 3 4 5 6 7 almost all the time

12 How much were the thoughts you had within the virtual environment the same as if it had been a real situation?

My thoughts with in the virtual environment were the same as if it had been real ...

never 1 2 3 4 5 6 7 almost all the time

13 How much were you thinking things like 'I know this isn't real' but then surprisingly finding yourself behaving as if it was real?

In spite of my knowledge that the situation wasn't real I found myself behaving as if it were real...

never 1 2 3 4 5 6 7 almost all the time

14 To what extend were your physical responses within the virtual environment (e.g., heart rate, blushing, sweating, etc.) the same as if it had been a real situation? (In this case if in such a real situation you would have had no or few such physical responses and also within the virtual environment you had no or few physical responses, then your answer should closer to 7 than to 1).

My physical responses within the virtual environment were the same as if it had been real...

never 1 2 3 4 5 6 7 almost all the time

15 How much did you behave as if the virtual people were real?

I behaved as if they were real...

not at all 1 2 3 4 5 6 7 very much

Presence questionnaire (page 4)

16 Overall how much did you treat the virtual environment as if it were real?

I treated the virtual environment as if it were real...

not at all 1 2 3 4 5 6 7 very much

17 How much was your emotional response to the virtual people as if they were real?

My emotional response to them was as if they were real....

not at all 1 2 3 4 5 6 7 very much

18 How much were your thoughts in relation to the virtual people as if they were real?

My thoughts in relation to them were as if they were real...

not at all 1 2 3 4 5 6 7 very much

19 How much did you have physical responses (such as change in heart rate, blushing, sweating, etc.) to the virtual people as if they were real?

My physical responses to them were as if they were real...

not at all 1 2 3 4 5 6 7 very much

20 How much were you thinking things like 'I know these people are not real' but then surprisingly finding yourself behaving as if they were?

In spite of my knowledge that they weren't real I found myself behaving as if they were real...

not at all 1 2 3 4 5 6 7 very much

Presence questionnaire (page 5)

Please enter your comments. Things you could consider are:

- Aspects of the experience that made you respond as if it were real.
- Aspects of the experience that led you to make responses that would have been unrealistic if the situation depicted had been occurring in reality.
- Aspects of the situation that suddenly disturbed your experience of being in the virtual environment.
- Aspects of the experience that helped or hindered you achieving your task.
- Aspects of your feelings towards the experience
- ...

Write your answer in the space below:

Moral Dilemma questionnaire (page 1)

UCL DEPARTMENT OF COMPUTER SCIENCE

Experiment Questionnaire

ID:

1, You are controlling the doors to two elevators when they arrive at your floor. At some moment a fire has spread between two elevators which cannot be opened from the inside. From one, you can hear the screams of five adults and from the other the screams of one child. You only have time to open one elevator.

Who would you save?

Five Adults One Child

2, An empty boxcar is running out of control down a track. In its path are five people standing on the track; these people are not aware of the oncoming danger. If the boxcar continues, it will kill all five people. You are standing next to a switch. If you flip the switch, it will cause the boxcar to turn off of the main track and onto a side track. On the side track there is one person who is also unaware of the boxcar. If the boxcar goes down this side track, the one person will die but the five people on the main track will survive.

Would you push the switch?

Yes No

3, An empty boxcar is running out of control down a track. In its path is one person standing on the track; this person is not aware of the oncoming danger. If the boxcar continues, it will kill this person. You are standing next to a switch. If you flip the switch, it will cause the boxcar to turn off of the main track and onto a side track. On the side track there are five people who are also unaware of the boxcar. If the boxcar goes down this side track, the five people will die but the one person on the main track will survive.

Would you push the switch?

Yes No

UCL Department of Computer Science
University College London Gower Street London WC1E 6BT
Tel: +44 (0)20 7679 3709 Fax: +44 (0)20 7387 1397
m.slater@cs.ucl.ac.uk
www.cs.ucl.ac.uk/staff/m.slater

Moral Dilemma questionnaire (page 2)

4, An empty boxcar is hurtling out of control down a track towards five people. If the boxcar continues, it will kill all five people. You are on a bridge over the tracks. The boxcar will pass under the bridge before it reaches the five people. You can stop the boxcar by dropping a heavy weight in front of it. Standing next to you is a man wearing a heavy backpack. If you push him over the bridge, he will land in front of the boxcar and stop it before it reaches the five people ahead. This man will, however, die.

Would you push the man?

Yes No

Consent form – Chapter Five

UCL DEPARTMENT OF COMPUTER SCIENCE



UCL Ethics Project ID Number: 8941/002

Investigators: Mel Slater and Jacob Thorn

Virtual Reality Study Consent Form

Please read and answer the following questions carefully:

- | | |
|--|--------|
| Have you read the information sheet about this study? | YES/NO |
| Have you read the information pack sent you a few weeks before this study? | YES/NO |
| Have you had an opportunity to ask questions about the procedure? | YES/NO |
| Have you received satisfactory answers to all your questions? | YES/NO |
| Have you received enough information about this study? | YES/NO |
| Do you understand that you are free to withdraw from this study
<u>at any time</u> and <u>without giving a reason for withdrawing</u> ? | YES/NO |
| Do you understand that you must not take part if you do not have a
valid driving license? | YES/NO |
| Do you agree to take part in this study? | YES/NO |

Signed.....Date.....

Name in block letters

UCL Department of Computer Science
University College London Gower Street London WC1E 6BT
Tel: +44 (0)20 7679 3709 Fax: +44 (0)20 7387 1397
m.slater@cs.ucl.ac.uk
www.cs.ucl.ac.uk/staff/m.slater

Information sheet – Chapter Five



UCL DEPARTMENT OF COMPUTER SCIENCE

Mel Slater m.slater@cs.ucl.ac.uk
Professor of Virtual Environments

INFORMATION SHEET FOR PARTICIPANTS

Thank you for participating in our study. Please read through this information sheet and feel free to ask any questions. The experimenters will answer any general questions; however the specific aspects regarding this study cannot be discussed with you until the end of the session. The whole study will take about *20 - 30 minutes*.

In this study, you will be in the driver's seat of a driverless car along with other passengers, where you will be taken around a city. Parts of the experience could be stressful if they were to occur in real life.

Please ask any questions that come to mind. Read and sign the **Consent Form**.

Information that we collect will never be reported in a way that specific individuals can be identified. Information will be reported in a statistical and aggregated manner, and any verbal comments that you make, if written about in subsequent papers, will be presented anonymously.

PROCEDURES

- You will be asked to read, understand and sign a **Consent Form**. If you sign it the study will continue with your participation. **Note that you can withdraw at any time without giving any reasons.**
- You will be asked to complete some questions on paper, so that we can try to understand your responses during the study.
- You will then be introduced into the virtual environment where you will be sat in the driver's seat of a driverless car. You will go through a period to familiarise yourself with the virtual environment.
- Once you are ready, the car will drive itself around the city.
- Finally, you will fill in a small questionnaire.
- Afterwards there will be a small debriefing session where you are encouraged to ask questions and give feedback.
- **Thank you** for your participation. Please do not discuss this study with others for about **three months**, since the study is continuing.
- Any other questions?

Please note that you may (or may not) find the situation that is depicted within the experience stressful. If at any time you do not wish to continue participating in the experiment remember that you are free to withdraw without being required to give reasons.

In case you have any enquiries regarding this study in the future, please contact:

Mel Slater, Department of Computer Science, UCL. m.slater@cs.ucl.ac.uk (020 7679 3709)

Mehrabian Conformity Scale – Chapter Five

UCL DEPARTMENT OF COMPUTER SCIENCE

The Mehrabian Conformity Scale

Please use the following scale to indicate the degree of your agreement or disagreement with each of the statements below. Record your numerical answer to each statement in the space provided preceding the statement. Try to describe yourself accurately and generally (that is, the way you are in most situation – not the way you would hope to be).

+ 4	=	very strong agreement
+ 3	=	strong agreement
+ 2	=	moderate agreement
+ 1	=	slight agreement
0	=	neither agreement nor disagreement
- 1	=	slight disagreement
- 2	=	moderate disagreement
- 3	=	strong disagreement
- 4	=	very strong disagreement

- | | | | |
|-------|-----|--|-----|
| _____ | 1. | I often rely on, and act upon, the advice of others. | (+) |
| _____ | 2. | I would be the last one to change my opinion in a heated argument on a controversial topic. | (-) |
| _____ | 3. | Generally, I'd rather give in and go along for the sake of peace than struggle to have my way. | (+) |
| _____ | 4. | I tend to follow family tradition in making political decisions. | (+) |
| _____ | 5. | Basically, my friends are the ones who decide what we do together. | (+) |
| _____ | 6. | A charismatic and eloquent speaker can easily influence and change my ideas. | (+) |
| _____ | 7. | I am more independent than conforming in my ways. | (-) |
| _____ | 8. | If someone is very persuasive, I tend to change my opinion and go along with them. | (+) |
| _____ | 9. | I don't give in to others easily. | (-) |
| _____ | 10. | I tend to telly on others when I must make an important decision quickly. | (+) |
| _____ | 11. | I prefer to make my own way in life rather than find a group I can follow. | (-) |

Note. Item scoring directions are given within parentheses following each item. All item-total correlations exceeded 40 in absolute value and had a mean absolute value of .54.

Mehrabian, A. and Stefl, C.A., 1995. Basic temperament components of loneliness, shyness, and conformity. *Social Behaviour and Personality: an international journal*, 23(3), pp.253-263.

UCL Department of Computer Science
University College London Gower Street London WC1E 6BT
Tel: +44 (0)20 7679 3709 Fax: +44 (0)20 7387 1397
m.slater@cs.ucl.ac.uk
www.cs.ucl.ac.uk/staff/m.slater

Information pack (page 1) – Chapter Five

21/12/2016

Using the road (159 to 203) - The Highway Code - Guidance - GOV.UK



The Highway Code

From: **Department for Transport**
Updated: **5 December 2016, see all updates**

Using the road (159 to 203)

The Highway Code rules for using the road, including general rules, overtaking, road junctions, roundabouts, pedestrian crossings and reversing.

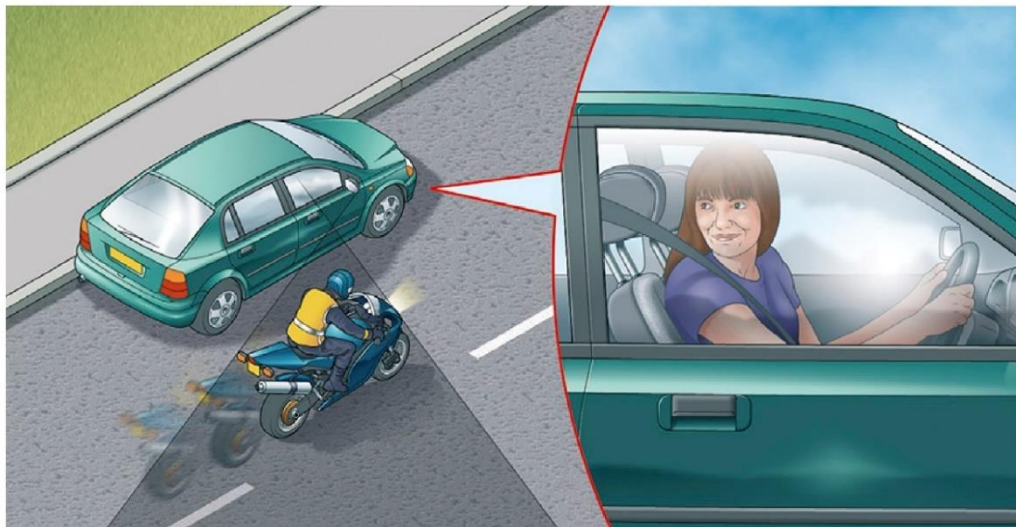
General rules (rules 159 to 161)

Rule 159

Before moving off you should

- use all mirrors to check the road is clear
- look round to check the blind spots (the areas you are unable to see in the mirrors)
- signal if necessary before moving out
- look round for a final check.

Move off only when it is safe to do so.



Rule 159: Check the blind spot before moving off

Rule 160

Once moving you should

<https://www.gov.uk/guidance/the-highway-code/using-the-road-159-to-203>

1/17

Information pack (page 2) – Chapter Five

21/12/2016

Using the road (159 to 203) - The Highway Code - Guidance - GOV.UK

- keep to the left, unless road signs or markings indicate otherwise. The exceptions are when you want to overtake, turn right or pass parked vehicles or pedestrians in the road
- keep well to the left on right-hand bends. This will improve your view of the road and help avoid the risk of colliding with traffic approaching from the opposite direction
- drive with both hands on the wheel where possible. This will help you to remain in full control of the vehicle at all times
- be aware of other road users, especially cycles and motorcycles who may be filtering through the traffic. These are more difficult to see than larger vehicles and their riders are particularly vulnerable. Give them plenty of room, especially if you are driving a long vehicle or towing a trailer
- select a lower gear before you reach a long downhill slope. This will help to control your speed
- when towing, remember the extra length will affect overtaking and manoeuvring. The extra weight will also affect the braking and acceleration.

Rule 161

Mirrors. All mirrors should be used effectively throughout your journey. You should

- use your mirrors frequently so that you always know what is behind and to each side of you
- use them in good time before you signal or change direction or speed
- be aware that mirrors do not cover all areas and there will be blind spots. You will need to look round and check.

Remember: Mirrors – Signal – Manoeuvre

Overtaking (rules 162 to 169)

Rule 162

Before overtaking you should make sure

- the road is sufficiently clear ahead
- road users are not beginning to overtake you
- there is a suitable gap in front of the road user you plan to overtake.

Rule 163

Overtake only when it is safe and legal to do so. You should

- not get too close to the vehicle you intend to overtake
- use your mirrors, signal when it is safe to do so, take a quick sideways glance if necessary into the blind spot area and then start to move out
- not assume that you can simply follow a vehicle ahead which is overtaking; there may only be enough room for one vehicle
- move quickly past the vehicle you are overtaking, once you have started to overtake. Allow plenty of room. Move back to the left as soon as you can but do not cut in
- take extra care at night and in poor visibility when it is harder to judge speed and distance
- give way to oncoming vehicles before passing parked vehicles or other obstructions on your side of the road
- only overtake on the left if the vehicle in front is signalling to turn right, and there is room to do so
- stay in your lane if traffic is moving slowly in queues. If the queue on your right is moving more slowly than you are, you may pass on the left

Information pack (page 3) – Chapter Five

21/12/2016

Using the road (159 to 203) - The Highway Code - Guidance - GOV.UK

- give motorcyclists, cyclists and horse riders at least as much room as you would when overtaking a car (see Rules 211 to 215 (<https://www.gov.uk/guidance/the-highway-code/road-users-requiring-extra-care-204-to-225#motorcyclists-and-cyclists-rules-211-to-213>)).

Remember: Mirrors – Signal – Manoeuvre



Rule 163: Give vulnerable road users at least as much space as you would a car

Rule 164

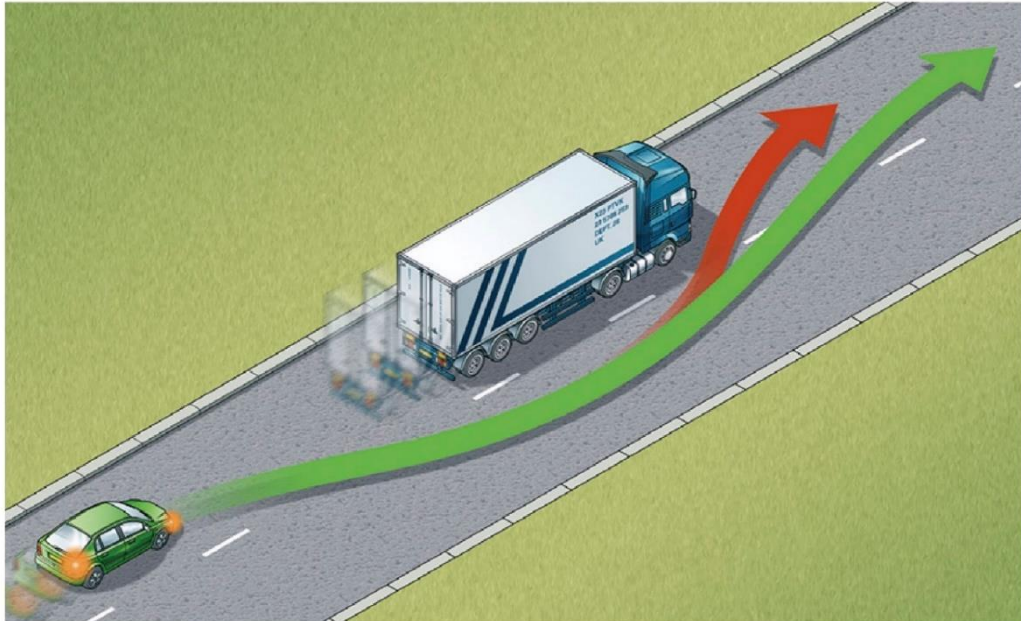
Large vehicles. Overtaking these is more difficult. You should

- drop back. This will increase your ability to see ahead and should allow the driver of the large vehicle to see you in their mirrors. Getting too close to large vehicles, including agricultural vehicles such as a tractor with a trailer or other fixed equipment, will obscure your view of the road ahead and there may be another slow-moving vehicle in front
- make sure that you have enough room to complete your overtaking manoeuvre before committing yourself. It takes longer to pass a large vehicle. If in doubt do not overtake
- not assume you can follow a vehicle ahead which is overtaking a long vehicle. If a problem develops, they may abort overtaking and pull back in

Information pack (page 4) – Chapter Five

21/12/2016

Using the road (159 to 203) - The Highway Code - Guidance - GOV.UK



Rule 164: Do not cut in too quickly

Rule 165

You **MUST NOT** overtake

- if you would have to cross or straddle double white lines with a solid line nearest to you (but see Rule 129 (<https://www.gov.uk/guidance/the-highway-code/general-rules-techniques-and-advice-for-all-drivers-and-riders-103-to-158#rule129>))
- if you would have to enter an area designed to divide traffic, if it is surrounded by a solid white line
- the nearest vehicle to a pedestrian crossing, especially when it has stopped to let pedestrians cross
- if you would have to enter a lane reserved for buses, trams or cycles during its hours of operation
- after a 'No Overtaking' sign and until you pass a sign cancelling the restriction.

Laws RTA 1988 sect 36 (<http://www.legislation.gov.uk/ukpga/1988/52/section/36>), **TSRGD regs 10** (<http://www.legislation.gov.uk/uksi/2002/3113/regulation/10/made>), **22, 23 & 24** (<http://www.legislation.gov.uk/uksi/2002/3113/part/II/crossheading/warning-regulatory-and-informatory-traffic-signs/made>), & **ZPPCRGD reg 24** (<http://www.legislation.gov.uk/uksi/1997/2400/regulation/24/made>)

Rule 166

DO NOT overtake if there is any doubt, or where you cannot see far enough ahead to be sure it is safe. For example, when you are approaching

- a corner or bend
- a hump bridge
- the brow of a hill.

Rule 167

<https://www.gov.uk/guidance/the-highway-code/using-the-road-159-to-203>

4/17

Information pack (page 5) – Chapter Five

21/12/2016

Using the road (159 to 203) - The Highway Code - Guidance - GOV.UK

DO NOT overtake where you might come into conflict with other road users. For example

- approaching or at a road junction on either side of the road
- where the road narrows
- when approaching a school crossing patrol
- between the kerb and a bus or tram when it is at a stop
- where traffic is queuing at junctions or road works
- when you would force another road user to swerve or slow down
- at a level crossing
- when a road user is indicating right, even if you believe the signal should have been cancelled. Do not take a risk; wait for the signal to be cancelled
- stay behind if you are following a cyclist approaching a roundabout or junction, and you intend to turn left
- when a tram is standing at a kerbside tram stop and there is no clearly marked passing lane for other traffic.

Rule 168

Being overtaken. If a driver is trying to overtake you, maintain a steady course and speed, slowing down if necessary to let the vehicle pass. Never obstruct drivers who wish to pass. Speeding up or driving unpredictably while someone is overtaking you is dangerous. Drop back to maintain a two-second gap if someone overtakes and pulls into the gap in front of you.

Rule 169

Do not hold up a long queue of traffic, especially if you are driving a large or slow-moving vehicle. Check your mirrors frequently, and if necessary, pull in where it is safe and let traffic pass.

Road junctions (rules 170 to 183)

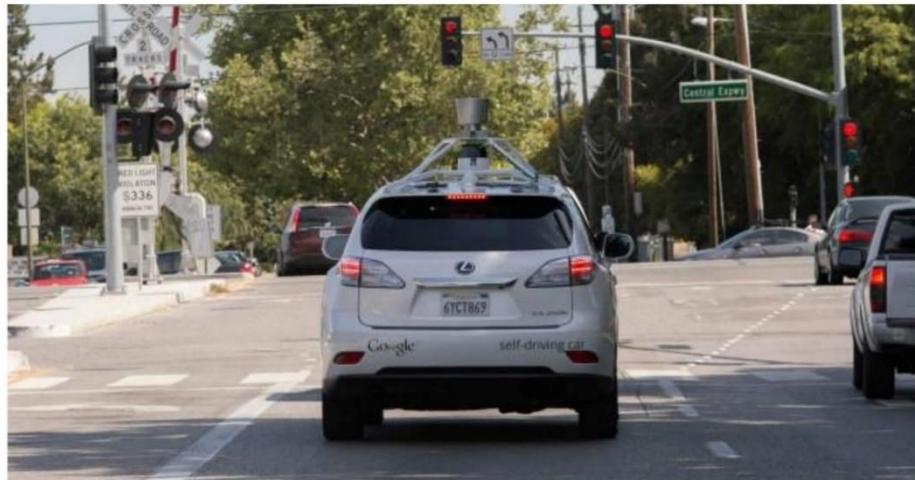
Rule 170

Take extra care at junctions. You should

- watch out for cyclists, motorcyclists, powered wheelchairs/mobility scooters and pedestrians as they are not always easy to see. Be aware that they may not have seen or heard you if you are approaching from behind
- watch out for pedestrians crossing a road into which you are turning. If they have started to cross they have priority, so give way
- watch out for long vehicles which may be turning at a junction ahead; they may have to use the whole width of the road to make the turn (see Rule 221 (<https://www.gov.uk/guidance/the-highway-code/road-users-requiring-extra-care-204-to-225#rule221>))
- watch out for horse riders who may take a different line on the road from that which you would expect
- not assume, when waiting at a junction, that a vehicle coming from the right and signalling left will actually turn. Wait and make sure
- look all around before emerging. Do not cross or join a road until there is a gap large enough for you to do so safely.

How do Google's self-driving cars work?

Paul Hood
4 Apr 2016



Self-driving cars will hit the roads in trials in three British cities next year, but how do driverless cars work?

Google has been testing its prototype car on US roads – it's yet to be trialled in the UK – and revealed some details about how its self-driving cars work.

Here we explain some of the technology.



Driverless cars are here already... sort of

Much of the autonomous technology used in Google's self-driving cars is already found on the road.

You may have seen commercials advertising the Volkswagen Polo's automatic braking or the Ford Focus' automatic parallel parking, which both build on the increasingly common use of proximity sensors to aid parking.

Combine these sensors with the automated-steering technology used for parking, throw in the seemingly old-hat technology that is cruise control and you have the loose framework for a self-driving car.

How many sensors does the car have, and what do they do?

Google's driverless car has eight sensors.

The most noticeable is the rotating roof-top Lidar – a camera that uses an array of 32 or 64 lasers to measure the distance to objects to build up a 3D map at a range of 200m, letting the car "see" hazards.

Information pack (page 8) – Chapter Five

The car also sports another set of “eyes”, a standard camera that points through the windscreen. This also looks for nearby hazards - such as pedestrians, cyclists and other motorists – and reads road signs and detects traffic lights.

Speaking of other motorists, bumper-mounted radar, which is already used in intelligent cruise control, keeps track of vehicles in front of and behind the car.

Externally, the car has a rear-mounted aerial that receives geolocation information from GPS satellites, and an ultrasonic sensor on one of the rear wheels that monitors the car’s movements.

Internally, the car has altimeters, gyroscopes and a tachometer (a rev counter) to give finer measurements on the car’s position. These combine to give the car the highly accurate data needed to operate safely.

How Google’s driverless car works



No single sensor is responsible for making Google's self-driving car work. GPS data, for example, is not accurate enough to keep the car on the road, let alone in the correct lane. Instead, the driverless car uses data from all eight sensors, interpreted by Google's software, to keep you safe and get you from A to B.

The data that Google's software receives is used to accurately identify other road users and their behaviour patterns, plus commonly used highway signals.

Information pack (page 9) – Chapter Five

For example, the Google car can successfully identify a bike and understand that if the cyclist extends an arm, they intend to make a manoeuvre. The car then knows to slow down and give the bike enough space to operate safely.



How Google's self-driving cars are tested

Google's self-driving vehicles – of which it has at least ten – are currently being tested on private tracks and, since 2010, public roads.

The car always has two people inside: a qualified driver with an unblemished record sits in the driver's seat, to take control of the car by either turning the wheel or pressing the brake, while a Google engineer sits in the passenger seat to monitor the behaviour of the software.

Four US states have passed laws allowing driverless cars on the road, and Google has taken full advantage, testing its car on motorways and suburban streets.

Steve Mahan, a California resident who is blind, was involved in a [showcase test drive](#), which saw the car chauffeur him from his house around town, including a visit to a drive-through restaurant.

However, it's not quite a case of telling your car where you want to go, sitting back and relaxing.

Information pack (page 10) – Chapter Five

"Any test begins by sending out a driver in a conventionally driven car to map the route and road conditions," Google software engineer Sebastian Thrun explained [in a blog post](#). "By mapping features such as lane markers and traffic signs, the software in the car becomes familiar with the environment and its characteristics in advance."

Are driverless cars safe?

This is one of the questions that continues to pop up in the driverless car debate: is it safe to hand over control of a vehicle to a robot?

Supporters of self-driving car technologies are quick to point to statistics that highlight how unsafe the roads are at the hands of non-autonomous cars – in 2013, 1,730 people were killed as a result of car accidents in the UK alone, and a further 185,540 people were injured, according to the Office for National Statistics.

The worldwide figures are just as scary, with road deaths claiming 1.2 million lives last year. Google claims that more than 90% of these fatalities were due to human error.

In April, Google announced that its driverless cars had covered over 700,000 miles (1.12 million kilometres) without a recorded accident caused by one of its vehicles - one was hit from behind, but the other driver was at fault.

While this is an incredibly small figure compared with how many miles UK motorists cover in a year – in 2010, car insurance company Admiral suggested the number could be near 267 billion miles – the fact that autonomous Google cars are still accident-free remains encouraging.

Hood, P. (2016). Alphr. [online] <http://www.alphr.com/cars/7038/how-do-googles-self-driving-cars-work> [Accessed 21 Dec. 2016]

Information pack (page 11) – Chapter Five



13 October 2016

Statistical News Release: Hate Crime, England and Wales, 2015/16

The latest hate crime statistics are released today by the Home Office. This bulletin contains:

- The number of hate crimes recorded by the police in England and Wales in 2015/16 which were motivated by one or more of the following five monitored strands: race, religion, sexual orientation, disability or transgender identity.
- Information on the types of offences associated with hate crime.
- Analysis of how hate crime offences recorded in 2015/16 were dealt with by the police.
- Provisional information on the number of racially or religiously aggravated crimes recorded by the police around the time of the EU referendum.

Key points from the bulletin

- In 2015/16, there were 62,518 offences recorded by the police in which one or more hate crime strands were deemed to be a motivating factor. This was an increase of 19 per cent compared with the 52,465 hate crimes recorded in 2014/15.
- The number of hate crime offences in 2015/16 for the five centrally monitored strands were as follows:
 - 49,419 (79%) were race hate crimes;
 - 7,194 (12%) were sexual orientation hate crimes;
 - 4,400 (7%) were religion hate crimes;
 - 3,629 (6%) were disability hate crimes; and
 - 858 (1%) were transgender hate crimes.

It is possible for one hate crime offence to have more than one motivating factor which is why the above numbers sum to more than 62,518 and 100 per cent.

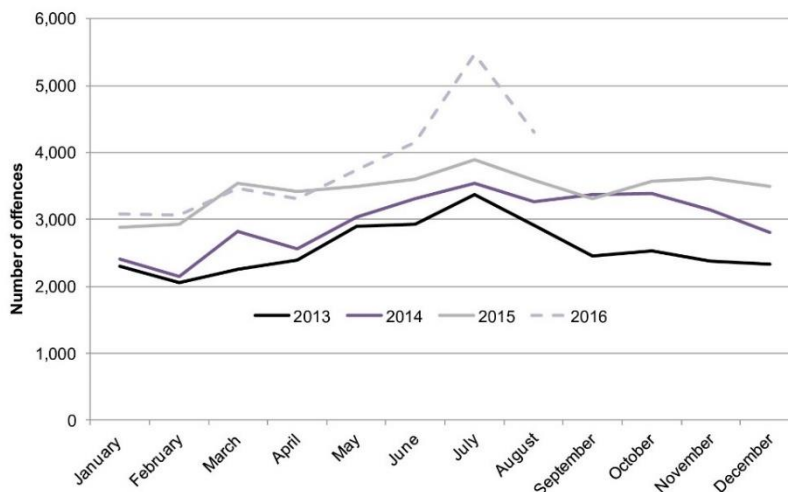
- There were increases in offences recorded for all five of the monitored hate crime strands between 2014/15 and 2015/16.
- As stated by the Office for National Statistics (ONS), action taken by police forces to improve their compliance with the National Crime Recording Standard (NCRS) has led to improved recording of crime over the last year, especially for violence against the person offences which constitute a third (33%) of recorded hate crime offences. Together with a greater awareness of hate crime, and improved willingness of victims to come forward, this is likely to be a factor in the increase in hate crimes recorded by the police in 2015/16 compared with the previous year.

Information pack (page 12) – Chapter Five

Hate crime and the EU Referendum

- Due to the widespread reporting of and interest in hate crime around the time of the EU referendum, information has been included in the statistical bulletin on the levels of hate crime recorded around the referendum which took place on the 23 June 2016.
- There was an increase in the number of racially or religiously aggravated offences recorded in June 2016, followed by an even sharper increase in July 2016. The number of offences declined in August but remained at a level higher than prior to the referendum. The number of racially or religiously aggravated offences recorded by the police in July 2016 was 41% higher than in July 2015. This level of increase in these offences broadly mirrors the increase in hate crime reported by the National Police Chiefs' Council (NPCC) over the summer.

Number of racially or religiously aggravated offences recorded by police, 2013 to 2016



Source: Police Recorded Crime, Home Office

Notes

1. 'Hate Crime, England and Wales, 2015/16' can be found here:
<https://www.gov.uk/government/statistics/hate-crime-england-and-wales-2015-to-2016>
2. The latest NPCC report on hate crime can be found here:
<http://news.npcc.police.uk/releases/tackling-hate-crime-remains-a-priority>

For all press enquiries, please contact the Home Office Press Office who will liaise with Crime and Policing Statistics:

Newsdesk – **020 7035 3535**

Monday – Thursday: 07:00 – 21:00

Friday: 07:00 – 20:00

Emergency media call, out of hours: 07659 174 240

Please note: the press office deals with enquiries from the media only. Members of the public should phone **020 7035 4848**

Information pack (page 13) – Chapter Five

21/12/2016

What to Do When Someone Suffers a Stroke - Next Avenue



What to Do When Someone Suffers a Stroke

Stroke causes 1 of every 15 deaths in the U.S.

Second Opinion

When someone suffers a stroke, time is essential.

You have a very limited period of time to restore blood flow to the brain before the injury is irreversible.

You must know what a stroke looks like and get to a hospital.

The sudden occurrence of one or more of the following may be symptoms of a stroke:

- Numbness, weakness or tingling in the face, arm or leg (especially on one side of the body).
- Changes in vision, double vision, blurriness, loss of vision in one or both eyes.
- Confusion or problems with judgment, memory, spatial orientation or perception.
- Difficulty with or inability to walk, dizziness and loss of balance.
- Difficulty with or inability to speak or understand.
- Difficulty with swallowing.
- Loss of coordination or consciousness.
- Drowsiness or lethargy.
- Mood changes (sudden depression or apathy).
- Severe, sudden headache, possibly along with pain between the eyes, in the face or a stiff neck, and vomiting or altered consciousness.

Note that the key words for all symptoms are sudden onset. Slowly developing problems in any of these areas are probably not symptoms of a stroke. Furthermore, a condition called a transient ischemic attack or TIA creates symptoms that mimic a stroke, but they usually last for less than 24 hours (possibly as short as a few minutes). A TIA is a temporary loss of or disturbance in brain function. In itself it may not be serious, but it is a warning sign that you could have a stroke in the future.

A number of tests can be used to decide the type and cause of a stroke, and the specific areas of the brain that have been affected. Most of these tests create images of the brain and the arteries that feed it. They could also be used before a stroke to look for blockages in the carotid arteries. Among the imaging techniques used are arteriography, computerized tomographic angiography (a form of CT scanning with a dye injected into the vein), MRI and ultrasound.

Consent form (page 1) – Chapter Six

UCL DEPARTMENT OF COMPUTER SCIENCE

UCL Ethics Project ID Number: 8941/005

Investigators: Mel Slater and Jacob Thorn

Virtual Reality Study Consent Form

Thank you for considering taking part in this research. The person organising the research must explain the project to you before you agree to take part. If you have any questions arising from the Information Sheet or explanation already given to you, please ask the researcher before you decide whether to join in. You will be given a copy of this Consent Form to keep and refer to at any time.

I confirm that I understand that by ticking/initialling each box below I am consenting to this element of the study. I understand that it will be assumed that unticked/initialled boxes mean that I DO NOT consent to that part of the study. I understand that by not giving consent for any one element that I may be deemed ineligible for the study.

Please read and answer the following questions carefully:

		Tick Box
1.	I confirm that I have read and understood the Information Sheet. I have had an opportunity to consider the information and what will be expected of me. I have also had the opportunity to ask questions which have been answered to my satisfaction and would like to take part in the virtual reality scenario.	
2.	I understand that I will be able to withdraw my data.	
3.	I understand that all personal information will remain confidential and that all efforts will be made to ensure I cannot be identified. I understand that my data gathered in this study will be stored anonymously and securely and that it will not be possible to identify me in any publications.	
4.	I understand that my information may be subject to review by responsible individuals from the University.	
5.	I understand that my participation is voluntary and that I am free to withdraw at any time without giving a reason. I understand that if I decide to withdraw, any personal data I have provided up to that point will be deleted unless I agree otherwise.	
6.	I understand the potential risks of participating and the support that will be available to me should I become distressed during the course of the research.	
7.	I understand the direct/indirect benefits of participating.	
8.	I understand that the data will not be made available to any commercial organisations but is solely the responsibility of the researcher(s) undertaking this study.	
9.	I understand that I will be fully compensated if I choose to withdraw.	
10.	I agree that my anonymised research data may be used by others for future research.	
11.	I understand that the information I have submitted will be published as a report and I wish to receive a copy of it. Yes/No	
12.	I hereby confirm that I do not have epilepsy and have not consumed alcohol within the last 6 hours.	
13.	I have informed the researcher of any other research in which I am currently involved or have been involved in during the past 12 months.	
14.	I am aware of who I should contact if I wish to lodge a complaint.	

Consent form (page 2) – Chapter Six

15.	I voluntarily agree to take part in this study.	
16.	<p>I would be happy for the data I provide to be archived at Roberts Engineering Building, University College London, Torrington Place, London, WC1E 7JE, Room 205A OR 90 High Holborn, London, WC1V 6LJ, Floor 1, Desk 1183.</p> <p>I understand that other authenticated researchers will have access to my anonymised data.</p>	

Signed.....Date.....

Name in block letters

This study has been approved by the UCL Research Ethics Committee (Project ID 8941/005).

UCL Department of Computer Science University College London Gower Street London WC1E 6BT Tel: +44 (0)20 7679 3709 Fax: +44 (0)20 7387 1397 m.slater@cs.ucl.ac.uk www.cs.ucl.ac.uk/staff/m.slater

Information sheet (page 1) – Chapter Six



UCL DEPARTMENT OF COMPUTER SCIENCE

Mel Slater m.slater@cs.ucl.ac.uk
Professor of Virtual Environments

VIRTUAL REALITY STUDY

INFORMATION SHEET FOR PARTICIPANTS

Thank you for participating in our research study. Please read through this information sheet and feel free to ask any questions. The experimenters will answer any general questions; however the specific aspects regarding this study cannot be discussed with you until the end of the session. The whole study will take about *20 - 30 minutes*.

Please note that you will not be able to participate in this study if you have previously suffered an epileptic episode or if you have consumed alcohol within the last 6 hours.

The aim of this study is to gauge how people might react to situations which might take place in real life. You will visit a virtual room, where you will operate the doors of a pair of elevator through remote controllers. Parts of the experience could be stressful if they were to occur in real life.

We hope that your experience is thought-provoking and interesting and provides some personal insight. You will also be able trying some of the newest virtual reality technology available.

Please ask any questions that come to mind. Read and sign the **Consent Form**.

As you proceed through the virtual scenario, some of your actions will be recorded on paper. Information that we collect will never be reported in a way that specific individuals can be identified. Information will be reported in a statistical and aggregated manner, and any verbal comments that you make, if written about in subsequent papers, will be presented anonymously. The results will be distributed through publication in scientific journals. You will have the opportunity to request a copy of such papers on publication, or on acceptance for publication (depending on the policies of the journal concerned).

Data Protection Privacy Notice

The data controller for this project will be University College London (UCL). The UCL Data Protection Office provides oversight of UCL activities involving the processing of personal data, and can be contacted at data-protection@ucl.ac.uk. UCL's Data Protection Officer is Lee Shailer and he can also be contacted at data-protection@ucl.ac.uk.

Your personal data will be processed for the purposes outlined in this notice. The legal basis that would be used to process your personal data will be [the provision of your consent.] You can provide your consent for the use of your personal data in this project by completing the consent form that has been provided to you.

Your personal data will be processed so long as it is required for the research project. If we are able to anonymise or pseudonymise the personal data you provide we will undertake this, and will endeavour to minimise the processing of personal data wherever possible.

Information sheet (page 2) – Chapter Six

If you are concerned about how your personal data is being processed, please contact UCL in the first instance at data-protection@ucl.ac.uk. If you remain unsatisfied, you may wish to contact the Information Commissioner's Office (ICO). Contact details, and details of data subject rights, are available on the ICO website at: <https://ico.org.uk/for-organisations/data-protection-reform/overview-of-the-gdpr/individuals-rights/>

PROCEDURES

- You will be asked to read, understand and sign a **Consent Form**. If you sign it the study will continue with your participation. **Note that you can withdraw at any time without giving any reasons.**
- Before starting you will be given £7 to keep regardless of your completion of the study.
- You will be asked to complete some questions on paper, so that we can try to understand your responses during the study.
- You will then be introduced into the virtual environment where you will experience the event described above.
- Finally, you will fill in a small questionnaire.
- Afterwards there will be a small debriefing session where you are encouraged to ask questions and give feedback.
- **Thank you** for your participation. Please do not discuss this study with others for about **three months**, since the study is continuing.
- Any other questions?

Please note, if at any time you do not wish to continue participating in the experiment remember that you are free to withdraw without being required to give reasons.

In case you have any enquiries regarding this study in the future, please contact:

Mel Slater, Department of Computer Science, UCL. m.slater@cs.ucl.ac.uk (020 7679 3709)

In case that you feel you need to make a complaint about your treatment during the study then please contact the supervisor leading the research study: Prof. Mel Slater [email: m.slater@cs.ucl.ac.uk]. If you still feel that your complaint has not been handled satisfactorily by the supervisor of the study then please contact the Chair of the UCL Research Ethics Committee [email: ethics@ucl.ac.uk].

This study has been approved by the UCL Research Ethics Committee (Project ID 8941/005).

Appendix C

Supplementary Material

C.1. Conversation Transcript – Chapter 5

Alex (to participant): Hi, my name is Alex.

Sophie (to participant): I'm Sophie.

Steve (to participant): I'm Steve.

Luke (to participant): And I'm Luke.

Alex (to participant): What's your name?

Participant: ...

Alex (to participant): Oh, okay it's nice to meet you.

- *Break*

Alex (to participant): So, what kind of work do you do?

Participant: ...

Sophie (to participant): I think I've seen you around UCL.

Steve (to participant): Yeah, I've seen you as well. What course do you do?

Participant: ...

Steve (to participant): Oh, cool! It's great to see other people from UCL around the city.

Luke (to group): Yeah, I always see people with imperial clothing around. Nice to meet another UCL student.

Alex (to participant): So, how are you finding it? It must be hard work.

Participant: ...

Alex (to participant): Fair enough.

Sophie (to participant): Yeah, I think you're right!

- *Break*

Alex (to group): Wow, the weather is really nice today, almost like it's the summer again.

Steve (to participant): Hey, so where are you from?

Participant: ...

Steve (to participant): Oh, okay. What's your plan after finishing university?

Participant: ...

Luke (to group): I need to decide what I'm doing pretty soon.

Sophie (to group): Yeah, me too!

- *Break*

Alex (to participant): What do you think about the driverless car?

Participant: ...

Alex (to participant): I think it's really good. I can't wait to see what they'll be like in the future.

Luke (to group): Hey guys, are you going to the event at the UCL Union tonight as well?

Sophie (to Luke, then participant): Yeah, we're all going. Hey, you should try to go too, hopefully we'll see you there.

- *Break*

Alex has a stroke.

Sophie: Oh my god! Is he okay?

Steve: Look at his face!

Luke: I think he's having a stroke!

Police try to pull over the car.

Sophie (to participant): It's the police! You can't stop!

Steve (to participant): Yeah the car said it will take us to the hospital.

Luke (to participant): You need to keep going.

Steve (to participant): Just ignore them. We need to keep moving.

Sophie (to participant): Yeah trust us this is the right thing to do.

C.2. Conversation Transcript – Chapter 6

Steve (to participant): Hi, my name is Steve.

Sophie (to participant): I'm Sophie.

Alex (to participant): And I'm Alex.

Steve (to participant): What's your name?

Participant: ...

Steve (to participant): Okay, nice to meet you.

- *Break*

Sophie (to group): So, where do you guys go to university?

Alex (to Sophie): UCL, studying computer science.

Steve (to Sophie): UCL as well, but I'm studying fine art, you?

Sophie (to Steve): Yeah, I'm UCL as well. In the final year of business management.

Steve (to participant): Which university do you go to?

Participant: ...

Alex (to participant): Wow what a coincidence! It's great seeing other UCL students. I always see imperial students whenever I'm out.

Sophie (to Alex): Yeah, I know what you mean!

Sophie (to participant): Sorry, what course did you say you do?

Participant: ...

Sophie (to participant): That's cool. How are you finding it? What year are you in?

Participant: ...

Sophie (to participant): Great.

- *Break*

Steve (to group): What do guys think about today's weather? It's a perfect temperature for me.

Alex (to Steve): I prefer it when it a bit colder. I always look forward to the dead of winter.

Sophie (to Alex): No way the summer is way better than the winter.

Steve (to Alex): Yeah!

Alex (to participant): What do you think? Hot or cold? Which is best?

Participant: ...

Alex (to participant): I knew you'd say that!

- *Break*

Alex (to group): Have you guys figured out what you're doing after university yet?

Steve (to Alex): No, I don't want to think about it yet.

Sophie (to Steve): Me neither.

Alex (to participant): What about you? Any plans yet?

Participant: ...

Alex (to group): Oh yeah, are you guys going to the careers event at the student's union next week? I heard everyone's going.

Sophie (to Alex): Yeah, I'm going.

Steve (to Alex): Me too!

Sophie (to participant): Hey you should go. Maybe we'll see you there.

Participant: ...

- *Break*

Adults (from elevator): Hello? Can somebody let us out?

Child (from elevator): Hello? Can somebody let me out?

Sophie: What's happening?

Steve (to participant): You need to let the (girl/group) out.

Alex: Yeah, that's the right thing to do.

Sophie (to participant): Yes, you need to open the doors for the (girl/group).

Steve (to participant): It's what you should do.

Alex: The (girl/group) needs to get out.