



universität  
wien

# MASTERARBEIT

Titel der Masterarbeit

## To kill or not to kill

The role of empathy and personal distress in moral decision making

Verfasserin

Mag. phil. Elisabeth Engl, Bakk. phil.

Angestrebter akademischer Grad

Master of Science (MSc)

Wien, Juni 2011

Studienkennzahl lt. Studienblatt: A 066 013

Studienrichtung lt. Studienblatt: Joint Degree Programme MEi:CogSci Cognitive Science

Betreuerin / Betreuer: Univ.-Prof. Dr. Claus Lamm

## **Acknowledgements**

I am immensely grateful to Dr. Jasminka Majdandžić, who offered me help and guidance at every step of our study. Without her ideas and insightful knowledge, this project would not have materialised the way it did. My thanks also go out to my supervisor, Prof. Dr. Claus Lamm, who not only offered crucial concepts and criticism, but to whom I could always come with questions. Last but not least, I thank my sister Magdalena for proof-reading yet another thesis, and a special someone for his great help with quantitative matters.

# Table of Contents

|   |    |
|---|----|
| <b>Chapter 1</b> Introduction .....   | 1  |
| <b>Chapter 2</b> Moral decision making: the state of research .....   | 2  |
| 2.1 Some general principles of moral decision making .....  | 2  |
| 2.2 Philosophical theories: the rule of reason over the passions? .....   | 4  |
| 2.3 Neuroscientific findings: Emotion, context, and humanisation in moral decision making<br>.....                                  | 8  |
| 2.3.1 Methods of studying moral decision making .....   | 8  |
| 2.3.2 Neuroscientific perspectives on the role of emotions in moral decision making ....  | 10 |
| 2.3.3 Explanations for the behavioural response pattern to moral dilemmas .....   | 16 |
| 2.3.4 Humanisation of the victim: a decisive factor? .....  | 18 |
| <b>Chapter 3</b> Using emotion ratings to gain insight into feelings of empathy and personal<br>distress: a behavioural study ..... | 23 |
| 3.1 Background.....   | 23 |
| 3.2 Hypotheses.....   | 26 |
| 3.3 Experimental design .....   | 27 |
| 3.3.1 Participants .....  | 27 |
| 3.3.2 Part I: Victim priming and dilemmas.....  | 28 |
| 3.3.3 Part II: Emotion rating scales: evaluating the period of decision making.....   | 32 |
| 3.3.3.1 Pilot study: exploring labels for emotions.....   | 34 |
| 3.3.4 Part III: Rating of victim attributes .....   | 36 |

|   |           |
|---|-----------|
| 3.3.5 Part IV: Graphical dilemma.....                     | 37        |
| 3.3.6 Individual difference measures in empathy.....      | 41        |
| 3.4 Data analysis .....                                   | 41        |
| <b>Chapter 4 Results.....</b>                             | <b>45</b> |
| <b>Chapter 5 Discussion .....</b>                         | <b>59</b> |
| <b>Chapter 6 Conclusion .....</b>                         | <b>66</b> |
| References.....   | 69        |
| Appendix A Victim priming stories.....                    | 74        |
| Appendix B Moral dilemmas.....                            | 79        |
| Appendix C The Interpersonal Reactivity Index (IRI) ..... | 81        |
| Appendix D Extended abstract.....                         | 83        |
| Appendix E Zusammenfassung .....                          | 86        |
| Appendix F Curriculum vitae .....                         | 89        |

# List of Figures

- Figure 1: The two most famous experimental moral dilemmas ..... 3
- Figure 2: The rational model of moral reasoning ..... 5
- Figure 3: Several fundamental views on the role of emotions in moral judgement.. ..... 12
- Figure 4: Haidt’s social intuitionist model of moral judgement ..... 14
- Figure 5: The moral brain as seen from neuroimaging studies ..... 15
- Figure 6: Activation in brain areas connected with emotional processing ..... 17
- Figure 7: The main brain areas associated with empathy ..... 20
- Figure 8: Human-graphical dilemma condition showing schematic drawings of humans in distress ..... 39
- Figure 9: Control (text-graphical) condition in which the schematic drawings have been replaced with text ..... 39
- Figure 10: Utilitarian decisions in humanised and neutral dilemma conditions ..... 45
- Figure 11: Personal distress in humanised and neutral dilemma conditions ..... 46
- Figure 12: Empathic concern in humanised and neutral dilemma conditions ..... 47
- Figure 13: Potential victim humanisation in humanised and neutral dilemma conditions ..... 48
- Figure 14: Dilemma difficulty in humanised and neutral dilemma conditions ..... 49
- Figure 15: Perceived difficulty of those humanised and neutral dilemmas that were decided on in a utilitarian way ..... 50
- Figure 16: Rating of potential victims in humanised and neutral dilemma conditions on scales of ‘humanness’ attributes ..... 51
- Figure 17: Rating of potential victims in humanised and neutral dilemma conditions on scales of similarity and connectedness ..... 54
- Figure 18: Rating of potential victims in humanised and neutral dilemma conditions according to whether participants’ opinions about them influenced their decisions ..... 55
- Figure 19: Personal distress ratings for graphical and text dilemma conditions ..... 56
- Figure 20: Empathic concern ratings for graphical and text dilemma conditions ..... 57

Figure 21: Potential victim humanisation ratings for graphical and text dilemma conditions .58

Figure 22: Dilemma difficulty ratings for graphical and text dilemma conditions .....59

## List of Tables

|  |    |
|--|----|
| Table 1: Correlations between participants' responses to different emotion ratings scales .....                            | 52 |
| Table 2: Correlations between participants' IRI scale values and their emotion rating values during decision moments ..... | 53 |

## **Abstract**

Experimental moral dilemmas often involve a decision whether to 'kill' one potential victim in order to save several others from certain death. The theory of utilitarianism proposes that most people will aim for maximum benefit for the greatest number of people. However, previous experiments show that in some dilemmas the potential victim is consistently saved, whereas in others the lone person is predominantly sacrificed. This disparity in utilitarian response patterns is suggested to result from differences in how strongly the potential victim is seen as a human being. Such 'humanisation' can be experimentally induced by referring to or reflecting about the hypothetical potential victim's state of mind. Humanised potential victims have indeed been previously shown to be sacrificed less often in moral dilemmas. In the current behavioural study, we examined the effect of humanisation on emotional aspects of moral decision making. It was hypothesised that the tendency to spare the potential victim is related to increased feelings of empathic concern and / or self-related feelings of distress that participants experience while deciding on a dilemma. The extent of involvement of these emotional complexes, however, remained to be determined. To this end, self-reported emotion rating scales were used to record ratings of empathic concern, personal distress, and successful humanisation. Moral judgements involving humanised as compared to neutral persons seemed to elicit higher levels of personal distress, rather than empathic concern. However, self-report scales should be complemented by psychophysiological measures in future studies, providing more objective measures of emotional responses.



## Chapter 1 Introduction

Moral behaviour is all about what we ought to do, but unfortunately there are many cases where there are no prescribed sets of right or wrong answers. These cases are called moral dilemmas. Contrary to what reasoning-based explanations predict, moral dilemmas are not always decided on by using utilitarianism, the concept of judging according to maximum benefit for the greatest number of people (Bartels, 2008). In experimental psychology, a typical moral dilemma involves hypothetical decisions whether to sacrifice the few in order to save the many (Cushman et al., 2006). However, whereas decision patterns for some moral dilemmas tend to follow utilitarian logic, they nearly uniformly disregard utilitarianism in others. What can account for these differences?

Recent evidence from neuroscience and moral psychology proposes that affectively laden intuitive responses, i.e. emotions, are at least as involved in moral judgement as reason. Two of the most important emotional concepts associated with moral judgement are empathy, i.e. sharing others' affective states (Singer & Lamm, 2009), and personal distress. In a previous behavioural and fMRI study, half of the potential victims were 'humanised', in other words described with a reference to his or her mental states, in a moral dilemma situation. FMRI data showed higher activation in brain areas associated with emotion processing and empathy as well as reduced utilitarianism in dilemmas involving such humanised victims (Majdandžić et al., in prep.). However, that study could not determine to what extent those differences can be attributed to empathic concern and personal distress, respectively.

The present study seeks to close that gap. The behavioural method of self-reported emotion ratings grouped into personal distress, empathic concern, and successful humanisation categories was adapted from Batson (1987) and applied to Majdandžić et al.'s paradigm. Participants were asked to make a decision on four different dilemmas and rate their emotions for each decision phase. The first hypothesis was that, as in Majdandžić et al.'s study, humanised potential victims would be sacrificed less often than neutral ones. Possibly because of using fewer dilemma items, this result could not be corroborated. However,

humanised dilemmas did elicit higher ratings of humanisation, confirming the effectiveness of our manipulation, and showed a trend towards increasing personal distress and more weakly, empathic concern. This provided an answer to our second, open, hypothesis which stated that some mixture of these emotions would be involved during the decision process. It follows that personal distress seems to be more strongly implicated than empathic concern.

A separate research question dealt with the differences in emotion ratings between a graphical dilemma condition using schematic representations of the hypothetical persons involved and a control, text-only, version of the dilemma. The graphical condition also shows an increase in personal distress, but not empathy.

Beyond evaluating the presented hypotheses, this thesis also seeks to point out the need for methodological refinement and interdisciplinary research. The findings from previous research, and therefore the scope of this thesis, lie at the crossroads of cognitive psychology, cognitive neuroscience, and philosophy. It will be shown that neuroscientific and psychological insights helped to modify earlier philosophical concepts. However, it will also be shown that experimental findings should always be questioned as to their validity, and a continuous improvement of experimental methods is called for in order to shed further light on the role of emotions in moral judgement.

## **Chapter 2 Moral decision making: the state of research**

### **2.1 Some general principles of moral decision making**

Making a moral decision is a multi-dimensional process. As we will see, automatic intuitions and conscious reasoning interact to let a human being arrive at moral judgements, which Haidt (2001, p.817) defines as

‘evaluations (good vs. bad) of the actions or character of a person that are made with respect to a set of virtues held to be obligatory by a culture or subculture’.

Moral decision making has most fruitfully been studied by depriving a person of easy and automatic decision making strategies; in other words, by using a moral dilemma. By definition, a dilemma does not have a unique, optimal solution. Instead, it can be described as choosing between two problematic outcomes. Two famous moral dilemmas that have in many variants been used in a plethora of experiments are the so-called *trolley* (or *switch*) and *footbridge* problems (Greene & Haidt, 2002). The basic version that both problems have in common involves telling the participant that an out-of-control trolley will hit and kill five people, unless they sacrifice the life of one other person. In the trolley dilemma, the participant can divert the trolley on another set of tracks by operating a lever. In that case, the trolley would just kill one person instead of five. In the footbridge version, the participant can decide to push a person onto the tracks, stopping the trolley from killing the five other people (no self-sacrifice is allowed for the participant).

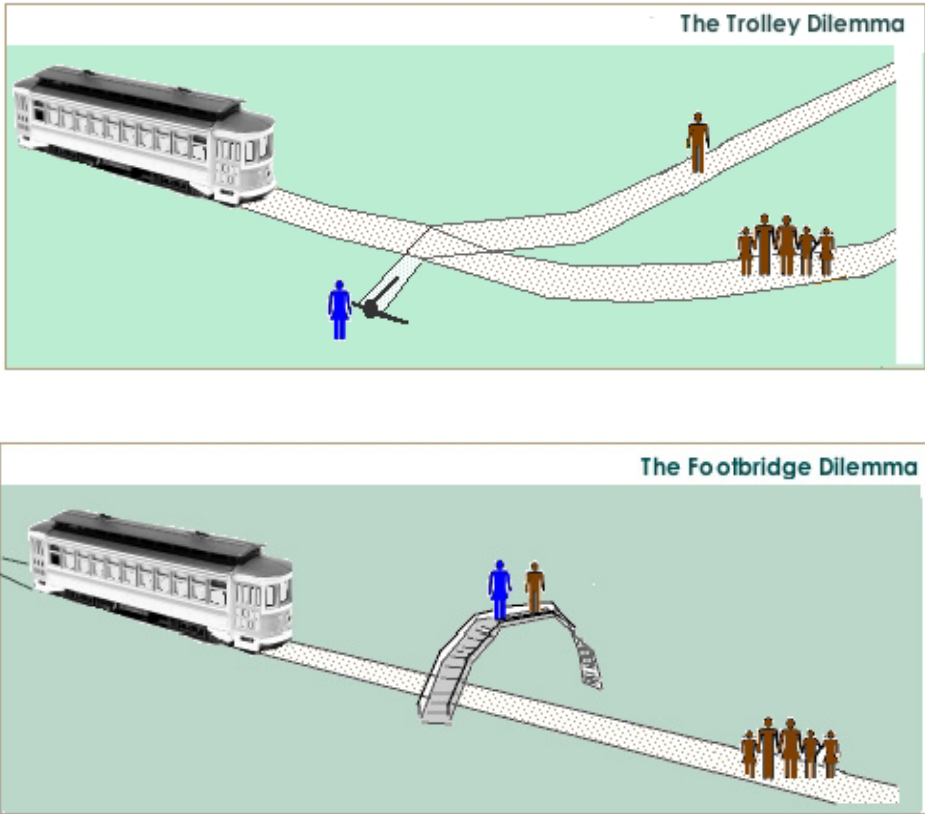


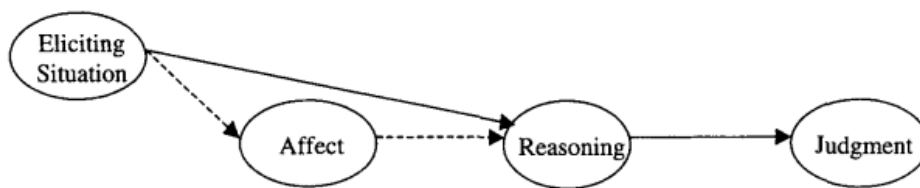
Figure 1: The two most famous experimental moral dilemmas (taken from [http://web.lemoyne.edu/~hevern/psy340\\_11S/lectures/psy340.12.1.emotions.html](http://web.lemoyne.edu/~hevern/psy340_11S/lectures/psy340.12.1.emotions.html), 9.5.2011)

Rational or 'utilitarian' thinking would imply that in both cases, sacrificing one person for the gain of a group of others would be the solution of choice. After all, a utilitarian choice is defined as 'the decision to perform an action directed toward achieving the greatest advantage and the minimum disadvantage for aggregate welfare' (Fumagalli et al., 2010). Thus, utilitarianism mandates promoting 'the greatest good for the greatest number' (Bartels, 2008, p. 382), where each person's welfare counts equally and 'good' is defined as the maximum amount of pleasure and happiness and the minimum amount of pain (Casebeer & Churchland, 2003).

However, countless experiments have shown that this rule of rationality is very often violated; and it seems not randomly so. When asked if it was morally acceptable to operate the lever in the trolley problem, around 80% people answer with yes (Cushman et al., 2006, Greene & Haidt, 2002). In contrast, when asked whether it was acceptable to push a large person onto the tracks to stop the trolley if that was the only option to save the other five people, only 25% of people respond that it is (Cushman et al., 2006) This discrepancy of moral judgement lies at the core of the central issue of moral philosophy: how do we evaluate a potential action according to its moral implications, and why do these implications differ for actions with similar outcomes?

## **2.2 Philosophical theories: the rule of reason over the passions?**

For a long time in Western history, philosophers have tended to view a priori rational reasoning as the dominant factor in moral decision making (Greene et al., 2001). It was supposed that, while emotions could come into play in the decision process, a moral judgement was primarily formed by weighing arguments relating to the different aspects of a problem, and then coming to a conclusion based on that reflection.



**Figure 2: The rational model of moral reasoning (from Haidt, 2001, p.815)**

In line with the importance of reason, Western philosophy has traditionally emphasised negative aspects of emotions<sup>1</sup>. Not looked upon kindly by the stoic and platonic philosophers, ‘the passions’ were seen as disturbances binding humans to the animalistic and to the materialistic, as opposed to the divine, world (Haidt 2001). However, cultivation and control of emotion did play a role in the so-called virtue theory, of which Plato and Aristotle were the first proponents. Virtue theory addresses moral questions in terms of what character traits people must possess, rather than merely what actions they must take, in order to achieve moral behaviour. The emphasis therefore lies on ‘knowing how’ to be virtuous (Casebeer & Churchland, 2003).

The suspicion of emotions continued with the rise of Christianity, where there was a tendency to view emotions as dangerous and a first step into the realm of sinfulness (Haidt, 2001). Additionally, the ideas of Ancient Greece were revived when philosophers of the Age of Enlightenment such as Descartes and Kant propagated reason as the only way to understand the world and, in Kant’s view, arrive at a sense of ethics (ibid.). However, one notable exception was Hume, who wanted reason to be ‘the slave of the passions’ (Hume, 1739/2007). Hume likened moral judgements to aesthetic ones in that both were derived from immediate, ‘built-in’ feelings of approval or disapproval towards good and bad acts,

---

<sup>1</sup> There are many definitions of the term ‘emotion’ that emphasise different aspects. Kleinginna and Kleinginna (1981, p.355) propose the following broad definition:

‘Emotion is a complex set of interactions among subjective and objective factors, mediated by neural-hormonal systems, which can (a) give rise to affective experiences such as feelings of arousal, pleasure/displeasure; (b) generate cognitive processes such as emotionally relevant perceptual effects, appraisals, labeling processes; (c) activate widespread physiological adjustments to the arousing conditions; and (d) lead to behavior that is often, but not always, expressive, goal-directed, and adaptive.’

respectively. Only emotion, in Hume's view, is a catalyst for action and goal selection. Using reason alone, humans may infer that a tragedy is about to happen, but they will only act to prevent it because they care and feel horrified, ashamed or disgusted.

19<sup>th</sup> century psychology also acknowledged the role of emotions in moral judgement. Freudian psychology asserts that post-hoc rationalization takes place to render unconscious desires socially acceptable (Freud, 1900 in Haidt, 2001). A few decades later, the completely different theoretical direction of behaviourism also disregarded moral reasoning and instead argued that moral behaviour consisted of the acts that a society happened to reward or punish (Skinner, 1971 in Haidt, 2001).

Moral psychology, however, then took another turn towards favouring reason. At the height of the 'cognitive revolution' of the 1960s that rejected behaviourism, the influential moral psychologist Kohlberg was interested in how moral judgement developed during childhood. He presented children at different stages of their development with moral dilemmas. While younger children acted egoistically, 'role taking', i.e. taking the perspective of other people, emerged as crucial in the development towards more sophisticated problem-solving techniques as the child matured (Kohlberg, 1971). Claiming that affect was indeed present in decision making, but could not distinguish between moral and immoral, Kohlberg attributed that capacity to reason alone. Since the 1980s, however, Kohlberg's influence has declined. Instead, theories such as the social intuitionist model (see section 2.3.2) that advocate a greater integration of affect have gained in importance (Greene & Haidt, 2002).

Apart from the debate about the importance of reason and emotion, moral philosophy has also been separated on another dimension. This is the issue of what constitutes a moral act, intention, or decision. First, the so-called consequentialism (of which utilitarianism is a variant) states that whether certain acts or intentions should be classified as good or bad should be entirely determined by the consequences they bring about (Alexander & Moore, 2007). The main goal for all consequentialists is to select the action that will result in the 'best' outcome. Of course, this requires consequentialists to specify what exactly constitutes 'good' outcomes,

which has been an issue of contested debate. The good for some means an increase in pleasure or happiness, while for others, the distribution of goodness is a part of goodness itself (ibid.).

On the other hand, deontology ('the study of duty'), a concept whose foundations can be traced back to Immanuel Kant, is fundamentally rule-based. Its proponents claim that there are cultural and evolutionary adaptive rules that all actions must adhere to in order to qualify as moral. This means that some actions, such as consciously harming an innocent person, are not allowed regardless of circumstances (Bartels, 2008). While consequentialism and deontology can result in the same action selection, consequentialists would hold that rules can be overturned if that results in a 'better' outcome.

These different takes on moral cognition do not necessarily have to exclude each other. In this author's opinion, the most plausible synthesis of the different theories results from series of behavioural studies conducted by Bartels (2008). His main point is that moral cognition is 'messy', i.e. influenced by context that can override deontological rules such as 'it is forbidden to kill intentionally', makes use of different simultaneous processes such as intuition and deliberation, and recruits representations of both deontological constraints and utilitarian considerations.

From a neuroscientific perspective, it is crucial to see that deontology and consequentialism place an emphasis on different cognitive processes. Therefore, the choice of philosophical framework can greatly influence neuroscientific findings. Greene et al. (2004) see fundamental rules of deontology, such as 'do not knowingly kill an innocent person', as evolutionary adaptive socio-emotional responses to moral violations. The abstract calculations that utilitarianism demands, however, could have only arisen after the evolution of those frontal lobe structures that support cognitive control (ibid.). For example, Moore et al. (2008) found that people who had greater working-memory capacity showed higher and more consistent levels of utilitarian responses in some moral dilemmas. Rules are not always unambiguous. They can conflict, such as when 'save lives' and 'do not kill' are both accepted

by an agent (Broeders et al., 2011). Due to environmental priming, conflicting rules may not be equally accessibly cognitively when a decision is pressing. For example, a dilemma can be phrased so that the concept of saving lives is emphasised ('the only way of saving five lives is to push the man off the bridge'), which results in the tendency to comply with the life-saving principle (Broeders et al., 2011).

## **2.3 Neuroscientific findings: Emotion, context, and humanisation in moral decision making**

### **2.3.1 Methods of studying moral decision making**

Moral decision making has now been studied from many different angles and with diverging methods. As we have seen in the previous chapter, philosophy first shaped the debate about moral judgement. Starting in the last century, a variety of other methods - mainly making use of moral dilemma paradigms - set out to challenge many philosophical assumptions, such as the role of emotions in moral judgement.

Behavioural experiments began in the 20th century and were the method of choice before brain imaging gave a new set of tools to researchers. An example is Kohlberg's (1971) work on moral development in children.

In addition, studies involving a lack of empathy and other emotions in pathologies such as autism and antisocial personality disorder or, as it was formerly called, psychopathy (e.g. Frith, 2001; Blair et al., 1997; Kiehl et al., 2001) as well as studies on patients with brain lesions provide another angle on moral behaviour. For example, Koenigs et al. (2007), building on studies by Damasio (1994), showed that patients with damage to the ventromedial prefrontal cortex, which is implicated in social emotion generation, exhibited very high levels of utilitarian responses in moral judgements. Recording the electrical activity of single neurons is also a method that, among other findings, showed how neurons in the



ventromedial prefrontal cortex are involved in controlling and correcting reward-related and punishment-related behaviour, and thus in emotion (Rolls, 2004).

Electroencephalogram (EEG) studies have also been used to link different evoked electrical potentials to differences in susceptibility to moral emotions. As an example, Bauer and Hesselbrock (1999) proposed a link between activity in the ventromedial prefrontal cortex (inferred by measuring the P300 event-related potential) and moral conduct as a result of testing children with self-control disorders.

Recently, a variety of functional magnetic resonance imaging (fMRI) studies have emerged researching moral decision making. Such neuroimaging studies are largely responsible for showing that brain areas associated with emotion are involved in moral decision making (Greene & Haidt, 2002).

Another method involves stimulating brain areas by using electrical current. As an example, Fumagalli et al. (2010) applied transcranial direct current stimulation (tDCS) over one of the main brain areas associated with rational decision making, the ventral prefrontal cortex (among others). Under some conditions, utilitarian decision making increased for both males and females (with a stronger increase in females) when the ventral prefrontal cortex was stimulated, suggesting that stimulating this area might increase rational behaviour in comparison to emotional processes.

It can be seen that contemporary research in moral decision making is an integration of different methods and is becoming truly interdisciplinary. Today, findings from evolutionary and cognitive psychology, cognitive neuroscience, and moral philosophy all add to the debate about what moral behaviour is and how it develops. Possibly the most important finding has been the vital role that emotions play in moral judgement, which will be discussed in the following section.

### **2.3.2 Neuroscientific perspectives on the role of emotions in moral decision making**

The current state of research in moral psychology and neuroscience proposes a solution to the riddle of the differing responses to the footbridge and trolley dilemmas: the role of intuitive affective responses. It is argued that, while reason does play a role in moral decision making, rational deliberation in decision making exists in conjunction with affective intuition and emotion (Greene & Haidt, 2002).

As we have seen, the different average responses to the footbridge and the trolley dilemmas show that rational utilitarianism is not the 'default mode' of decision making, but just one of several possibilities that depend on the situational context.

Haidt's 'social intuitionist model' (2001), synthesising evidence from evolutionary biology, neuroscience, and cognitive psychology, provides compelling evidence that moral reasoning is not the deciding factor that leads to moral judgement, but rather a post-hoc construction formed after a judgement has been reached. Haidt calls his model intuitionist because he sees moral judgement as a kind of automated reaction formed through evolutionary ('killing others is bad') or cultural ('one should follow the commands of God') processes, akin to perception (Haidt, 2001). In contrast to effortful and slower reasoning, intuitions are characterised by quick, mostly unreflected cognitive responses to a stimulus (Kahnemann, 2003). One strand of research regards intuitions as being formed by reinforcing bodily emotional experiences triggered by experiences in the world, until thinking about something alone elicits the emotional response ('somatic marker theory'; Damasio, 1994). Embodied cognition, i.e. the notion that our bodily circumstances and interactions with the world shape our perception and cognition of it (Wilson, 2002), is also seen by some to play a part in forming moral intuitions (Haidt, 2001). Sensory and motor processes, according to this view, provide the framework and constraints for the mind's interactions with the environment.

For instance, concepts such as the association that pure and clean food is good because it is safe to eat, but must be guarded against contamination, can translate into the world of moral purity when such a concept is attributed to children or young adults (Haidt, 2001). However,

moral intuitions are not only learned, but are evolutionary (selective) adaptations of social species that regulate sharing, authority, and reciprocal altruism (Greene & Haidt 2002; Haidt, 2001). Different cultures then reinforce different aspects from a great pool of potential moral behaviour according to a complex web of traditions and living circumstances. Some might emphasise the qualities of individual autonomy, while others stress community ethics. Children then acquire these values mostly implicitly and by imitating their peers and, to a lesser extent, parents (Haidt, 2001).

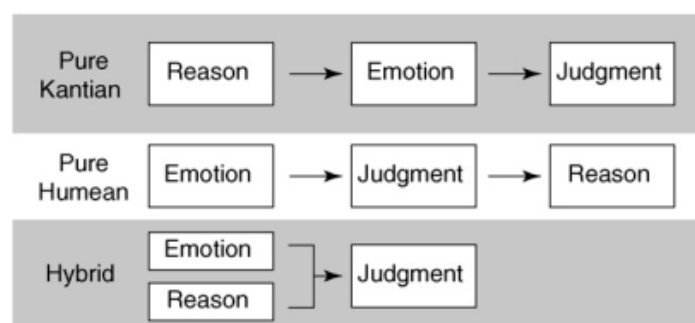
Intuitive reactions are accompanied by affective experiences, and Haidt follows previous research that sees those emotions as the direct cause of moral judgement (Kagan, 1984 in Haidt, 2001). Haidt also views the process of forming a judgement as a social experience in which post-hoc rationalisations are used to justify one's intuitions and influence those of others (Haidt, 2001). However, it is important to emphasise that it is not Haidt's intention to replace one extreme with another. He considers two processes - reasoning and intuition – to be involved in moral judgement, one of which has been overstated in its importance in previous moral philosophy and psychology. What Haidt questions is the causality of reasoning in moral judgement. He demonstrates this point with experiments in which offensive, but harmless actions such as cleaning a toilet with the national flag were judged as universally morally wrong, even when the participants acknowledged that no harm was done (Haidt et al., 1993 in Haidt, 2001). For Haidt, quick intuitions that arise in the conscious mind without an awareness of their underlying processes form the core of decision making and are followed by slower, at least partly conscious rationalisations and reflections.

This view is supported by a variety of research showing that large parts of social cognition are automated, unconscious processes (Bargh & Chartrand, 1999; Moll et al., 2003). Eisenberg (2000) agrees and sees embarrassment, guilt and shame, which are centred on the perception of self, as well as empathy as the main 'moral' emotions. Moll et al. (2003) cite guilt, gratitude and compassion as examples of prosocial moral emotions, whereas he views contempt, indignation and xenophobia as moral emotions that act to promote group dissolution and social reorganisation.

However, there are voices of dissent disagreeing with the notion that emotions are necessary for moral judgement. Huebner et al. (2008) argue that emotions accompany moral judgements, but do not necessarily constitute them. Instead they supposedly only point our attention to the morally salient stimuli in our environment and motivate action. Emotions follow from ‘a fast, unconscious process that operates over causal-intentional representations’ that is supposed to cause moral judgements (Huebner et al., 2008, p.5). This author thinks that the question of whether emotions constitute moral judgements or are merely related to them is worth pursuing. Huebner et al. point out correctly that the precise moment at which the effects of emotions occur is unknown. However, no alternative model has been established plausibly enough (Huebner et al. merely sketch the outline of a new theory) to warrant dismissing emotions as a probable constituent of moral judgement.

Haidt acknowledges that in some cases, a conscious effort of reasoning can override intuitive reactions and alter moral judgement. This fits to the findings of Greene et al. (2007) that in cases where people refuse to push a man off a footbridge in order to save five lives, more automatic emotional processes are involved, whereas utilitarian decisions seem to make more use of reasoning processes (as indicated by increased response time).

Thus, moral judgement arises out of hybrid factors. A comparison of this view with the philosophical takes on moral judgement discussed in section 2.2 is seen in figure 3:



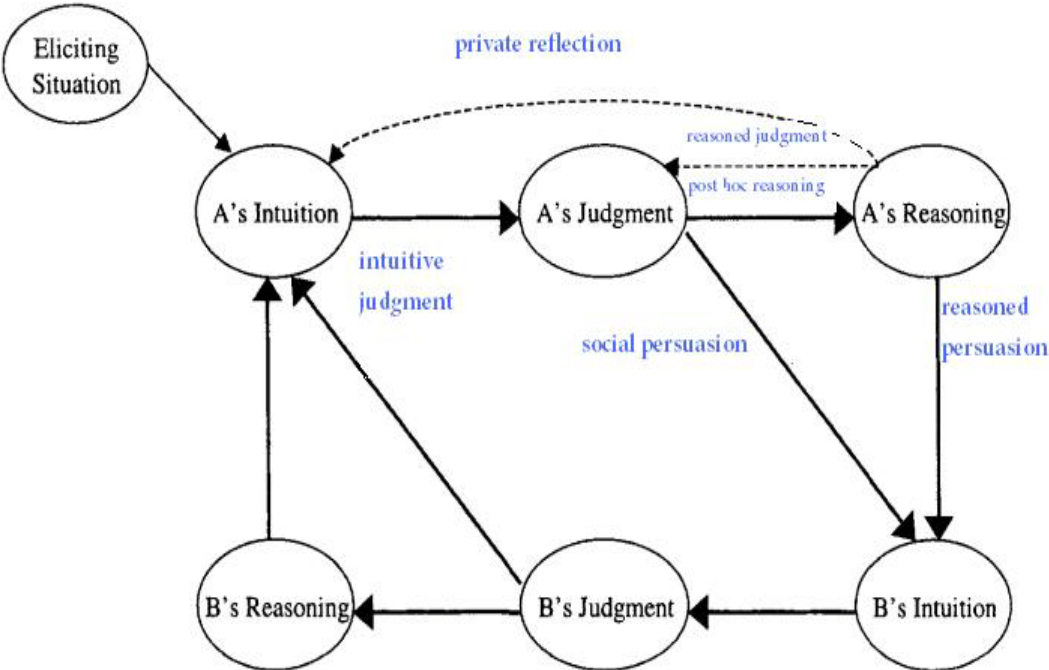
**Figure 3: Several fundamental views on the role of emotions in moral judgement. The ‘hybrid’ model could be seen as Haidt’s (2001) and Bartels’ (2008) concepts of an intertwined process of emotion and reason (graphics adapted from Huebner et al., 2008, p.2).**

The main real causative effect of reason seems to be social, i.e. the possibility of influencing the intuitions of others by making moral judgements (Haidt, 2001). Another, but according to Haidt not primary, factor that changes initial intuitions are subsequent intuitions that arise mainly through perspective taking. Putting oneself into others' shoes enables another emotional reaction that can compete with the primary intuition; a selection is then made either by following the strongest intuition or let reason act as a judge (Haidt, 2001). Just one of many examples for automated moral judgement in a social context is the 'I agree with people I like' heuristic (Chaiken, 1980). Basically, if a person you like makes a moral judgement about another person, and there is no conflict for you regarding previous positive feelings for that other person, this will affect your own intuition and you will generally side with the friend. Only when conflicting intuitions (say a prior friendship with the person who is judged) come into play, effortful and conscious reasoning takes place (Haidt, 2001).

Even then, reasoning is mostly driven by motivations (ibid.). It takes the role of striving to attain coherence with previous intuitions and the already established cultural worldview. Additionally, reason is suggested to be mainly useful for justifying current social goals, for example agreeing with people one likes, rather than searching for truth (Chen & Chaiken, 1999).

In any case, moral actions seem to correlate more with emotion than reasoning (Haidt, 2001). Psychopaths are prime examples of well-functioning reasoning abilities joined with deficits in affective capacity. They understand the harm that actions can cause to others and know which actions are considered wrong by cultural consensus, but can execute them nevertheless because they are not burdened by attachment to others, sympathy, or remorse (Blair, 1995). Patients with damage to the ventromedial prefrontal cortex exhibit similar emotional unresponsiveness and indecisiveness when making real-life decisions (Damasio, 1994). Even in healthy participants, factors such as stereotyping can influence decision making. Cikara et al. (2010) concluded that potential victims that are associated with 'low-competence' and 'low-warmth', such as homeless people or drug addicts, are hypothetically sacrificed significantly more often than others. This is a reminder that moral valuations do not

exist in a vacuum, but are context-dependent; the context therefore has to be controlled for in any experiment involving moral decision making.

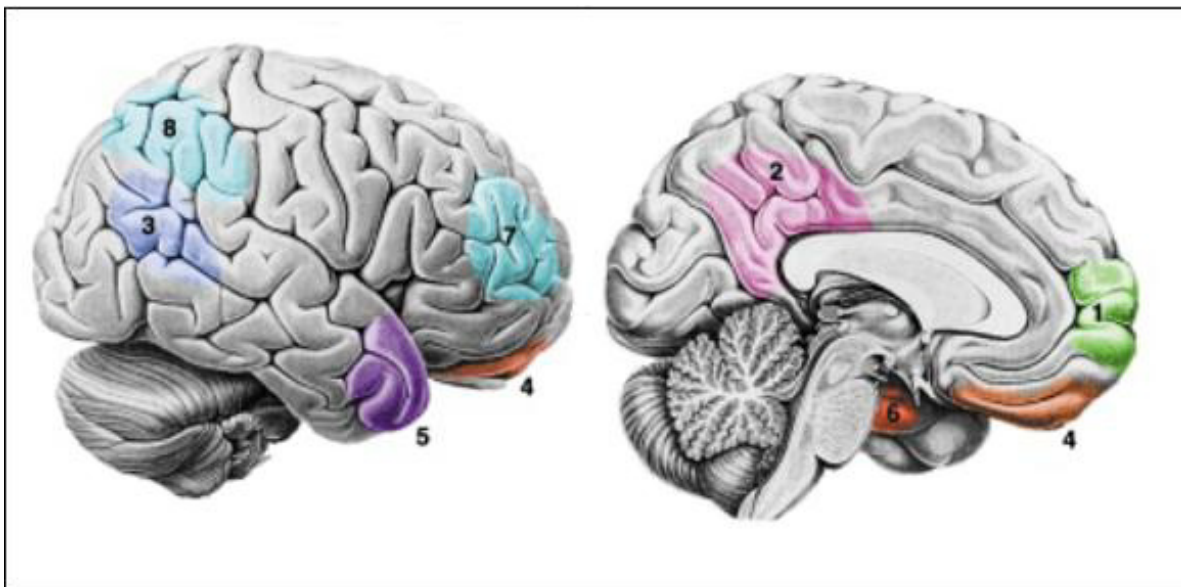


**Figure 4: Haidt's social intuitionist model of moral judgement (adapted from Haidt, 2001, p.815)**

How, then, are affective intuitions implemented in the brain? Moral-emotional processing is a distributed system. There is no brain area that is specifically and exclusively 'moral' (Casebeer & Churchland, 2003). Rather, areas associated with taking the perspective of others (theory of mind) and the processing of all emotions can contribute to moral judgement to varying extent (Greene & Haidt, 2002; Moll et al., 2005).

Therefore, a great variety of brain areas are now thought to be involved in the very different aspects and functions of moral cognition. A selection of the main areas is shown in figure 5; their most relevant functions for moral judgement are explained below (outline taken from Greene & Haidt, 2002):

1. Medial frontal gyrus: integration of emotion into decision making, attributing mental states to others (Farrow et al., 2001; Frith, 2001)
2. Posterior cingulate, precuneus, retrosplenial cortex: integration of emotion, imagery, and memory (Fletcher et al., 1995)
3. Superior temporal sulcus, inferior parietal lobe: representations of intentional movements, involved in several aspects of theory of mind such as taking the self-perspective (Gallagher & Frith, 2003)
4. Orbitofrontal/ventromedial frontal cortex: representation of reward/punishment value, control of inappropriate behaviour
5. Temporal pole: linking emotion to segregated and processed perceptual inputs, theory of mind (Olson et al., 2007)
6. Amygdala: assessment of reward/punishment value (rapid and predominantly negative)
7. Dorsolateral prefrontal cortex and
8. Parietal cortex: working memory, other non-emotional cognitive functions



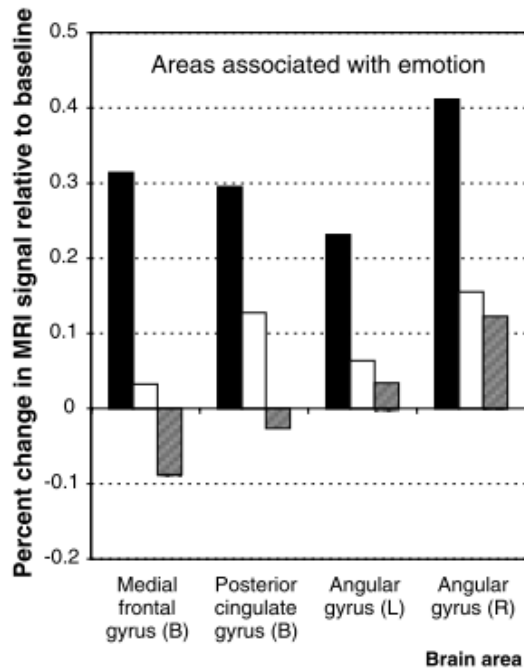
**Figure 5: The moral brain as seen from neuroimaging studies. 1=medial frontal gyrus, 2=posterior cingulate, precuneus, retrosplenial cortex, 3=superior temporal sulcus, inferior parietal lobe, 4=orbitofrontal ventromedial frontal cortex, 5=temporal pole, 6=amygdala, 7=dorsolateral prefrontal cortex, 8=parietal lobe (Greene & Haidt, 2002, p.521).**

### **2.3.3 Explanations for the behavioural response pattern to moral dilemmas**

The previous overview indicates that intuitions and emotions, interacting with conscious reasoning, are responsible for the differences in responses to moral dilemmas. But what does this mean in a concrete sense; why does sacrificing a person for the greater good seem right for some moral dilemmas, but not for others? For instance, why is it not considered morally acceptable to leave a dying person by the roadside because of the fear that the car's leather seats might be damaged, whereas spending money on luxury goods rather than donating to charity is (Greene, 2003)? Greene argues that the difference lies in 'up-close-and-personal' versus impersonal situations. As human cognition is a mixture of domain-specific emotional and social responses that have developed through evolution and domain-independent reasoning capacities, humans can distinguish between personal and impersonal situations (Greene & Haidt 2002). 'Personal' for Greene and Haidt means that the moral violation causes serious bodily harm to a specific person in such a way that the harm inflicted on that person is an act of an agent's will, not just a shift of an existing threat from another party ('ME HURT YOU'). According to these criteria, using a switch to reroute a trolley can be thought of as 'editing', rather than 'authoring', harm. Therefore, the sense of agency that is involved in pushing a person onto a set of tracks is lacking and the trolley dilemma would not be called personal.

Functional magnetic resonance imaging (fMRI) studies corroborate the view that there is a difference between personal and impersonal situations. Brain areas such as the angular gyrus, the medial frontal gyrus, and the posterior cingulate gyrus are activated on a higher level in 'personal' moral dilemmas as compared to 'impersonal' and altogether 'non-moral' ones (Greene et al., 2001). Although not part of the classical 'emotion' networks in the brain, fMRI research has implicated those areas in emotion processing (Maddock, 1999; Teasdale et al., 1999).





**Figure 6: Activation in brain areas connected with emotional processing (measured with fMRI) for ‘moral-personal’, ‘moral-impersonal’, and ‘non-moral’ dilemmas (Greene et al., 2001, p.2106)**

Furthermore, giving the ‘morally appropriate’ answer to a situation such as the footbridge dilemma that was characterised as ‘personal’ elicited reaction times (the time to make a decision) up to two seconds longer than giving the same answer to an ‘impersonal’ dilemma. Greene et al. (2001) credit this to emotional interference. Consequently, when a participant chose to respond that sacrificing the person in a ‘personal’ dilemma was morally inappropriate, thus giving an answer more congruent with emotional processing, reaction times plunged. Reaction times for ‘impersonal’ dilemmas were higher when sacrificing the victim was judged inappropriate, suggesting that this was an emotionally incongruent decision.

Dilemmas are thus not all equally likely to engage emotional processing and therefore influence moral judgement in different ways. However, as Greene et al. also acknowledge, this distinction between ‘personal’ and ‘impersonal’ may be useful, but is still rigid and lacks refinement in regard to which factors really engage emotional involvement.

Cushman et al. (2006) follow Greene’s argumentation and state that ‘personalness’, meaning close and direct contact with the victim, changes utilitarian decisions in the direction of a

preference to save the victim. The most recent integration of this previous research is Greene et al.'s (2009) notion of 'personal force', meaning that an agent directly impacts a victim with his or her muscles or a rigid object that directly transfers the agent's muscular force. Pushing somebody with one's own hands or a stick, for example, would be considered personal force, whereas pulling a trigger would not (Greene et al., 2009; Moore et al., 2008). Physical contact is not the same as personal force, as Greene et al. demonstrated in a variant of the footbridge dilemma that used a pole instead of hands to push the victim. This paradigm resulted in the same effect as the variant in which hands were used. As the pole increased the distance between agent and victim, the concept of personal force is also not identical to spatial proximity to the victim. When Greene et al. asked subjects if they found it morally acceptable to push the victim from the footbridge in those different variants of the dilemma, no effect regarding spatial proximity or physical contact was found; any significant difference in responses was due to personal force, which made sacrificing the victim less appealing (Greene et al., 2009). An effect of intentional harm, i.e. whether a victim is sacrificed in order to save others or died merely as a side effect of some other action, has also been proposed as a determinant of response differences, first by Aquinas (2006, in Greene et al., 2009). This effect was indeed measured in conjunction with the concept of personal force, but not independently (Greene et al., 2009).

#### **2.3.4 Humanisation of the victim: a decisive factor?**

Majdandžić et al. (in prep.) hypothesise that these different aspects of 'personalness' can be summed up as a concept of humanisation, i.e. how strongly a potential victim is perceived as a human being. Perceived humanisation would then account for the astonishing difference in utilitarian moral judgements in such paradigms as the trolley and the footbridge dilemmas. Majdandžić et al.'s idea fits with the 'ME HURT YOU' model of Greene and Haidt (2002), who argue that there must be a 'you', i.e. an individual that is unmistakably experienced as such by the agent, in any moral dilemma to make it personal.

To test this hypothesis, Majdandžić et al. used a moral dilemma paradigm consisting of four blocks of six dilemmas each. The dilemmas were identical to those used in the current study and followed the well-established pattern of letting participants choose between not harming one person at the expense of several others' lives, or sacrifice that person in order to save the others' lives (see appendix B). 'Humanisation', i.e. stimulating the participant to see the potential victims as human beings with thoughts and feelings (mental states), was induced by presenting brief priming stories<sup>2</sup> of the potential victims that implicitly referred to their mental states. An episode of the potential victims' lives was described, either by referring to their mental states (humanised condition) or by mentioning factual information, such as a victim's occupation, daily routine, or possessions, only (neutral condition). Subsequently, participants had to answer two multiple-choice questions that required them either to take the perspective of the person described (humanised condition) or not (neutral condition). Thus, two (eight in total) victim priming stories and questions were presented to the 40 male participants before each block of six moral dilemmas (24 in total), in which the primed potential victims could either be saved or sacrificed.

Majdandžić et al. were interested in two measurements. First, behavioural utilitarian response patterns for the dilemmas were recorded. It was hypothesised that potential victims in humanised conditions would be saved more often than those in neutral conditions. Second, fMRI was used to assess activations in response to dilemmas with humanised potential victims. In humanised conditions, stronger involvement of areas associated with emotional conflict was predicted.

Majdandžić et al. found that participants whose potential victims had been humanised indeed made significantly fewer ( $t(39) = 2.27, p = 0.028$ ) utilitarian decisions. During the response phase of the questions about the victims, brain areas associated with mentalising<sup>3</sup>, such as the precuneus, the temporal pole, and the temporoparietal junction, also showed higher

---

<sup>2</sup> The victim priming stories, like the dilemmas, were identical to those used in the current study and can be found in appendix A.

<sup>3</sup> Imagining the mental states of other people, also called 'mentalising' or 'theory of mind', means making (mostly automatic) inferences to other people's desires, beliefs, long-term dispositions, and short-term emotional states in order to anticipate their actions (Frith & Frith, 2006).

activation if the victims were humanised (Majdandžić et al., in prep.). This is a confirmation of the claim that the victims were perceived as ‘human beings with thoughts and feelings’.

Additionally, increased activity was found in brain areas associated with emotional conflict, emotion regulation, and emotion reappraisal during the decision phases in humanised conditions. More specifically, those areas were the bilateral anterior cingulate cortex, extending into ventromedial prefrontal cortex and orbitofrontal cortex, the bilateral anterior insula / inferior frontal gyrus, the bilateral precuneus / posterior cingulate cortex, the bilateral supplementary motor area, the right posterior insula / superior temporal gyrus extending into the temporal pole, and the bilateral anterior midcingulate cortex (Majdandžić et al., in prep.). However, some of these areas, most notably the dorsal anterior cingulate cortex and the anterior insula, are not only involved in emotion processing, but also in the generation and processing of empathy (Lamm et al., 2011 & 2007; Morrison et al., 2007; Singer et al., 2004)<sup>4</sup>.

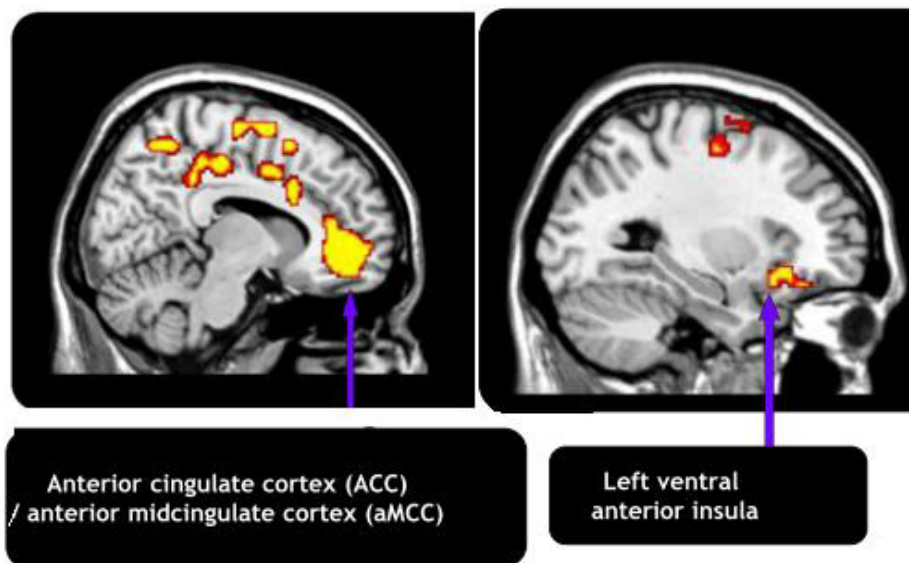


Figure 7: The main brain areas associated with empathy (adapted from Majdandžić et al. (in prep.))

<sup>4</sup> However, different emotions are processed in different neural systems. For instance, disgust seems to activate preferentially the anterior insula, whereas fear seems to preferentially activate the amygdala (Carr et al., 2003).

Majdandžić concluded that these results indicate a stronger emotional conflict during decisions about humanised victims. However, the results do not clarify if higher activation levels in the complex network of brain areas described above reflect increased empathy or increased personal distress, as many different emotional processes are associated with those areas. This clarification was the subject of the current study that will be described in the following chapter.

First, however, it is necessary to define the concept of empathy and differentiate it from the related term of perspective taking that plays a role in the humanisation process. What is meant by personal distress and how it can be seen as separate from empathy will become clear in chapter 3.

Perspective taking is not an emotional reaction. It is the cognitive ability to consider other points of view and allows an individual to anticipate the behaviour and reactions of other people (Davis, 1983). It can be divided into two strands, namely imagining how another person feels and how you would feel in another situation (Batson et al., 1997).

Empathy is different from perspective taking in that it can be characterised as an emotional response. However, the two concepts are related because other-centred perspective taking is supposed to *induce* feelings of empathy, which in turn could lead to a more intense perception of the other person as a human being. Batson et al. (1997) found that imagining how the other feels produces empathy and in turn altruistic motivation, while imagining one's own feelings in the place of the other person produces empathy, but also personal distress and therefore egoistic motivation. Oswald (1996) conducted a study in which cognitive and affective perspective taking were brought about by instructing participants to pay attention to either the thoughts or the feelings of a stimulus person. Participants in the affective perspective-taking condition then reported higher levels of empathic arousal than control participants and offered help more frequently, which is in line with the concept that the goal of empathy is reducing someone else's suffering.

Although no complete consensus has been reached on how to define empathy, three core components form the basis of this concept. First, another person's emotions have to be

recognised by an agent (via facial expressions, body language, behaviour, or verbally) either automatically or intentionally, without confusing oneself with the other. Second, there has to be a mechanism of affective responsiveness, i.e. sharing and internalising the other's emotions. And third, the empathic kind of perspective taking, i.e. imagining how the other person feels, has to be employed. Here, processes of regulation emotions play a crucial role so that, for example, the agony felt by a person is not shared to the same extent by another person watching (Decety & Jackson, 2004). Therefore, empathy can be seen as a comprehension of another's emotional state and a similarity in affective response to that person's (expected) feelings (Eisenberg, 2000). In other words, empathy means understanding what another person is feeling and what that person's experience feels like (Singer et al., 2004). Observing or imagining another person's emotional state automatically - i.e. without that observation rising to consciousness - activates a representation of that state in the observer and elicits the associated somatic responses (Eisenberg, 2000).

The neural structures that are thus activated largely overlap with those that are active when the same emotion is experienced first-hand (Singer & Lamm, 2009).

It would be simplistic, however, to reduce empathy to primitive matching of perceptual input to an affective response. In contrast, empathy is a complex, changing process that is not just streamlined in the bottom-up direction of an automatic and direct match between perceptions and actions. On the contrary; emotion regulation and control and the conscious appraisal of situational context are all involved in modulating empathy from the top (meaning the prefrontal and cingulate cortices) down (Decety & Lamm, 2006).

For instance, in a study that combined behavioural data with fMRI, Lamm et al. (2007) found that the capacity to differentiate the self and one's own emotions from those of another person, as well as the opportunity for cognitive appraisal, influence whether observing another person in pain leads to empathic concern or personal distress. While self-other differentiation was compared in two groups of participants, one group was told to imagine themselves in the suffering person's place and another was instructed to imagine the feelings of the other person. Cognitive appraisal means the conscious knowledge of the outcome for a

suffering person will be negative or positive. Those participants who imagined the feelings of that person exhibited higher ratings of empathic concern. When imagining themselves being in the same situations, measurements of personal distress were higher, as they were when participants were told that the outcome for the suffering person was unfavourable.

The concept of humanisation induced by perspective taking has been used in previous research to test levels of empathy. Harris and Friske (2006) found that 'humanised' persons elicited higher levels of empathic concern in participants. Empathic concern, in turn, has been found to result from taking the other's perspective (Oswald, 1996; Davis et al., 1996; Lamm et al., 2008) and to lead to a desire to reduce the other person's suffering (Batson et al., 1987). Majdandžić et al. (in prep.) thus hypothesised and found, as mentioned before, that humanised potential victims would be sacrificed less often in moral dilemmas than those who were not humanised, and that increased activity in brain areas related to emotional conflict would be measured using fMRI.

## **Chapter 3 Using emotion ratings to gain insight into feelings of empathy and personal distress: a behavioural study**

### **3.1 Background**

The current study builds on the one by Majdandžić et al., and Majdandžić was also crucially involved in the present study. This study seeks to explain the emotional background of why 'humanised' victims are sacrificed less often than neutral ones. In Majdandžić et al.'s paper it was assumed that humanisation could be induced by perspective taking. Humanisation would then enable feelings of empathic concern. However, it could also be true that self-related personal feelings of distress that were caused by a greater awareness of the potential victims' humanness (in contrast to the group of other victims that had not been described with a reference to their mental states) caused the effect of saving the humanised potential victims at the expense of other people. Due to the taboo of killing a human being, the

potential victims would have been assigned a higher value if humanisation took place, which would result in a greater internal conflict and thus feelings of stress, anger, and / or frustration. Therefore, both empathic concern and personal distress could have arisen out of greater humanisation of the primed potential victim.

The differentiation between empathic concern and self-related feelings was first described by Batson (1987). According to Batson, perceiving a suffering person may invoke two distinct emotional processes. One is empathic concern which in turn results in an altruistic motivation to end the suffering. This contrasts with the traditional view that all motivation evoked by emotional arousal is directed towards reducing that arousal, even if that means reducing the suffering of another person in the case of empathy (Hull 1952 in Batson et al., 1987). According to that view, empathy and personal distress would not be qualitatively distinct as they are both motivated by a reduction in arousal. On the other hand, Batson sees empathy as a set of 'other-focused' emotions and describes it as feeling 'sympathetic, moved, compassionate, tender, warm, softhearted' (Batson et al., 1987, p.26). Empathic concern varies individually, depending on experience with that particular situation and its respective perception, and different relationships to the suffering person. Additionally, other factors such as dispositional differences to empathy as measured in the Interpersonal Reactivity Index of Davis (1983) and pathologies modulate the capacity for empathy (Batson et al., 1987).

The other emotional process likely to be involved in moral judgement is personal distress, a complex of rather self-focused emotions such as feeling 'alarmed, upset, worried, disturbed, distressed, troubled' (Batson et al., 1987, p.26). Other researchers agree with this broad definition, although they do not necessarily use the same range of adjectives. In Eisenberg's view, personal distress is a 'self-focused, aversive, affective reaction to the apprehension of another's emotion (e.g. discomfort or anxiety)' (Eisenberg, 2000, p.672). An example would be feeling anxious when viewing another person who is sad.

Batson conducted principal components analysis on scales of those self-reported emotions ('alarmed', 'sympathetic', ...) that subjects rated on a 7-point scale when they watched a person in distress (Batson et al., 1987). The results of his analysis yielded two different



principal components, one for empathy and one for personal distress. This seems to refute the notion that there is such a thing as 'empathic distress'.

It has been mentioned that empathy and personal distress also differ in their motivations. In a personal distress situation, it is not another person's suffering, but one's own negative arousal that should be contained (Batson et al., 1987). However, what personal distress and empathy have in common is their relation to perspective taking. Perspective taking has not only been associated with eliciting empathy, but also with inducing personal distress. In a study by Lamm et al. (2008), it was found that among participants watching videos of a person in pain, those who were instructed to imagine themselves in the same situation exhibited physical signs of pain (the tightening of muscles above the eyes), while those who were told to imagine the other person's feelings did not display this physical reaction.

Other studies corroborated this effect (Archer et al., 1981). However, Batson maintains that the separation of empathy and personal distress is still somewhat ambiguous. Wording plays a large role in this qualification: in some of his six studies, the terms 'grieved' and 'worried' seem to be associated more closely with empathy than personal distress. Accounting for the two different motivations of empathy and personal distress, however, did yield a more conclusive separation of the two emotional categories. It was found that when escape from a dilemma situation was sufficiently easy, egoistic motivation (reducing one's level of negative arousal) did not lead to increased helping, whereas altruistic motivation (reducing the other's suffering) did (Batson et al., 1987). Another study found that empathically aroused individuals help even when no-one will know if they choose not to help (Fultz et al., 1986).

Other researchers also conclude that the separation between empathy and personal distress is incomplete. In Eisenberg's view, an empathic response can turn into either sympathy (i.e. feelings of sorrow or concern for another person that arise out of comprehension of the other's state of mind), personal distress, or some combination of both through cognitive processing (Eisenberg, 2000).

However, while this author agrees that empathy and personal distress sometimes overlap, Batson's experimental results suggest that the distinction between the two complexes of emotions is meaningful.

### **3.2 Hypotheses**

In the current study, we aimed to replicate the behavioural results of Majdandžić et al. (in prep.) as well as to clarify to what extent personal distress and empathic concern are present during moral judgement formation.

Therefore, when making a moral decision involving a potential victim that has been primed in a humanised way, a significant majority of participants should

1. sacrifice them less often than neutral victims, thus confirming Majdandžić et al. (in prep.)
2. show significantly higher levels of feelings of empathy towards that person and / or higher levels of personal distress<sup>5</sup>
3. take indirect note of the victims' humanisation by professing on a greater scale to understand them and by giving them higher ratings of similarity to themselves.

---

<sup>5</sup> It has to be pointed out that no reasonable claim for the greater role of personal distress or empathic emotions could be made before conducting the experiment. It is probable that both kinds of emotions are present and mixed to an extent unknown before the current study. Therefore, this hypothesis is phrased in an open way.

## 3.3 Experimental design

### 3.3.1 Participants

54 male volunteers participated in the current study. The decision to test only males was made for two reasons. First, Majdandžić et al.'s experiment also involved only males; doing the same makes the results of the two studies more comparable. Second, a number of studies on moral decision making assert that there are gender differences in moral decision making. Testing both males and females as one group would therefore result in a greater sample variance in an experimental context.

Females score higher on emotion ratings such as 'upset' and 'sympathy' than men (Skoe et al., 2011). Furthermore, they are found to display somewhat more of a 'care' and males more of a 'justice' orientation (Jaffee & Hyde 2000). Care-based orientation in this context means that interpersonal relationships, guided by empathy and altruism, dominate decision making. On the other hand, a justice-based orientation stresses adherence to order and rules (Robertson et al., 2007).

However, these findings have been debated for several decades. Some (mostly behavioural) studies dispute the claims of gender differences, whereas other (mostly imaging) studies support them (Harenski et al., 2008). In any case, though, males tend to provide more utilitarian answers in 'personal' moral dilemmas (Fumagalli et al., 2010).

Roughly half of the participants in the current study were psychology students, the others either students from a variety of other areas or university graduates. All were German speakers at native level, as the experiment was conducted in German. The median age was 22, with a range from 19 to 57 (this was the only person above the age of 35, however). Some psychology students received credits for a class as compensation, and all subjects entered a draw of 25, 50, and 75 euros.

In this behavioural study of about 35-45 minutes' duration, participants were seated individually in front of a computer screen inside a soundproof room. The experiment was

programmed in Presentation. To answer questions on the screen, they were instructed to press certain keys (for 'yes' and 'no' answers) or move a slider (for emotion rating scales) to their desired position. They were told beforehand that the experiment was about moral decision making and that they should try to remember the people introduced to them via short stories (i.e., remember which stories were told about persons A, B, ...). With the exception of a few questions for emotion assessment that required an answer within 60 seconds, participants themselves could determine the speed of text reading. However, they were instructed not to deliberate too long over each decision, but to go with whatever felt right for them. Participants were also instructed to try to imagine themselves as really being in a described situation, even though this could prolong decision time.

### **3.3.2 Part I: Victim priming and dilemmas**

As in Majdandžić et al.'s study (in prep.), participants were first presented with stories about two males who would later serve as potential victims in separate moral dilemmas. One of the stories implicitly required perspective taking, meaning that the reading participant was confronted with the fictitious person's wishes, thoughts, and feelings. As explained in the previous chapter, this is meant by 'humanisation' of the potential victim. The other story did not require perspective taking. Instead, a person would be described as owning a bicycle or running a company without explicit reference to his state of mind. At the end of each story, two questions were asked that reinforced the 'perspective taking versus non perspective taking'-divergence. It was not obvious what the 'right' answers to these questions were and participants could not find them in the texts. Instead, the questions were meant to make participants delve into the mental states of the potential victims (humanised condition) or concern themselves with factual knowledge (neutral condition).

A sample victim priming story and sample questions for each condition are described below:

### **Person J [neutral]**

J is a 23-year-old man who breeds dogs. His eldest dog, which he has had since he was a child himself, is already twelve years old and nearly blind. On the morning of that day, she ran into the street and overlooked a car due to her bad eyesight. She was struck by a car and badly hurt one of her paws. When J takes her to the veterinarian, he is told that the dog will not be able to use her paw at all in the future. He finds it difficult to decide whether to keep her alive.

He just does not know whether she could ever have a life worth living again. Besides, an operation would be extremely expensive. On the other hand, he is very attached to this dog and thinks that she could possibly live a decent life with only three paws. She looks at him devotedly, and he feels desperate. He strokes her head and tries to make a decision, fast.

How, do you think, will J decide?

1. He will put the dog down
2. He will consent to the dog's operation

J feels very guilty. Why do you think he feels that way?

1. Because he thinks that he has not watched his blind dog properly
2. Because the price of the operation is a factor in the decision whether to put down the dog.

### **Person K [neutral]**

K is a 25-year-old, affluent farmer. He owns 200 cows, a large pasturage and 250 hectares of forest. His occupation is cattle breeding and he has organised his farm so that the animals can be outside throughout the year. The large pasture is cleverly partitioned so that the animals can, without any hassle, be led to new rich and green meadows. Part of the meadows lies fallow, so that grass can be harvested twice a year to ensure there is enough hay for the animals in winter.

The 250 hectares of forest are cultivated according to the newest principles of forestry, and cleared areas are reforested in time. Person K does not do the machine-intensive forest work himself, but contracts it out to specialised companies. Both branches of his enterprise are highly profitable and this kind of operational management leaves K enough time to follow other interests.

In your opinion, why are machines suitable for forest work so expensive?

1. Because they are high-tech products and cannot be mass-produced
2. Because they are very big and have to be portable nonetheless

In your opinion, which kind of forest work is more machine-intensive?

1. Felling timber
2. Transporting timber

Of course, the personality and occupation of the potential victims described in these stories may have a huge influence on identification with the victim. For that reason, we used a pool of eight different stories, four for each condition, from which a total of four was selected randomly for each subject. In addition to the plots above, the humanised victims could be a mountaineer, a person dining in a restaurant, or an action film fan; those in the neutral condition could be a cyclist, supermarket manager, or a photographer (for full victim priming stories, see appendix A). One of four male photographs was also presented along with each victim primer as a memory aid, as the subjects were supposed to remember the identity of each potential victim throughout the experiment. The photographs were also randomly selected for each participant.

After being primed with stories about one humanised and one neutral potential victim, each subject was presented with two moral dilemmas, one featuring the humanised primed potential victim, the other the neutral one. This order was randomised. A story was presented as a text, and the subject had to decide whether to

1. sacrifice the primed person, which would save a group of other people who had not been previously introduced, or
2. leave the primed person unharmed, which the subject was told would result in the death of the group of other people<sup>6</sup>.

After completing the two dilemmas, a second block consisting of two more victim primers - one humanised, one neutral - and two more dilemmas involving the newly primed potential victims was presented, resulting in a total of four dilemmas each subject had to decide on. This was a significant reduction from the 24 dilemmas that were presented to each subject in Majdandžić et al.'s study. While statistical power may be higher if more items are tested, this reduction was nevertheless necessary because of the different aim of the present study. Here, the subjects were expected to conjure up detailed emotional recollections about the moment of decision for each dilemma. After making 24 decisions, it would be very unlikely that the subjects would have been able to differentiate between their respective memories of their feelings and thoughts; those moments would likely be quite indistinguishable from one another.

Similarly to the victim priming stories described above, the setting of a moral dilemma story - for example, the number of other people that could be saved, whether they are children or adults, far away or nearby, killed by animals or an avalanche - hugely influences the decisions made. Therefore, the connection between the primed victims and the dilemmas was randomised, meaning that each potential victim drawn randomly from the victim pool could be assigned to any dilemma story in order to avoid measuring only the effects of different story types rather than contrasting humanised and neutral victims. However, the randomisation was adapted so that the same number of humanised and neutral trials for each dilemma story was presented over all subjects.

---

<sup>6</sup> In dilemmas in which, according to Majdandžić et al.'s pretests, participants nearly uniformly decided to sacrifice the potential victim or the group of other people, 'certain death' was sometimes rephrased to include the possibility of serious injuries. This was done in order to allow a more even distribution of utilitarian decisions.

The following is one of the four dilemma stories each subject was shown (the other three dilemmas can be found in appendix B). To the question at the end, each subject had to press keys on the computer keyboard for 'yes' or for 'no'.

### **Mountain rescue**

You are a member of the mountain rescue team. Together with a colleague, you are on a mission to look for victims of an avalanche. Two buried people have been localised; however, they are situated far apart. You and your colleague try to dig out one victim each. The person that you are trying to retrieve is person X; he is alive, because he is moving his foot, but seriously hurt. In this moment, a second avalanche occurs and buries your colleague.

If you hurry over to the other site immediately and start digging, you can save your colleague and the other victim. Person X, who you have been digging out, will then certainly die. If you continue digging out person X, your colleague and the other victim will die for sure. However, person X will then be able to live.

Will you run to the other site?

### **3.3.3 Part II: Emotion rating scales: evaluating the period of decision making**

In the second part of the experiment, participants were asked about their emotions during the moment of decision of each dilemma. By letting the participants rate their own perceived emotions after deciding, it was hoped that the difference in intensity of their ratings would be a hint to whether feelings of empathic concern or of personal distress contribute more to the process of decision making.



Three sets of three questions each had to be answered for each decision phase separately. As a memory aid, the pictures of the potential victims were shown alongside a few keywords from the respective dilemma and victim priming stories. Subjects were asked to respond by adjusting a slider shown on the screen to the position between 'not / none at all' and 'very much' that they preferred. For analysis purposes, this position was then recorded as a value between 0 and 100. The questions were in randomised order and were designed to test three issues, after Batson (1987):

1. Self-related emotions (related to personal distress)

How frustrated did you feel?

How torn did you feel?

How stressed did you feel?

2. Emotions relating to the potential victim (i.e. empathic concern)

How strongly did you imagine the thoughts and feelings of this person?

How much compassion did you feel for this person?

How moved were you during the moment of decision?

3. The extent and nature of humanisation of the potential victim<sup>7</sup>

To what extent did you see this person as a human being, rather than a means to an end?

How responsible did you feel for this person's well-being?

To what extent did you see this person as a human being with needs, desires, and feelings?

In addition, each subject was asked to rate (using the slider) the perceived difficulty of deciding on the dilemma.

---

<sup>7</sup> This third set of questions not only served the purpose of testing the efficiency of the humanisation manipulation, but also aimed to establish if humanisation had an influence on factors that could reduce participants' propensity for utilitarian responses, such as feeling responsible for another person.

While the order of the questions was randomised in the study, in the analysis the values for the ten questions were combined into those three groups, plus a fourth that contained only the difficulty rating.

### **3.3.3.1 Pilot study: exploring labels for emotions**

To clarify what adjectives we should adapt from Batson (1987), i.e. what wording was most comprehensible to participants, a pilot study testing six volunteers was conducted. All participants were female, which does not present a problem as the focus did not lie on decision making itself, but just an evaluation on how useable a range of different basic emotions and their wordings was for letting participants express their feelings.

In the pilot study, participants were presented with a very similar experimental design as in the main experiment. After being primed with two potential victim stories - one humanised, one neutral - each participant was asked to make decisions about eight dilemma stories featuring either of the two potential victims. After their decisions, they had to rate their emotions in the moment of each decision, again as in the main experiment. The scale used was slightly cruder and ran from 'not at all' (0) to 'very much' (10), where participants had to select any integer between those values. The main difference was that participants only had to rate their emotions for two randomly selected dilemmas, as the choice of emotions was wider. Participants were asked to rate their levels of feeling confused, angry, frustrated, torn, and agitated. They also had to rate how strongly they imaged the potential victims visually, how strongly they imagined thoughts and feelings of the potential victims, how much compassion they had for each potential victim, how much active interest they felt towards the situation, how worthy of protection the potential victims were, whether they saw the potential victims as means to an end, and whether the potential victims reminded the participants of a person they knew in real life.

From the first five ratings, which are descriptions of self-related emotions, we selected the ones with the three highest mean ratings (torn, confused, and frustrated, whose mean ratings

were above 6.5 out of 10). To obtain a greater relevance for measuring personal distress, the term 'confused' was later exchanged with 'stressed'.

The pilot study was explorative as it was not known from the outset what the emotional descriptions participants most identified with were. Batson's (1987) choice of wording is not set in stone, but rather dependent on the context of his experiments. For that reason, the 'thinking aloud' technique was used in addition. This method relies on the concept that when participants are left undisrupted rather than having to explain themselves to another person, language can be a verbalisation of the parts of thought that are 'inner speech' without altering thought processes (Ericsson & Simon, 1998). With no further limitations, participants were thus instructed to speak aloud what they thought and felt during each process of decision making, without deliberating too long. These - often disconnected and associative - snippets of thought were then recorded, transcribed, and checked for recurring phrases and patterns.

Rational, utilitarian deliberations clearly dominated over emotional descriptions. At times, every straw of rationality was grasped, seemingly in order to find and justify a solution ('This guy ran over a stop signal, he made a mistake!', 'My priority is always the person in trouble I encounter first.'). Altogether, two main lessons were learnt from this study. First and perhaps most obviously, all participants professed that introspection was a hard feat to accomplish. Even when answering only two sets of emotion ratings, it was remarkable that all participants took much less time on the second set (this is an observation, as time was not measured exactly). This author's interpretation of this behaviour is that after completing one set of emotion ratings, participants subsequently did not ponder their emotions as thoroughly and possibly followed patterns that they had established previously. Recalling specific emotional reactions for eight recent situations, therefore, could be asking too much. For that reason, we limited the dilemma situations of the main study to four. The second lesson learnt was the observation that some participants had trouble with the great differentiation of self-related emotions. For example, two participants professed that 'confused' and 'torn' were indistinguishable to them; another two did not know what 'agitated' or 'taking an active

interest in the situation' were supposed to mean. We therefore limited the choice of emotions to those who did not seem to present a problem of comprehension.

### **3.3.4 Part III: Rating of victim attributes**

After reporting their emotion ratings for each dilemma's moment of decision, a more detailed exploration of the perceived humanisation of the primed potential victims followed. Participants were asked to rate all potential victims concerning different traits. Again, a slider was used that participants could set anywhere between 'not at all' and 'very much', which was translated to a value between 0 and 100 for analysis purposes. All victims were shown on a single screen, each with an individual slider, and the respective question was shown on the top of the page. The ratings for each victim had to differ, at least very slightly, in order to nudge the participants towards really differentiating between their ratings of the victims.

In randomised order, participants were asked to rank potential victims on the following scales:

#### 1. Ratings of perceived humanisation

How alive did this person seem to you?

How tangible did this person seem to you?

How human did this person seem to you?

How abstract did this person seem to you? (Rating reversed)

How interchangeable (with another person) did this person seem to you? (Rating reversed)

#### 2. Ratings of perceived similarity to oneself, i.e. a human being, and of influence on decision

How connected do you feel to this person?

How well do you understand this person?

How well do you think you know this person?

How similar do you consider this person to yourself?

How much did your opinion of this person influence your decision?

The second group of questions had already been included in Majdandžić et al.'s study (in prep.), where these attributes were rated significantly higher for humanised conditions. They were used in the current study to confirm that the manipulation of humanising potential victims had a comparable effect in both studies.

### **3.3.5 Part IV: Graphical dilemma**

After the main part of the experiment described above, the final section of the study explored a separate research question. 46 of the 54 participants in the main experiment participated in this last section. Here, the idea was to explore whether the form of dilemma presentation plays a role in the emotions that the participants profess to feel during their decisions. While the dilemma situations were presented as texts in the main study, in the pilot study this author sometimes witnessed a process of just making a decision for decision's sake in order to escape the pressing dilemma (a typical comment: 'So I cannot just push back the gorillas [and save everyone]? No? Ok...well then I have to drown him, apparently. I have to act and drown him. I don't like that...next scenario, please!'). This effect could possibly be minimised, and thus emotion ratings enhanced, if the confrontation with the victim were more direct.

As a first step towards a possible future comparison of the influence of more 'direct' presentation modes, this experiment used graphical schematic representations of the dilemma situation and the human characters that it featured in addition to a text (human-graphical condition). As in the main experiment, participants were first primed with a potential victim, followed by a dilemma story, a required decision, and emotion ratings. Here, the research question was whether a schematic graphical representation of the potential victims featured in the dilemma story would elicit higher emotional involvement in participants than a scenery-and-text graphical condition without any representation of humans (the premise being, as in the whole experiment, that higher emotional involvement was reflected in higher levels of emotion ratings by each subject). The main hypothesis was that a graphical human,

especially facial, representation would lead to a more direct and less abstract emotional involvement compared to dilemmas without representations of humans. This should be true even if the representation is very schematic, as can be seen from research in face perception (e.g. Sinha et al., 2006). The schematic human drawings used in this condition are inspired by Ekman's classification of emotions, in this case especially fear, in the context of facial expressions (Ekman & Friesen, 1975) as well as the so-called Chernoff faces which are crude abstractions of facial features whose expressions are nevertheless usually well recognised as expressing certain emotions (Morris et al., 2000).

The following story was used as a setup for the dilemma:

### **Lifeguard**

You are a lifeguard and currently in the process of rescuing a swimmer, person M, whose legs have been badly hurt by a speed boat. While you are swimming back to shore with person M in your arms, you suddenly notice two children nearby who are in danger of drowning. You are too far away from the shore for anyone to notice your hand signals. It is also impossible to keep all three people above water - the overall weight would be too heavy, and you need at least one arm to swim. If you push the injured person M away, you can save the two children. Person M will then drown. If you swim to shore with person M alone, the two children will die. Person M will then be saved.

Will you push the injured M away?

Half of the 46 participants were presented with a schematic drawing of one dilemma situation and the potential victims that appeared in it, accompanied by the text in order to ensure that all participants associated the same conditions with the dilemma situation (only one person or another group of people can be saved; a decision has to be made, the people featured in the dilemma cannot save themselves on their own). This is shown in figure 8:

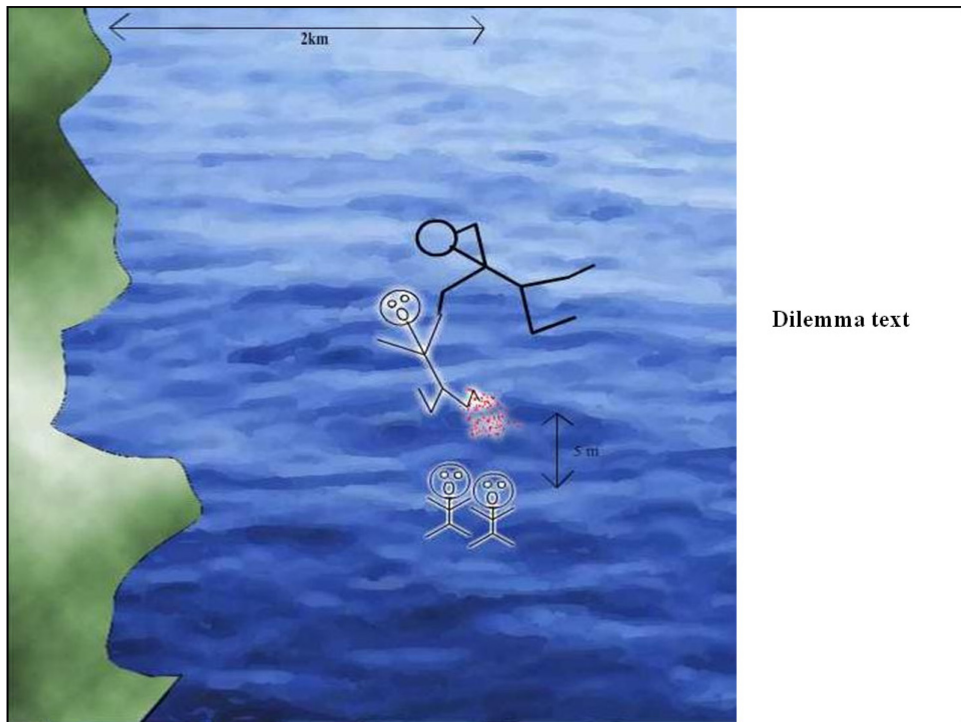


Figure 8: Human-graphical dilemma condition showing schematic drawings of humans in distress

The other half, i.e. 23 participants, received a control image in which the seaside scenery was displayed next to the text, but instead of schematic representations of the 'human beings' in the story, they would remain represented in text form in the picture (see figure 9).

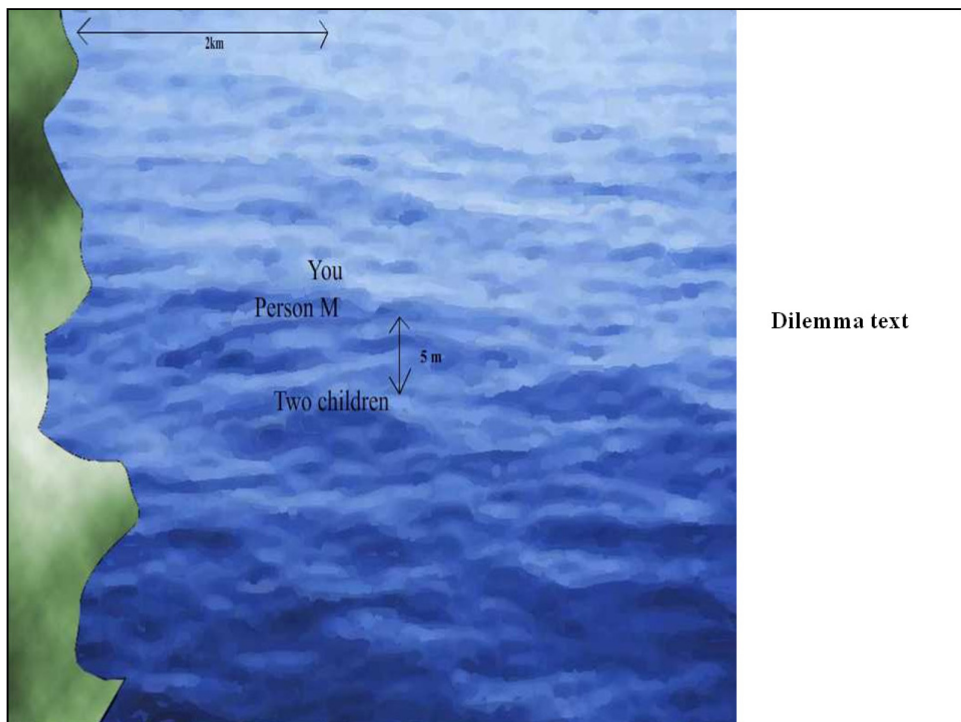


Figure 9: Control (text-graphical) condition in which the schematic drawings have been replaced with text

For the sake of making the conditions under which the human-graphical dilemma and its text-graphical control condition were presented as equal as possible, the identity of the potential victim was not randomised, but left constant. Any differences in emotion ratings should be traceable to a difference in graphical humanisation. Thus, all other factors should remain the same in order to ensure comparability. It would not matter, therefore, if the potential victim were described in a humanised or neutral way. We chose an in-between style, as can be seen from the questions following the priming story: the first question is not related to the person's state of mind, the other is.

### **Person M**

M is a 24-year-old man who works as a furniture remover. He works at a large company and organises about four relocations per week. Because he lifts heavy boxes all day, he is very strong. It is also his job as a furniture remover to pack the boxes himself. Concerning the allowed maximum weight of the boxes, he is bound by strict rules. However, M and his colleagues often ignore the rules because they are paid a fixed rate per relocation and finish earlier if they carry fewer boxes. On one occasion he exaggerated his load a bit and had to stay home for a week because of back pain.

What do you think is the allowed maximum weight per box for a furniture remover?

1. The allowed maximum weight is 20 kilograms.
2. The allowed maximum weight is 25 kilograms.

Why do you think M would rather finish earlier than carry lighter boxes?

1. Because he is young and strong and likes to be challenged.
2. Because he is actually rather lazy and looks forward to lounging on his couch at home.



### **3.3.6 Individual difference measures in empathy**

A supplementary point of interest in the main experiment was to look for a generalisation of empathy and distress ratings, i.e. to see if a participant's emotion rating scales for the moment of decision (state emotions) would correlate with more general self-reported emotional scales (trait emotions). In order to assess individual traits in empathy that can be correlated with each participant's decision-period emotion ratings, we used the self-reported Interpersonal Reactivity Index (IRI), a scale of four seven-item dimensions of empathy developed by Davis (1983, see appendix C). Subjects are asked to answer items such as 'Sometimes I don't feel very sorry for other people when they are having problems' on a 5-point scale (1 = does not describe me well, 5 = describes me very well). The IRI acknowledges the notion that empathy is a multi-factor concept of several relatively independent components. The first scale, which is supposed to measure perspective-taking, contains questions about seeing things from other people's point of view. Items on the fantasy scale assess the extent of identification with characters in fictional situations. The empathic concern scales measures a participant's general levels of concern and compassion for others who undergo negative experiences, while the personal distress scale measures the personal feelings of anxiety and discomfort that result from observing another's negative experience (Davis, 1983). Davis selected the items for the four subscales by factor analysis, and internal consistency for the four scales seems to be high with a range from 0.68 to 0.79 (Davis & Franzoi, 1991 in Schutte et al., 2001).

### **3.4 Data analysis**

Data were analysed using SPSS Statistics 17.0. First, each participant's percentage of utilitarian responses was determined for humanised and neutral dilemmas, respectively. As defined in section 2.1, a utilitarian choice means that the primed person is sacrificed in order to save the greater number of other people. As each subject made two decisions for neutral

and humanised dilemmas, respectively, this value could either be 0 (if the subject never decided in a utilitarian way, i.e. always saved the primed potential victim), 50 (if one decision was utilitarian, the other was not), or 100 (if the subject sacrificed the primed potential victim both times). The ratings for humanised dilemmas and those for neutral dilemmas were then combined for all subjects and a paired t-test was applied to test for differences in the means of the two groups. This and all other t-tests taken in this study were conducted as one-tailed tests, as we had a specific expectation about the direction of the effects. For example, we expected a higher mean for utilitarian decisions in neutral dilemma conditions than in humanised ones. Also, because of the results of Majdandžić et al. (in prep.), we expected that humanised conditions would elicit higher emotional and difficulty ratings.

As a next step, the ratings of participants' emotions during the decision phase (part II of the experiment) were analysed. The ten emotion ratings (see section 3.3.3) were combined into four scales, representing personal distress (average rating over questions 1-3), empathic concern (average rating over questions 4-6), humanisation (average rating over questions 7-9), and dilemma difficulty (rating of question 10). These four values were obtained for humanised and neutral dilemma conditions separately, allowing us to compare them across these conditions.

First, a repeated measures analysis of variance (ANOVA) was conducted in order to test for equality of means of the three rating scales that most directly measured self- and other-directed emotions: personal distress, empathic concern, and victim humanisation. The variables were separated according to manipulation (humanised and neutral conditions). Then, one-tailed paired t-tests were applied for each scale to test for differences in group means for humanised and neutral conditions.

Additionally, the relative difficulty of utilitarian decisions in humanised dilemmas was compared to the one in neutral dilemmas by means of a one-tailed paired t-test.

It was then determined whether the responses of each person for the different emotion rating scales correlated. To give an example, a high correlation between the personal distress and the empathic concern categories would mean that individual participants' high response

levels for the personal distress category generally correspond to high response levels for the empathic concern category. While not directly related to providing an answer to our research questions, testing this correlation could provide an insight into the real independence of these scales. A very high correlation for all groups would question this independence and suggest that instead, there are some participants who generally tend to give very high or low ratings regardless of the questions asked and others that tend to steer clear of extremes. We tested for correlations between ratings on personal distress and empathic concern, personal distress and humanisation, and empathic concern and humanisation; each of these categories was separated into humanised and neutral conditions.

Correlations were also determined between each participant's ratings on the perspective taking, empathic concern, and personal distress scales of the Interpersonal Reactivity Index (IRI) and the emotion rating categories during the moment of decision. For this part of the analysis only, emotion rating categories were not separated into humanised and neutral conditions. Instead, the combined average was used in order to compare 'moment-of-decision' state emotions with those ratings, or trait emotions, that participants thought they exhibited generally. These correlations are important because while the IRI scale supposedly measures a person's traits or general propensities for different components of empathy, it is not self-evident that during the moment of decision, these levels will be reflected. For example, it does not follow logically from very high general empathic IRI ratings that a given person will stick to that pattern in any momentary situation. It could well be that measures for general tendencies of empathic concern and empathic concern relating to four concrete decisions are two separate beasts. A weak correlation would therefore not necessarily mean that either of the two measures is fundamentally wrong. However, a strong correlation between, for instance, the two measures of empathic concern would lend more credibility to the claim that the questions asked in the emotion ratings were indeed a measure of empathy (if one accepts the premise that the IRI scale is a reliably tested measure for this category). Besides correlating the IRI values for personal distress and empathic concern with the value of personal distress and empathic concern during decision periods, the IRI value for

perspective taking was correlated with empathic concern during decision periods. Moreover, the IRI 'fantasy' value related to identification with fictional characters was correlated with moment-of decision humanisation and empathy. The three IRI scales of personal distress, empathic concern, and perspective taking were also each correlated with the perceived dilemma difficulty during the decision phase, averaged over all dilemmas.

Similarly to part II, the analysis of part III, the rating of victim attributes, was done by grouping the ten questions into three categories for analysis, which were all again separated into humanised and neutral dilemma categories. The first five attributes (alive, tangible, human, abstract, interchangeable) were combined into a measure of perceived humanness of a potential victim. The ratings for abstraction and interchangeableness were inverted to ensure comparability (because a smaller value for abstraction and interchangeableness means greater humanisation), meaning that any value between 0 and 100 was subtracted from 100. This way, a counter-value of the negative humanness was attained. For example, if a participant ranked a potential victim as 20 out of 100, this would be reflected as 80 in the analysis. The second group of concepts can be thought of as perceived similarity to oneself and therefore another human being (see section 3.3.4). In this author's view, combining an assessment of the perception of being connected to, feeling similarity to, understanding, and knowing another person can be considered to be another - indirect - measure of humanisation. The question relating to the extent of influence that a participant's opinion about the potential victim exerted on the decision was treated separately.

For all those groups, differences between means of humanised and neutral conditions were compared using one-tailed paired t-tests.

Part IV, the human-graphical versus text-graphical dilemma condition, was analysed using one-tailed independent samples t-tests. This method was selected because the means of two different populations were compared and the mean emotional and difficulty ratings were expected to be higher for the human-graphical dilemma group than for the text-graphical dilemma group. To check if the variances in the respective two groups were equal, Levene's test for equality of variances was used ( $p > 0.05$  implied equal variances, see Field (2009)). As

the questions asked were the same as in part II of the experiment, the grouping for analysis purposes was identical.

## Chapter 4 Results

The outcome of the decision making test does not replicate Majdandžić et al.'s findings that humanised victims were sacrificed less often than neutral ones. The difference between the mean percentage of utilitarian decisions in humanised and neutral dilemmas was not significant ( $t(53)=-0.151$ ,  $p=0.440$ , mean (M) (humanised)=64.81, M(neutral)=65.74, standard error of the mean (SE) (humanised)=5.23, SE(neutral)=4.14).

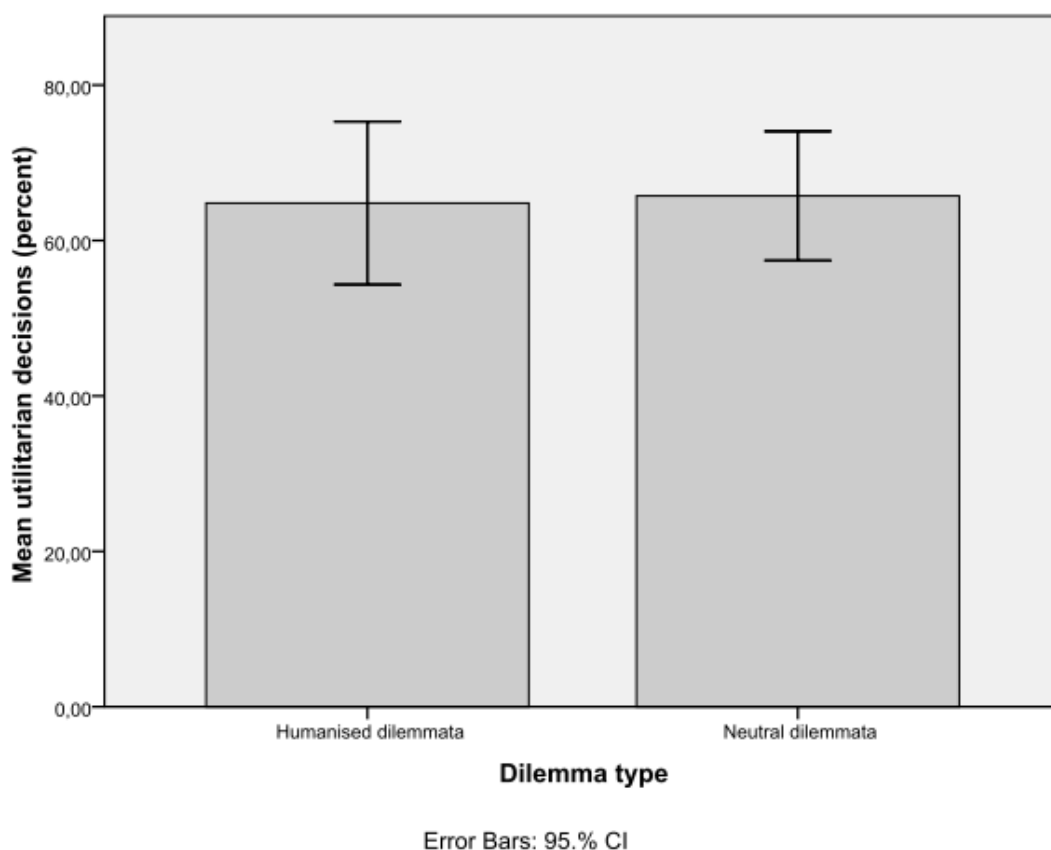


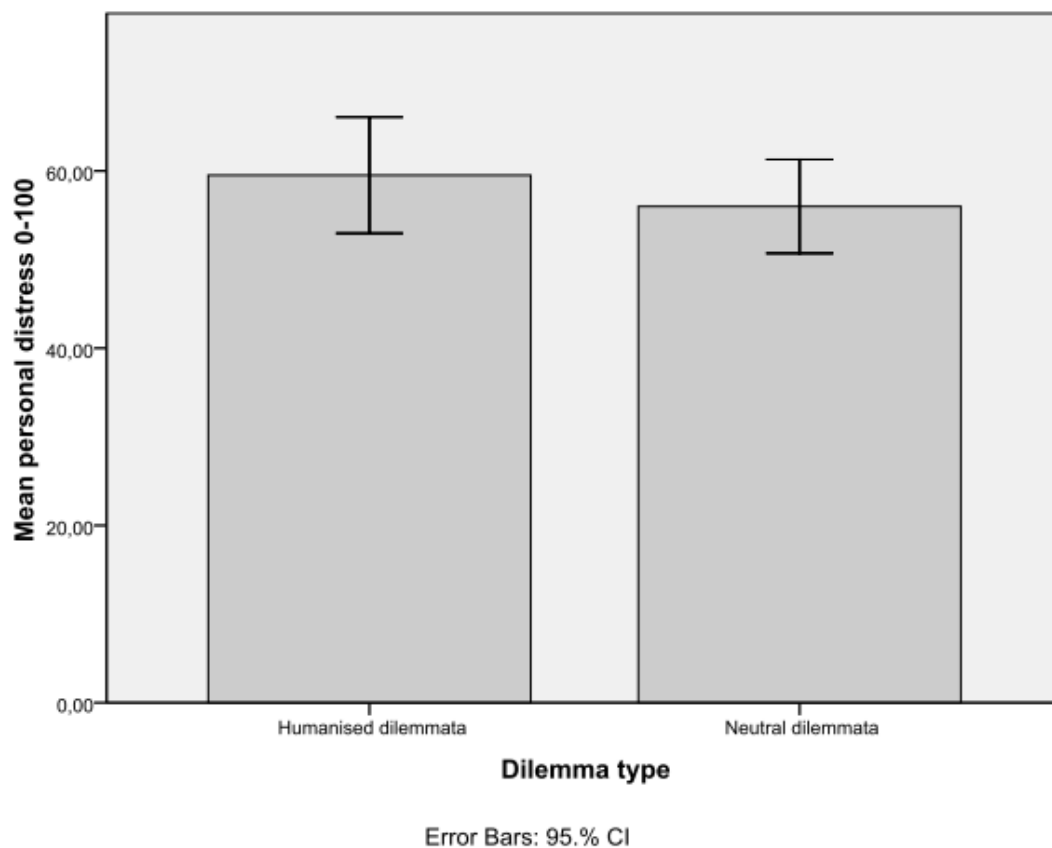
Figure 10: Utilitarian decisions in humanised and neutral dilemma conditions

The repeated measures ANOVA showed that the variances of the three emotion rating scales differ according to Mauchly's test of sphericity ( $p=0.006$ ). Therefore, sphericity, or the equality of variances of the differences between separate emotion scales, cannot be

assumed (Grieve, 1984). Using the Greenhouse-Geisser correction, this test supports the view that there is an effect of humanisation manipulation on the emotion rating scales ( $p=0.037$ ).

Moreover, paired t-tests between the separate emotion ratings categories for humanised and neutral conditions show significance, or at least trends, for some aspects.

The difference in means between personal distress levels for humanised and neutral dilemma conditions is borderline significant. Slightly higher levels of personal distress were recorded for humanised conditions ( $t(53)=1.627$ ,  $p=0.055$ ,  $M(\text{humanised})=59.52$ ,  $M(\text{neutral})=56.00$ ,  $SE(\text{humanised})=3.27$ ,  $SE(\text{neutral})=2.64$ ).



**Figure 11: Personal distress in humanised and neutral dilemma conditions**

Empathic concern, however, showed a considerably weaker trend towards being higher in humanised dilemmas ( $t(53)=1.304$ ,  $p=0.099$ ,  $M(\text{humanised})=56.43$ ,  $M(\text{neutral})=53.84$ ,  $SE(\text{humanised})=2.55$ ,  $SE(\text{neutral})=2.47$ ).

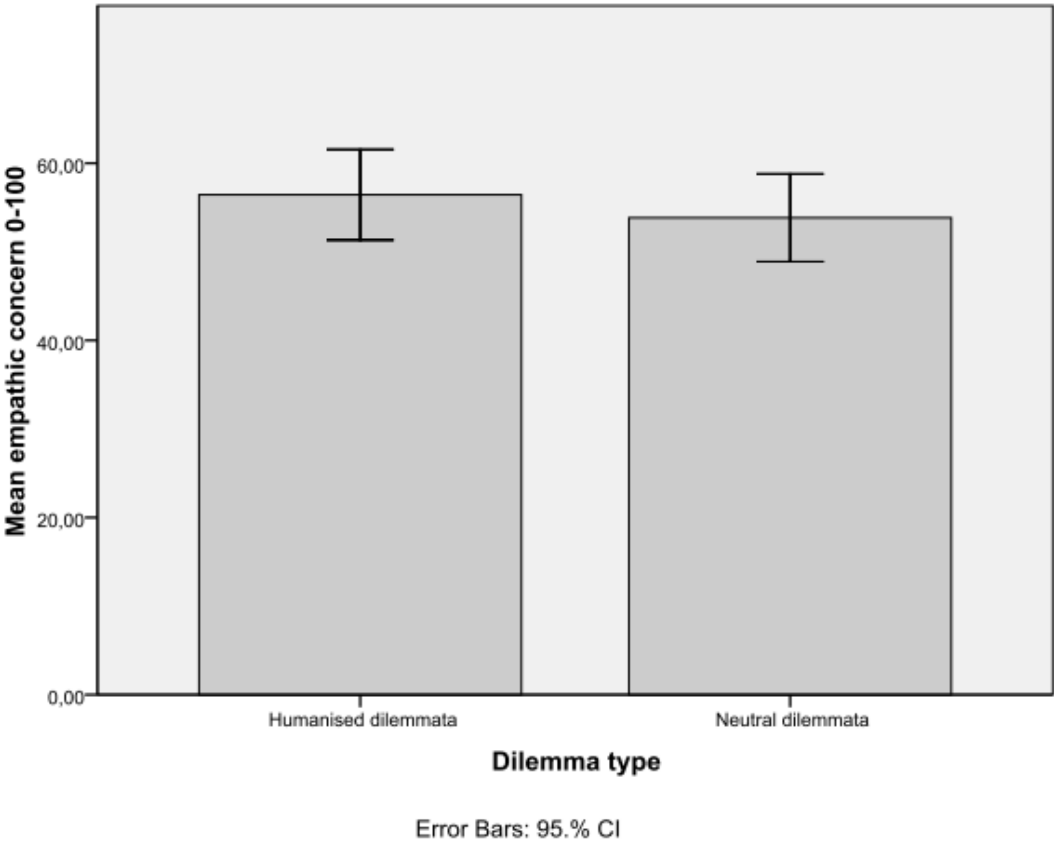


Figure 12: Empathic concern in humanised and neutral dilemma conditions

A much clearer (and significant) result could be recorded for humanisation: potential victims were perceived as human beings more strongly in humanised than in neutral dilemma conditions ( $t(53)=2.361$ ,  $p=0.011$ ,  $M(\text{humanised})=60.65$ ,  $M(\text{neutral})=56.27$ ,  $SE(\text{humanised})=2.54$ ,  $SE(\text{neutral})=2.51$ ).

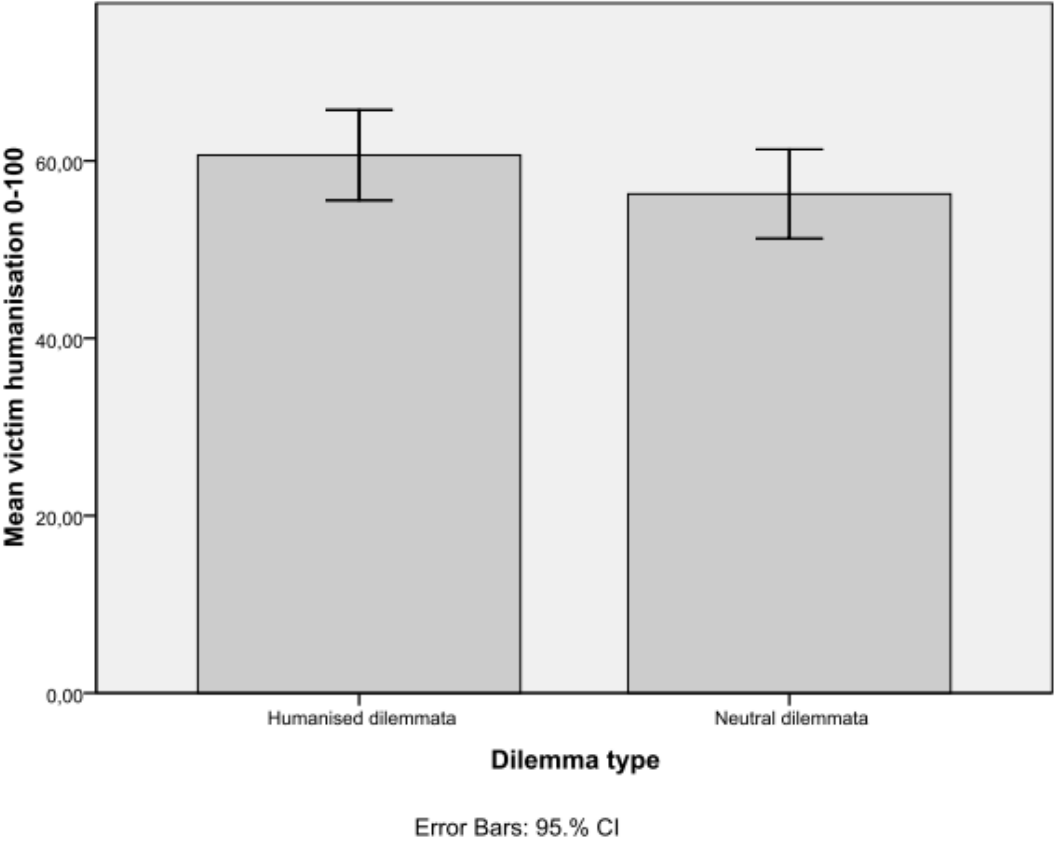


Figure 13: Potential victim humanisation in humanised and neutral dilemma conditions



Finally, participants did not perceive making decisions in humanised conditions as more difficult than in neutral dilemmas ( $t(53)=-0.304$ ,  $p=0.382$ ,  $M(\text{humanised})=59.67$ ,  $M(\text{neutral})=60.70$ ,  $SE(\text{humanised})=3.71$ ,  $SE(\text{neutral})=2.88$ ).

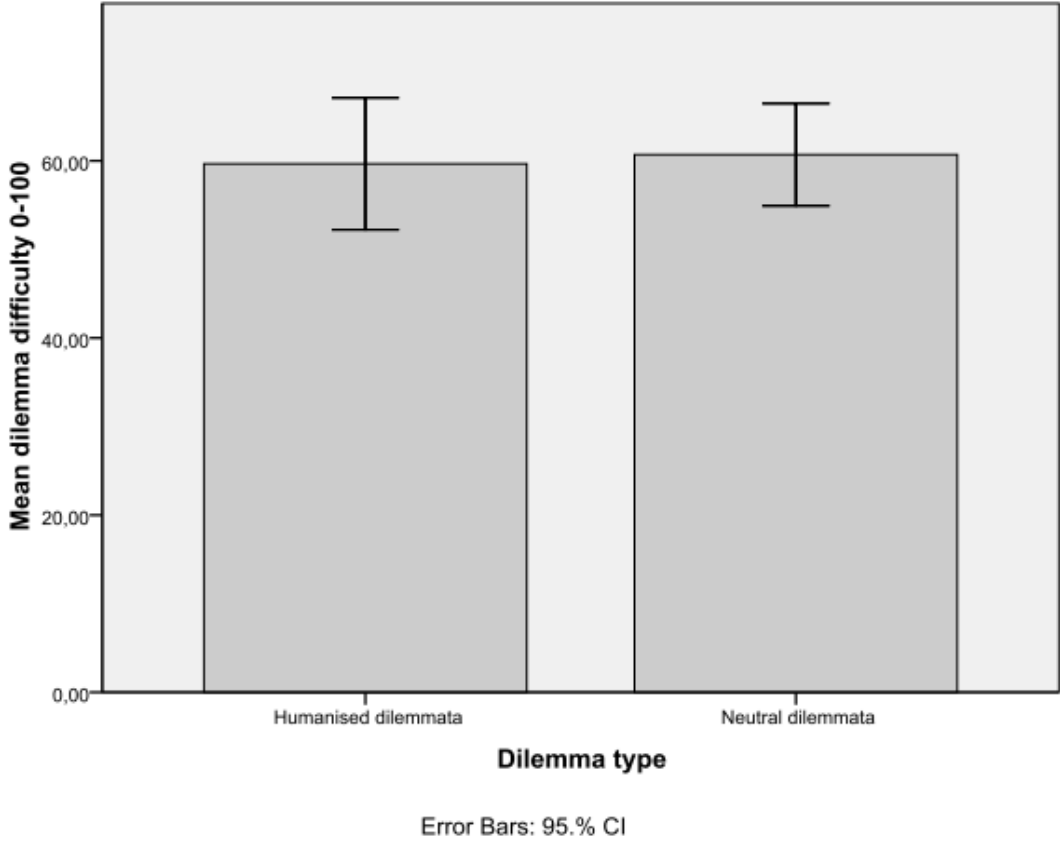
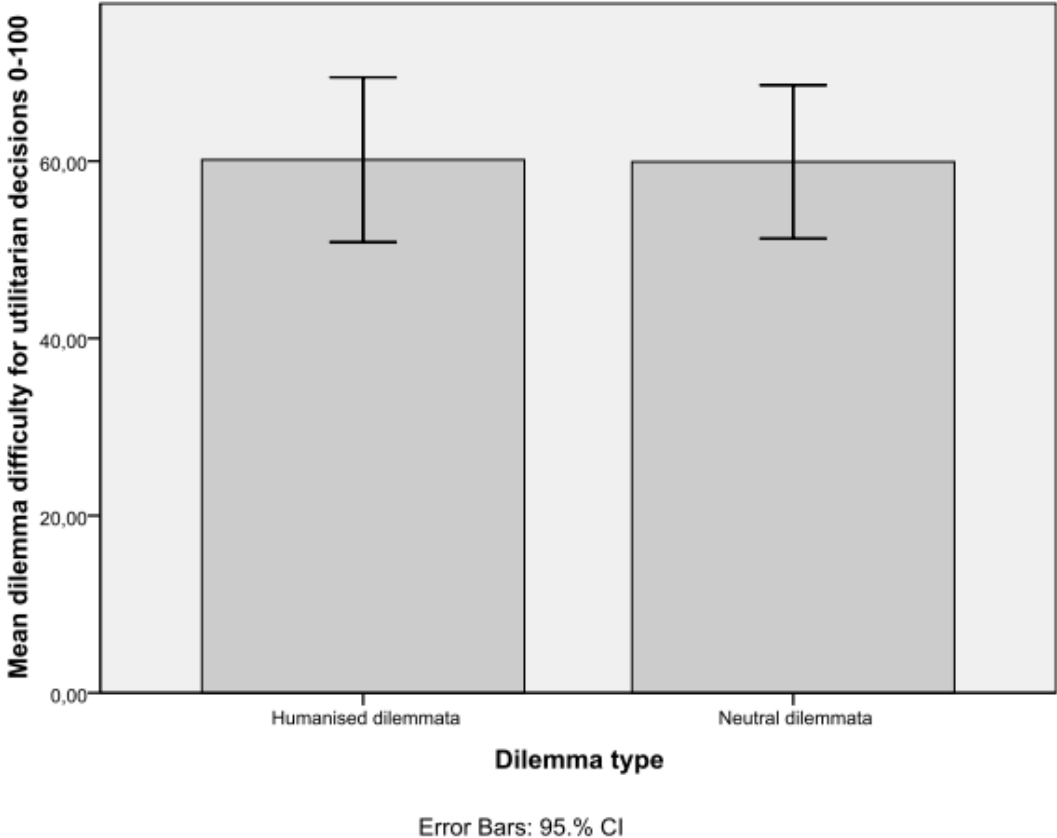


Figure 14: Dilemma difficulty in humanised and neutral dilemma conditions

In the same direction, while we expected that deciding to sacrifice a humanised potential victim would be more of a struggle in humanised than in neutral dilemma conditions, the result (see figure 15) shows an equal level of difficulty ( $t(40)^8 = 0.056$ ,  $p=0.478$ ,  $M(\text{humanised})=60.15$ ,  $M(\text{neutral})=59.92$ ,  $SE(\text{humanised})=4.60$ ,  $SE(\text{neutral})=4.28$ ).



**Figure 15: Perceived difficulty of those humanised and neutral dilemmas that were decided on in a utilitarian way**

<sup>8</sup> There are only 40 degrees of freedom for this paired t-test because only 41 participants decided in a utilitarian way on at least one humanised and at least one neutral dilemma.

Rating the victims on scales of different attributes that assessed their humanness (see section 3.3.4) showed a very clear result. When humanised, the potential victims were rated significantly higher on such items as 'alive', 'tangible', etc. ( $t(53)=2.203$ ,  $p=0.016$ ,  $M(\text{humanised})=61.35$ ,  $M(\text{neutral})=56.94$ ,  $SE(\text{humanised})=1.89$ ,  $SE(\text{neutral})=1.98$ ).

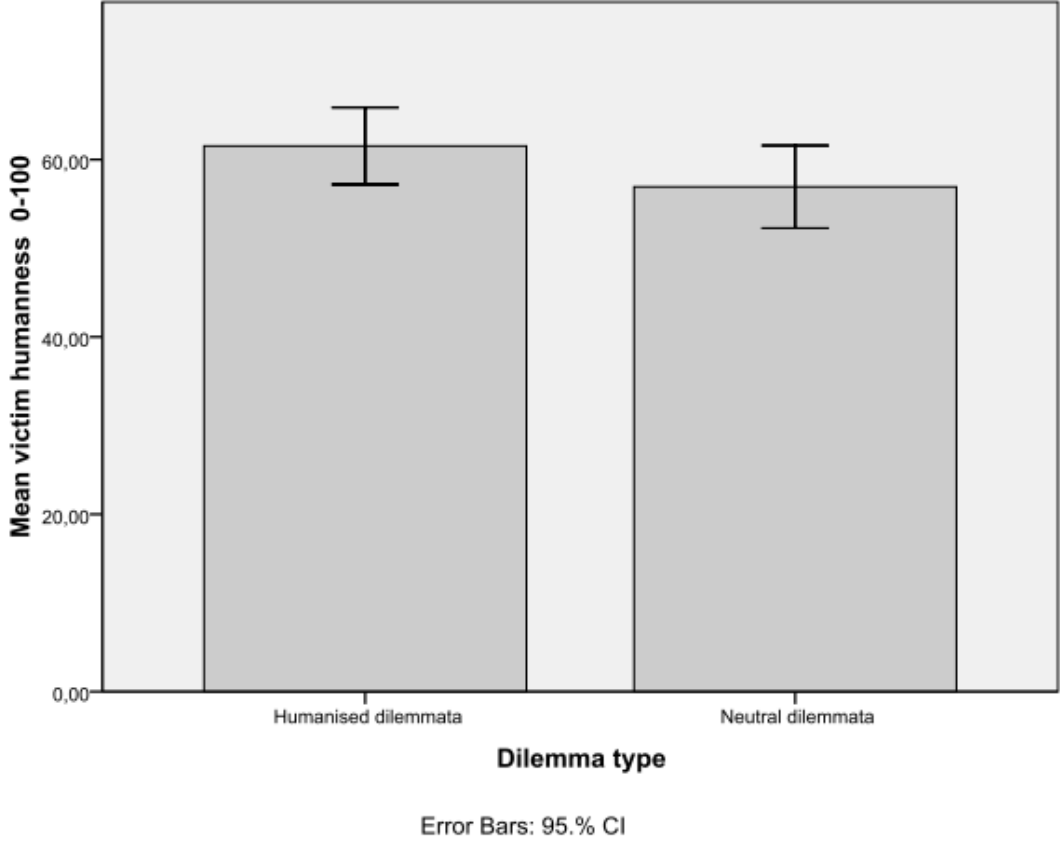


Figure 16: Rating of potential victims in humanised and neutral dilemma conditions on scales of 'humanness' attributes

As for the correlations of participants' answers to the different emotion ratings scales, all correlations are positively strong and highly significant (see table 1). This means that participants who tended to give high ratings for personal distress also tended to do this for empathic concern and humanisation, and those who gave high ratings for empathic concern mirrored this behaviour for humanisation.

| <b>Correlation</b>  | <b>Pearson's correlation coefficient</b> | <b>P-value (one-tailed)</b> | <b>Significance threshold (Bonferroni-corrected)</b> |
|---|--|-----------------------------|--|
| Personal distress- empathic concern (humanised condition) | r=0.64                                   | p<0.001                     | p=0.025  |
| Personal distress- empathic concern (neutral condition)   | r=0.61                                   | p<0.001                     | p=0.025  |
| Personal distress- humanisation (humanised condition)     | r=0.49                                   | p<0.001                     | p=0.025  |
| Personal distress- humanisation (neutral condition)       | r=0.41                                   | p=0.001                     | p=0.025  |
| Empathic concern- humanisation (humanised condition)      | r=0.69                                   | p<0.001                     | p=0.025  |
| Empathic concern - humanisation (neutral condition)       | r=0.66                                   | p<0.001                     | p=0.025  |

**Table 1: Correlations between participants' responses to different emotion ratings scales (sample size: 54)**

On the other hand, with one exception (see below), no significant correlations were found between participants' trait measures of personal distress, empathy, and perspective taking

(IRI scale) and their averaged state (decision phase) ratings of personal distress, empathy, and dilemma difficulty. As the correlation coefficients are each near zero and their p-values greatly above significance threshold levels, the variables vary in different ways and no correlation can be detected. Just one correlation - between the IRI value for fantasy, i.e. identification with fictional characters, and moment-of-decision humanisation - is highly significant. On a different note, overall scores on the IRI questionnaire did not differ between the participants of the current study and Majdandžić et al.'s fMRI participants, which confirms that the samples are comparable on these traits.

| <b>Correlation IRI scale value with moment-of-decision value</b> | <b>Pearson's correlation coefficient</b> | <b>P-value (one-tailed)</b> | <b>Significance threshold (Bonferroni-corrected)</b> |
|--|--|-----------------------------|--|
| Personal distress - personal distress                            | r=0.02                                   | p=0.434                     | p=0.050  |
| Empathic concern - empathic concern                              | r=-0.02                                  | p=0.451                     | p=0.017  |
| Perspective taking - empathic concern                            | r=-0.08                                  | p=0.292                     | p=0.017  |
| Personal distress - dilemma difficulty                           | r=0.04                                   | p=0.392                     | p=0.017  |
| Empathic concern - dilemma difficulty                            | r=-0.07                                  | p=0.313                     | p=0.017  |
| Perspective taking - dilemma difficulty                          | r=0.01                                   | p=0.459                     | p=0.017  |
| Fantasy - empathic concern                                       | r=0.13                                   | p=0.184                     | p=0.017  |
| Fantasy - humanisation   | r=0.44                                   | p<0.001                     | p=0.050  |

**Table 2: Correlations between participants' IRI scale values and their emotion rating values during decision moments (sample size: 54)**

The rate of connectedness and perceived similarity to the potential victim was also very significantly higher in humanised dilemma conditions than in neutral ones ( $t(53)=3.535$ ,  $p<0.001$ ,  $M(\text{humanised})=50.93$ ,  $M(\text{neutral})=43.49$ ,  $SE(\text{humanised})=2.33$ ,  $SE(\text{neutral})=1.79$ ).

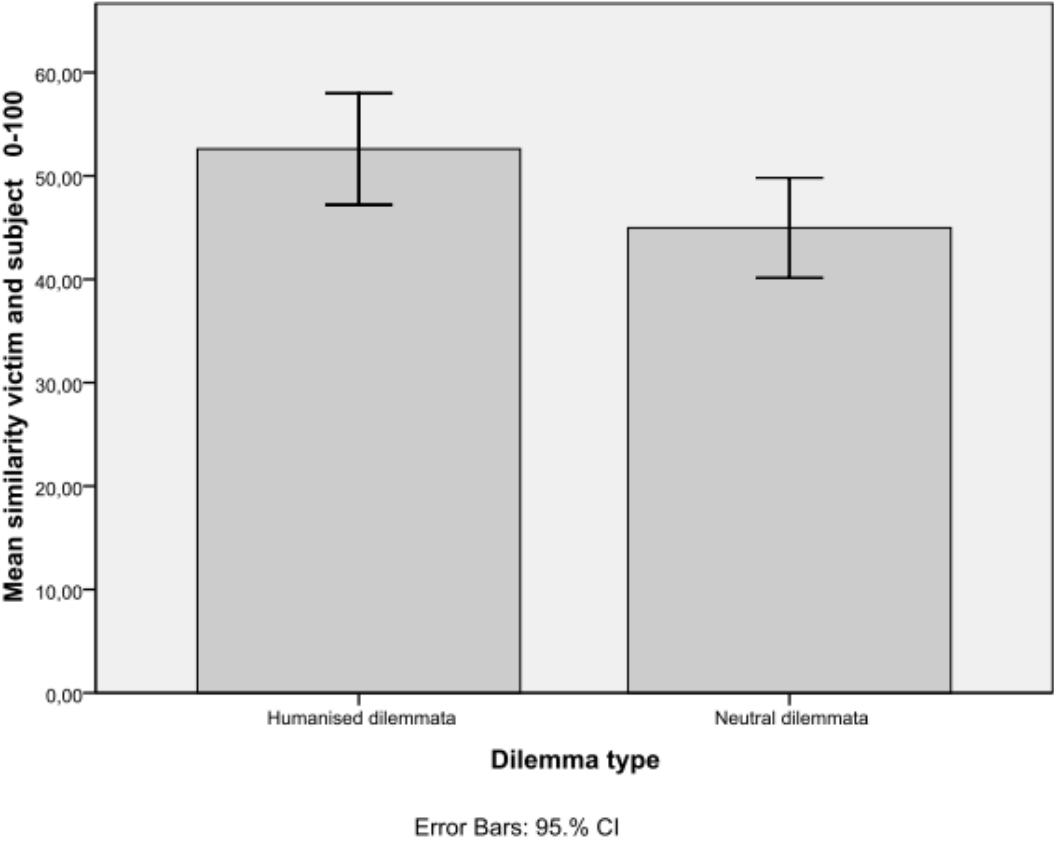


Figure 17: Rating of potential victims in humanised and neutral dilemma conditions on scales of similarity and connectedness

Furthermore, there was also a highly significant difference between humanised and neutral dilemmas in regard to participants' statements to what extent their opinion about the potential victims had influenced their decisions ( $t(53)=3.511$ ,  $p<0.001$ ,  $M(\text{humanised})=31.70$ ,  $M(\text{neutral})=22.97$ ,  $SE(\text{humanised})=4.14$ ,  $SE(\text{neutral})=2.87$ ).

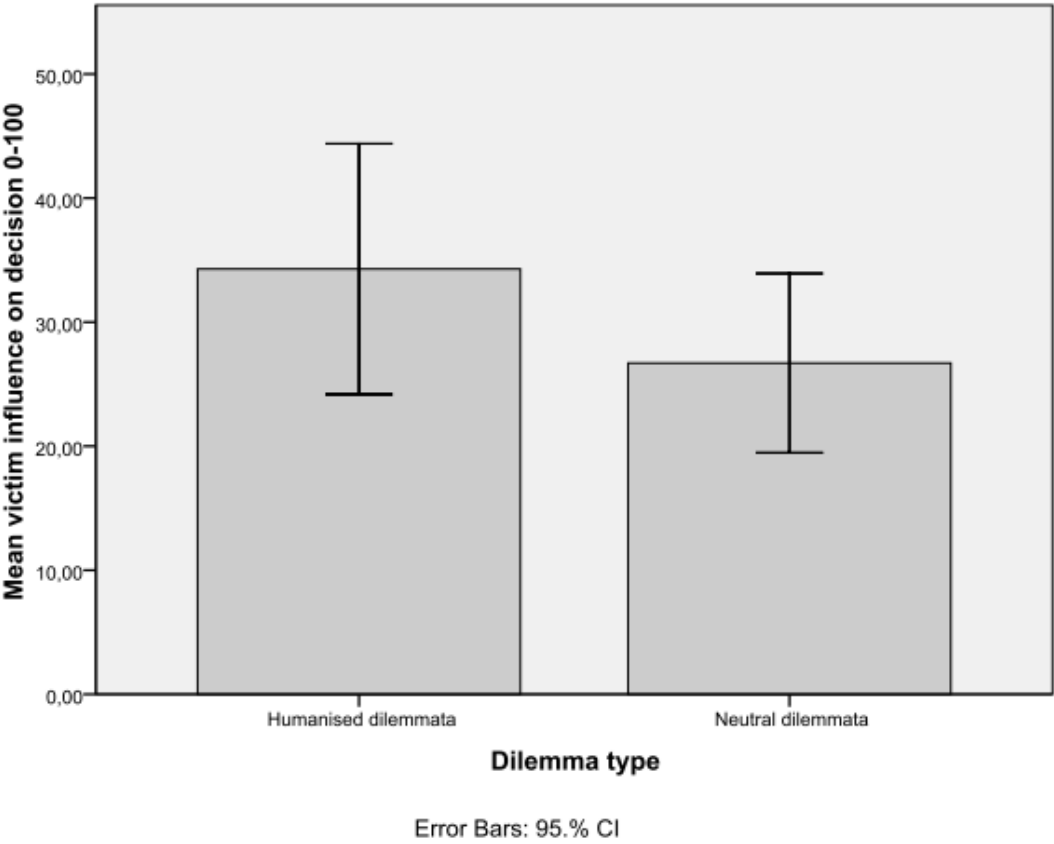


Figure 18: Rating of potential victims in humanised and neutral dilemma conditions according to whether participants' opinions about them influenced their decisions

The results for the separate question of how emotion ratings differed in a human-graphical and a text-graphical dilemma condition show higher levels of personal distress in human-graphical dilemmas. However, this result is only on the verge of being significant ( $t(44)=1.611$ ,  $p=0.057$ ,  $M(\text{human-graphical})=70.11$ ,  $M(\text{text-graphical})=58.97$ ,  $SE(\text{human-graphical})=5.21$ ,  $SE(\text{text-graphical})=4.54$ ). For the personal distress condition and all other independent t-tests in this section, Levene's test for equality of variances shows that the variances of the two groups are equal ( $p>0.05$ ).

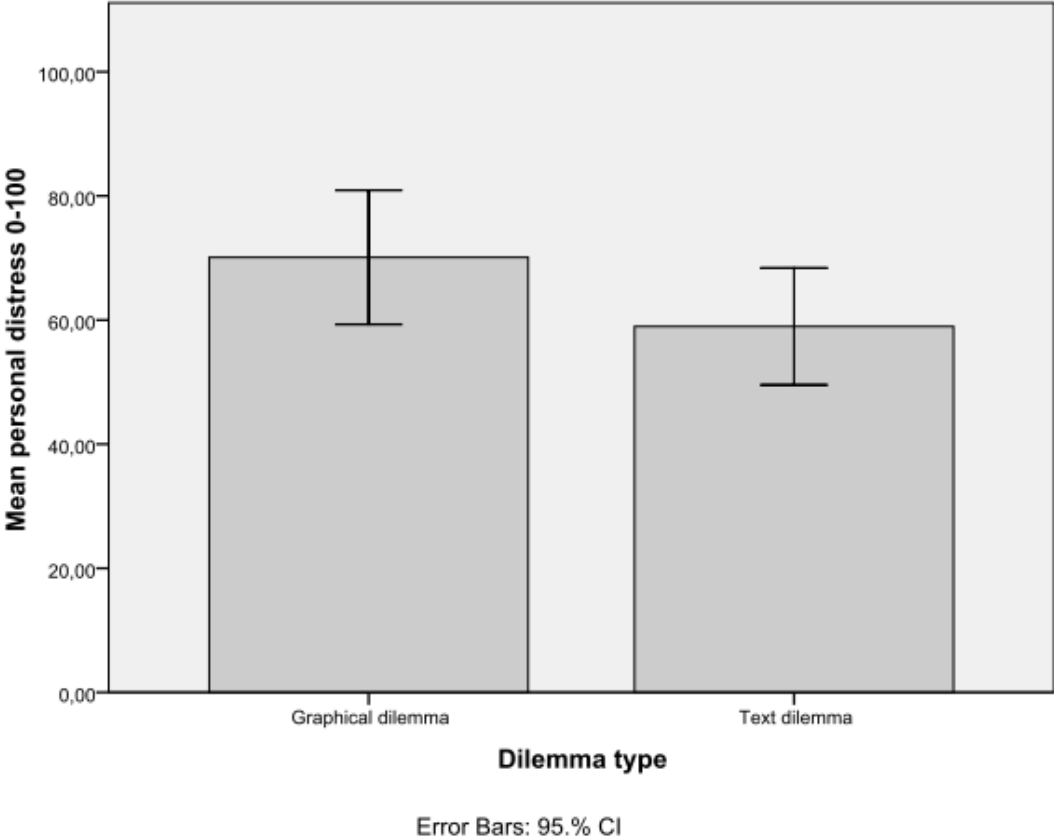


Figure 19: Personal distress ratings for human-graphical and text-graphical dilemma conditions



In contrast to personal distress ratings, empathic concern ratings do not differ in human-graphical and text-graphical dilemma conditions ( $t(44)=-0.032$ ,  $p=0.487$ ,  $M(\text{human-graphical})=63.00$ ,  $M(\text{text-graphical})=63.22$ ,  $SE(\text{human-graphical})=5.28$ ,  $SE(\text{text-graphical})=3.93$ ).

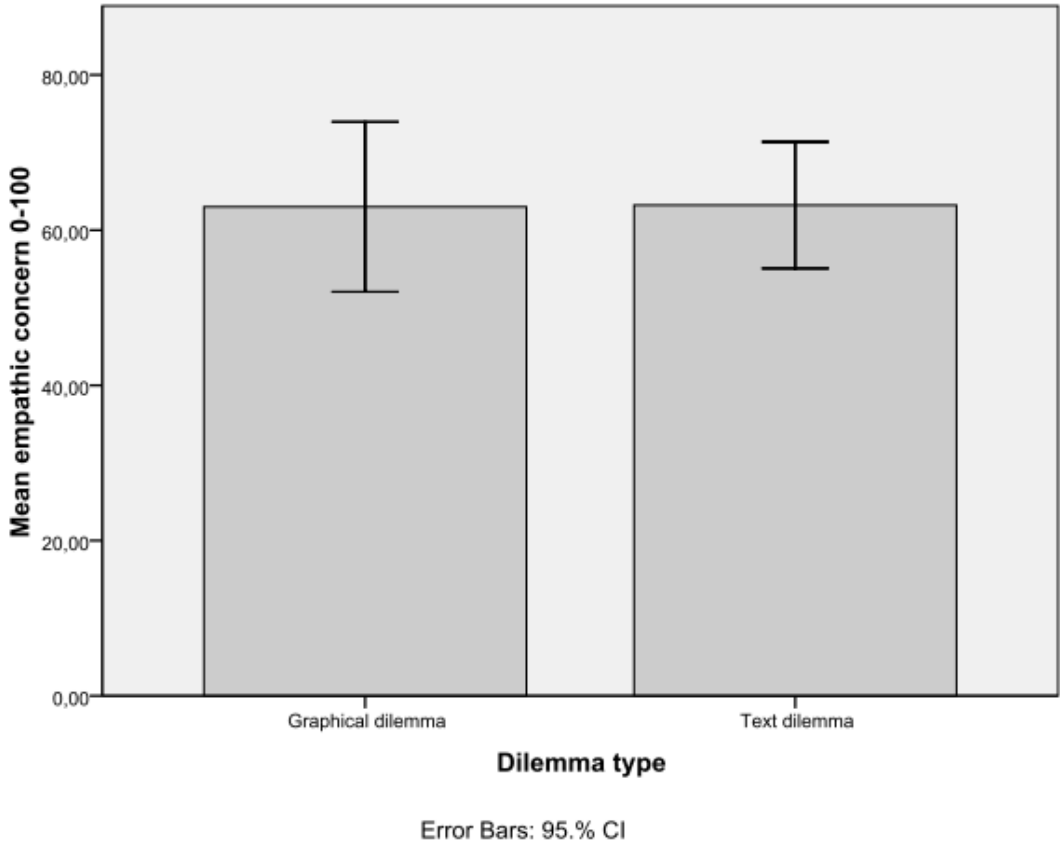


Figure 20: Empathic concern ratings for human-graphical and text-graphical dilemma conditions

Potential victim humanisation, however, is also on the same level in the human-graphical and text-graphical dilemmas ( $t(44)=0.317$ ,  $p=0.376$ ,  $M(\text{human-graphical})=63.45$ ,  $M(\text{text-graphical})=65.33$ ,  $SE(\text{human-graphical})=4.49$ ,  $SE(\text{text-graphical})=3.87$ ).

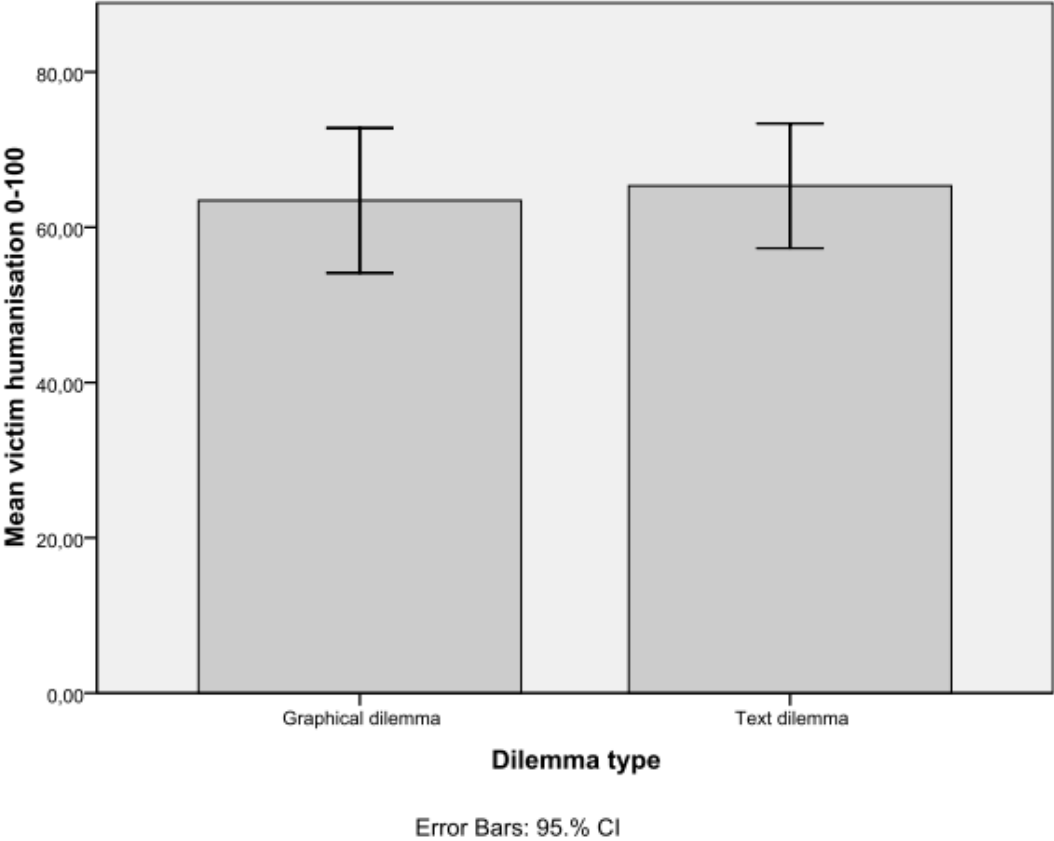


Figure 21: Potential victim humanisation ratings for human-graphical and text-graphical dilemma conditions

Finally, the human-graphical dilemma condition was perceived as significantly more difficult to decide on ( $t(44)=1.781$ ,  $p=0.041$ ,  $M(\text{human-graphical})=68.49$ ,  $M(\text{text-graphical})=50.49$ ,  $SE(\text{human-graphical})=7.33$ ,  $SE(\text{text-graphical})=6.96$ ).

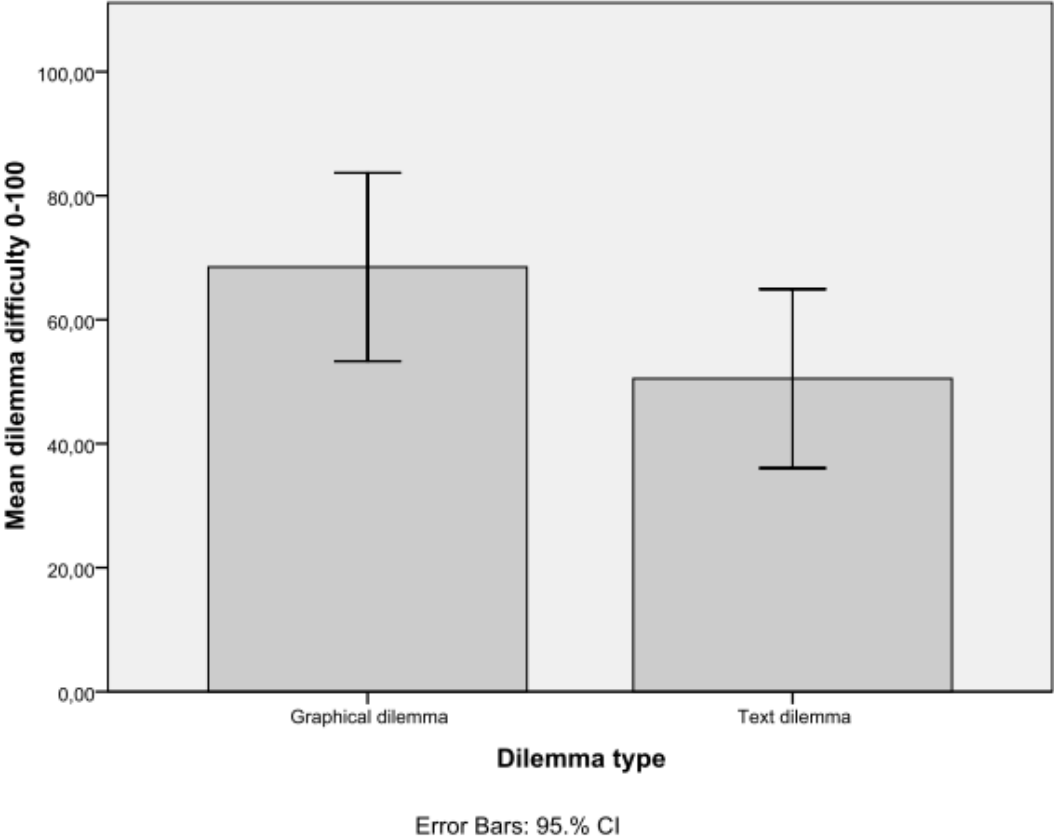


Figure 22: Dilemma difficulty ratings for human-graphical and text-graphical dilemma conditions

## Chapter 5 Discussion

As we will see, the findings presented in the previous section can shed some light on the question of which emotions are involved in moral decision making.

First, however, it has to be discussed why we were not able to replicate Majdandžić et al.'s findings that humanised potential victims were sacrificed less often than neutral ones. Our findings do not automatically disprove that result: as only two humanised and two neutral dilemmas were used in the present experiment (in contrast to 24 used in Majdandžić et al.'s study), each participant could only attain one of three possible utilitarian values for each of those two categories, as described in section 3.4. This results in quite crude possibilities for jumps in variance. It could be that the effect of humanisation influencing the decision outcome is too weak to be measured in a paradigm that uses just four dilemmas. Limiting the number of dilemmas was a tradeoff that was made in the current study in order to allow participants to remember each moment of decision as accurately as possible. Thus it is possible that Majdandžić et al.'s finding could be replicated using a greater number of items.

In the emotion ratings, we found a trend towards higher levels of feelings of personal distress in humanised dilemma conditions. The difference in group means of empathic concern, however, was more clearly insignificant.

This could mean that empathic concern is less important in making a moral decision than personal distress and that the greater activation found in the anterior insula and dorsal anterior and medial cingulate cortex in humanised dilemmas (Majdandžić et al., in prep.) is probably due to greater levels of personal distress rather than empathic concern. This interpretation can arguably be extended to the more general processes of moral cognition. Thus, it could explain the tendency to leave the potential victim in the footbridge dilemma unharmed, but to sacrifice that person in the trolley dilemma. The potential victim in the footbridge dilemma can be interpreted as being perceived more strongly as a human being, which would lead to increased levels of personal distress and, to a lesser extent, empathic

concern. The heightened sense of harming a human being in the footbridge dilemma would then prevent a greater number of participants from sacrificing that person.

However, on a methodological level, caution is advised when relying on the method of self-reported emotions alone. It is an unknown factor whether some emotions are prone to higher social desirability than others. A participant in a study knows that his or her responses will be evaluated and possibly judged by someone, even if researchers try to avoid this impression. The differences in means might therefore be related to different assumed social approval ratings for different emotional responses. To measure this effect, ratings of the current experiment would have to be compared with approval ratings in other studies.

An additional method of separating the effects of empathy and personal distress would be testing participants with high-functioning autism in comparison to neurotypical subjects. As autistic people are reported to have deficits in empathy, but probably not in self-related feelings of personal distress (Baron-Cohen & Wheelwright, 2004), a difference in decision responses could then be attributed to differences in empathic concern. Moreover, a useful future approach would be to combine self-reported emotion ratings with other methods of assessing emotional arousal, for example skin conductance and heart rate measurements. This would alleviate disadvantages of self-reported emotions such as delays between making a decision and reporting the emotions felt during that phase and the potentially different social desirability of reporting various emotions.

As expected, the extent of perceived victim humanisation was found to be higher in humanised than in neutral dilemmas. This shows that our way of describing a person's state of mind in the victim priming stories does indeed lead to greater awareness of humanness and greater feelings of responsibility. It is, however, unexpected that there was no difference in the perceived difficulty of humanised and neutral dilemmas and that making a utilitarian decision was perceived as equally difficult in humanised and neutral dilemmas. We hypothesised that perceiving the potential victim more strongly as a human being would make it harder for participants to decide whether to sacrifice the potential victim, especially if

the decision eventually taken was to sacrifice the person. Other conditions such as dilemma story, order, and photographs of the potential victim were randomised and should therefore leave no effect, i.e. could not have been the cause of elevating difficulty in neutral dilemmas or reducing it in humanised conditions. It can therefore only be concluded that increased perception of humanness was at best a very minor factor in judging the difficulty of a dilemma.

Part III, the more detailed assessment of potential victim humanness, showed that humanised potential victims were seen as more 'real' and 'alive' human persons. It can therefore be concluded that the method of referring to a potential victim's state of mind in a priming story does indeed alter a participant's perception of that victim. Therefore, humanisation can be seen as a useful concept for explaining differences in utilitarian decision making and could possibly extend to the famous examples of the trolley and footbridge dilemmas. In humanised conditions, the potential victim is seen as more alive, tangible, and less abstract - in other words, the victim is seen to a greater extent as a human being. Correspondingly, participants rated their connectedness and similarity to the humanised victims significantly higher than to neutral victims. Participants also indicated that their opinion of the humanised potential victims influenced their decisions to a greater extent than their opinion about neutral potential victims. However, it has to be stated that participants gave low overall ratings of perceived victim influence, which means that they saw the identity and characterisation of the potential victims as only a minor factor when making their decisions. However, it is questionable whether this self-reported causality is reliable. As Bargh and Chartrand (1999) emphasise, most cognition operates outside of consciousness, and people are not generally very good at knowing what factors led to their judgements. This author would therefore exclude such a measure in further studies.

The strong correlations between participants' responses to the different emotion rating scales imply the existence of 'high raters', 'medium raters', and 'low raters', i.e. participants who tend to stay within a chosen range of values for their answers and either differ in their ability

or willingness to report emotions or experience them in different intensities. Batson (1987) suggests an additional reason for a strong correlation between distress and empathic responses: people could on a general level differ in their receptiveness to perceive a victim's need, which could result in similar rating tendencies for both emotions.

The Interpersonal Reactivity Index, on the other hand, seems to measure something else than we did in this experiment, as the correlations between participants' IRI values for their 'trait' self-reported levels of personal distress, empathy, and perspective taking and their averaged values of personal distress, empathy, and dilemma difficulty ratings in the phase of decision making are close to zero. The one exception is the IRI fantasy scale (identification with fictional characters), which correlates very strongly ( $p < 0.001$ ) with the perceived humanisation of the potential victim during the moment of decision, and to a lesser extent ( $p = 0.184$ ) with moment-of-decision empathy.

This leaves us with several possible interpretations. First, there could be almost no correlations between the IRI values and decision phase measurements because the two scales measure two completely different constructs. While the 'state' ratings for personal distress ask for an assessment of how torn, frustrated, and stressed participants were when they made a decision in a particular dilemma, examples of the 'trait' IRI scale measurements are how ill-at-ease a participant generally feels in emergency situations (which is related to emotion control) or the extent of helplessness in emotional situations. Batson et al. (1987) also point out this distinction and argue against calling both measurements 'empathy'. This interpretation does not need to hold for each of the four IRI subscales. It could be possible that the fantasy scale and the humanisation scale each measure how strongly a fictional person is perceived as a human, which would explain the exceptionally strong positive correlation.

Second, while introspection is always questionable, rating emotions during decisions that were made a few minutes in the past is different from assessing one's general behaviour. The latter might tempt participants to give a rating that favours more what a participant would like to be like than what he or she actually is like. In this author's opinion, it is debatable and

at that point undecided to what extent the two measures could be compared - in future studies, a factor analysis of the items used for emotion ratings in the current study is advisable.

A third interpretation of the lack of correlations between the IRI scale and decision phase measurements has more fundamental implications for the theory that empathy and personal distress can really be separated clearly into 'other-related' and 'self-related' emotions, as Batson (1987) claims. This interpretation is only true if we do assume that the IRI scale and our moment-of-decision items measure similar concepts. In that case, we could argue that the overlap between empathy and personal distress is too strong to test the two concepts separately. It was shown in section 2.3.4 that even when experiencing empathy, i.e. an 'other-related' emotion, brain areas that normally process one's own emotions are recruited. This seems logical, as one's own emotions are the templates for the emotions that we expect someone else to feel. It could also be possible that the different items of personal distress and empathy that Batson proposes and which we modelled our items on are not phrased and defined clearly enough. More precise wording of emotion rating items should be striven for in future experiments. For example, in the current study we assumed that being moved signified empathic concern. However, we cannot be sure what participants understood under the term 'moved'. This should be clarified beforehand, perhaps by using the 'thinking aloud' technique. Moreover, adjectives for describing personal distress could be obtained by presenting participants with a highly stressful paradigm that does not involve another person suffering and therefore is devoid of empathic content. Such a paradigm might for instance involve hypothetical calamities happening to the participant him- or herself. Letting the participant describe how he or she felt after being exposed to such personal distress could yield additional insight into the words associated with personal distress. Similarly, paradigms for empathy without personal distress could be constructed. Participants could for example be asked to imagine the feelings of another person in a happy situation and then report their feelings towards that person. It is debatable whether such reports could be useful to rate



empathy in stressful situations, but it might still be worthwhile to explore the language that participants use when they empathically feel with another person.

Another possibility to further evaluate the viability of emotion rating scales would be if instead of hypothetical dilemmas, participants had to rate their emotions when thinking back to dilemma situations that they experienced in real life (as in Skoe et al., 2011). However, real-life dilemmas are unique and recollections of them cannot be controlled. Therefore, this method should only be an additional means to evaluate the plausibility of emotion ratings.

In the case of the graphical dilemma, we found a trend towards higher levels of personal distress for the human-graphical than for the text-graphical condition. Means for empathic concern, however, showed no difference. This suggests that graphical depictions of human features do not induce higher levels of empathy than a presentation without representations of humans, while higher levels of personal distress are elicited. However, the same limitations as for the emotion ratings in the main study hold for this experiment.

The result that is least clear and defies initial expectations is that humanisation was not rated higher in the human-graphical than in the text-graphical dilemma. The reason that this finding is surprising is that it does not fit the result that higher ratings of personal distress, and also difficulty (see figure 22) were elicited by the human-graphical presentation mode. If these differences are not due to greater perceived humanness of the potential victims, and all other conditions were randomised, it is unknown what causes them. In any case, in future experiments the graphical human stimuli should be more carefully constructed so that one can be more certain, for example, that their facial expressions signify distress.

Another difference in comparison to the main experiment was that the human-graphical dilemma was judged as being significantly harder to decide on than the text-graphical dilemma. This is what we expected, but there is more than one possible interpretation of this effect. Viewing distressed faces could automatically elicit distress without a conscious, cognitive difference in perceived humanness (this is called emotional contagion; see Barsade, 2002). In that case, our manipulation of humanisation would not be rendered invalid.

However, the possibility remains that consciously perceived humanness accounts for the difference in difficulty. This would mean that our methods for capturing humanisation have to be refined, as no effect of humanisation was found in the graphical experiment. For instance, whether feeling responsible for a person is related to perceiving them as human is a question that should be checked before including this item in a group measure. In any case, the greater reported difficulty is in line with the result that higher levels of personal distress were elicited in the human-graphical than in the text-graphical condition. After all, personal distress is likely a result of not being able to favour one solution over another, i.e., a greater behavioural conflict.

## **Chapter 6 Conclusion**

Moral decision making is an interaction of automated intuitions and conscious reasoning. The view that emotions are constituents of moral judgement is contrary to traditional philosophical views, but has been established in a variety of neuroscientific and behavioural experiments.

In the current study, we set out to examine what kind of emotions contributed to the fewer utilitarian responses and different patterns of brain activity that Majdandžić et al. (in prep.) found in dilemmas in which potential victims were humanised, i.e. described with a reference to their mental states. Behavioural and fMRI data implicated two components as possible constituents of that difference that, due to the fact that they recruit overlapping brain areas, could not be distinguished using fMRI. First, increased perception of a person as a human being could have led to greater awareness that killing a human is wrong, and thus to increased personal distress for the participant making that decision. Second, greater humanisation could have led to feelings of empathy via perspective taking that was induced by imagining the other's state of mind, resulting in a desire in the participant to reduce the suffering of the potential victim.

We modelled our paradigm on the previous study by Majdandžić et al. (in prep.). First, we primed participants with some hypothetical persons described in humanised and neutral

ways. Those individuals were then used as potential victims in dilemma stories in which they could either be sacrificed in order to save a greater number of people, or left unharmed at the expense of the others' lives. Participants then had to rate their emotions during the decision phase in the categories personal distress, empathic concern, and humanisation of the victim (as well as assess dilemma difficulty). The differentiation between personal distress and empathy was adapted from Batson (1987). As a further measurement of humanisation, participants had to rate each potential victim on scales of human-like features and similarity to oneself, i.e. a human being.

Finally, we introduced a human-graphical condition as a new form of dilemma presentation using schematic drawings of the humans featured in a dilemma story. It was hypothesised that the human-graphical condition would elicit higher emotion ratings than a dilemma presented in text form (with control 'scenery' graphics), as the faces present in the human-graphical condition should increase perceived humanisation.

We were not able to replicate Majdandžić et al.'s finding that humanised potential victims were sacrificed less often than neutral ones, but the results of the two studies are difficult to compare because of the limited number of dilemmas used in the current study. We did find higher ratings of personal distress and victim humanisation, but not of empathic concern, for humanised dilemmas. This suggests that personal distress plays a greater role than empathic concern in moral judgement, although empathic concern does seem to be involved as well. Humanised potential victims were also rated as more human-like and similar to the participants. This confirms that our manipulation of victim priming stories was effective. In the graphical dilemma condition, higher levels of personal distress were recorded in the human-graphical than in the text-graphical form, but ratings of empathic concern did not differ. Furthermore, mean ratings of personal distress exceeded those of empathic concern. These results further corroborate the view that in moral decision making, empathic concern is not as strong a component as personal distress.

The finding that humanisation in the human-graphical dilemma was not stronger than in the text-graphical dilemma is something of a puzzle, but can possibly be explained by the

emotional contagion elicited by viewing distressed faces. A more fundamental problem which should be addressed in future experiments is the question of adequate wording. For instance, it is unclear whether participants understand the term 'being moved' in the same way as the experimenters and whether the adjectives used to denote personal distress and empathy can exclusively be attributed to those concepts. Using supplementary psychophysiological methods of measuring emotional arousal such as skin conductance could also assist in separating personal distress from empathy more clearly. In such a way, our central finding, namely that victim humanisation leads to a greater increase in participants' feelings of personal distress, rather than empathic concern, could further be strengthened.

## References

- Alexander, L. & Moore, M., 2007. Deontological ethics. *Stanford Encyclopedia of Philosophy*. Available at: <http://plato.stanford.edu/entries/ethics-deontological/>.
- Archer, R.L. et al., 1981. The role of dispositional empathy and social evaluation in the empathic mediation of helping. *Journal of Personality and Social Psychology*, 40(4), p.786-796.
- Bargh, J.A. & Chartrand, T.L., 1999. The unbearable automaticity of being. *American Psychologist*, 54(7), p.462-479.
- Baron-Cohen, S. & Wheelwright, S., 2004. The empathy quotient: an investigation of adults with Asperger syndrome or high functioning autism, and normal sex differences. *Journal of Autism and Developmental Disorders*, 34(2), p.163–175.
- Barsade, S.G., 2002. The ripple effect: Emotional contagion and its influence on group behavior. *Administrative Science Quarterly*, 47(4), p.644–675.
- Bartels, D.M., 2008. Principled moral sentiment and the flexibility of moral judgement and decision making. *Cognition*, 108(2), p.381–417.
- Batson, C.D. et al., 1997. Perspective taking: Imagining how another feels versus imaging how you would feel. *Personality and social psychology bulletin*, 23(7), p.751-758.
- Bauer, L.O. & Hesselbrock, V.M., 1999. P300 decrements in teenagers with conduct problems: implications for substance abuse risk and brain development. *Biological Psychiatry*, 46(2), p.263–272.
- Blair, R.J.R., 1995. A cognitive developmental approach to morality: Investigating the psychopath. *Cognition*, 57(1), p.1–29.
- Blair, R.J.R. et al., 1997. The psychopathic individual: A lack of responsiveness to distress cues? *Psychophysiology*, 43(2), p.192-198.
- Broeders, R. et al., 2011. Should I save or should I not kill? How people solve moral dilemmas depends on which rule is most accessible. *Journal of Experimental Social Psychology*.
- Carr, L. et al., 2003. Neural mechanisms of empathy in humans: a relay from neural systems for imitation to limbic areas. *Proceedings of the National Academy of Sciences of the United States of America*, 100(9), p.5497-5502.
- Casebeer, W.D. & Churchland, P.S., 2003. The neural mechanisms of moral cognition: A multiple-aspect approach to moral judgement and decision-making. *Biology and Philosophy*, 18(1), p.169–194.

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), p.752-766.

Chen, S. & Chaiken, S., 1999. The heuristic-systematic model in its broader context. *Dual-process theories in social psychology*, p.73–96.

Cikara, M. et al., 2010. On the wrong side of the trolley track: Neural correlates of relative social valuation. *Social Cognitive and Affective Neuroscience*, 5(4), p.404-413.

Cushman, F. et al., 2006. The role of conscious reasoning and intuition in moral judgement. *Psychological science*, 17(12), p.1082-1089.

Damasio, A.R., 1994. *Descartes' error: Emotion, reason, and the human brain*. New York: Grosset/Putnam

Davis, M.H., 1983. Measuring individual differences in empathy: Evidence for a multidimensional approach. *Journal of Personality and Social psychology*, 44(1), p.113-126.

Decety, J. & Jackson, P.L., 2004. The functional architecture of human empathy. *Behavioral and Cognitive Neuroscience Reviews*, 3(2), p.71-100.

Decety, J. & Lamm, C., 2006. Human empathy through the lens of social neuroscience. *TheScientificWorldJOURNAL*, 6, p.1146–1163.

Eisenberg, N., 2000. Emotion, regulation, and moral development. *Annual review of psychology*, 51(1), p.665–697.

Ekman, P. & Friesen, W.V., 1975. *Unmasking the face: A guide to recognizing emotions from facial clues*. Oxford: Prentice-Hall

Ericsson, K.A. & Simon, H.A., 1998. How to study thinking in everyday life: Contrasting think-aloud protocols with descriptions and explanations of thinking. *Mind, Culture, and Activity*, 5(3), p.178–186.

Farrow, T.F.D. et al., 2001. Investigating the functional anatomy of empathy and forgiveness. *Neuroreport*, 12(11), p.2433-2438.

Field, A., 2009. *Discovering statistics using SPSS*. London: SAGE publications

Fletcher, P.C. et al., 1995. The mind's eye—precuneus activation in memory-related imagery. *Neuroimage*, 2(3), p.195–200.

Frith, C.D. & Frith, U., 2006. The neural basis of mentalizing. *Neuron*, 50(4), p.531–534.

Frith, U., 2001. Mind blindness and the brain in autism. *Neuron*, 32(6), p.969–979.

- Fultz, J. et al., 1986. Social evaluation and the empathy–altruism hypothesis. *Journal of Personality and Social Psychology*, 50(4), p.761-769.
- Fumagalli, M. et al., 2010. Brain switches utilitarian behavior: does gender make the difference? *PloS ONE*, 5(1), p.219–226.
- Gallagher, H.L. & Frith, C.D., 2003. Functional imaging of ‘theory of mind’. *Trends in Cognitive Sciences*, 7(2), p.77–83.
- Greene, J., 2003. From neural ‘is’ to moral ‘ought’: what are the moral implications of neuroscientific moral psychology? *Nature Reviews Neuroscience*, 4(10), p.846–850.
- Greene, J.D. et al., 2009. Pushing moral buttons: The interaction between personal force and intention in moral judgement. *Cognition*, 111(3), p.364–371.
- Greene, J.D. et al., 2008. Cognitive load selectively interferes with utilitarian moral judgement. *Cognition*, 107(3), p.1144–1154.
- Greene, J.D. et al., 2004. The neural bases of cognitive conflict and control in moral judgement. *Neuron*, 44(2), p.389–400.
- Greene, J.D. et al., 2001. An fMRI investigation of emotional engagement in moral judgement. *Science*, 293(5537), p.2105-2108.
- Greene, J. & Haidt, J., 2002. How (and where) does moral judgement work? *Trends in cognitive sciences*, 6(12), p.517–523.
- Grieve, A., 1984. Tests of sphericity of normal distributions and the analysis of repeated measures designs. *Psychometrika*, 49, p.257-267.
- Haidt, J., 2001. The emotional dog and its rational tail: A social intuitionist approach to moral judgement. *Psychological review*, 108(4), p.814-834.
- Harenski, C.L. et al., 2008. Gender differences in neural mechanisms underlying moral sensitivity. *Social Cognitive and Affective Neuroscience*, 3(4), p.313-321.
- Harris, L.T. & Fiske, S.T., 2006. Dehumanizing the lowest of the low. *Psychological Science*, 17(10), p.847-853.
- Huebner, B. et al., 2009. The role of emotion in moral psychology. *Trends in Cognitive Sciences*, 13(1), p.1–6.
- Hume, D., 1739/2007. *A treatise of human nature: a critical edition*. Oxford: The Clarendon press
- Jaffee, S. & Hyde, J.S., 2000. Gender differences in moral orientation: A meta-analysis. *Psychological Bulletin*, 126(5), p.703-726.

- Kahneman, D., 2003. A perspective on judgment and choice: Mapping bounded rationality. *American psychologist*, 58(9), p.697-720.
- Kiehl, K.A. et al., 2001. Limbic abnormalities in affective processing by criminal psychopaths as revealed by functional magnetic resonance imaging. *Biological Psychiatry*, 50(9), p.677–684.
- Kleinginna, P.R. & Kleinginna, A.M., 1981. A categorized list of emotion definitions, with suggestions for a consensual definition. *Motivation and emotion*, 5(4), p.345–379.
- Koenigs, M. et al., 2007. Damage to the prefrontal cortex increases utilitarian moral judgements. *Nature*, 446(7138), p.908–911.
- Kohlberg, L., 1971. *Stages of moral development*. Available at: <http://info.psu.edu.sa/psu/math/Stages%20of%20Moral%20Development%20According%20to%20Kohlberg.pdf>.
- Lamm, C. et al., 2007. The neural substrate of human empathy: effects of perspective-taking and cognitive appraisal. *Journal of cognitive neuroscience*, 19(1), p.42–58.
- Lamm, C. et al., 2011. Meta-analytic evidence for common and distinct neural networks associated with directly experienced pain and empathy for pain. *Neuroimage*, 54(3), p. 2492-2502.
- Lamm, C. et al., 2008. Perspective taking is associated with specific facial responses during empathy for pain. *Brain research*, 1227, p.153–161.
- Maddock, R.J., 1999. The retrosplenial cortex and emotion: new insights from functional neuroimaging of the human brain. *Trends in Neurosciences*, 22(7), p.310–316.
- Majdandžić, J. et al., Humanized perception of potential victims during moral decision making recruits empathy-related brain areas (*in prep.*).
- Moll, J. et al., 2003. Morals and the human brain: a working model. *Neuroreport*, 14(3), p.299-305.
- Moll, J. et al., 2005. The neural basis of human moral cognition. *Nature Reviews Neuroscience*, 6(10), p.799–809.
- Moore, A.B. et al., 2008. Who shalt not kill? Individual differences in working memory capacity, executive control, and moral judgement. *Psychological science*, 19(6), p.549–557.
- Morris, C.J. et al., 2000. Experimental analysis of the effectiveness of features in Chernoff faces. *Proceedings of the International Society for Optical Engineering*, p. 12–17.
- Morrison, I. et al., 2007. The sight of others' pain modulates motor processing in human cingulate cortex. *Cerebral Cortex*, 17(9), p.2214-2222.



- Naqvi, N. et al., 2006. The role of emotion in decision making. *Current Directions in Psychological Science*, 15(5), p.260-264.
- Olson, I.R. et al., 2007. The enigmatic temporal pole: a review of findings on social and emotional processing. *Brain*, 130(7), p.1718-1731.
- Oswald, P.A., 1996. The effects of cognitive and affective perspective taking on empathic concern and altruistic helping. *The Journal of social psychology*, 136(5), p.613–623.
- Robertson, D. et al., 2007. The neural processing of moral sensitivity to issues of justice and care. *Neuropsychologia*, 45(4), p.755–766.
- Rolls, E.T., 2004. The functions of the orbitofrontal cortex. *Brain and Cognition*, 55(1), p.11–29.
- Schutte, N.S. et al., 2001. Emotional intelligence and interpersonal relations. *The Journal of Social Psychology*, 141(4), p.523–536.
- Singer, T. & Lamm, C., 2009. The social neuroscience of empathy. *Annals of the New York Academy of Sciences*, 1156(1), p.81–96.
- Singer, T. et al., 2004. Empathy for pain involves the affective but not sensory components of pain. *Science*, 303(5661), p.1157-1162.
- Sinha, P. et al., 2006. Face recognition by humans: Nineteen results all computer vision researchers should know about. *Proceedings of the IEEE*, 94(11), p.1948–1962.
- Skoe, E.E.A. et al., 2002. The role of reported emotion in real-life and hypothetical moral dilemmas. *Personality and Social Psychology Bulletin*, 28(7), p.962-973.
- Teasdale, J.D. et al., 1999. Functional MRI study of the cognitive generation of affect. *American Journal of Psychiatry*, 156(2), p.209-215.
- Wilson, M., 2002. Six views of embodied cognition. *Psychonomic Bulletin & Review*, 9(4), p.625–636.

## Appendix A Victim priming stories

### Person A<sup>9</sup>

A is a 25-year-old man who has been an enthusiastic mountaineer since his youth. Above all, he loves the feeling of freedom that he has when he is in the mountains and feels the wind on his face. The chilly breeze causes him to experience a close bond to the mountains, to an extent that he has to watch that his excitement does not get him into trouble: once before, he got lost during an ascent under conditions of poor visibility, which was a traumatic adventure for him.

Whenever he thinks of that incident, he feels ashamed: his girlfriend got very worried about him and urged him to abstain from such tours in the future – but he didn't take her seriously and ridiculed her sorrows. Today, however, conditions are good: it is sunny, and after a demanding ascent he rests on the peak and enjoys the view. As he notices the fresh, cold breeze and enjoys the silence while the sun warms his face, he feels a smile on his face, and he knows that he will never give up mountaineering.

Why do you think A feels ashamed when he thinks back to his mountain adventure?

1. Because he now thinks he was too vain at that time and wanted to prove himself without any regard to his girlfriend.
2. Because he was not able to reach the peak despite his climbing experience.

Do you think A is sure that he will never get into such trouble again?

1. Yes, because he thinks he is now older, calmer, and more prudent.
2. No, because he knows that even under unfavourable conditions, his excitement will be so overwhelming that he will still accept risks very easily.

---

<sup>9</sup> Potential victims for this study were selected from a pool of victims A-J

## Person B

B is a 27-year-old man who often goes cycling on weekends. He mostly takes long trips with his bike that is especially suited for rocky undergrounds due to its thick tyres with a deep profile. He mostly does cross-country biking and then crosses a hilly area. For this reason his bike has 24 gears. When it rains, he often shortens his trips and only uses asphalt roads.

In those situations it is better to pump up the tyres a bit more. He also has another bike which he uses only to cycle to work and which he recently had to replace because it was already quite old. His new bike is nearly identical; the only difference is that its colour is now blue instead of green.

Why do you think it is necessary to pump up the tyres when cycling on an asphalt road?

1. Because then the tyres will have less rolling resistance.
2. Because then the tyres can compensate for the unevenness of the ground.

How many gears do you think most city bicycles have?

1. Nowadays most city bicycles have 21 or more gears.
2. Most city bicycles definitely have fewer than 21 gears.

## Person D

D is a 28-year-old man who likes eating out at good restaurants. This time, he ordered shrimps – he felt that this choice fitted nicely with the restaurant and his own taste. However, when a deliciously smelling lasagna dish is served at the table next to his, he suddenly doesn't feel like eating his shrimps anymore. Before he calls the waiter to the table to change his order, he ponders for a moment whether it would be embarrassing to change to a simple lasagna without even tasting the shrimps.

However, when the lasagna is put in front of him he is happy about his decision. He starts eating so hastily that he burns his tongue. But he doesn't care - the aromatic smell of melting cheese and fresh oregano pleases him much more than the shiny shrimps.

Why do you think that D ordered shrimps in the first place?

1. Because he has eaten shrimps often before and did not know that another dish could take his fancy.
2. Because he was embarrassed to order a simple dish in a restaurant.

What do you think D's plans are after having eaten his lasagna that quickly?

1. He will order another lasagna and eat it more slowly, as he found the first portion too small.
2. He will order some ice cream to cool his burnt tongue.

### **Person E**

E is a 25-year-old man who works in a supermarket as a storage manager. Mostly, he works in the evenings from 5 pm. Before he starts, he eats a sausage roll and drinks a glass of coke in the company cafeteria. His job consists of assigning duties to his employees and finding solutions to the problems that sometimes come up.

For example, on one occasion there was a problem in the meat storage room: a short circuit had caused a defect in the cooling system. The only solution was to transport the meat into the refrigeration room for vegetables. Additionally, E called an electrician to repair the circuit. The electrician arrived shortly after the incident, and therefore E was able to transport the meat back into the meat storage room that same evening.

Why do you think the short circuit happened in the meat storage room, as opposed to the vegetable storage room?

1. This was a coincidence - the short circuit could also have happened in the vegetable storage room.
2. Because the meat storage room uses up a lot more electricity due to the much lower temperature.

How long do you think meat may be stored outside the refrigeration room in a supermarket?

1. Not longer than 10 minutes.
2. The allowed time depends on the outside temperature.

### **Person G**

G is a 24-year-old man who likes watching action movies. For that purpose he even built a home cinema featuring a large screen and a red armchair. He really enjoys watching a movie before going to bed – he appreciates being on his own and not having to bother with the constant noise in a normal cinema. He is very proud of his own cinema. One day he wants to watch a movie that he has not seen for a long time.

When he has just started the film and made himself comfortable on his red armchair with a bit of popcorn, he notices that the tape does not contain the movie that he wanted, but one of his childhood videos. Wistfully he watches himself playing soccer with his brother. When he was a child he often played with him, but now it has been a long time since he last saw his brother, who lives abroad. Suddenly he feels like calling him and chatting about the past.

Why do you think that G has not called his brother for such a long time?

1. Because he has so much to tell him that he doesn't know where to start.
2. Because of his bad conscience; he finds it difficult to call his brother because he has not called him for such a long time.

Why do you think that G likes to watch action movies so much?

1. Because he recognises himself in athletic and adventurous actors.
2. Because he gets bored and falls asleep easily when he watches other types of movies.

### **Person H**

H is a 25-year-old man who often takes photographs. Mostly he takes sports photos, but sometimes he also takes pictures for magazine documentaries. His camera is specially equipped for fast shots and was very expensive. The sports photos that he takes are mostly soccer pictures because he often attends soccer games. Sometimes he sells the pictures to magazines and newspapers - that way, he earns quite a bit of extra money.

On one occasion, one of his pictures even appeared on page 1 of a newspaper. It was the picture of a famous soccer player who was just being attacked by a player from the opposing team. In the past, he always used slide film; a few years ago, however, he changed to a digital camera. This made processing and printing photos much quicker and cheaper. Unfortunately, this development also resulted in more competition from other photographers, which is why photography has always remained a hobby for him.

Why do you think fast-shooting cameras are so expensive?

1. Because for fast shots a lot of light and therefore a big lens is necessary.
2. Because fast shots require special electronic components.

How many slide films do you think are still sold nowadays?

1. Only 25% of the sales volume that was reached five years ago.
2. Only 12.5% of the sales volume that was reached five years ago.

## **Appendix B Moral dilemmas**

### **Pop concert**

You are in the audience of a pop concert and are standing on the first balcony. Suddenly a part of the second balcony above you collapses and falls onto five members of the audience who are standing nearby. The victims are all badly injured and must be retrieved quickly; otherwise they will bleed to death. In order to rescue the victims you have to lift part of the debris as quickly as possible. This would cause the debris to fall even farther, namely onto the ground floor. On this spot stands person X - he cannot hear you, and time is short. If you lift the balcony debris and let it fall down, the five members of the audience will survive. Person X on the ground floor will then be struck to death. If you do nothing, the five people under the debris will soon die, but nothing happens to person X.

Will you lift the debris?

### **Zoo**

You are a zoo employee and are doing an inspection round. Suddenly you see that the water trench around the gorilla enclosure is only half full. The gorilla colony is acting very agitatedly. The leader has already crossed the water trench, seized and injured a visitor, person X, and dragged him into the water trench where he lies unconscious. The other gorillas are approaching the water trench and attempt to reach other visitors. They have lost interest in the unconscious person X. You can only push the leader and the other gorillas back into the enclosure and prevent other visitors from getting hurt if you fill up the water trench quickly using a high pressure hose. The unconscious person X will then drown in the water trench. If you do not act, the gorillas will cross the water trench and, in their aggressive mood, kill or hurt several visitors. Person X in the water trench will then presumably be saved.

Will you operate the high pressure hose?

### **Flap bridge**

You control a flap bridge and have just opened the bridge. A boat, in which one of the boatmen, person X, is sitting on the mast beam, is passing under the bridge. Suddenly you see that a car, whose driver apparently overlooked a stop signal, is already on the tilted bridge. If you do not act, the car will soon be falling off the bridge which is tilting further and further. This would result in the death of or at least severe injuries for the five passengers. Although you cannot stop the bridge from opening, you can change the bridge's direction by turning a lever, so that the bridge closes again. This would prevent the car from falling off the bridge, and its passengers would be saved. However, person X, who is sitting on the mast beam and cannot leave the boat in time, would be crushed by the bridge and would at least be severely hurt. If you do not act, the car with its five passengers will fall off the bridge, but the boatsman, person X, will be saved.

Will you turn the lever to change the bridge's direction?



## Appendix C The Interpersonal Reactivity Index (IRI)

(see Davis, 1983)

The following statements inquire about your thoughts and feelings in a variety of situations. For each item, indicate how well it describes you by choosing the appropriate letter on the scale at the top of the page: A, B, C, D, or E. When you have decided on your answer, fill in the letter on the answer sheet next to the item number. READ EACH ITEM CAREFULLY BEFORE RESPONDING. Answer as honestly as you can. Thank you.

ANSWER SCALE:

1 (does not describe me well) - 2 - 3 - 4 - 5 (describes me very well)

1. I daydream and fantasize, with some regularity, about things that might happen to me. (FS)
2. I often have tender, concerned feelings for people less fortunate than me. (EC)
3. I sometimes find it difficult to see things from the 'other guy's' point of view. (PT) (-)
4. Sometimes I don't feel very sorry for other people when they are having problems. (EC) (-)
5. I really get involved with the feelings of the characters in a novel. (FS)
6. In emergency situations, I feel apprehensive and ill-at-ease. (PD)
7. I am usually objective when I watch a movie or play, and I don't often get completely caught up in it. (FS) (-)
8. I try to look at everybody's side of a disagreement before I make a decision. (PT)
9. When I see someone being taken advantage of, I feel kind of protective towards them. (EC)
10. I sometimes feel helpless when I am in the middle of a very emotional situation. (PD)
11. I sometimes try to understand my friends better by imagining how things look from their perspective. (PT)
12. Becoming extremely involved in a good book or movie is somewhat rare for me. (FS) (-)
13. When I see someone get hurt, I tend to remain calm. (PD) (-)
14. Other people's misfortunes do not usually disturb me a great deal. (EC) (-)

15. If I'm sure I'm right about something, I don't waste much time listening to other people's arguments. (PT) (-)
16. After seeing a play or movie, I have felt as though I were one of the characters. (FS)
17. Being in a tense emotional situation scares me. (PD)
18. When I see someone being treated unfairly, I sometimes don't feel very much pity for them. (EC) (-)
19. I am usually pretty effective in dealing with emergencies. (PD) (-)
20. I am often quite touched by things that I see happen. (EC)
21. I believe that there are two sides to every question and try to look at them both. (PT)
22. I would describe myself as a pretty soft-hearted person. (EC)
23. When I watch a good movie, I can very easily put myself in the place of a leading character. (FS)
24. I tend to lose control during emergencies. (PD)
25. When I'm upset at someone, I usually try to "put myself in his shoes" for a while. (PT)
26. When I am reading an interesting story or novel, I imagine how I would feel if the events in the story were happening to me. (FS)
27. When I see someone who badly needs help in an emergency, I go to pieces. (PD)
28. Before criticizing somebody, I try to imagine how I would feel if I were in their place. (PT)

NOTE: (-) denotes item to be scored in reverse fashion PT = perspective-taking scale FS = fantasy scale EC = empathic concern scale PD = personal distress scale

Scoring: 1 = 0, 2 = 1, 3 = 2, 4 = 3, 5 = 4

Except for reversed-scored items, which are scored:

1 = 4, 2 = 3, 3 = 2, 4 = 1, 5 = 0

## **Appendix D Extended abstract**

Moral dilemmas are characterised by the absence of a unique, optimal solution. Therefore, they are ideal for experimental designs measuring emotional conflict in moral behaviour. In experimental psychology, dilemmas usually feature a potential victim whose life can either be saved at the expense of several people's lives, or sacrificed so that the others can live. Surprisingly, dilemmas are not always decided on by following utilitarianism, i.e. the notion that the majority should be saved at the expense of the minority. Some dilemmas are typically solved in a utilitarian way by 80% of participants, others only by 25% (Cushman et al., 2006). Previous research attempted to explain this difference by evaluating the directness of the physical confrontation with or spatial proximity to the potential victim (Greene et al., 2001). Majdandžić et al. (in prep.) sought to unify these factors with the concept of 'humanisation', i.e. how strongly a potential victim is perceived as a human being with thoughts and feelings. Humanisation was induced by the stimulation of a participant to take the potential victim's perspective by referring to his or her state of mind in a short story. In a mixture of behavioural and fMRI experiments, Majdandžić et al. found that humanised potential victims were sacrificed less often than neutral ones. Additionally, decisions involving humanised potential victims elicited higher activations in brain regions associated with emotional conflict as well as empathy (among others, the dorsal anterior cingulate cortex and the anterior insula).

In accordance with previous research (Cushman et al., 2006; Greene et al., 2001; Haidt, 2001), Majdandžić et al. were able to demonstrate that emotions, not only reason as proposed by utilitarianism, play a role in moral decision making. The strongest candidates were self-related feelings of personal distress and other-related empathic concern. Empathy (or understanding what another person feels and what that feels like) could have been induced by perspective taking in the case of humanised potential victims. However, humanisation could also have led to increased awareness that a human being was about to be 'killed', which would likely have resulted in heightened feelings of personal distress. The

extent of the involvement of personal distress and empathy in moral decision making could not be determined using fMRI, as both categories of emotions utilise overlapping neural structures (Singer & Lamm, 2009).

This behavioural study seeks to close that gap. 54 male participants were tested (gender differences in moral decision making have been documented; see Skoe et al., 2011). As in Majdandžić et al.'s study, participants were first primed with humanising or neutral stories about individuals. They later featured as potential victims in four moral dilemmas that were presented to participants as a text. After the participants had decided on the dilemmas, they were asked about their empathic (e.g. 'How strongly did you imagine the thoughts and feelings of person X?') and self-related aversive (e.g. 'How stressed did you feel?') emotions. Participants rated their perceived emotions with a slider that translated to values from 0 to 100. The differentiation between personal distress and empathy and the terms associated with those concepts had first been described by Batson (1987) and were further developed in an explorative pilot study that let participants talk freely about the emotions they felt while making a moral decision. Furthermore, potential victims were rated on scales such as 'alive', 'tangible', or 'abstract' in order to assess whether humanisation had been constructed successfully.

Our main result is that moral judgements involving humanised as opposed to neutral potential victims seem to enhance mainly feelings of personal distress, and to a lesser extent empathic concern. This could mean that the motivation behind moral behaviour is not purely altruistic, but includes an element of self-related stress relief. Humanisation was found to be induced successfully. However, self-reporting scales should be subject to further testing and enhanced by biophysical methods of assessing emotional arousal.

In order to explore the influence of less abstract modes of dilemma presentation, a graphical dilemma was included at the end of the experiment. This dilemma featured schematic drawings of human faces in comparison to a control dilemma that consisted only of scenery and text. Personal distress ratings were significantly higher for the human-graphical dilemma,

suggesting that more direct modes of presentation could influence participants' emotional involvement in experimental settings.

## Appendix E Zusammenfassung

Moralische Dilemmata sind dadurch charakterisiert, keiner einzig korrekten Lösung zugänglich zu sein. Alle möglichen Entscheidungen können von verschiedenen Menschen als angebracht oder unangebracht gewertet werden. In der experimentellen Psychologie werden vor allem solche (hypothetischen) Szenarien verwendet, bei denen das Leben einer Person geopfert werden kann, um mehrere andere Personen zu retten (Eisenberg, 2000). Der Utilitarismus, eine der dominierenden Theorien der Entscheidungsfindung, würde nun voraussagen, dass mehrere Menschenleben Vorrang vor einem einzigen hätten und sich Menschen nach diesem rationalen Grundsatz richten. Überraschenderweise kamen mehrere Studien zum Ergebnis, dass das keineswegs der Fall ist - manche Dilemmata werden von durchschnittlich 80 % der Versuchsteilnehmer utilitaristisch gelöst, andere nur von 25 % (Cushman et al., 2006). Dieses Rätsel wurde schließlich dadurch erklärt, dass in manchen Situationen die Konfrontation mit der Person, die geopfert werden sollte, physisch direkter ist (Greene et al., 2001). Majdandžić et al. (in Vorbereitung) hingegen nannten als übergeordneten Faktor das Ausmaß, inwieweit diese Person als Mensch mit Gedanken und Gefühlen wahrgenommen wurde. In einer Mischung aus Verhaltens- und fMRI-Experimenten gelangten Majdandžić et al. zu dem Ergebnis, dass solche ‚humanisierten‘ Personen seltener geopfert wurden als neutral beschriebene. Außerdem wurden bei der Entscheidung über solche als ‚human‘ wahrgenommenen potentielle Opfer Gehirnregionen, die mit der Verarbeitung von Emotionen sowie mit Empathie assoziiert werden, stärker aktiviert.<sup>10</sup>

Majdandžić et al. konnten mit diesen Daten (übereinstimmend mit anderen Studien, siehe Cushman et al., 2006; Greene et al., 2001; Haidt, 2001) zwar erklären, dass emotionale Prozesse in das Treffen moralischer Entscheidungen involviert sind. Ob jedoch eher empathische oder selbstbezogen-aversive Emotionen eine größere Rolle spielen, konnte nicht bestimmt werden, da diese unterschiedlichen Prozesse überlappende neurale

---

<sup>10</sup> Beispiele für solche Regionen sind der anteriore und posteriore cinguläre Cortex und die anteriore Insel.

Strukturen nutzen (Singer & Lamm, 2009). Empathie ergibt sich aus der Möglichkeit, die Perspektive einer ‚humanisierten‘ Person im Vergleich zu einer Person, die ohne Referenz zu ihren Gedanken und Gefühlen beschrieben wurde, leichter einzunehmen. Dieselbe Humanisierung könnte jedoch auch verstärkt das Gefühl hervorrufen, einen *Menschen* zu opfern und somit eine Stärkung der selbstbezogenen Stress- und Aversionsgefühle bewirken. Diese Studie versucht, die Frage zu klären, ob empathische oder selbstbezogen-aversive Gefühle eine größere Rolle in moralischen Entscheidungsprozessen spielen, wobei vor Beginn der Studie noch keine Präferenz in eine der beiden Richtungen angegeben werden konnte. Der experimentelle Aufbau und die Beschränkung auf männliche Versuchspersonen (bei emotionalen Entscheidungsprozessen gibt es Differenzen zwischen Männern und Frauen; siehe Skoe et al., 2011) wurden von Majdandžić et al. übernommen, allerdings ist das hier beschriebene Experiment eine reine Verhaltensstudie. Zuerst lasen die 54 Versuchspersonen je eine kurze Geschichte über eine Person, die entweder ‚humanisiert‘ oder neutral beschrieben wurde. Humanisierung wurde durch Beschreibung der Gefühle und Gedanken der Person erzielt, während eine neutrale Beschreibung sich auf Beschreibungen auf Sachebene beschränkte. Diese Personen spielten danach die Rolle als potentiell Opfer in vier moralischen Dilemmata, die den Versuchspersonen als Text präsentiert wurden. Nachdem die Versuchspersonen ihre Entscheidung (Person opfern oder nicht) bekanntgegeben hatten, wurden sie mit verschiedenen Fragebogenmaßen nach ihren empathischen (z.B. ‚Wie stark stellten Sie sich die Gedanken und Gefühle dieser Person vor?‘) und selbstbezogen-aversiven Emotionen (z.B. ‚Wie gestresst fühlten Sie sich während der Entscheidung?‘) während des Entscheidungsprozesses befragt. Ein solches Schema war erstmals von Batson et al. (1987) entwickelt worden und wurde im Rahmen dieser Studie mit einer explorativen Pilotstudie, in der Versuchspersonen assoziativ und spontan über ihre Emotionen in solchen Entscheidungssituationen sprachen, erweitert. Außerdem wurden die potentiellen Opfer verschiedenen Begriffen wie lebendig oder abstrakt zugeordnet, um ein weiteres Maß für die erfolgreiche Humanisierung der Personen zu erlangen.

Das Hauptergebnis der Studie ist, dass Humanisierung der potentiellen Opfer einen Einfluss auf die Gefühlssituation während moralischen Entscheidungen ausübt. Selbstbezogen-aversive Gefühle wurden in humanisierten Dilemmata höher erlebt als bei der Entscheidung über neutrale Dilemmata, während sich empathische Emotionen nicht änderten. Diesen wird daher eine geringere Rolle im Prozess der Entscheidungsfindung zugestanden. Dieses Ergebnis impliziert auch, dass neben Altruismus auch der Wunsch nach einer Verringerung des eigenen Stressniveaus zur Motivation bei moralischen Entscheidungen beitragen kann. In zukünftigen Studien sollte jedoch jedenfalls die Validität der emotionsbezogenen Fragebögen durch zusätzliche Methoden wie Messung der elektrodermalen Aktivität und der Herzfrequenzvariabilität überprüft werden.

Um einen ersten Schritt in Richtung größerer emotionaler Direktheit bei der Präsentation von moralischen Dilemmata zu setzen und zu erheben, inwieweit die Präsentationsform eines moralischen Dilemmas Einfluss auf die Intensität der von den Versuchspersonen angegebenen Gefühle hat, wurde am Ende der Studie ein grafisches Dilemma präsentiert, bei dem die handelnden Personen in ihrer Notlage schematisch dargestellt werden. Es wurde ein Anstieg der von den Versuchspersonen wahrgenommenen selbstbezogen-aversiven Gefühle festgestellt, was eine größere emotionale Involviertheit bei Präsentationsarten mit verringertem Abstraktionsgrad nahelegt.



## **Appendix F Curriculum vitae**

**E-mail:** elisabeth\_engl@hotmail.com

**Date of Birth:** 30<sup>th</sup> October, 1985

**Place of Birth:** Linz, Austria

### **CURRENT GRADUATE EDUCATION**

---

**10/2009-06/2011 MSc in Cognitive Science, University of Vienna**

### **PREVIOUS GRADUATE/UNDERGRADUATE EDUCATION**

---

**10/2004-06/2009 Diploma (Magister) in Political Science, University of Vienna**

Degree with distinction. Thesis topic: Sexual violence in a new war - a case study of the conflict in the Democratic Republic of the Congo.

**10/2007-09/2009 BA in Sinology, University of Vienna**

Degree with distinction. In 2008, I also spent six months at the Shanghai Conservatory of Music (Chinese and piano courses).

**10/2004-01/2008 Diploma in classical piano (concert division), University of Music, Vienna**

First diploma 2006 (with distinction), second in January 2008.

**09/2003-07/2004 First year of Bachelor of Music in classical piano, Royal College of Music, London**

**06/2003 Matura (school leaving certificate), Europagymnasium Auhof, Linz, Austria**

Starting in 2000, I was also a full-time student of piano and conducting at the Bruckner Conservatory in Linz and spent six months at Oxford High School in 2001.

### **PREVIOUS RESEARCH EXPERIENCE**

---

**09-12/2010 (20 hrs/week): Department of Cognitive Sciences, Budapest University of Technology and Economics**

Supervisor: Dr. Gyula Kovács. I independently planned, conducted, and analysed face processing and cross-modal perception experiments using psychophysical methods, EEG, and eye tracking.

**07-08/2010 (full time): The Sinha Laboratory for Vision Research, Massachusetts Institute of Technology**

Supervisor: Dr. Pawan Sinha. I worked in a small group planning, conducting, and analysing face perception experiments using EEG.

**02-06/2010 (20 hrs/week): Department of Biochemistry and Molecular Biology, Center for Brain Research, Medical University of Vienna**

Supervisor: Dr. Werner Sieghart. In this project, I investigated aspects of the distribution of GABA<sub>A</sub> receptors in rodents displaying high anxiety-related behaviour (HAB).

**WORK EXPERIENCE**

---

**2000-2008 Piano soloist**

Concerts (solo and ensemble) in Europe, Asia, and the US. Winner of several soloist competitions.

**1999-2008 Private piano teacher**

I also organized music outreach programmes in schools and held a piano master class at the Instituto Superior de Arte, Havana, in 2007.