



Terms and Conditions of Use of Digitised Theses from Trinity College Library Dublin

Copyright statement

All material supplied by Trinity College Library is protected by copyright (under the Copyright and Related Rights Act, 2000 as amended) and other relevant Intellectual Property Rights. By accessing and using a Digitised Thesis from Trinity College Library you acknowledge that all Intellectual Property Rights in any Works supplied are the sole and exclusive property of the copyright and/or other IPR holder. Specific copyright holders may not be explicitly identified. Use of materials from other sources within a thesis should not be construed as a claim over them.

A non-exclusive, non-transferable licence is hereby granted to those using or reproducing, in whole or in part, the material for valid purposes, providing the copyright owners are acknowledged using the normal conventions. Where specific permission to use material is required, this is identified and such permission must be sought from the copyright holder or agency cited.

Liability statement

By using a Digitised Thesis, I accept that Trinity College Dublin bears no legal responsibility for the accuracy, legality or comprehensiveness of materials contained within the thesis, and that Trinity College Dublin accepts no liability for indirect, consequential, or incidental, damages or losses arising from use of the thesis for whatever reason. Information located in a thesis may be subject to specific use constraints, details of which may not be explicitly described. It is the responsibility of potential and actual users to be aware of such constraints and to abide by them. By making use of material from a digitised thesis, you accept these copyright and disclaimer provisions. Where it is brought to the attention of Trinity College Library that there may be a breach of copyright or other restraint, it is the policy to withdraw or take down access to a thesis while the issue is being resolved.

Access Agreement

By using a Digitised Thesis from Trinity College Library you are bound by the following Terms & Conditions. Please read them carefully.

I have read and I understand the following statement: All material supplied via a Digitised Thesis from Trinity College Library is protected by copyright and other intellectual property rights, and duplication or sale of all or part of any of a thesis is not permitted, except that material may be duplicated by you for your research use or for educational purposes in electronic or print form providing the copyright owners are acknowledged using the normal conventions. You must obtain permission for any other use. Electronic or print copies may not be offered, whether for sale or otherwise to anyone. This copy has been supplied on the understanding that it is copyright material and that no quotation from the thesis may be published without proper acknowledgement.

The Cognitive Psychology of Moral Reasoning

Eoin Gubbins

PhD

University of Dublin, Trinity College

2014

This research was funded by a Government of Ireland scholarship from the Irish

Research Council for Humanities and Social Sciences

Declaration

I declare that this thesis has not been submitted as an exercise for a degree at this or any other university and it is entirely my own work.

I agree to deposit this thesis in the University's open access institutional repository or allow the Library to do so on my behalf, subject to Irish Copyright Legislation and Trinity College Library conditions of use and acknowledgement.

Signed: *Eoin Gubbins*

Date: *13/02/14*



Thesis 10356

Acknowledgments

First I would like to express my gratitude to my supervisor, Ruth Byrne, for her knowledge, advice, and generosity with her time. I could not have hoped to have had a better supervisor. I would also like to thank my internal appraisers, Shane O'Mara and Lorraine Swords for their thoughtful advice and comments on my work throughout the three years of my studies, along with everyone in the School of Psychology and the Institute of Neuroscience at Trinity College Dublin who gave me helpful comments on my research.

I would like to thank Phil Johnson-Laird for collaborating with myself and my supervisor on the experiments reported in Chapter 2, and for welcoming me during my research visit to Princeton University. Special thanks also go to Max Lotstein for programming Experiment 1.

Thanks also to Jessica Stanley, Gemma Gordon, and Mary Parkinson for helping with inter-rater reliability coding; Iseult Cremen for data entry and for help with data collection; and Laura Deighan and Hannah Jellings for help with data collection.

I am very grateful to the Irish Research Council for the Humanities and Social Sciences for supporting my research. I am also thankful for support from the Graduate Studies Office at Trinity College Dublin and the Graduate School at Princeton University for the funding I received from them.

Thanks are also due to reviewers at *Thinking & Reasoning* for their comments on an article based on the experiments reported in Chapter 3; to the audience at the conference of the European Society for Cognitive Psychology in San Sebastian, Spain in September 2011 for their comments on a poster based on Experiment 4; to the audience at the International Conference on Thinking in London, England in July

2012 for their comments on a talk based on Experiment 5; and to reviewers for the CogSci conference in Berlin, Germany in August 2013, as well as the audience at CogSci in Berlin for their comments on a poster based on Experiments 1 and 3.

Thanks to my parents for their love and support; to my sister Sarah, for being inspirational; and to the rest of my family. Thanks to Hazel McCarthy, Deirdre Twomey, Róisín McManus, Emmett McNickle, Darina Errity, Mary Parkinson, Jessica Stanley, and Sonja Delmonte for being there with me. Thanks to Aoife Clarke, Bridget Finnegan, Catherine Rock, Jean Herlihy, Fionnuala Malone, Melissa Daly, Declan Ahern, Owen Power, Podge O’Caoinleáin, Nigel O’Callaghan, Helen Nolan, Colm O’Gríofa, Finola Casey, and Aodhnait O’Neill, for forcing me to enjoy myself. Finally, I would like to thank Marie Boland for generally being the best.

Table of Contents

	Page
Chapter 1 Introduction	16
Heuristics and Moral Consistency	19
Emotions and Reasoning in Moral Judgment	26
Elevation, Counterfactuals and Moral Action	41
Aims of the Present Research	44
Chapter 2 Consistency in moral judgments	
Introduction	47
Experiment 1	54
Experiment 2	64
Experiment 3	74
General Discussion	89
Chapter 3 Information about Emotions Modulates Reasoning about Moral Dilemmas	
Introduction	97
Experiment 4	104
Experiment 5	116
General Discussion	131
Chapter 4 Elevation and counterfactual thinking	
Introduction	138
Experiment 6	151
Experiment 7	164
Experiment 8	175
General Discussion	187

Chapter 5 Discussion

	Moral Consistency	195
	Emotion as Information in Moral Reasoning	202
	Elevation, emulation and counterfactual thoughts	210
	Implications for understanding moral psychology	216
References		224
Appendices	Appendix 2.1 – Experimental Materials	236
	Experiments 1 and 2	
	Appendix 2.2 – Consent and Debriefing	238
	Experiment 1	
	Appendix 2.3 – Complete criteria for	240
	consistency for Experiments 1 and 2	
	Appendix 2.4 – Experiment 2 Consent and	241
	Debriefing	
	Appendix 2.5 – Criteria for Consistency and	244
	Inconsistency in Experiment 3	
	Appendix 2.6 – Experiment 3 Materials	245
	Appendix 2.7 – Experiment 3 Consent and	247
	Debriefing	
	Appendix 3.1 – Experiment 4 Materials	249
	Appendix 3.2 – Experiment 4 Consent and	251
	Debriefing Forms	
	Appendix 3.3 – Experiment 5 Materials	254
	Appendix 3.4 – Experiment 5 Consent and	258
	Debriefing Forms	
	Appendix 4.1 – Materials for the Good condition	261
	in Experiment 6	
	Appendix 4.2. – Experiment 6 Exclusion Criteria	264
	Appendix 4.3 – Experiment 6 Inter-rater reliability	265
	Appendix 4.4. – Experiment 7 exclusion criteria	267
	Appendix 4.5. – Experiment 7 Inter-rater reliability	268
	Appendix 4.6. – Experiment 8 Exclusion criteria	269

Appendix 4.7. – Experiment 8 Inter-rater reliability 271

List of Tables and Figures

Table/Figure	Title	Page
Table 2.1	The morally good, morally bad, and neutral conjunctions for Experiment 1, and the participants' mean judgments of the first conjunct, <i>A</i> , the second conjunct, <i>B</i> , and their conjunction, <i>A&B</i> , in Experiment 1.	58
Figure 2.1	The mean number of triples in the three categories of consistency and violation depending on whether the conjunction was first or last in a triple of judgments.	61
Table 2.2	The moral, immoral and neutral conjunctions for Experiment 2, and the participants' mean judgments of the first conjunct, <i>A</i> , the second conjunct, <i>B</i> , and their conjunction, <i>A&B</i> , in Experiment 2 on the coarse scale, and fine scale (in parentheses).	68
Figure 2.2	The mean number of triples in the three categories of consistency and violations depending on whether the conjunction was first or last in a triple of judgments, and on whether a fine scale or coarse scale was used.	70
Figure 2.3	The mean number of consistent triples for Experiment 2 for moral and immoral contents, depending on whether the conjunction was first or	71

last in a triple of judgments for both fine and coarse scales.

Table 2.3	Matched, Higher Moral, and Higher Immoral materials, with the average ratings for the moral and immoral conjuncts and the conjunction in Experiment 3.	82
Figure 2.4	Mean number of consistent, moral inconsistent, and immoral inconsistent triples for conjunction-first and last triples.	84
Figure 2.5	The mean proportion of triples in the three categories of consistency and violation depending on whether the conjunction was first or last in a triple of judgments, and on whether the materials were Matched, Higher Moral or Higher Immoral.	85
Figure 3.1	Percentages of decisions to act in Experiment 4 for personal and impersonal dilemmas in the non-directed, emotion-directed and reason-directed conditions.	109
Figure 3.2	Percentages of justifications featuring emotive content for personal and impersonal dilemmas in the non-directed, emotion-directed and reason-directed conditions in Experiment 4.	111
Figure 3.3	Percentages of main types of reasons for personal and impersonal dilemmas in	112

	Experiment 4.	
Figure 3.4	Mean latencies for the emotion sentences in Experiment 5 for personal and impersonal dilemmas.	124
Figure 3.5	Mean latencies for the decision sentences in Experiment 5 for personal and impersonal dilemmas for the decision to act and not to act, prefaced by strong and mild emotions.	126
Figure 3.6	Percentages of judgments that the decision was appropriate in Experiment 5.	127
Figure 3.7	Mean latencies for judgments about the appropriateness of the decision in Experiment 5.	128
Figure 4.1.	Proportions of counterfactuals as downward and subtractive, and intentions to change, as well as general rather than specific intentions to change, for good and bad memories in Experiment 6.	159
Figure 4.2.	Proportions of counterfactual precursors (participants' ratings on the 0-10 scale for how expected and intended the behaviour in their memory was, and proportions of participants whose responses were categorized as actions, or as specific) for good and bad conditions in Experiment 6.	161
Figure 4.3.	Proportions of results for mood before recording the memory and mood after	162

(these are based on the ratings on the 0-10 scale participants provided), and of positive single words to describe the participants' feelings, for good and bad conditions for Experiment 6.

Figure 4.4. Proportions of counterfactuals as downward 169
and as subtractive, and intentions to change,
as well as general rather than specific intentions
to change, for good and bad memories in remote and recent
conditions in Experiment 7.

Figure 4.5 Proportions of counterfactual precursors 171
(participants' ratings on 0-10 scale for how
expected and intended the behaviour in their
memory was, and proportions of participants
whose responses were categorized as actions,
or as specific) for good and bad memories, in
remote and recent conditions, in
Experiment 7.

Figure 4.6. Proportions of results for mood before recording 173
the memory and mood after (these are based on
the ratings on the 0-10 scale participants provided),
and of positive single words to describe the
participants' feelings, for good and bad memories
in remote and recent conditions for
Experiment 7.

Figure 4.7.	Proportions of counterfactuals as downward and as subtractive, and intentions to change, as well as general rather than specific intentions to change, for good and bad memories in self-focused and other-focused conditions in Experiment 8.	181
Figure 4.8.	Proportions of counterfactual precursors (participants' ratings on 0-10 scale for how expected and intended the behaviour in their memory was, and proportions of participants whose responses were categorized as actions, or as specific) for good and bad memories, in self-focused and other-focused conditions, in Experiment 8.	184
Figure 4.9.	Proportions of results for mood before recording the memory and mood after (these are based on the ratings on the 0-10 scale participants provided), and of positive single words to describe the participants' feelings, in good and bad memories, in self-focused and other-focused conditions for Experiment 8.	185

Summary

In this thesis, the results of eight experiments are presented investigating the cognitive processes involved in moral reasoning. Two major themes were explored, involving important contemporary debates in moral psychological research. The first of these themes was the role of domain-general cognitive processes in moral judgment. The second was the role of emotions as information in moral reasoning. These issues were explored over the course of three experimental series.

The first series of experiments showed that people are sometimes inconsistent when they make numeric moral judgments. Participants sometimes judged, for example, “*A man attacked a security guard*” to be immoral to the degree of -70, “*A man stole money from a security guard*” to be immoral to the degree of -40, and “*A man attacked a security guard and stole money from him*” to be immoral to the degree of -70, even though two immoral acts should be worse than one. Consistency in these judgments was also influenced by irrelevant factors such as whether participants judged the pair of actions first, or the individual actions first. People were also sometimes inconsistent when making mixed-valence moral judgments. For example, participants sometimes judged “*A person hit his son*” to be immoral to the degree of -60, “*A person took his sister’s child into his home after his sister died*” to be moral to the degree of +60, but judged their conjunction, “*A person hit his son, and the next week took his sister’s child into his home after his sister died*” to be immoral to the degree of -30. This suggests that people give far more influence to immoral items in mixed-valence moral judgments than they do when judging the actions individually. Similar fallacies occur in judgments about non-moral matters, and so it is argued that these results imply that cognition about moral and non-moral matters involve the same processes.

In the second series of experiments (Experiments 4 and 5), the role of emotions as information in moral judgments was investigated. The results demonstrated that, when prompted to do so, participants could produce emotional and reasoned justifications equally for dilemmas that tend to produce strong emotional reactions, such as those involving direct physical contact, and those that do not, such as those involving only indirect contact. They also demonstrated that participants were primed to read quickly about protagonists in moral dilemmas making decisions that were incongruent with the emotions they had read that the protagonist had experienced. The experiments provide some support for theories that emphasise dual-processes of emotion and controlled cognition in moral judgment, but also suggest a more complex picture in which emotions can act as information that is integrated into people's moral reasoning.

The third series of experiments (Experiments 6-8) investigated the emotion of moral elevation, and how people think and reason about the kind of memories that lead to this emotion. Elevation is the feeling of uplift that often accompanies witnessing, or hearing about, people doing morally good things for others. The experiments provided evidence suggesting that the kind of memories that prompt elevation are also the kind of memories that prompt the generation of counterfactual alternatives. The experiments also demonstrated that participants who witnessed someone doing something good for someone else tended to think about changing their behaviour, but that they did so in a general way, rather than a specific way (i.e. with no precise plan). It is speculated that the generation of counterfactual alternatives, and general intentions to change behaviour, may play a role in why people who have experienced elevation often have difficulty when they try to emulate good behaviour.

It is suggested that the results of the experiments reported in this thesis provide evidence that domain-general cognitive processes, such as heuristic mental shortcuts and the generation of counterfactual alternatives, are influential in thinking about moral issues, as they are in thinking about non-moral issues. Furthermore, the results indicate that information about emotions may be integrated into reasoning about moral matters, and may prompt people to form inferences about what alternatives they consider. These results have implications for how people make moral judgments and decisions, as well as for how people behave morally.

Chapter 1 Introduction

People regularly make moral judgments and decisions. They evaluate whether behaviour is moral or immoral, and to what degree. Subjects of such judgments include people in public life, such as politicians and celebrities; people in their lives, such as family and friends; and people from history, such as war criminals and spies. In 2013, Edward Snowden, a former National Security Agency and Central Intelligence Agency employee, leaked classified information about the online surveillance network the United States and other countries have built, because of his views about the importance of the privacy (Greenwald, MacAskill & Poitras, 2013). How did he balance the obligation he felt towards revealing behaviour that, in his view, was morally wrong, with the obligation he felt to protect his country? How do people make judgments of such actions? Does he seem honourable and praiseworthy for uncovering electronic surveillance by governments on a scale many may be uncomfortable with? Does he seem treasonous for betraying the trust placed in him by his government? Or are people's judgments of his behaviour complexly influenced by each of these actions?

Such judgments are important both because of the breadth of their occurrence, and the depth of their impact. People judge the morality of their family and friends, but also that of governments and corporations. People also often discuss and debate the morality and immorality of different behaviour. Such judgments may affect voting, friendships, and purchasing decisions, among other behaviours of societal import. That moral judgments can influence such behaviours suggests that understanding the cognitive psychological processes that influence and produce them is of great importance.

This thesis attempts to clarify the operation of some of these processes, and also to address some theoretical and normative debates related to them. In the first series (Chapter 2, Experiments 1-3), the issue of whether moral judgments are similar to, or different from, non-moral judgments was investigated. When people make judgments under uncertainty, their judgments often rely on irrelevant factors. For example, when judging probabilities, they are led by features of descriptions that make conjunctions more representative of a category than an individual event to commit a conjunction fallacy, judging the probability of two events to be no less probable than either individually (e.g. Tversky & Kahneman, 1983). Mental shortcuts such as this are often referred to as heuristics, and they sometimes lead people into error, as in the example above. Some researchers have suggested that people make similar errors in moral judgment (e.g. Sunstein, 2005; Baron, 2010). Others have suggested that moral judgments are importantly different from non-moral judgments, and resistant to the effects of context in a way that non-moral judgments are not (e.g. Dwyer, Huebner & Hauser, 2010). Experiments 1-3 in Chapter 2 addressed this debate by testing whether a similar pattern of errors occurred in moral as in non-moral judgments. It is argued that the results of Experiments 1-3 support the suggestion that intuitive, unconscious processes influence moral decision-making, and that it is therefore similar to decision-making in other domains.

Another debate relating to the role of intuitive judgment in moral reasoning is whether emotional reactions to moral scenarios prompt intuitive moral judgments, to which people do not have conscious access. It has been demonstrated empirically that emotions sometimes occur along with judgments that do not minimise harm for the greatest number of people (e.g. Greene, Sommerville, Nystrom, Darley & Cohen, 2001), that people sometimes are sometimes unable to rationally justify their moral

judgments (e.g. Cushman, Young & Hauser, 2006), and that emotional priming influences moral judgments differently depending on how moral dilemmas engage emotional processes (e.g. Valdesolo & DeSteno, 2006). These results have been taken to indicate that emotions sometimes lead people to make unconscious, intuitive moral judgments (e.g. Haidt, 2001; Greene et al., 2001).

However, cognitive control, working memory, and conscious reasoning also influence moral reasoning (e.g. Greene, Morelli, Lowenberg, Nystrom & Cohen, 2008; Moore, Clark & Kane, 2008; Paxton, Ungar & Greene, 2012). The second series of experiments, presented in Chapter 3 (Experiments 4 and 5), investigated the role of emotions as information in moral judgments, and how people reason about emotional information. This research adopted the widely tested paradigm of employing dilemmas that contrast direct and indirect harm (e.g. Greene et al. 2001; Greene, Nystrom, Engell, Darley & Cohen, 2004; Moore, Lee, Clark & Conway, 2011), and asked people to justify their own decisions in moral dilemmas, and to make judgments of other people who had made decisions in moral dilemmas. The experiments tested whether people differentially provided emotional and reasoned justifications depending on the instructions they were given, as well as whether they integrated information about another person's emotions into their reasoning when they were judging someone else.

Whereas the second series of experiments investigated the role of emotion in moral judgments in general, the third series tested predictions about a particular moral emotion. This emotion is called elevation, which is the feeling of moral uplift people often experience when seeing or hearing about someone doing something very good for someone else (e.g. Haidt, 2000). Elevation is thought to increase altruistic and pro-social behaviour (e.g. Algoe & Haidt, 2009; Cox, 2010). The experiments reported in

Chapter 4 tested hypotheses about the cognitive processes that may underlie the desire to change one's behaviour for the better, and particularly whether the generation of counterfactual alternatives to reality is likely to influence such desires.

This introduction reviews many of the phenomena that have shaped modern psychological theories of moral thought. It is structured in order to address key issues and debates relevant to the subsequent empirical chapters. First, in the section entitled 'Heuristics and Moral Consistency', the introduction focuses on the question of whether intuitive, unconscious heuristics influence moral judgments, and the implications of this question for theories of moral judgment. Second, in the section entitled 'Emotions and Reasoning in Moral Judgment', the subject of the introduction is how conscious, unconscious and emotional processing interact in producing moral judgments. Third, in the section entitled 'Elevation, Counterfactuals and Moral Action', the introduction focuses on literature related to the moral emotion of elevation, which arises in response to witnessing moral exemplars, and is thought to prompt people to behave well morally. Finally, the aims of the present research are discussed.

Heuristics and Moral Consistency

Are the psychological processes that influence moral judgment similar to, or different from, those that influence judgment in other domains? This question has implications for both psychological and normative theories of moral judgment. As noted above, if moral judgments are resistant to the tendency to rely on the kind of heuristics that arise in non-moral judgments, it would suggest moral judgments are domain-specific, rather than domain-general. In turn, this would imply that psychologists should focus their efforts on identifying the specifically moral features

of moral judgment. In contrast, if moral judgments are subject to the same kinds of contextual effects, psychologists should attempt to identify the domain-general processes that operate in moral reasoning.

Experiments 1-3 used a task similar to one used to demonstrate inconsistency in probability judgments, in which participants are asked to make judgments of two individual events, and their conjunction. In experiments testing judgments of the probability of conjoined events, participants often judged that the conjunction of two events was more probable than at least one of the events individually. For example, participants sometimes judged that the probability that President Obama would be re-elected (at the time of testing, prior to the November 2012 U.S. Presidential election) was 60%, that the probability that the economy would recover was 45%, and that the probability of both occurring was 55% (Khemlani, Lotstein & Johnson-Laird, 2012). Such judgments violate the principle that the probability of two events occurring together cannot be greater than the probability of either occurring individually.

In moral judgment, the analogous principle is that two moral events should be more moral than either individually, and vice versa for immoral events. Participants in Experiments 1 and 2 were asked to judge how moral or immoral it was, for example, for someone to take her sister's child into her home after her sister died; for someone to scrimp and save to pay for the child to go to University; and for someone to do both of these things. If their numeric rating of the conjunction was greater than their rating of each conjunct (i.e. indicating that the conjunction of actions was more moral than either action individually), their judgments were consistent. An example of a consistent judgment would be if they rated the woman's action of taking her sister's child into her home to be moral to the degree of +50, the action of saving for the child to go to University to be moral to the degree of +70, and the action of taking the child

in *and* saving for the child to go to University to be moral to the degree of +90. For other patterns of judgments, they were inconsistent. For example, if they judged the action of taking the child into her home to be moral to the degree of +50, the action of saving for the child to go to University to be moral to the degree of +70, and the action of taking the child in *and* saving for the child to go to University to be moral to the degree of +60, their judgment was inconsistent.

Participants in Experiment 3 were asked to make judgments of mixed-valence items. For example, they were asked to judge how moral or immoral it was for someone to give a teaching job to his nephew over better-qualified candidates; for someone to donate money to a charity that helped to fund cancer research; and for someone to do both of these things. Judgments were consistent if they followed another simple principle: numeric ratings of mixed-valence conjunctions should not give more influence to either the moral or immoral item than they were given individually, and therefore should be additive. An example of a consistent judgment in this task would be rating the man's action of giving a teaching job to his nephew over a better qualified candidate to be immoral to the degree of -30, the action of donating money to a charity to fund cancer research to be moral to the degree of +20, and the pair of actions to be immoral to the degree of -10 (i.e. $-30 + 20 = -10$). In many other domains, it has been found that negative stimuli outweigh positive stimuli (e.g. Baumeister, Bratslavsky, Finkenauer & Vohs, 2001). If moral judgments followed this pattern, it would provide further evidence that the same psychological processes influence moral and non-moral cognition.

Moral grammar and domain-specific moral judgments. The theory of a universal moral grammar has been a prominent domain-specific account of moral

judgment in recent years. This theory makes an analogy between the linguistic idea of generative grammar, and moral judgment (e.g. Mikhail, 2007, 2009, 2011; Hauser, 2006; Dwyer et al., 2010). This perspective suggests that people possess an innate, unconscious knowledge of a variety of legal rules, and a readiness to automatically represent human behaviour in legal terms (e.g. Mikhail, 2007; 2009).

Some initial evidence in support of a moral grammar has been taken to suggest the universality and innateness of certain basic features of moral judgment. For instance, at a young age, children display evidence of the ‘side-effect effect’; that is, like adults, their judgments of intention depend on the valence of the action (e.g. Leslie, Knobe & Cohen, 2006). Specifically, children, like adults, judge unintended but foreseen behaviour as intentional when an outcome is bad, but not when it is good (Leslie et al., 2006). Such results have been interpreted as indicating that children have a level of moral knowledge incommensurate with the level of explicit moral instruction they could have been given at such a young age, which in turn may suggest that certain features of moral judgment are innate (e.g. Mikhail, 2007; Dwyer et al., 2010).

Another principle of the universal moral grammar theory, namely that certain legal principles are unconsciously represented in the mind and automatically acted upon, is supported by different strands of evidence. Some moral judgments that closely resemble key legal principles (such as the fact that harm caused by direct contact is impermissible, whereas harm not caused by direct contact is permissible) appear to be made almost universally (e.g., Greene et al., 2001; Mikhail, 2007). Furthermore, people can only sometimes justify such judgments by appeal to moral principles (e.g., Cushman et al., 2006), suggesting that the judgments arise unconsciously (e.g., Hauser, 2006). Taken together, evidence such as this has been

interpreted as suggesting that moral judgments are innate and often unconscious intuitions, perhaps related to legal principles such as the prohibition against harm (e.g. Mikhail, 2009; Dwyer et al., 2010).

Adding to the idea that moral and non-moral cognition may be dissimilar, some evidence suggests that moral judgments are somewhat resistant to the kind of heuristic mental strategies that are often found in judgments about non-moral matters. For instance, when participants make judgments of clearly forbidden actions (such as killing five people with no benefit), and clearly obligatory actions (such as saving five people at no cost), the order in which the dilemmas are received rarely influences judgments (e.g. Dwyer et al., 2010). This has been interpreted as evidence that moral judgments are usually not influenced by domain-general heuristics. There is, however, some evidence that this may not be the case, and some studies that contradict this view are discussed in the next section.

Moral heuristics and domain-general moral judgment. In research on non-moral cognition, the heuristics and biases approach has been used to conceptualise how people seem to stray from optimal behaviour. Heuristics are intuitive mental short cuts that may be adaptive in certain contexts, but in others lead to mistakes (e.g. Tversky & Kahneman, 1983; Kahneman & Tversky, 1984; Kahneman, 2011). It has been argued that such heuristics may also lead to mistakes in moral judgments (e.g. Sunstein, 2005; Baron, 2010).

One important piece of evidence for this approach can be taken from a study in which people were asked to choose between different government policy responses to a spreading disease. The disease is expected to kill 600 people. When the options were described in terms of ‘saving’ lives, people chose an option that saved 200

people over an option that gave a one-third chance that everyone would be saved, and a two-thirds chance that nobody would be saved. When the same options were described as 'letting people die', their preferences were reversed (Kahneman & Tversky, 1984). There does not seem to be a rational basis for switching preferences because of a simple reframing of an issue such as this. The influence of framing on decision-making has been demonstrated in moral and non-moral domains (e.g. Kahneman & Tversky, 1984). Another study, using similar scenarios, demonstrated that participants faced with a choice between policy options that were either optimal on a utilitarian conception of morality (meaning they maximised how many people lived), or biased away from a utilitarian conception, varied their decisions depending on which of several different questions they were asked (Baron, 2010). For example, presented with a choice, to be made by a government agency, between (a):

An outbreak of viral meningitis will affect 100,000 people all starting on the same day. A vaccine will prevent all of these cases.

and (b):

Another epidemic of viral meningitis will affect 110,000 people, one at a time, over the course of year. Another vaccine will prevent all of these cases.

participants tended to choose (b), the optimal choice in that it involves the most people being saved, more often when they were asked which choice "would make you trust the government agency more, so that you would be more willing to see it have authority over decisions like this in the future?". However, they chose (a) more often

when they were asked which program “seems more ethical to you?” (Baron, 2010). This seems to be another case in which a seemingly morally irrelevant reframing of the task changes participants’ moral preferences.

The conceptualisation of moral judgments as being vulnerable to heuristics – a term that implies that they are sometimes in error in the same way that judgments of facts are sometimes in error – may be criticised from a philosophical point of view. Different people might defend different theories of what is right and wrong. While some might agree with a utilitarian conception of moral goodness (as adopted by e.g. Sunstein, 2005; Baron, 2010), according to which behaviour is moral when it attempts to maximise good outcomes for everyone, others might have different definitions of what constitutes moral behaviour (e.g. Waldmann, Nagel & Wiegmann, 2012).

However, the idea that mental shortcuts influence and sometimes bias people’s moral judgments does not necessarily depend on any person’s particular moral viewpoint. Instead, it suggests that, whatever one’s moral commitments, intuitive processes such as those that operate in non-moral judgments are likely to lead people into moral judgments that are not defensible even on their own moral viewpoint (e.g. Sunstein, 2005). It has also been suggested that moral reasoning is in fact no different from deontic reasoning (that is, reasoning about concepts related to obligation and permission; Bucciarelli, Khemlani & Johnson-Laird, 2008), and can be explained by psychological theories of deontic reasoning (e.g. Bucciarelli & Johnson-Laird, 2005). This perspective suggests that moral judgments are likely to be neither complete nor consistent. Therefore, as in the moral heuristics approach, it indicates that people will make errors in their moral judgments.

These viewpoints suggest an equivalence between moral cognition and cognition in the non-moral domain. If there is an innate moral grammar, it should be

free from the heuristic influences the domain general approaches discussed above imply (e.g. Dwyer et al., 2010). The experiments in Chapter 2 test this notion by investigating whether heuristics that occur in non-moral judgments also occur in the context of moral judgment. If similar effects do not occur in moral judgments, it will bolster the domain-specific perspective on moral cognition by suggesting that moral judgments are different from judgments about non-moral matters. On the other hand, if similar effects occur, it will strengthen the idea that the same processes operate in cognition about moral and non-moral issues.

Unconscious intuitions such as those involved in the heuristic judgments discussed above have been found to play a particular role in moral judgments when a scenario elicits an emotional reaction. Emotional reactions often seem to trigger intuitive processing, sometimes leading people to make moral judgments without a strong idea of the reasons behind them. The role of emotion in conscious and unconscious moral reasoning is discussed in detail in the next section, particularly as it relates to the empirical work reported in Chapter 3.

Emotions and Reasoning in Moral Judgment

The experiments reported in Chapter 3 (Experiments 4 and 5) examined how people make judgments of the acceptability of other people's decisions in moral dilemmas, and whether emotions may be used as information in such judgments. The dilemmas used in these experiments have been used to explore the distinction between the perceived morality of harms carried out directly and indirectly. This distinction is often illustrated with reference to a pair of similar dilemmas. The first of these, illustrating an option to harm using indirect action, is known as the trolley problem:

You are at the wheel of a runaway trolley quickly approaching a fork in the tracks. On the tracks extending to the left is a group of five railway workmen. On the tracks extending to the right is a single railway workman. If you do nothing the trolley will proceed to the left, causing the deaths of the five workmen. The only way to avoid the deaths of these workmen is to hit a switch on your dashboard that will cause the trolley to proceed to the right, causing the death of the single workman.

This is an example of an impersonal, indirect dilemma, and most people say that it is appropriate to push the switch in this problem in order to save the lives of the five men (e.g. Greene et al., 2001). A different result occurs when people respond to the trolley problem's companion dilemma, involving an option to harm directly, which is called the footbridge problem:

You are on a footbridge over the railway tracks towards which a runaway trolley is quickly approaching. On the tracks beyond the footbridge is a group of five railway workmen. If you do nothing the trolley will proceed on the tracks, causing the deaths of the five workmen. The only way to avoid the deaths of these workmen is to push a nearby stranger off the bridge so that his large body will stop the trolley, causing the death of the stranger.

This is an example of a personal, direct dilemma, and most people say it is inappropriate to push the man in this scenario (e.g. Greene et al., 2001). The question

for philosophers and psychologists has been why two dilemmas with identical consequences for action – that is, one man dead and five men saved – should prompt such different patterns of responses. One possible reason is that personal, direct dilemmas seem to prompt a greater degree of emotional activation than impersonal, indirect dilemmas (e.g. Greene et al., 2001; Valdesolo & DeSteno, 2006). This may occur because the thought of physically pushing a person to their deaths is more emotionally salient than the thought of pushing a switch (e.g. Greene et al., 2001)

People frequently make judgments about people who have experienced moral dilemmas. For instance, when the child sexual abuse scandal involving Pennsylvania State University former American football assistant coach Jerry Sandusky became public knowledge in 2011, much attention was focused on graduate assistant Mike McQueary, who knew about the abuse, and told his head coach, Joe Paterno, but did not go to the authorities (e.g. Peele, 2011). Although McQueary told his superior, many argued that he did not do the right thing when he tried to solve his dilemma between protecting children from future abuse, and displaying loyalty to his coach. McQueary said he regretted his failure to help more, claiming he was “Severely shocked, flustered, hastened, frantic” (Wetzel, 2012). Information like this may give people pause in judging behaviour to be immoral. Even if people think McQueary’s behaviour was morally wrong, this may be mitigated somewhat by the information he provided about his emotionally confused state.

The experiments reported in Chapter 3 tested this idea in two ways. Experiment 4 examined whether people considered emotions to be important elements of justifications for their behaviour in moral dilemmas, and differently for personal and impersonal dilemmas. Experiment 5 tested whether people took information about the emotions of actors in moral dilemmas into account when thinking about the

behaviour of these actors, differently for personal and impersonal dilemmas. The following sections discuss the progressing understanding of the psychological roles of reasoning and emotion in moral judgment.

Cognitive developmental theory. Early work on moral reasoning emphasised the role of conscious reasoning to an extent that is now considered somewhat unrealistic. For years, psychological research on moral thought was characterised by the assumption of cognitive-developmental theory that moral development progressed with age through an increasing understanding of concerns related to justice (e.g. Colby, Kohlberg, Gibbs, Lieberman, Fischer & Saltzstein, 1983). This theory emphasised controlled cognition, and suggested that people progress through a series of stages of moral reasoning, learning to deal with moral problems in increasingly sophisticated ways.

According to this theory, development was marked by a changing understanding of concepts related to justice. For example, at the earliest stage of reasoning according to cognitive developmental theory, ‘having a right’ was understood to mean that a person has power or authority over something or someone (e.g. Kohlberg, 1973). However, at the highest stage of moral reasoning, the same concept was understood to mean that all people have a right to be treated fairly by all other people (e.g. Kohlberg, 1973). At each stage, participants progressed in their understanding of concepts related to justice such as ‘having a right’ and ‘obligation’ in ways that they did not express at the previous stage.

One of the dilemmas used in interviews that helped to provide evidence of people’s progress through the stages was the Heinz dilemma:

In Europe, a woman was near death from a very bad disease, a special kind of cancer. There was one drug that the doctors thought might save her. It was a form of radium for which a druggist was charging ten times what the drug cost him to make. The sick woman's husband, Heinz, went to everyone he knew to borrow the money, but he could only get together about half of what it cost. He told the druggist that his wife was dying, and asked him to sell it cheaper or let him pay later. But the druggist said, "No, I discovered the drug and I'm going to make money from it." So Heinz got desperate and broke into the man's store to steal the drug for his wife.

According to the cognitive-developmental theory, progress through the stages and levels of moral reasoning involved taking the perspectives of others and combining these points of view (e.g. Kohlberg, 1973). Participants at different ages should provide judgments of Heinz's behaviour that demonstrate a progressively more sophisticated (at least according to the cognitive-developmental theory) understanding of Heinz's wife's right to life, and how it supersedes the druggist's property rights, as well as how her right to life may generate an obligation on Heinz's part to steal the drug (Kohlberg, 1973). These predictions were supported by longitudinal evidence. Older participants' justifications of their judgments in dilemmas like the one above were more sophisticated, and participants rarely reverted to justifications representative of an earlier stage of reasoning (Colby et al., 1983).

However, more recent research has undermined the idea that people consistently make moral judgments through conscious perspective taking and

reasoning about concepts relating to justice. This research (e.g. Haidt, Bjorklund, & Murphy, 2000; Wheatley & Haidt, 2005; Cushman et al., 2006) suggests that emotions and unconscious inferences can and do influence at least some moral judgments, and that therefore conscious moral reasoning, as tested by the Heinz dilemma, may not be the only way in which people come to moral decisions. Furthermore, if judgments take place outside of consciousness, the reasons people provide for their moral judgments may sometimes be post-hoc justifications, rather than genuine guides to how they made their decision. These ideas are discussed in the next section.

Emotions, intuitions, and moral reasoning. Evidence for the role of unconscious emotional processing in moral judgments can be seen when people make judgments of people who have committed moral violations. People often instantly say that such actions are wrong, and subsequently are unable to justify their judgments, indicating that, rather than being the cause of the judgment, their reasoning is post-hoc (e.g. Haidt et al., 2000). The following scenario helps to illustrate this phenomenon:

Julie and Mark are brother and sister. They are traveling together in France on summer vacation from college. One night they are staying alone in a cabin near the beach. They decide that it would be interesting and fun if they tried making love. At the very least it would be a new experience for each of them. Julie was already taking birth control pills, but Mark uses a condom too, just to be safe. They both enjoy making love, but they decide not to do it

again. They keep that night as a special secret, which makes them feel even closer to each other. What do you think about that? Was it OK for them to make love?

Most people judge that the siblings' behaviour was wrong (Haidt, 2001). When asked to justify this judgment, they often point out the dangers of inbreeding, until it is pointed out that the siblings used two forms of birth control (Haidt, 2001). They also often point out that the siblings might experience some emotional harm, until it is pointed out that the experience actually made them feel closer to one another (Haidt, 2001). Once their objections are exhausted, they tend not to retract their judgment, but rather continue to insist that the behaviour was wrong, even though they cannot justify this judgment (e.g. Haidt, 2001). Findings like this raise the question of whether some moral judgments may arise outside of consciousness.

The social intuitionist model (e.g. Haidt, 2001; Haidt & Bjorklund, 2008) suggests that judgments like these are made by the intuitive system, and influenced by immediate, emotional reactions to the behaviour involved. In the case of the scenario above, for example, people may experience a strong flash of negative emotion, perhaps revulsion, when they read about the incestuous siblings, which may lead to an instant reaction that their behaviour is morally wrong (Haidt, 2001). The social intuitionist model includes six processes, with authors emphasising that the first four, which focus on the intuitive system and on social reasoning, are the greatest influences on moral judgment, and that the latter two, which allow a limited role for private reasoning, are used relatively rarely. These processes are as follows:

1. *Intuitive judgment*: Moral judgments arrive in consciousness automatically and without effort, as the result of moral intuitions.
2. *Post hoc reasoning*: After making an intuitive moral judgment, people carry out an effortful search for reasons to support that judgment, in spite of the fact that the judgment may not have been made on the basis of those reasons.
3. *Reasoned persuasion*: People use verbal moral reasoning to persuade others of their beliefs. This type of persuasion is hypothesised to be at least partially affective in character; that is, it is suggested that people will appeal to emotions in attempting to convince others.
4. *Social persuasion*: People are often influenced by the moral judgments of others, whether their true beliefs, or merely their expressed beliefs are influenced.
5. *Reasoned judgment*: People are capable – albeit rarely – of using conscious reasoning to come to a conclusion about moral matters.
6. *Private reflection*: People are capable, in the course of thinking about a moral issue, of triggering a new moral intuition – perhaps through perspective taking.

Diverse evidence supports the idea that intuitions, and particularly emotional intuitions, play an important role in moral judgment. For instance, participants who had been hypnotically primed to experience a flash of disgust when they read a normatively irrelevant word (e.g. *take* or *often*) made greater judgments of disgust and moral condemnation for a person described in a morally neutral scenario that featured

their disgust trigger word, compared to scenarios that did not feature that word, despite being unable to justify their judgment (Wheatley & Haidt, 2005). This finding supports the idea that emotions arising outside of conscious awareness can be the cause of moral judgments.

Other emotions can also influence moral actions. For example, elevation, an emotion described as the sense of warmth and uplift people often experience after witnessing someone else do something good *for* someone else (e.g. Haidt, 2000), plays a role in inspiring moral action. This emotion has been found to encourage pro-social behaviour (e.g. Algoe & Haidt, 2009, Cox, 2010), and being primed with this emotion, or with general positive affect, influences judgments differently in certain moral dilemmas (e.g. Strohminger, Lewis & Meyer, 2011; Valdesolo & DeSteno, 2006). Furthermore, as discussed above, people are sometimes unable to justify their decisions in moral dilemmas (e.g. Cushman et al., 2006). Studies like these also support the idea that some moral decisions are not made through conscious reasoning, and are often influenced by emotions.

Some researchers have argued that the social intuitionist model over-emphasises intuitions over conscious reasoning. For instance, it has been suggested that although people may make most moral judgments intuitively, prior reasoning may inform moral intuitions (e.g. Pizarro & Bloom, 2003). The types of scenarios used experimentally may also play an important role in the kind of processes engaged in by participants. For example, evidence in favour of the cognitive-developmental theory was derived from experiments using balanced dilemmas (such as the Heinz dilemma discussed above), and questions that invited participants to tease apart relevant distinctions (Monin, Pizarro & Beer, 2007). In contrast, the scenarios that have provided evidence for intuitionism in moral judgment often involve emotionally

salient moral violations such as the incest example above, or eating a pet (Monin et al., 2007). In the next section, theories are discussed that take into account data on unconscious intuitions, and also conscious reasoning and controlled cognitive processes in moral judgment.

Dual-process theory and the personal-impersonal distinction. One of the experimental methods that has been most illustrative in drawing out the sometimes contrasting effects of emotions, intuitions and reasoning in moral decision making has been the use of moral dilemmas. These have been particularly useful in demonstrating what has come to be known as the personal-impersonal distinction. This distinction has been prominent in the public discourse in Ireland in 2013. As part of the public debate over the legalisation of abortion, pro-life campaigners have argued that abortion is wrong because it involves the *direct* killing of the innocent (e.g. Minihan & Cullen, 2013). Direct action often seems to be prohibited in a way that indirect action is not, an effect that psychologists have carried out a great deal of research on.

As discussed above, this idea has been supported experimentally, and the distinction may be based in differential emotional engagement for different scenarios. Moral dilemmas involving harm committed via direct personal contact appear to involve greater emotional processing, and to be judged less permissible, than dilemmas that do not (e.g. Greene et al., 2001). This research used fMRI to demonstrate that personal dilemmas such as the footbridge problem (i.e. those involving direct, personal contact, such as pushing the man) prompt a greater degree of emotional activation (e.g. in the medial frontal gyrus, posterior cingulate gyrus and angular gyrus) than do impersonal dilemmas such as the trolley problem (i.e. those involving only indirect, impersonal contact, such as pushing the switch). It has been

hypothesised that this greater degree of emotional engagement in personal dilemmas results in participants being less likely to act in personal than impersonal dilemmas (e.g. Greene et al., 2001). This finding of differential emotional involvement in these dilemma types has led to subsequent research examining this distinction, and exploring the relative influences of unconscious emotional and controlled cognitive processes in moral decision-making.

In the trolley and footbridge scenarios, when people judge saving the five men and sacrificing the one to be acceptable, they are making a utilitarian judgment (i.e. a choice with the aim of maximising benefits by saving five lives, and minimising bad consequences by only killing one person, e.g. Mill, 1863/2007). When they judge not sacrificing the one and allowing the five to die to be acceptable, they are making a deontological judgment (that is, a choice with the aim of not committing harm, e.g. Kant, 1788/2002). Some moral dual-process theorists (e.g. Greene et al., 2004) have suggested that controlled, cognitive processes make key contributions that influence people to make utilitarian choices, while emotional, intuitive processes contribute to deontological choices. As evidence for this, it has been demonstrated that when people made utilitarian choices in personal moral dilemmas (i.e. pushing the man off the bridge in order to save the five men), greater activity was observed in brain areas involved in cognitive control (e.g. the anterior dorsolateral prefrontal cortex) than for non-utilitarian choice (i.e. not pushing the man) (e.g. Greene et al., 2004).

The same study also found differences between difficult and easy personal moral dilemmas. Difficulty was defined by relatively long and short reaction times (difficult dilemmas requiring long reaction times, and indicating greater conflict). Difficult personal moral dilemmas contrast an extremely emotionally aversive action with a strong utilitarian reason to carry out such an action (e.g. a dilemma that

involved the need to suffocate your crying child to avoid detection by enemy soldiers, who would kill you and everyone in your village if they find you). Easy personal moral dilemmas, on the other hand, still involve an emotionally aversive, direct action, but contrast it with a less compelling reason to carry out that action, and therefore result in less conflict and quicker reaction times (e.g. a dilemma in which a teenage mother considers killing her unwanted baby). Making judgments of difficult personal dilemmas led to greater activation in the dorsolateral prefrontal cortex (which is associated with abstract reasoning and cognitive control), and in the anterior cingulate cortex (which is associated with processing conflict) compared to easy personal dilemmas (Greene et al., 2004).

Another study demonstrated that cognitive load selectively interferes with utilitarian moral judgment, again implicating cognitive control as an influential process in utilitarian moral judgment (e.g. Greene et al., 2008). In this study, participants responding to high conflict personal dilemmas under load (i.e. completing a separate, cognitively engaging but morally irrelevant task while making their judgment) took longer than participants not under load to make utilitarian judgments. Since cognitive control should be interfered with under load, the authors interpreted this result as implicating cognitive control in making such judgments. There was no such difference under load for non-utilitarian judgments.

Further evidence in favour of dual-processes of emotional and controlled cognition in moral judgment comes from studies of people with neurological deficits in brain areas associated with emotion (e.g. patients with ventromedial prefrontal damage). Such patients were more likely than controls to approve of direct, personal moral violations such as pushing the man in the footbridge problem (the utilitarian response) (e.g. Ciaramelli, Muccioli, Làdavas, & di Pellegrino, 2007; Koenigs, Young,

Adolphs, Tranel, Cushman, Hauser, & Damasio, 2007). This indicates that emotions may play a causal role in deontological judgments, such as refraining from pushing the man. In contrast, these patients made similar judgments of impersonal and non-moral dilemmas to control participants, indicating that they have the same knowledge of moral principles as such controls, but not the same prohibition on acting in personal dilemmas (e.g. Koenigs et al., 2007).

Domain-general dual-process theories help to explain how, although people are able to reason through difficult problems (in non-moral matters), they still fall victim to errors such as the heuristics discussed in the section on moral consistency (e.g. De Neys, 2006; Kahneman, 2011; see Evans, 2008, for a review). These theories often conceptualise cognition as occurring through an intuitive System 1, which carries out the majority of processing quickly and unconsciously, and a slower and more deliberative System 2, which can carry out more demanding computations. In the domain of moral judgment and decision-making, dual-process theories have focused on the contrast between conscious reasoning and emotional intuitions. Until recently, positive evidence for conscious reasoning in moral judgment was relatively rare. The following section discusses some evidence for this process.

Moral reasoning. Much of the evidence discussed so far has focused on how emotions and unconscious inferences influence moral judgment. However, other studies have provided evidence that conscious reasoning about moral problems does take place. One such study investigated the role of individual differences in working memory capacity on moral judgment. This variable reflects variation in cognitive control. Individual differences in working memory capacity predicted differences in judgments of appropriateness of personal moral violations, similar to pushing the man

in the footbridge problem, when the harm could not be avoided (Moore et al., 2008). This result suggests that deliberative reasoning about whether or not the harm was inevitable was involved in higher working memory capacity participants' increased judgments of appropriateness. Higher working memory capacity was also associated with longer reaction times in judging personal killing to be appropriate (Moore et al., 2008).

Utilitarian responses (i.e. those that maximise good for the greatest number of people, and which are associated with controlled cognitive processes) to moral dilemmas have often been taken to be reasoned, rather than intuitive (e.g. Greene et al., 2001, 2004). These responses require people to overcome any aversion to harming someone that they might experience, and decide that the balance of lives saved is worth the harm caused. Participants induced to be more reflective by completing the Cognitive Reflection Test (developed by Frederick, 2005) were more likely to give utilitarian responses to moral dilemmas than those who had not (Paxton et al., 2012). In the same experiment, high trait reflectiveness (that is, the natural tendency people have to be more or less reflective), as measured by the Cognitive Reflection Test, was related to utilitarian responding. This suggests that participants who produce more reflective responses on the Cognitive Reflection Test also tend to consider the consequences in moral dilemmas and override an intuitive, emotional tendency to respond in a non-utilitarian fashion. In another experiment, the authors found that by varying both argument strength and deliberation time, they were able to influence acceptance of incestuous behaviour (Paxton et al., 2012). That is, participants were more accepting of incestuous behaviour when they read a strong argument justifying it than a weak argument, but only when participants were given extra time, encouraging them to deliberate. Without the extra time, there was no difference in

how effective strong and weak arguments were, suggesting that conscious deliberation was involved in the decision to judge the incestuous behaviour as acceptable. The study therefore demonstrated that conscious moral reasoning occurs, but that it depends on various contextual factors. These results suggest that information about different features of moral scenarios is integrated into people's judgments of the scenarios.

Emotional information in moral judgment. Different dilemmas and scenarios appear to engage emotional processes to different extents. Many experiments ask participants whether someone's moral behaviour was right or wrong. The experiments discussed in Chapter 3 instead focused on the dilemma that person was facing, what thoughts and feelings they experienced, and how information about those thoughts and feelings may have influenced participants' judgments of people's actions. The experiments tested whether information about emotion was included in people's justifications of their own behaviour, and also whether people integrated information about the emotions *other* people experienced into their judgments of those people's actions. These questions were explored using personal and impersonal dilemmas. Since these dilemma types differentially prompt emotional processing, it is possible to compare how people differentially use, and are influenced by, emotional information.

The evidence that emotions play a role in moral judgments is powerful; another role they may have is in motivating moral action (e.g. Dwyer et al., 2010). The next section discusses a particular emotion that appears to play a role in motivating good behaviour: moral elevation. The experiments in Chapter 4

(Experiments 6-8) investigated this emotion, and also how people think and reason about morally good, charitable or honourable behaviour they have seen or heard about.

Elevation, Counterfactuals and Moral Action

While a great deal of attention in the moral psychological literature has focused on the role of emotions generally, some emotions are likely to have specific moral functions. For instance, participants induced to feel disgust through hypnosis, or with unpleasant odours or film clips, appear to produce more severe moral judgments than controls (e.g. Wheatley & Haidt, 2005; Schnall, Haidt, Clore & Jordan, 2008).

Positive emotions may also have different effects from negative ones such as disgust. Most of the research discussed in this chapter so far has focused on moral violations, and on morally aversive dilemmas in which actions have both moral and immoral outcomes. Moral elevation, on the other hand, is an emotion people often feel when they witness very morally good behaviour – that is, someone doing something very good, honourable or charitable for someone else (e.g. Haidt, 2000; Algoe & Haidt, 2009).

An important effect of this emotion is that people who have experienced it appear to be motivated to carry out morally good actions themselves. For instance, nursing mothers who had been primed by an elevating video produced more nurturing behaviour (such as hugging) than mothers primed with an amusing video (Silvers & Haidt, 2008). This indicates that seeing someone do something good for someone else may increase feelings of closeness. Elevated mothers in the same study also produced more milk, as measured by nursing pads, suggesting that elevation may be linked to

oxytocin release (Silvers & Haidt, 2008), which is itself associated with empathy for strangers and subsequent generosity (e.g. Barraza & Zak, 2009).

Experimental results vary with regard to how broadly the pro-social effects of elevation are applied. One study found that elevated participants helped with different kinds of tasks from the ones they had been elevated in (Schnall, Roper & Fessler, 2010), indicating that elevation can inspire helping that is different from simple emulation. However, another study demonstrated increases in volunteerism in participants elevated by participating in a trip designed to allow them to participate in helping others for a prolonged period. In this study, therefore, only helping of a similar kind increased (Cox, 2010).

Elevation also influences moral judgments, and differently from other forms of positive affect. While mirth (induced by listening to recordings of comedians) increases tolerance for deontological violations like pushing the man off the bridge in the footbridge scenario, elevation (induced by listening to recordings of inspiring stories) decreases such tolerance (e.g. Strohminger et al., 2011). As discussed earlier, experiments such as these appear to support theories of moral judgment that incorporate a causal role for emotional responses in motivating moral judgments (e.g. Haidt, 2001; Greene et al., 2008; Bucciarelli et al., 2008).

The experiments reported in Chapter 4 tested how people think about the types of behaviour that lead to elevation, and how thinking about that behaviour may lead one to behave well morally oneself. Although the emotional experience of elevation may be a key component, it is hypothesised that a set of cognitive processes may be vital in transitioning from hearing about or seeing someone do something exceptionally good, to deciding to do something similar oneself. Important features of behaviour that leads to elevation seem to be that they often involve unusually good

actions, controllable by the person doing them. For example, the participants in the experiments reported in Chapter 4 produced their own memories of someone doing something good for someone else, including examples of people saving other peoples' lives, caring for dying people with empathy and kindness, and donating organs to help unknown others.

Thinking about unusual, controllable actions such as these may drive people to think about how things might have been different (e.g. Kahneman & Tversky, 1982; McCloy & Byrne, 2000; Roese, 1997). Specifically, it is hypothesised that counterfactual thoughts, or thinking about what might have been, may be triggered when one sees someone else deliberately behaving in an extremely moral manner. In turn, it is suggested that thinking about such how things might have been different may be a component of why people do not always follow through on their intentions to emulate morally good behaviour. This hypothesis was tested by investigating the type of counterfactuals people produce in response to remembering extremely good actions (i.e. imagining a better or worse world, adding a new feature to the world or removing a feature from the world), how people think about changing their behaviour after witnessing moral exemplars, and also what people think about different features of the behaviour in the memory itself (e.g. whether the behaviour was expected or unexpected, intentional or unintentional, an action or an inaction). The experiments were designed to test whether counterfactual alternatives are involved in the formation of intentions to change their own behaviour after witnessing morally good behaviour.

Aims of the Present Research

This chapter has discussed modern psychological theories, and many of the empirical findings that influenced their development. The research reported in this thesis aims to test and extend different features of these theories. Chapter 2 presents three experiments (Experiments 1, 2, and 3) investigating consistency in numeric moral judgment. These experiments tested whether similar inconsistencies emerge in judgments of morality as emerge in judgments of probability, such as the failure to judge two events to be less probable than the events individually. The results of these experiments relate to the debate over whether moral judgments are different from non-moral judgments.

Chapter 3 reports two experiments (Experiments 4 and 5) that examined the role of emotion in moral judgment, and specifically how people reasoned about other peoples' emotions as they face moral dilemmas. Since emotions influence moral judgments (e.g. Haidt et al., 2000; Greene et al., 2001; Valdesolo & DeSteno, 2006), it was hypothesised that knowledge of other peoples' emotions during moral dilemmas would influence how think and reason about their decisions. These experiments explored the relation of knowledge about other peoples' emotions to the personal-impersonal distinction, and whether people reasoned about emotional information differently for different dilemma types. As personal dilemmas are thought to engage emotional areas of the brain to a greater extent than impersonal dilemmas (e.g. Greene et al., 2001), these dilemma types may also result in differential engagement with other peoples' emotions. These experiments are important with regard to social moral reasoning, and judgments such as these could be especially relevant to legal judgments, given that people on trial are often given a chance to

speak in their own defence. They also provide insight into the expectations people have of how emotions can lead to moral action.

The process of how emotions may lead to moral action is the primary focus of Chapter 4, which discusses three experiments (Experiments 6, 7 and 8) related to the moral emotion of elevation. This emotion is elicited when people see or hear about someone else doing something good for someone else, and people often express a desire to emulate such good behaviour after feeling elevated. The experiments reported in Chapter 4 test the cognitive processes of how people move from feeling elevated, to wishing to emulate good behaviour, and particularly focus on how the generation of counterfactuals may play a role in this transition.

This thesis aims to investigate psychological processes with real-world import for moral reasoning. The experiments are also discussed with regard to their implications for everyday moral reasoning, and for prescriptive theories about what people *should* do. As discussed at the beginning of this introduction, the pervasiveness of moral judgments, and their potential influence on judgments that determine governments, friendships and consumer choices, among many other decisions, makes understanding how they are reached vital. People sometimes consider their moral judgments to be almost as objectively true as scientific facts (Goodwin & Darley, 2008), in spite of the broad diversity in such judgments. The experiments reported here investigate how these judgments, which are given such importance by those making them, are reached.

In Chapter 2, the question of whether people can make consistent moral judgments is investigated. As mentioned above, this question is discussed in terms of empirical research on the topic, and also in terms of its relevance to the debate over whether the processes involved in moral reasoning are the same as, or different from

those involved in reasoning about other matters. The implications of moral consistency for everyday reasoning, and for the application of normative theories of moral behaviour, are also discussed.

Chapter 2 Consistency in Degrees of Morality

In 2012, there was a public outcry in Ireland when a man convicted of sexual assault walked free from court, while a man who had committed tax fraud related to garlic importing was sentenced to a six year custodial sentence by the same judge (e.g. Shanahan & Gallagher, 2012). Similarly, in sports, fans often rage at perceived inconsistencies in the treatment of players from different teams by referees and citing bodies. For instance, in 2013 French rugby fans were outraged when Irish player Paul O'Connell was not banned for kicking an opponent in the head, while French player Jerome Fillol was banned for 14 weeks for spitting in an opponent's face (e.g. Kinsella, 2013). People seem to notice clear distinctions between the moral wrongness of different actions, and such asymmetries in punishment can cause a strong negative reaction.

These examples illustrate something about morality that most people accept implicitly: some morally good actions are better than other morally good actions, and some morally bad actions are worse than other morally bad ones. It is better to save a person from a fire than to donate five euro to charity, and it is worse to sexually assault someone than to commit tax fraud related to garlic importing. While these statements might seem obvious, they demonstrate something important about morality: that is, there are degrees of morality, and people are able to order events according to how moral they are.

In turn, it can be inferred that to carry out two good acts is better than to carry out just one of those acts. For example, it would be better to save a person from a fire *and* donate five euro to charity, than to carry out only one of these actions. These judgments of conjoined moral actions, and how they relate to judgments of individual moral actions, are the subjects of this chapter. Are people consistent when they make

these kinds of judgments? Commonsense suggests that two bad actions are worse than one, but do people make judgments in this way?

The experiments reported in this chapter test how people make judgments of paired and individual moral and immoral actions, and particularly, whether they can do so consistently. Past research on judgment of groups of moral and immoral actions has tested how judgments of pairs (and larger groups) of moral and immoral actions compare to pre-tested averages (e.g. Birnbaum, 1972, 1973; Risky & Birnbaum, 1974). This work was undertaken to identify a model that would best describe the distribution of responses compared to pre-tested averages of the individual items. The first of these studies indicated that judgments of pairs of immoral actions were best described by both the average scale value of the items, and the range of the items, and not by an additive or averaging model (Birnbaum, 1972). This model suggests that the greater the range of the items in a set, the lower the judgment of morality will be. The second and third of these studies demonstrated that immoral actions had more influence in mixed-valence moral judgments than moral action (e.g. Birnbaum, 1973; Risky & Birnbaum, 1974). The goal of the studies in this chapter, in contrast, was to investigate consistency between judgments of pairs of moral actions, and judgments of the individual actions. Thus, participants' judgments of the conjunction of a pair of moral, immoral, or mixed-valence items were compared to their judgments of the individual items to assess whether their judgments were consistent.

The ways in which moral judgments are combined have important implications for what is unique about such judgments, and what parallels exist between moral decision-making and decision-making in other domains. The experiments reported in this chapter test whether moral judgments follow similar patterns to non-moral judgments. The next sections discuss the occurrence of a

particular type of error in judgments of probability, and how the experiments reported in this chapter examine whether these error occur in judgments of morality.

Conjunction Fallacies in Non-Moral Judgments

It has been observed that participants sometimes do not follow the standard principle of conjunctions of probabilities. This principle is that probability of a conjunction of events cannot be greater than the probability of either of the conjunction's constituent events, since the probability that both events occur together is included in the probability set of just one of them occurring. Certain factors can make the occurrence of two events seem more probable than the occurrence of just one (e.g. Tversky & Kahneman, 1983). For example, when presented with descriptions of a person who possesses the kind of traits one might expect of someone who is active in the feminist movement, participants often judged that it was more likely that the person was active in the feminist movement *and* a bank teller, than that they were just a bank teller (e.g. Tversky & Kahneman, 1983). This is in spite of the fact that the possibility of the person being a feminist *and* a bank teller is included in the possibility set of her being a bank teller, and so the pair of events cannot be more probable than her being a bank teller. The violation of consistency here has been referred to as the conjunction fallacy (e.g. Tversky & Kahneman, 1983).

Similar effects have been found when people are asked to make judgments of the probability of unique events (e.g. Khemlani et al., 2012). For example, when participants were asked to judge (before the 2012 US elections) what was the probability that President Obama would be re-elected, and also the probability that the economy would recover, they often rated the conjunction as being more probable than one or both conjuncts. These results provide further evidence that people sometimes

commit conjunction fallacies when estimating probabilities. The next section discusses whether similar results occur in judgments of morality.

Heuristics and Moral Judgment

The question of whether such fallacies also occur in moral judgment relates to the larger issue of whether reasoning about moral matters is unique and domain-specific, or similar to reasoning about non-moral matters. As discussed in Chapter 1, the universal moral grammar theory suggests that moral judgment is domain-specific and context-independent (e.g. Dwyer et al., 2010). This means that a specialised mechanism is involved in moral judgment, and moral judgment should not be influenced by heuristic effects such as those found in judgments of probability, or in financial judgments (e.g. Kahneman & Tversky, 1984; Thaler & Benartzi, 2004; Thaler & Sunstein, 2008), among other domains. It has been demonstrated that certain judgments, such as the prohibition against deliberate harm, are made almost universally (e.g. Greene et al., 2001; Mikhail, 2007), and that children make moral judgments that are similar to the judgments of adults from a very young age (e.g. Leslie et al., 2006), which, among other evidence, may indicate that our most basic moral thoughts are innate, universal, and based on a so-called ‘moral grammar’.

However, it has been demonstrated empirically that moral reasoning is subject to some of the same processes that influence reasoning in other domains. As discussed in Chapter 1, framing effects can change preferences for government policies even when the results of each policy remain the same (e.g. Kahneman & Tversky, 1984; Baron, 2010). Similar results occur in studies of risky financial gambles; for example, a simple reframing of options can reverse the choices people make (Kahneman & Tversky, 1984). Irrelevant affect also influences judgment in moral dilemmas (e.g.

Valdesolo & DeSteno, 2006; Strohminger et al., 2011), as it does decisions with nothing to do with morality, such as judgments of the likability of neutral stimuli (e.g. Winkielman, Zajonc and Schwarz, 1997). These studies, and others like them, suggest that the same processes are involved in moral reasoning as are in reasoning about other domains. The experiments reported here provide a further test of this question.

Experimental Task

The experiments discussed in this chapter test whether numeric judgments of conjunctions of moral actions are subject to similar heuristic influences as those present in numeric judgments of the probability of conjunctions of events. If they are, this report will add to the body of evidence demonstrating similarities between moral and non-moral judgment.

The task used in these experiments involved making judgments of moral and immoral actions, individually and in conjunction. As an example, consider how moral or immoral you would judge the following action to be, on the following scale, from most immoral (-100), through neither moral nor immoral (0), to most moral (+100):

A man hit his son

Most immoral	Neither moral nor immoral	Most moral
-100	0	+100

And now make a judgment of the following action:

A man slept with his wife's friend

Most immoral	Neither moral nor immoral	Most moral
-100	0	+100

And finally, consider the two actions together:

A man hit his son, and slept with his wife’s friend

Most immoral	Neither moral nor immoral	Most moral
-100	0	+100

Participants were asked to make judgments of triples like the one above: two individual actions, and the conjunction of those actions. This task allowed for the investigation of consistency in moral judgment in the same way that past studies (e.g. Khemlani et al., 2012) have investigated consistency in probability judgments. Participants’ judgments of the conjunction of two actions were compared to their judgments of individual actions in order to assess consistency.

The three experiments test two different kinds of consistency. Experiments 1 and 2 test consistency of same-valence judgments (i.e. two moral or two immoral actions), while Experiment 3 tests consistency for mixed-valence judgments (i.e. one moral and one immoral judgment). In Experiments 1 and 2, consistent responses were defined as follows: two moral actions should be judged to be more moral than either individually, and two immoral actions should be judged to be more immoral than either individually. In these experiments, a triple of judgments was categorised as consistent if the judgment of the conjunction was greater than the judgments of the individual actions. For example, consider a participant who rates sleeping with one’s wife’s friend to be immoral to the degree of -50, hitting one’s son to be immoral to the degree of -70, and the conjunction of the two to be immoral to the degree of -95. Their judgment would be categorised as consistent. Otherwise, the triple was categorised as inconsistent (see Appendix 2.3 for full criteria of consistency and inconsistency for same-valence judgments).

In Experiment 3, a triple of judgments was categorised as consistent if the judgment of the conjunction was equal to the sum of the judgments of the individual actions. For example, consider a participant who rates embezzling funds to be immoral to the degree of -50, and intervening to stop a fight to be moral to the degree of +40, and the conjunction of the two to be immoral to the degree of -10. Such a judgment would be rated as consistent, but any judgment that gave more influence to either the moral or immoral action than indicated by the sum of the two was categorised as inconsistent.

A theory of numeric moral judgment has been developed, according to which two mental processes are involved in judgments such as these (Gubbins, Byrne & Johnson-Laird, 2013). According to this theory, an intuitive process constructs a representation of whether the action is moral or immoral, and to roughly what degree. Subsequently a deliberative component maps this representation onto a numeric scale. The task used in the experiments reported here required participants not only to make the fast intuitive judgment of whether and to what degree the action was moral or immoral, but also the deliberative judgment of how moral or immoral the action was on a numeric scale. According to the theory, these processes (i.e. the mapping of intuitive representations of morality onto a numeric scale, and the process of combining judgments of actions) are likely to be quite inexact, and may introduce error in judgments of conjunctions (Gubbins et al., 2013). The error resulting from both of these processes may manifest itself in inconsistencies.

Experiments 1 and 2 tested whether people combine judgments of same-valence (i.e. two good or two bad) actions consistently – that is, whether they judge the conjunction of two moral actions to be more moral than either individually, and the conjunction of two immoral actions to be more immoral than either. These

experiments also tested whether the placement of the conjunction in the triple of judgments influences consistency. It was hypothesised that if participants judged two individual actions, and then their conjunction, it would be easier to make consistent judgments compared to if they judged the conjunction followed by the individual components of the conjunction.

Experiment 2 also tested whether participants would make more consistent judgments on a coarser rather than a fine scale. It was hypothesised that a coarse scale might be more similar than a fine scale to the kind of scale people make moral judgments on in real life. It was also hypothesised that such a scale would result in fewer inconsistencies because a small change in participants' ratings due to error introduced by the processes discussed above would be less likely to make enough of a difference to render the triple inconsistent.

Experiment 3 tested whether moral or immoral items have a greater influence in conjunctions of mixed-valence (i.e. good *and* bad) actions, or if people can make consistent judgments of these actions. Since previous research indicated that negative stimuli are stronger than positive stimuli in many domains (e.g. Baumeister et al., 2001), and also in morality (e.g. Risky & Birnbaum, 1974), it was hypothesised that the immoral item would have greater influence than the moral item in mixed-valence conjunctions.

Experiment 1

The aim of this experiment was to investigate consistency in same-valence numeric moral judgments, as well as the question of whether the position of the conjunction in a triple influences consistency. Due to the error introduced by the processes of mapping moral judgments onto a numeric scale, and by combining moral

judgments, it was hypothesised that a reliable proportion of triples would be inconsistent. It was also hypothesised that, if the conjunction is judged at the end of the triple, rather than at the beginning, inconsistencies should be reduced. Reading the conjunction last should allow participants to use more sophisticated mathematical strategies, such as a rough form of addition, when making their judgments of the conjunction. This result also arose in a similar experiment on judgments of conjunctions of probabilities, and so if it emerged in moral judgments, it would provide another example of similarity between moral and non-moral judgment.

A typical trial in the experiment appeared as follows on the participant's computer screen, one judgment after another:

A man gave blood so that more would be available to help people who had been in accidents. Please rate the degree to which the action was moral (from 1 up to +100) or immoral (from 1 down to -100) or neither (0)

Most immoral	Neither moral nor immoral	Most moral
-100	0	+100

A man donated money to a charity that helped to fund cancer research. Please rate the degree to which the action was moral (from 1 up to +100) or immoral (from 1 down to -100) or neither (0)

Most immoral	Neither moral nor immoral	Most moral
-100	0	+100

A man gave blood so that more would be available to help people who had

been in accidents, and donated money to a charity that helped to fund cancer research. Please rate the degree to which the action was moral (from 1 up to +100) or immoral (from 1 down to -100) or neither (0)

Most immoral	Neither moral nor immoral	Most moral
-100	0	+100

Participants rated the morality, or immorality, of the events in each sentence on a numerical slider scale from -100 to 0 to +100.

Method

Participants

Participants were recruited using the Amazon Mechanical Turk website, and took part in the experiment online. They were compensated with \$0.40 per participant. Thirty-six participants took part in the experiment. As well as moral and immoral items, participants also judged neutral fillers. These were designed to be morally neutral, and so participants should have judged them to be 0 (neither moral nor immoral). Ratings of these items were therefore used as an *a priori* criterion for participation: participants were eliminated if their average rating of the neutral items was greater than 20 out of 100. Seven participants were eliminated because of this criterion, and one participant was eliminated for reporting errors on the website. The 28 remaining participants were 16 women, and 12 men, ranging in age from 19 to 64 years. The principal results remained statistically significant when the seven rejected participants are included in the analyses.

Design and Materials

Participants acted as their own controls and evaluated the same 20 sets of three sentences (see Appendix 2.1 for complete materials used in Experiments 1 and 2). They carried out one block of ten triples with the conjunction at the beginning of the triples ($A\&B, A, B$), and one block of ten triples with the conjunction at the end of the triples ($A, B, A\&B$). One group of participants carried out the conjunction-first block first, and another group of participants carried them out in the opposite order, and the participants were assigned alternately to these two groups. Each block consisted of four triples of moral actions, four triples of immoral actions, and two triples of morally neutral actions. The order of the conjuncts, A , then B , remained constant – only the position of the conjunction, first or last, varied from one block to the other.

Procedure

Participants read a paragraph of instructions (see Appendix 2.2 for consent and debriefing for Experiment 1), and then carried out two practice trials of one morally good triple and one morally bad triple. They then carried out the experiment proper. The key instruction was: “Please rate the degree to which the action was moral (from 1 up to +100) or immoral (from 1 down to -100) or neither (0)”. Finally, they typed a comment on how they made their judgments, and read a debriefing page.

Results

Table 2.1 presents the conjunctions from the set of 20 triples, and the mean numerical judgments for each conjunct and the conjunction.

Table 2.1: The morally good, morally bad, and neutral conjunctions for Experiment 1, and the participants' mean judgments of the first conjunct, *A*, the second conjunct, *B*, and their conjunction, *A&B*, in Experiment 1. The judgments were made on a scale from -100 (most immoral), through 0 (neither moral nor immoral).

<i>Contents</i>	<i>A</i>	<i>B</i>	<i>A&B</i>
<i>Moral Contents</i>			
A woman worked for many years to raise money for research into kidney disease, and donated one of her kidneys to a friend of hers, saving her friend's life.	72	88	90
The staff of a hospital banded together to subsidize the construction of a badly needed burns unit in their hospital, and worked together to save the victims of a terrorist bomb.	69	70	81
A woman took her sister's child into her home after her sister died, and scrimped and saved to pay for the child to go to University.	69	73	80
A man intervened to help stop a fight, and tended to the wounds of a stranger injured in the fight.	55	65	74
A man gave blood so that more would be available to help people who had been in accidents, and donated money to a charity that helped to fund cancer research.	61	57	73
A young man helped an old woman to cross the street, and he volunteered at a home for the elderly.	55	62	63
A husband rejected an opportunity to cheat on his wife, and took his wife for a romantic walk.	62	32	61
The director of a factory hired an equal number of able-bodied and disabled people, and paid them equally.	52	54	61
<i>Immoral Contents</i>			
A babysitter beat the infant in her care, and starved her so that the child became malnourished.	-91	-90	-95
The headmaster of a school physically and sexually abused one of his pupils.	-74	-84	-92
A violent bully terrorized the playground and beat up a younger girl with a hammer for no apparent reason.	-66	-91	-92
The owner of a bar forced the foreign girl who was working for him to accept a salary below minimum wage, and to have sex with him.	-53	-85	-88
The financial director of a company embezzled funds for many years, and covered his embezzlements by framing his colleague.	-72	-81	-84
A husband slept with his wife's friend, and left his wife, stealing her car.	-66	-68	-83
A man insulted a colleague for no reason, and made it look as though she had committed a crime so that she lost her job.	-36	-78	-79
A man attacked a security guard and stole money from him.	-65	-66	-68
<i>Neutral Contents</i>			
A woman finished her dinner at eight o'clock on a Tuesday evening, and, having watched a television program, prepared to go to bed at eleven in the evening.	1	1	1
A woman wrote an email to a colleague to arrange to meet for lunch in order to discuss a new sales strategy, and made a phone call to a hotel in a city she would be visiting to enquire about the availability of their rooms.	2	1	2
A man bought a newspaper in a corner shop beside the train station, and posted an important document to a business associate in another city.	-2	4	4
A man sat on the couch in his apartment on a wet, rainy evening in the city, and read a collection of detective stories quickly because he was very interested in what would happen next.	1	1	1

There was no order effect of which block participants received first in the ratings between conjunction-first and conjunction-last, $M = 63.65$, $SD = 19.49$, and $M = 68.78$, $SD = 22.89$ for the moral actions, Mann-Whitney, $U = 78$, $z = .729$, $p = .466$, and $M = -77.97$, $SD = 17.9$, and $M = -75.26$, $SD = 15.85$, for the immoral actions, Mann-Whitney, $U = 78.5$, $z = .706$, $p = .48$, and so their data have been collapsed for the remaining analyses.

Consistency and Inconsistency

Trials in which the conjunction was rated as greater than its conjuncts in morality or immorality were classified as consistent; trials in which the conjunction was not rated higher than one of the conjuncts were classified as one violation of consistency; and trials in which the conjunction was not rated as higher than either conjunct were classified as two violations of consistency (unless all three ratings were equal and less than 100, in which case they were classified as a single violation¹). A detailed explanation of how triples were categorised as consistent or inconsistent can be found in Appendix 2.3. The results for five individual triples were removed from the analysis because their ratings were of different valences than they should have been – for example, a moral item was given an immoral rating, such as -30, or vice versa. All results are two-tailed and use an alpha level of .05.

Responses were consistent rather than inconsistent at a greater than chance level, binomial test, $z = -9.97$, $p < .001$, $r = .47$. Furthermore, 24 of the 28 participants made more consistent than inconsistent judgments, with 1 tie, Binomial test, $z = 3.85$, $p < .001$, $r = .73$. However, compared to if there were no tendency towards

¹ Only 2% of trials, other than those with neutral contents, elicited at least one judgment of 0, which signified a morally neutral event. If both conjuncts were judged as 0 but the conjunction had a rating greater than 0, the trial also counted as two violations. If conjuncts were rated as 100 and the conjunction was not, the trial also counted as two violations.

inconsistency, violations of consistency occurred reliably in judgments of moral actions (on 30% of trials), Wilcoxon, $z = 4.136$, $p < .001$, $r = .55$, and in judgments of immoral actions (22% of trials), Wilcoxon, $z = 4.05$, $p < .001$, $r = .54$. As noted earlier, the patterns were also reliable when the seven participants who were excluded from the analysis were included.

Individuals varied significantly in their ability to make consistent judgments, Friedman test, $\chi^2 (27, n = 11) = 54.263$, $p < .005$. There was no difference between any of the moral contents in terms of consistency, Friedman test, $\chi^2 (25, N = 7) = 4.529$, $p = .717$. There was a non-significant difference between some of the immoral contents in terms of consistency, Friedman test, $\chi^2 (26, n = 7) = 13.87$, $p = .054$. Four participants made only consistent judgments, often by relying on the extreme ends of the scales, whereas the least able participant was consistent for only five out of the 16 triples with a moral or immoral content. This result may have arisen due to individual differences in relying on intuition (e.g. Kahneman, 2011).

Participants rated conjunctions as most moral (+100) or most immoral (-100) on 37% of trials. These extremes contributed to consistency, because triples in which the conjunction was judged as 100 counted as consistent. If these triples are excluded, inconsistency rises from 26% overall, to 42% overall.

Position of Conjunction and Morality

Figure 2.1 presents the mean number of triples that were consistent, that committed a single violation of consistency, and that committed a double violation of consistency, depending on whether the conjunction was first or last in a triple of judgments. Judgments were consistent more often when the conjunction came last (M

= 6.21, *SD* = 1.6) than when it came first (*M* = 5.39, *SD* = 1.85), Wilcoxon, *z* = 2.465, *p* < .01, *r* = 0.33.

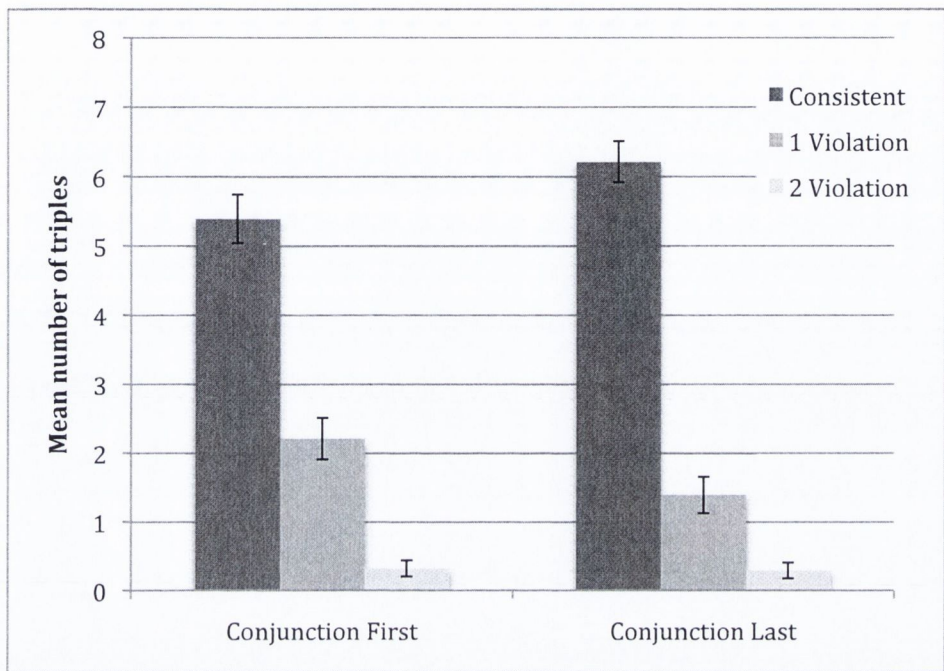


Figure 2.1: The mean number of triples in the three categories of consistency and violation depending on whether the conjunction was first or last in a triple of judgments.. Error bars are standard error of the mean.

An unexpected result was that the participants made more consistent responses for immoral actions (*M* = 6.14, *SD* = 1.67) than for moral actions overall (*M* = 5.46, *SD* = 1.73), Wilcoxon, *z* = 2.068, *p* < .05, *r* = .28.

Discussion

Experiment 1 established that participants were generally consistent for numeric moral judgments, but also that they generated a significant amount of inconsistencies, relative to what one might expect if there had been no tendency

towards inconsistency. While participants violated consistency on 26% of trials for morality, they did so for 56% of trials for probability (Khemlani et al., 2012). This suggests that judgments of pairs of moral actions are easier for participants to make consistently than judgments of the probability of pairs of events. It may be that the normative rule for probability (i.e. that two events are *not* more probable than one) is counter-intuitive, whereas the rule for morality (i.e. that two good actions *are* better than one) is more straightforward. An alternative interpretation of this result is that moral judgments are, in some respect, less susceptible than non-moral judgments to contextual effects such as heuristics. However, there remained a significant tendency to be inconsistent compared to what would occur if people always tended to be consistent, and, as the next section discusses, the contextual effect of the placement of the conjunction was also significant. These results emphasise the similarities between judgments of probability and morality.

The experiment demonstrated that, as in judgments of probability, consistency improves when the conjunction is read at the end of a triple, rather than at the beginning. This result showed that violations of consistency were reduced if individuals had already made numerical judgments of the conjuncts, perhaps because, as suggested earlier, having made these judgments allowed them to use more sophisticated numerical strategies such as addition for the conjunction. This result of improved consistency when the conjunction is read last also suggests a similarity between moral and non-moral judgments, and demonstrates the sort of order effect that, according to the moral grammar theory of moral judgment, should not occur (e.g. Dwyer et al., 2010).

What do these results mean for everyday moral reasoning? Consistency in this experiment meant that a pair of moral actions was judged to be more moral than either

of its constituent actions, and a pair of immoral actions was judged to be more immoral than either of its constituent actions. The presence of inconsistency in these judgments suggests that sometimes, people may not judge the actions of people who do multiple moral things to be as good as they should. Similarly, people may not judge the actions of people who do multiple immoral things to be as bad as they should.

Furthermore, as discussed above, people were less consistent when the conjunction was the first item judged, rather than the last. This suggests that people were more likely to raise their judgments of pairs of moral or immoral actions after having judged the constituent actions already, than they were to lower their judgments of individual moral or immoral actions after having judged the conjunction. In other words, a person who judged two moral items individually was subsequently more likely to provide a higher rating of morality for their conjunction. On the other hand, a person who had already judged their conjunction was less likely to judge the constituent items to be less moral than the conjunction.

In practical terms, this result suggests that when people make judgments about multiple good or bad actions before judging actions individually (i.e., akin to the conjunction-first condition), they may be more likely to make inconsistent judgments. For example, they may not always adjust downward their judgments of new, individual good or bad things they hear about. This implies that hearing about one person doing two good things, or two bad things, may influence judgments of future times people hear about one of the same actions again – they may sometimes judge an individual action to be as bad as the pair.

Possible explanations for the result that people are more consistent for immoral than moral items are that people may concentrate more when judging

immoral actions, or that they may simply care more about immoral actions. These suggestions are in keeping with the finding that negative-valence items have greater influence than positive-valence items across many domains (e.g. Baumeister et al., 2001). It may be that people perceive the outcomes from immoral actions to be so bad that they are worth paying special attention to.

Experiment 1 demonstrated that people are generally consistent in numeric moral judgment, but sometimes inconsistent, and that consistency varies depending on the position of the conjunction. Experiment 2 tested whether these results would replicate, and also whether consistency would be improved when participants used a coarser scale, which was hypothesised to be more similar to the kinds of scales people likely use in everyday life.

Experiment 2

People seem to have clear ideas about what is right and wrong, and what is more right or wrong than other things, but as Experiment 1 shows, questions like *how* right or wrong something is may be more difficult to answer. It seems likely that the fine-grained scale used in Experiment 1 allows for far more precision than people actually use in everyday moral judgments. The differences in morality between different intuitively constructed representations of judgments such as those discussed above are likely to be much more coarse. In Experiment 2, participants made judgments not only on the finely grained scale from -100 to +100 used in Experiment 1, but also a scale from -7 (most immoral) through 0 (neither moral nor immoral) to +7 (most moral).

Coarse and fine scales have different fundamental features, which have implications for moral consistency. Specifically, if the process of mapping an intuitive

moral judgment onto a numeric scale introduces error, then the task used in Experiment 1 should be easier using a coarse scale, because there are fewer inconsistent options for participants to choose from, and therefore less room for the error introduced to affect consistency. A slight jitter in mappings to a coarse scale should not tend to change categories from a consistent assignment to an inconsistent assignment, whereas such a change is more likely with a fine scale. It was hypothesised, therefore, that the use of coarse scales in daily life might hide people's more flagrant inconsistencies, and that this would result in improved consistency on the coarse scale compared to the fine scale in this experiment.

The aim of this experiment was to replicate and extend the results of Experiment 1 by testing the prediction that instances of inconsistency should be reduced for a coarse scale (1-7) compared to a fine one (1-100). Participants rated the morality of the events in each sentence either on the fine slider scale from -100 to +100 used in Experiment 1, with -100 anchored as immoral, +100 as moral and 0 anchored as neither, or the coarse, Liker-type button scale discussed above, from -7 to +7, with -7 anchored as immoral, +7 as moral and 0 anchored as neither.

Method

Participants

Participants were Trinity College Dublin undergraduate psychology students compensated with course credit. Thirty-two participants took part in the experiment – 24 women and 8 men. Participants ranged in age from 17 to 32 years, with a mean age of 19.5 years. Similarly to Experiment 1, the decision was made *a priori* to exclude any participant whose average rating of the neutral items (two for each scale) was

higher than 1.5 (coarse scale) and 20 (fine scale); in fact no participant fit those criteria.

Design and Materials

Participants acted as their own controls and evaluated the same 20 sets of three sentences as the previous experiment. The materials were presented on Survey Gizmo (www.surveygizmo.com). They completed 20 triples, half with the fine scale, and half with the coarse scale. Within each of these blocks, half of the triples had the conjunction at the beginning, and half had the conjunction at the end. The order of the conjuncts, *A*, then *B*, remained constant; only the placement of the conjunction changed. Half of the participants completed the fine scale first, and half the coarse scale first. The order of the blocks of conjunction-first and conjunction-last within the two blocks of fine scale and coarse scale triples was counterbalanced. Contents were assigned to each block using random numbers generated by the website www.random.org, with the constraint that each block had two moral triples, two immoral triples, and one neutral triple.

Procedure

Participants read a paragraph of instructions, and gave their consent to participate by clicking to continue (see Appendix 2.4 for consent and debriefing). They then carried out the experiment. The key instruction for the finely grained scale was: “Please rate the degree to which the action was moral (from +1 to +100) or immoral (from -1 to -100) or neither (0)”. The instruction for the coarse scale was: “Please rate the degree to which the action was moral (+7) or immoral (-7) or neither (0)”. Each sentence in each triple was rated on a separate page.

Results

Order Effects

Table 2.2 presents the conjunctions from the set of 20 triples, and the mean numerical judgments for each conjunct and the conjunction for the coarse and fine scales in Experiment 2. There were no order effects, as established by the following 8 tests carried out using a Bonferroni correction of $p = .006$ (.05 divided by 8 tests). There was no order effect of whether participants completed a conjunction-first or conjunction-last block first for the fine scale, $M = 58.44$, $SD = 20.82$, and $M = 57.14$, $SD = 22.39$ for ratings of the moral actions on the fine scale, Mann-Whitney test, $U = 126.5$, $z = .057$, $p = .955$, so their data was collapsed over order. There was also no order effect of whether participants completed a conjunction-first or conjunction-last block first for the coarse scale, $M = 4.77$, $SD = 1.13$, and $M = 5.01$, $SD = 1.29$ for ratings of the moral actions on the coarse scale, Mann-Whitney test, $U = 109$, $z = .717$, $p = .473$, so their data was collapsed over order. There was also no order effect of whether participants completed a conjunction-first or conjunction-last block first for the fine scale, $M = -61.7$, $SD = 14.96$ and $M = -61.41$, $SD = 21.99$ for ratings of the immoral actions on the fine scale, Mann-Whitney test, $U = 123$, $z = .188$, $p = .851$, so their data was collapsed over order. There was also no order effect of whether participants completed a conjunction-first or conjunction-last block first for the coarse scale, $M = -5.09$, $SD = .81$ and $M = -4.94$, $SD = 1.25$ for ratings of the immoral actions on the coarse scale, Mann-Whitney test, $U = 121.5$, $z = .245$, $p = .806$, and so their data are collapsed for the remaining analyses.

Table 2.2: The moral, immoral and neutral conjunctions for Experiment 2, and the participants' mean judgments of the first conjunct, *A*, the second conjunct, *B*, and their conjunction, *A&B*, in Experiment 2 on the coarse scale, and fine scale (in parentheses).

<i>Contents</i>	<i>A</i>	<i>B</i>	<i>A&B</i>
<i>Moral Contents</i>			
A woman organized a charity event to raise money for research into kidney disease, and donated one of her kidneys to a friend of hers.	4.3 [62]	5.7 [84]	6.3 [89]
The staff of a hospital subsidized the construction of a badly needed burns unit, and worked together to save the victims of a terrorist bomb.	5.2 [54]	5.9 [55]	6.2 [70]
A woman took her sister's child into her home after her sister died, and paid for the child to go to University.	5.3 [65]	5.4 [48]	6.1 [72]
A young man helped an old woman to cross the street, and he gave his neighbours money after their house burned down.	3.8 [46]	4.4 [56]	5.7 [65]
A man intervened to help stop a fight, and organized Christmas gifts for children at an orphanage.	3.6 [51]	5 [62]	5.6 [76]
A woman gave blood so that more would be available to help people who had been in accidents, and donated money to a charity that helped to fund cancer research.	4.8 [54]	4.7 [49]	5.6 [69]
The director of a factory hired an equal number of able-bodied and disabled people, and stopped to help a motorist whose car had broken down by the side of the road.	4.3 [57]	3.8 [47]	4.9 [62]
A husband rejected an opportunity to cheat on his wife, and took his wife for a romantic walk.	3.2 [49]	.9 [26]	3.2 [50]
<i>Immoral Contents</i>			
A babysitter beat the infant in her care, and stole her employer's jewellery.	-6.8 [-82]	-5 [-55]	-6.8 [-88]
The headmaster of a school gave a teaching job to his nephew over better-qualified candidates, and sexually abused one of his pupils.	-3.6 [-53]	-6.8 [-92]	-6.7 [-93]
A man slept with his wife's friend, and hit his son.	-5.3 [-67]	-5.5 [-70]	-6.3 [-85]
A woman parked in the space reserved for disabled employees, and made it look as though a colleague had committed a crime so that he lost his job.	-3.8 [-35]	-6.1 [-70]	-6.2 [-75]
The owner of a bar forced the foreign girl who was working for him to accept a salary below minimum wage, and knowingly sold alcohol to a twelve-year-old.	-4.8 [-63]	-4.8 [-55]	-5.4 [-68]
The financial director of a company embezzled funds, and humiliated her assistant.	-4.1 [-49]	-3.8 [-45]	-5 [-61]
A man attacked a security guard, and stole money from his employer.	-4.2 [-49]	-4.7 [-49]	-4.9 [-69]
A school child terrorized the playground, and cheated in an exam.	-3.7 [-32]	-2.8 [-19]	-4.1 [-41]
<i>Neutral Contents</i>			
A woman finished her dinner at eight o'clock on a Tuesday evening, and prepared to go to bed at eleven in the evening.	0 [0]	0.1 [0]	0.1 [0]
A man bought a newspaper in a corner shop beside the train station, and posted an important document to a business associate in another city.	0 [7]	0.1 [10]	0.1 [7]
A woman wrote an email to a colleague to arrange to meet for lunch, and made a phone call to a hotel to enquire about the availability of their rooms.	0.1 [3]	0.1 [0]	0 [1]
A man sat on the couch in his apartment on a wet, rainy evening in the city, and read a collection of detective stories quickly because he was very interested in what would happen next.	-0.1 [4]	0 [7]	0 [7]

Furthermore, whether they completed a block of triples with the coarse scale first, or a block of triples with the fine scale first, there was no order effect in participants' ratings $M = 54.85$, $SD = 18.17$ and $M = 60.72$, $SD = 61$, $SD = 24.23$ for the moral actions on the fine scale, Mann-Whitney test, $U = 106.500$, $z = .810$, $p = .418$, so their data was collapsed over order. There was also no order effect for completing the coarse or fine block first in participants' ratings $M = 4.69$, $SD = 1.2$ and $M = 5.08$, $SD = 1.2$ for the moral actions on the coarse scale, Mann-Whitney test, $U = 106.500$, $z = .811$, $p = .417$, so their data was collapsed over order. There was also no order effect for completing the coarse or fine block first in participants' ratings $M = -57.45$, $SD = 16.88$ and $M = -65.67$, $SD = 19.67$ for the immoral actions on the fine scale, Mann-Whitney test, $U = 102.500$, $z = .961$, $p = .336$, so their data was collapsed over order, or in their ratings, $M = -4.68$, $SD = .97$ and $M = -5.34$, $SD = 1.03$ for the immoral actions on the coarse scale, Mann-Whitney test, $U = 76.000$, $z = 1.962$, $p = .05$, so again, their ratings were collapsed over order.

Consistency

The same criteria for consistency were followed as in Experiment 1. Six individual triples were removed from the analysis because they contained judgments of an inappropriate valence (e.g. -30 for a moral item, or +50 for an immoral item), and two were removed because of administrative error.

As can be seen from Figure 2.2, participants produced consistent rather than inconsistent triples at a greater than chance level, binomial test, $z = 7.65$, $p < .001$, $r = .39$, replicating Experiment 1. Furthermore, 20 of 32 participants made more consistent than inconsistent judgments, with 4 ties, binomial test, $z = 2.08$, $p < .05$, $r = .37$. This result holds for the fine scale (22 of 32 participants made more consistent

than inconsistent judgments, with 5 ties, binomial test, $z = 3.08, p < .005, r = .54$), replicating Experiment 1, but not for the coarse scale (19 of 32 participants made more consistent than inconsistent judgments, with 3 ties, Binomial test, $z = 1.49, p = .14$). As in Experiment 1, compared to if there was no tendency for participants to respond inconsistently, violations of consistency occurred reliably, on 33% of trials, Wilcoxon, $z = 4.871, p < .001, r = .61$.

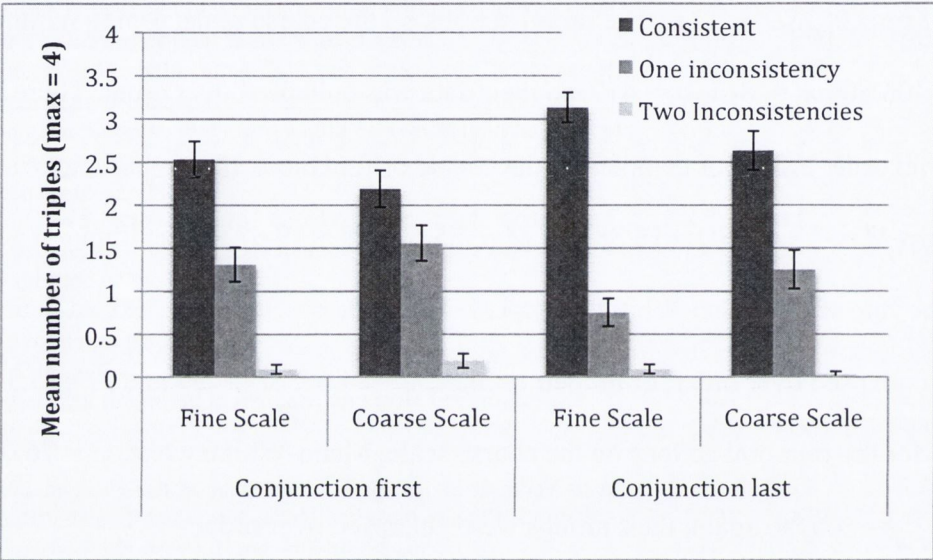


Figure 2.2: The mean number of triples in the three categories of consistency and violations depending on whether the conjunction was first or last in a triple of judgments, and on whether a fine scale or coarse scale was used. Error bars are standard error of the mean.

As in Experiment 1, participants varied significantly in their ability to make consistent judgments, Friedman, $\chi^2 (10, n = 31) = 78.005, p < .001$. Unlike in Experiment 1, participants made consistent judgments more often for some of the moral contents, Friedman, $\chi^2 (7, n = 28) = 20.539, p < .01$; as in Experiment 1, this was also the case for some of the immoral contents, Friedman, $\chi^2 (7, n = 29) = 18.543, p < .05$ (although this result was not significant for Experiment 1).

Scale, Morality and Position of Conjunction

Contrary to the prediction made regarding the potential influence of scale on consistency, judgments were consistent more often for the fine scale (-100 to +100) than the coarse scale (-7 to +7), Wilcoxon, $z = 2.162, p < .05, r = .27$.

An unexpected result in Experiment 1 was that the participants made more consistent responses for immoral actions than for moral actions. In Experiment 2, there was no significant difference in consistent judgments between moral and immoral contents overall, Wilcoxon, $z = .386, p = .70$.

As in Experiment 1, judgments were consistent more often when the conjunction came last than when it came first, Wilcoxon, $z = 3.040, p < .005, r = .38$.

The relationship between the three variables for judgments of consistency is depicted in Figure 2.3.

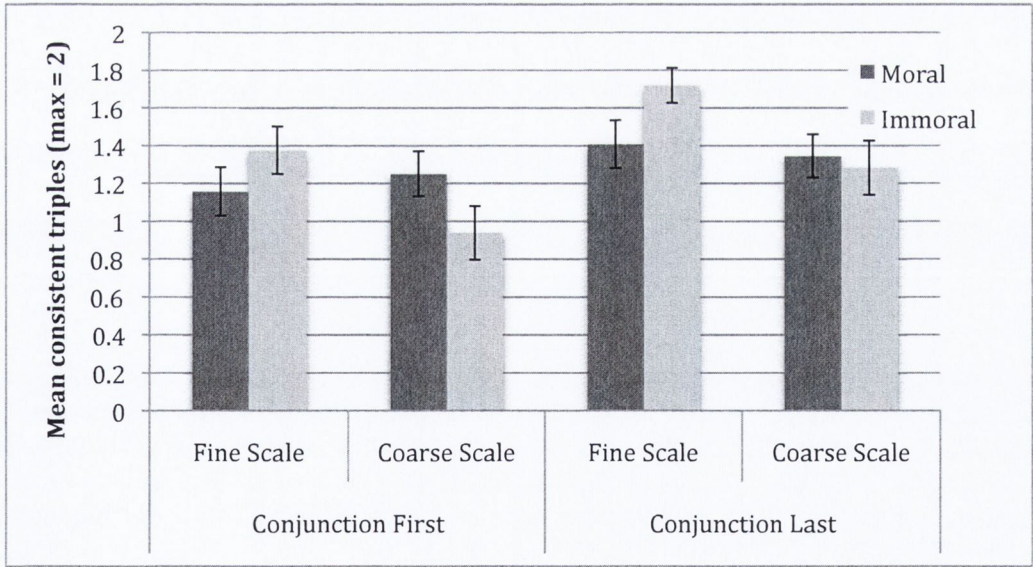


Figure 2.3 The mean number of consistent triples for Experiment 2 for moral and immoral contents, depending on whether the conjunction was first or last in a triple of judgments for both fine and coarse scales. Error bars are standard error of the mean.

There was no significant interaction of morality and position of conjunction, Wilcoxon, $z = 1.311$, $p = .190$. There was also no significant interaction of scale and position of conjunction, Wilcoxon, $z = .652$, $p = .515$. However, there was a significant two-way interaction of scale and morality, Wilcoxon, $z = 2.325$, $p < .05$, $r = .29$. This interaction was decomposed using four Wilcoxon Signed Rank tests, with a Bonferroni corrected alpha level of $p < .01$. Participants made significantly more consistent judgments for immoral contents when using the fine scale than when using the coarse scale, Wilcoxon, $z = 2.952$, $p < .005$, $r = .37$. There was no difference in consistent judgments for moral contents when using the fine scale compared to the coarse scale, Wilcoxon, $z = .053$, $p = .957$. People were somewhat more consistent using the fine scale for immoral ($M = 3.09$, $SD = 1.06$) than for moral contents ($M = 2.56$, $SD = 1.24$), although this result was not significant, Wilcoxon, $z = 1.992$, $p = .046$, $r = .25$. There was no difference in consistency between moral and immoral contents for the coarse scale, Wilcoxon, $z = 1.664$, $p = .096$.

Discussion

This experiment replicated the major results of Experiment 1 for judgments on a fine (-100 to +100) scale, and extended them to judgments on a coarse (-7 to +7) scale. Participants were again generally consistent, and more consistent than chance, but there was also a reliable tendency to violate consistency compared to if there were no tendency to respond inconsistently (i.e. zero violations). Furthermore, participants were again more consistent when the conjunction was presented at the end of the triple, rather than at the beginning. As in Experiment 1, these results support the contention that moral judgments are domain general, and subject to some of the same heuristic influences as non-moral judgments.

The manipulation of scale resulted in the opposite effect to that predicted: people were more consistent when using the fine scale overall. However, this effect only occurred for immoral items, and not for moral items. Similarly, participants were somewhat (although not significantly) more consistent for immoral than moral items for the fine scale, replicating the result from Experiment 1 (where only the fine scale was used, and participants were also more consistent for immoral items), but there was no difference in consistency between moral and immoral items for the coarse scale. Furthermore, although a majority of people made more consistent than inconsistent judgments for the coarse scale, the result was not significant.

It is possible that the fine scale, with its ability to allow participants to provide a precise judgment, may have encouraged people to try to be more precise in their judgments. The coarse scale, meanwhile, only allowed for rough judgments, which may have encouraged participants to take less care in their judgments. As for the question of why the effect only occurs for immoral items, as discussed above, there is a pervasive tendency for negative stimuli to have a more powerful influence on overall judgment than positive stimuli (e.g. Baumeister et al., 2001). Perhaps people may care more about, or alternatively, concentrate more for immoral than moral items due to a desire to avoid negative outcomes. These observations may provide an explanation for why the effect of scale is more powerful for immoral items: the fine scale may encourage precision, and, as discussed in the section on Experiment 1, consistency may be perceived to be more important for immoral items, or concentration may be greater for these items, because of a desire to protect against immoral actions and negative outcomes.

Although the effect of scale was in the opposite direction from that predicted, the result was nonetheless significant in terms of whether moral cognition is domain-

general or domain-specific. Experiments 1 and 2 demonstrated that consistency improved when the conjunction was judged at the end rather than the beginning of a triple, and when participants used a fine scale as opposed to a coarse scale. The influence of the position of the conjunction on consistency directly parallels an effect from research on consistency in probability, while a similar but opposite effect of scale occurs in probability judgments, in which consistency is improved by using a coarse scale as opposed to a fine scale (e.g. Khemlani et al., 2012). These results seem incompatible with the view that moral judgments are made without being influenced by the kinds of mental shortcuts that influence non-moral judgments (e.g. Dwyer et al., 2010).

The first two experiments discussed in this chapter tested whether people are able to make consistent judgments of two actions of the same valence. These experiments demonstrated effects of position of conjunction, and of immoral items. The third experiment in this series examined whether such effects occur when people make judgments of mixed-valence moral actions: that is, when people do both good and bad things.

Experiment 3

People are not usually uniformly virtuous or lacking in virtue. Most people do both good and bad things, and deciding which is more important in forming an overall impression of morality is a complex task. Renowned cyclist Lance Armstrong has been roundly criticised for using performance-enhancing drugs to win many cycling titles, but he also used his fame to raise millions of dollars for cancer research (e.g. Belsky, 2012). As discussed at the beginning of this chapter, rugby player Paul O'Connell was condemned by many fans for recklessly (though accidentally) kicking

Leinster's Dave Kearney in the head (e.g. Peavoy, 2013), but has also been involved in helping charities in his native Munster (e.g. Bielenberg, 2013). How do people make judgments of these mixed moral and immoral actions? Does one cancel the other out? Do people sometimes say a good action outweighs a bad action, and vice versa, or does one side persistently have a greater influence than the other?

Commonsense suggests that the latter is the case, and that immoral items are likely to have a pervasive effect. This idea is supported by experimental evidence from non-moral domains such as emotion and learning, among others, that negative-valence items are more important to combined judgments than positive valence items (e.g. Baumeister et al., 2001). Past studies also indicate that immoral items have a stronger influence than moral items on judgments of how morally commendable or reprehensible a person who commits multiple moral and immoral acts is (e.g. Risky & Birnbaum, 1974). In this latter study, participants were asked to judge the morality of carrying out an entire set of items, and they tended to judge a person to be immoral if they committed an immoral deed, even if several moral deeds were also attributed to that person.

The research question in Experiment 3, reported here, was different in that it focused on whether or not participants can make judgments of pairs of mixed-valence actions that are consistent with their own judgments of the individual actions. In Experiment 3, the method from Experiments 1 and 2 was used to allow the comparison of participants' judgments of individual actions with their judgments of conjunctions. This comparison provided a stronger test of the effect of the dominance of immoral items in the judgment of multiple moral actions than that reported by Risky and Birnbaum (1974), since participants' judgments of the individual events were made either just before, or just after, their judgments of the conjunction, and

their judgments of the conjunctions were assessed in comparison with *their own* judgments of the individual items, rather than a pre-tested sample. The following is an example of their task:

A person donated one of her kidneys to a friend of hers. Please rate the degree to which the action was moral (from +1 to +100) or immoral (from -1 to -100) or neither (0)

-100 0 +100

A person beat the infant in her care. Please rate the degree to which the action was moral (from +1 to +100) or immoral (from -1 to -100) or neither (0)

-100 0 +100

A person donated one of her kidneys to a friend of hers, and the next week beat the infant in her care. Please rate the degree to which the actions were moral (from +1 to +100) or immoral (from -1 to -100) or neither (0).

-100 0 +100

Given that participants produced inconsistent judgments in the previous experiments, it was predicted that participants would produce inconsistent judgments, and that the conjunctions would give greater influence to the immoral items than to the moral items. Again, given the results of the previous experiments, it was predicted that the position of the conjunction would influence consistency, and specifically that participants would be more consistent when they made their judgment of the conjunction last.

Triples in which the conjunction equalled the sum of the conjuncts were categorised as consistent (see Appendix 2.5 for the full criteria of consistency and inconsistency). Four categories of possible moral inconsistency were identified. These were cases in which the conjunction was judged to be more moral than the sum of the participant's judgments of the individual items would indicate it should have been. Consider the example above, in which the conjunction was, '*A person donated one of her kidneys to a friend of hers, and the next week beat the infant in her care*'. The four types of moral inconsistency were: (i) the conjunction was better than either conjunct (e.g. *A person donated one of her kidneys to a friend of hers* = 30, *A person beat the infant in her care* = -10, *A person donated one of her kidneys to a friend of hers, and the next week beat the infant in her care* = 50); (ii) the conjunction was as good as the good conjunct (e.g. *A person donated one of her kidneys to a friend of hers* = 50, *A person beat the infant in her care* = -20, *A person donated one of her kidneys to a friend of hers, and the next week beat the infant in her care* = 50); (iii) the conjunction was good, but not as good as the good conjunct (e.g. *A person donated one of her kidneys to a friend of hers* = 50, *A person beat the infant in her care* = -30, *A person donated one of her kidneys to a friend of hers, and the next week beat the infant in her care* = 40); and (iv) the conjunction was 0, eliminating a tendency to rate the immoral item higher in the conjuncts (e.g. *A person donated one of her kidneys to a friend of hers* = 50, *A person beat the infant in her care* = -70, *A person donated one of her kidneys to a friend of hers, and the next week beat the infant in her care* = 0). Each of these categories were instance where the conjunction was rated to be more moral than the sum of the conjuncts indicated it should have been, and they were combined into a single category of moral inconsistent triples.

A corresponding four immoral inconsistent categories were identified and combined into a single category of immoral inconsistent triples; these were triples in which the ratings of the conjunction was more immoral than the sum of the ratings of the conjuncts would suggest it should have been, in corresponding ways to the moral inconsistent categories above. It should be noted that the criterion for consistency in this experiment was somewhat more stringent than in Experiments 1 and 2, in that consistency required addition. This additive criterion was impossible in the first two experiments, since adding judgments of the two conjuncts precisely might have exceeded the limits of the scale; however, addition was always possible in Experiment 3.

This experiment also used three different kinds of moral materials, constructed based on ratings of the materials in Experiment 2. These were matched materials (in which the moral and immoral items had been rated close to equivalently moral or immoral in Experiment 2), higher moral materials (in which the moral item had been rated more moral than the immoral item had been rated immoral), and higher immoral materials (in which the immoral item had been rated more immoral than the moral item had been rated moral). The purpose of using these different materials was to test whether the hypothesised influence of the immoral item over the moral item would occur not only when the immoral item was much more immoral than the moral item was moral, but also when the items were moral or immoral to a similar degree. Using these different materials also permitted the investigation of whether the moral item might have more influence than the immoral item when it had been rated much more moral than the immoral item was immoral.

Method

Participants

Participants were Trinity College Dublin undergraduate psychology students compensated with course credit. Twenty participants took part in the experiment, and they were 13 women and 7 men. Participants ranged in age from 19 to 62 years, with a mean age of 31 years. The same *a priori* criterion of excluding any participants whose average rating of the neutral items was greater than 20 was adopted as in Experiments 1 and 2. One participant was excluded on these criteria, leaving 19 participants in the sample.

Design and Materials

As discussed in the previous section, materials were constructed based on the average scores of items on the fine scale in Experiment 2. Three sets of materials were constructed: Matched materials (materials in which the average rating of the moral and immoral items was close, e.g. +84 and -82); Higher Moral materials (materials in which the moral conjunct had on average been rated as more moral, e.g. +54, than the immoral conjunct had been rated immoral, e.g. -19); and Higher Immoral materials (materials in which the immoral conjunct had on average been rated more immoral, e.g. -70, than the moral conjunct had been rated moral, e.g. +51). There were eight triples in the Matched materials, four in the Higher Moral materials, and four in the Higher Immoral triples; the difference in the number of materials was due to administrative error, and was dealt with by analysing proportions of triples, rather than the raw number of triples. Full materials, along with each item's average rating in Experiment 2, can be seen in Appendix 2.6.

A within-subjects design was used. Participants completed one block of triples with the conjunction at the start of the triple, and one block of triples with the conjunction at the end of the triple, and which block they completed first was alternated. Each block contained four Matched triples, two Higher Moral triples, two Higher Immoral triples, and two Neutral triples. Whether participants received the moral or immoral conjunct first was balanced so that within each block of conjunction-first and last, half of the Matched, Higher Moral and Higher Immoral triples had the moral conjunct first, and half had the immoral conjunct first. The experiment was carried out online using Survey Gizmo (www.surveygizmo.com). Participants were asked to rate how moral or immoral the action or actions were on a Likert-type scale with ten-point differences between buttons, from -100 (most immoral), through 0 (neither moral nor immoral) to +100 (most moral). The scale was labelled with written anchors at the three points mentioned above (i.e. most immoral, neither moral nor immoral, most moral), and each button had a numeric label (e.g. -90, -80, -70, etc).

Procedure

Participants read a paragraph of instructions, and then carried out two practice trials. They then carried out the experiment proper. The key instruction was: “Please rate the degree to which the action was moral (from +1 to +100) or immoral (from -1 to -100) or neither (0)”². Finally, they typed a comment answering if they experienced

² Due to an administrative error, 7 participants received questions in the conjunction-first block in which they were asked to rate the morality of the “actions” for the single action conjuncts, and to rate the morality of the “action” for the conjunction of actions. Comparisons of the responses of participants who received the error in these instructions, and those of the participants who did not, revealed no significant differences in terms of consistent, moral inconsistent or immoral inconsistent triples overall, when the conjunction was read first and last, or for any of the materials, Matched, Higher Moral, or Higher Immoral.

any website errors, and read a debriefing page (see consent and debriefing, Appendix 2.7).

Results

Three individual triples were removed in cases where participants judged a moral conjunct to be immoral (e.g. where an item should have received a moral judgment, such as +30, participants made an immoral judgment, such as -30), or vice versa. The materials used in Experiment 3, as well as the mean ratings for each item, can be seen in Table 2.3.

Table 2.3. Matched, Higher Moral, and Higher Immoral materials, with the average ratings for the moral and immoral conjuncts and the conjunction in Experiment 3.

<i>Contents</i>	<i>Moral</i>	<i>Immoral Conjunction</i>	
<i>Matched Conjunctions</i>			
A person beat the infant in her care, and the next week donated one of her kidneys to a friend of hers.	80	-87	-50
A person stole money from his employer, and the next week helped an old woman to cross the street.	39	-68	-46
A person slept with his wife's friend, and the next week organized a charity event to raise money for research into kidney disease.	54	-63	-42
A person forced the foreign girl who was working for him to accept a salary below minimum wage, and the next week organized Christmas gifts for children at an orphanage.	54	-64	-41
A person embezzled funds, and the next week intervened to help stop a fight	47	-63	-38
A person hit his son, and the next week took his sister's child into his home after his sister died.	65	-61	-39
A person gave a teaching job to his nephew over better-qualified candidates, and the next week donated money to a charity that helped to fund cancer research.	49	-50	-18
A person attacked a security guard, and the next week paid for his dead sister's child to go to University.	63	-55	0
<i>Higher Moral Conjunctions</i>			
A person terrorized their workplace canteen, and the next week subsidized the construction of a badly needed burns unit.	62	-64	-29
A person humiliated her assistant, and the next week gave her neighbours money after their house burned down.	61	-52	-12
A person cheated in an exam, and the next week gave blood so that more would be available to help people who had been in accidents.	60	-51	1
A person parked in the space reserved for disabled employees, and the next week helped save the victims of a terrorist bomb.	74	-38	35
<i>Higher Immoral Conjunctions</i>			
A person sexually abused one of his pupils, and the next week rejected an opportunity to cheat on his wife.	37	-86	-80
A person made it look as though a colleague had committed a crime so that he lost his job, and the next week hired an equal number of able-bodied and disabled people.	17	-81	-69
A person knowingly sold alcohol to a twelve-year-old, and the next week took his wife for a romantic walk.	4	-67	-54
A person stole her employer's jewellery, and the next week stopped to help a motorist whose car had broken down by the side of the road.	52	-56	-34

There was no order effect of whether participants completed the conjunction-first block or the conjunction-last block first in participants' ratings for the good items, $M = 57.82$, $SD = 18.61$ and $M = 45.65$, $SD = 19.43$ respectively, Mann-Whitney $U = 30$, $z = 1.225$, $p = .220$, so the results were collapsed over order. There was also no order effect of whether participants completed the conjunction-first block or the conjunction last block in participants' ratings for the bad items, $M = -70$, $SD = 20.05$ and $M = -$

57.67, $SD = 23.14$ respectively, Mann-Whitney $U = 32.5$, $z = 1.021$, $p = .307$, so the results were collapsed over order.

Consistency

Overall, participants made more immoral inconsistent judgments ($M = 9.26$, $SD = 3.33$) than moral inconsistent judgments ($M = 3.05$, $SD = 2.5$), Wilcoxon, $z = 3.424$, $p < .005$, $r = .55$, or consistent judgments ($M = 3.53$, $SD = 2.44$), Wilcoxon, $z = 3.292$, $p < .005$, $r = .53$. There was no difference between moral inconsistent and consistent judgments, Wilcoxon, $z = .655$, $p = .512$.

Consistency and Position of Conjunction

The same pattern of results occurred for conjunction-first and conjunction-last triples, as illustrated by Figure 2.4. For conjunction-first triples, participants made more immoral inconsistent ($M = 4.68$, $SD = 1.63$) than moral inconsistent ($M = 1.63$, $SD = 1.3$) judgments, Wilcoxon, $z = 3.370$, $p < .005$, $r = .55$, or consistent judgments ($M = 1.63$, $SD = 1.21$), Wilcoxon, $z = 3.348$, $p < .005$, $r = .54$. There was no difference between moral inconsistent and consistent judgments, Wilcoxon, $z = .053$, $p = .958$.

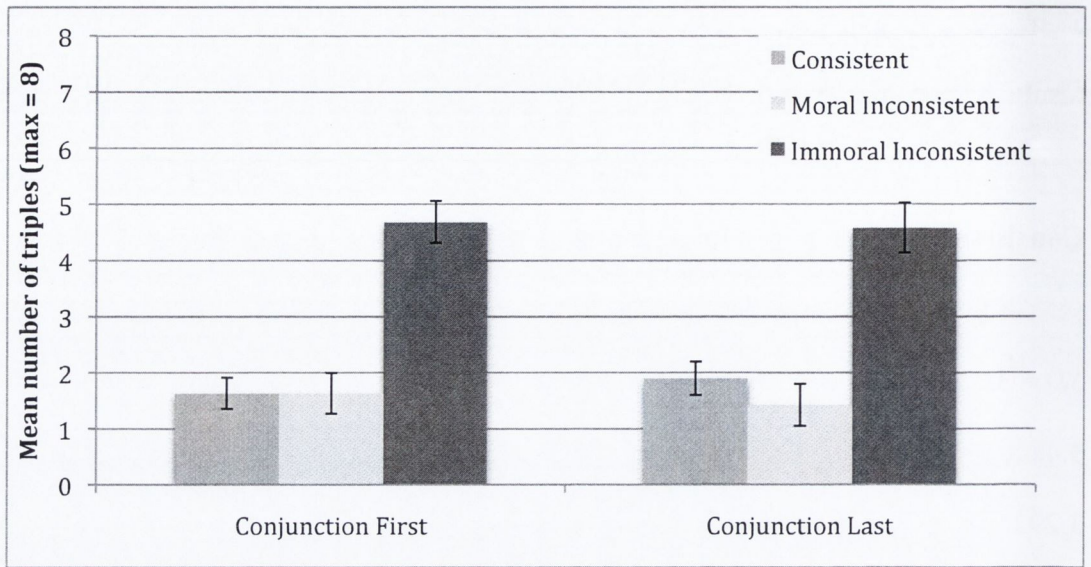


Figure 2.4. Mean number of consistent, moral inconsistent, and immoral inconsistent triples for conjunction-first and last triples. Error bars are standard error of the mean.

For conjunction-last triples, participants again made more immoral inconsistent judgments ($M = 4.58$, $SD = 1.92$) than moral inconsistent judgments ($M = 1.42$, $SD = 1.64$), Wilcoxon, $z = 3.111$, $p < .005$, $r = .51$, or consistent judgments ($M = 1.89$, $SD = 1.59$), Wilcoxon, $z = 2.929$, $p < .005$, $r = .48$. Again, there was no difference between moral inconsistent and consistent judgments, Wilcoxon, $z = 1.029$, $p = .303$.

There were also no differences between conjunction-first and last for immoral inconsistent judgments, Wilcoxon, $z = .208$, $p = .835$, for moral inconsistent judgments, Wilcoxon, $z = .356$, $p = .722$, or for consistent judgments, Wilcoxon, $z = .744$, $p = .439$. Immoral inconsistent judgments were by far the most common overall, and position of conjunction, unlike in Experiments 1 and 2, had no influence on the pattern of judgments generated by participants.

Consistency for Matched, Higher Moral and Higher Immoral triples

The relationship between the material types for both conjunction-first and last triples can be seen in Figure 2.5.

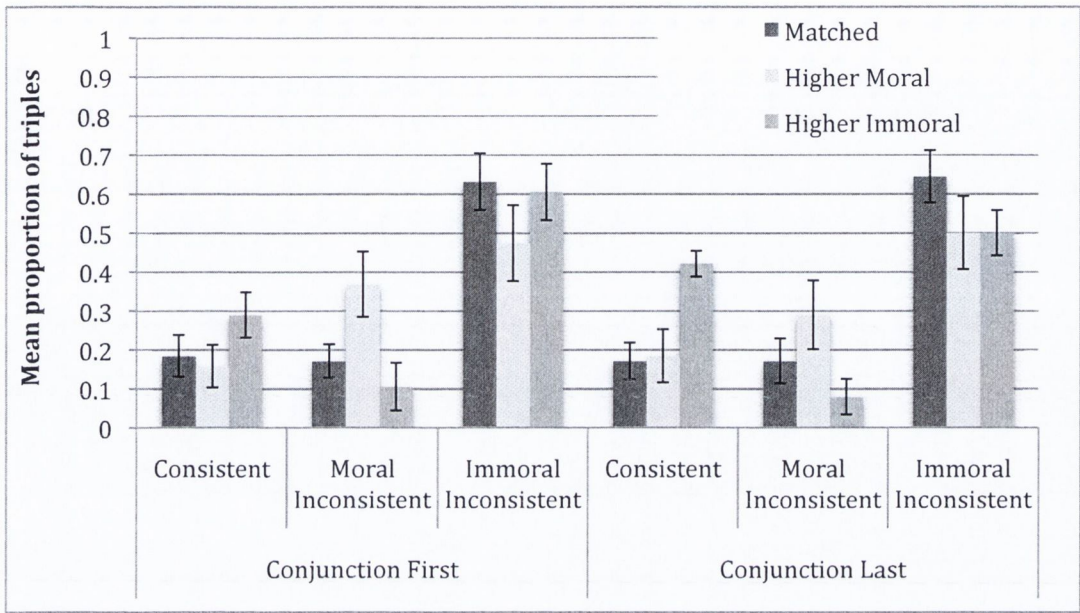


Figure 2.5: The mean proportion of triples in the three categories of consistency and violation depending on whether the conjunction was first or last in a triple of judgments, and on whether the materials were Matched, Higher Moral or Higher Immoral. Error bars are standard error of the mean.

Matched triples. For Matched triples, the most common result was for responses to be judged immoral inconsistent. A higher proportion of Matched triples were judged to be immoral inconsistent ($M = .64$, $SD = .29$) than moral inconsistent ($M = .17$, $SD = .18$), Wilcoxon, $z = 3.277$, $p < .005$, $r = .53$, as well as than consistent ($M = .18$, $SD = .19$), Wilcoxon, $z = 3.104$, $p < .005$, $r = .5$. There was no difference between the proportions of moral inconsistent and consistent judgments, Wilcoxon, $z = .362$, $p = .717$.

Higher Moral triples. A higher proportion of Higher Moral triples were judged to be immoral inconsistent ($M = .49$, $SD = .35$) than were judged to be consistent ($M = .17$, $SD = .22$), Wilcoxon, $z = 2.375$, $p < .05$, $r = .39$; but not than moral inconsistent ($M = .33$, $SD = .3$), Wilcoxon, $z = 1.15$, $p = .25$. There was also no difference between the proportions of moral inconsistent and consistent triples, Wilcoxon, $z = 1.422$, $p = .155$.

Higher Immoral triples. A higher proportion of Higher Immoral triples were judged to be immoral inconsistent ($M = .55$, $SD = .24$) than moral inconsistent ($M = .09$, $SD = .24$), Wilcoxon test, $z = 3.002$, $p < .005$, $r = .49$, and than consistent ($M = .36$, $SD = .21$), Wilcoxon test, $z = 1.999$, $p < .05$, $r = .32$. A greater proportion of Higher Immoral triples were also judged to be consistent than moral inconsistent, Wilcoxon test, $z = 2.382$, $p < .05$, $r = .39$.

Differences between material types. There was no difference between the proportion of immoral inconsistent judgments in Matched and Higher Moral triples, Wilcoxon, $z = 1.592$, $p = .111$, or in Matched and Higher Immoral triples, Wilcoxon, $z = .1261$, $p = .207$, or in Higher Moral and Higher Immoral triples, Wilcoxon, $z = .604$, $p = .546$. This lack of differences is characteristic of the prevalence of immoral inconsistent judgments throughout the dataset, for each material type.

A greater proportion of Higher Moral triples were judged moral inconsistent than Matched triples, Wilcoxon, $z = 1.955$, $p = .051$, two-tail, $r = .32$, and than Higher Immoral triples, Wilcoxon, $z = 2.063$, $p < .05$, $r = .33$. A somewhat greater proportion of Matched than Higher Immoral triples were judged moral inconsistent, although this difference was not significant, Wilcoxon, $z = 1.912$, $p = .056$.

A greater proportion of Higher Immoral triples were consistent than were Matched triples, Wilcoxon, $z = 2.556, p < .05, r = .41$, and than were Higher Moral triples, Wilcoxon, $z = 2.658, p < .01, r = .43$). There was no significant difference between the proportions of Matched and Higher Moral triples that were judged to be consistent, Wilcoxon, $z = .182, p = .856$.

Discussion

The primary result for mixed-valence moral judgments was that the immoral item outweighed the moral item, and judgments were immoral inconsistent, rather than either moral inconsistent or consistent. This means that judgments of the conjunction of a moral and immoral action were rated as more immoral than the sum of the participants' judgments of the individual items. So even when the items were rated very similarly (in Experiment 2), when participants heard that someone beat the infant in her care, and the next week donated one of her kidneys to a friend of hers, it was generally the former, rather than the latter, that had the greater influence on their judgment. Furthermore, the way that this happened (i.e. immoral inconsistent) suggests that, when combining moral and immoral judgments, the moral item is disproportionately lacking in influence when compared to the same participant's judgment of that item individually.

The same results occurred for conjunction-first and conjunction-last triples, suggesting that, unlike the tasks in Experiments 1 and 2, position of conjunction does not influence moral judgments for the task of judging triples of mixed moral valence. This suggests that the effect of position of conjunction on consistency is dependent on the two items being judged being of the same valence (i.e. two moral or two immoral), and does not extend to mixed-valence judgments.

For Matched materials, participants made immoral inconsistent judgments more than either consistent or moral inconsistent judgments; however, for Higher Moral triples, the difference between immoral inconsistent and moral inconsistent judgments was eliminated. That is, when participants made judgments of someone who does something *very* moral and something *slightly* immoral, there was no significant tendency for the immoral item to overwhelm the moral item. So, for example, when participants read that a person cheated in an exam, the negative judgment they had of this action could be sometimes be offset by the knowledge that the next week the person gave blood so that more would be available to help people who had been in accidents. However, there was also no significant tendency for the moral item to be more influential in judgments of the conjunction than the immoral item, even though these items were constructed based on pre-tests suggesting that these moral items were more moral than the immoral items were immoral.

In contrast, for Higher Immoral triples, judgments were rarely moral inconsistent, as both immoral inconsistent and consistent responses were significantly more common. The former effect indicates that, unlike with Higher Moral triples, a very immoral item usually outweighs a slightly moral item.

Overall, the results of this experiment emphasise that acting immorally is perceived to be worse than acting morally is perceived to be good. This is the case even when participants' own judgments of individual moral and immoral actions are used as the baseline against which consistency is measured. Only when people judge a very moral action and a slightly immoral action does the immoral action fail to have a greater influence. While this last result provides some hope that those who carry out immoral actions can be redeemed in the minds of others, it also underlines that such redemption is often an uphill battle.

General Discussion

The experiments reported in this chapter employ cognitive tasks in the moral domain that often yield inconsistency in non-moral domains. The research question was whether people could make judgments of the morality of pairs of actions that were consistent with their judgments of the actions individually. In this section the results of the experiments are discussed, and particularly how they relate to the theoretical debate over whether cognition about moral matters is similar to cognition about non-moral matters.

Experiment 1 demonstrated that participants could combine pairs of same-valence moral judgments consistently, but that they also reliably produce inconsistent judgments. So, for example, when participants were asked how good it is for someone to donate a kidney to save a life, and how good it is to work for many years to raise money for research into kidney disease, they usually successfully judged the pair of actions to be better than either individually, but also sometimes failed to do so. When people were inconsistent, they judged a pair of moral actions to be no better than one of the actions individually. In these cases, the results indicated that participants sometimes judged the morality – or immorality – of the pair of actions to come primarily from one of the actions, and that the other contributed relatively little, if anything. This result suggests that people's judgments can sometimes be relatively insensitive to multiple moral events, and only one moral or immoral action may be factored into people's judgments.

The main novel finding of Experiment 1 is that people produce more consistent judgments when the conjunction is judged after the conjuncts. This difference seems important to everyday moral judgments. Inconsistent judgments

were those in which the conjunction was not judged to be better (in the case of moral triples) or worse (in the case of immoral triples) than the individual moral or immoral actions that comprised it. This result therefore implies that if people hear about someone, for example, attacking a man *and* stealing from him, they are more likely to judge subsequent actions to be at least as immoral as those immoral things together, compared to if they heard the separate bad actions first. Likewise, if people hear about someone intervening to help stop a fight *and* tending to the wounds of someone injured in a fight, they are more likely to judge subsequent individual actions to be at least as moral as those moral things together, than they would if they heard the separate good actions first.

Although the experiments tested judgments of actions, rather than judgements of people, it is tentatively suggested that these results could have important implications for moral judgments of people who have consistently carried out moral and immoral actions in the past. The overall presence of inconsistency in Experiments 1 and 2 suggests that people may sometimes receive a ‘free pass’ when they commit multiple moral wrongs, in that they may not be judged any more harshly for multiple wrongs than for individual wrongs. Similarly, people who do multiple good things sometimes may not be praised any more than people who do individual good things. However, the fact that inconsistency occurred more when the conjunction was read first also suggests that, when people hear about multiple moral wrongs, their subsequent judgments of individual items may be more inclined to be as negative as if the judgment was of multiple wrongs. A similar effect may occur when people hear about multiple good actions before hearing about individual good actions.

As well as replicating the principal results of Experiment 1, Experiment 2 demonstrated that, contrary to prediction, participants are more consistent using a fine

scale than a coarse scale. It is possible that the fine-grained scale encouraged people to be more precise because of the many options it makes available, while the coarse scale, with its relatively few options, discouraged people from precision. This effect occurred significantly for immoral items, but not for moral items. This may illustrate a heuristic tendency to focus more, or to care more, when dealing with immoral items than with moral, perhaps in order to avoid negative outcomes. Thus, if people are driven by both the fine scale *and* immoral contents to be more precise and careful when making their judgments, this may explain why participants were more consistent for the fine scale, but only significantly so for immoral items.

Experiment 3 demonstrated the strong effect of an immoral item when judging triples of mixed moral and immoral actions. Participants tended to provide judgments of triples in which the conjunction was judged to be more immoral than the sum of participants' judgments of the individual actions. This type of immoral inconsistency was more common than the corresponding type of moral inconsistency (wherein the conjunction was judged to be more moral than would be suggested by the sum of participants' judgments of the individual actions), and than consistency (wherein judgment of the conjunction equalled the sum of the judgments for the actions individually). This pattern occurred for the dataset overall, for conjunction-first and last triples, for Matched triples, and for Higher Immoral triples. These results suggest that immoral actions are more influential in mixed-valence moral judgments than moral actions. These results are similar to effects of negative stimuli in non-moral domains (e.g. Baumeister et al., 2001), which suggests that cognition about moral matters is similar to cognition about non-moral issues.

Perhaps the most important result from Experiment 3 was that judgments of mixed-valence conjunctions tended to be disproportionately influenced by the

immoral item even when participants made judgments of the individual moral and immoral actions themselves. Even with the recent knowledge of how moral and immoral they think the actions are individually, participants still failed to weigh the moral item appropriately in their judgments of the conjunction. It seems that when people heard about someone carrying out both moral and immoral actions, the value they put on those actions when they occurred individually was relatively unimportant: the immoral item mattered more.

Experiment 3 also demonstrated that the presence of a highly moral action (i.e. for Higher Moral materials) decreases the influence of the immoral action somewhat. This result indicates that in spite of the general tendency of immoral items to outweigh moral items, very moral actions were still valued highly. It may be inferred that it *is* possible for people to redeem their immoral deeds by carrying out extremely moral actions. However, when people read about someone doing something very immoral and something only somewhat moral (i.e. in Higher Immoral materials), or when they read about someone doing similarly moral and immoral things (i.e. in Matched materials), they very rarely thought the moral item outweighed the immoral item. Unless the actor's immoral behaviour was almost trifling, it was very difficult to overcome with moral actions. For instance, the backlash against Lance Armstrong since the revelation of his use of performance enhancing drugs (after years of denials) has been intense and spirited, and does not appear to have been ameliorated much by the knowledge of the good he did in generating money to fund cancer research during the course of his career.

These experiments represent only a beginning, along with some previous research (e.g. Baron, 2010) in terms of identifying what domain-general processes might operate in moral judgment. It is possible that future experiments might utilise

the framework provided by other types of heuristics to test whether they occur when people make moral judgments, and also whether there are contextual influences to which moral judgments are, in fact, immune. The results of the experiments suggest that a variety of domain-general processes may be implicated in moral reasoning, and future research might help to explain other moral phenomena by identifying such processes.

Domain-general or domain-specific?

All three of the experiments in this chapter presented effects mitigating consistency in numeric moral judgments that are similar to effects found in non-moral judgments. The types of inconsistency observed in Experiment 1, as well as the fact that consistency was improved when the conjunction was judged at the end of the triple rather than the beginning, were similar to effects observed in similar experiments on the probabilities of unique events (Khemlani et al., 2012). These results demonstrated for the first time a type of conjunction fallacy in moral judgments. The same effects replicated in Experiment 2, along with the novel observation that participants were more consistent when they made judgments on fine, rather than coarse scales, in spite of the fact that coarse scales provided fewer opportunities for inconsistency. Furthermore, participants were more consistent for the fine scale overall, and this effect also occurred for immoral items, but not for moral items – an effect that also has parallels in non-moral domains such as learning, where negative stimuli often have a stronger influence than positive stimuli, as discussed above (e.g. Baumeister et al., 2001).

This influence of the immoral was also found in Experiment 3, in which mixed-valence moral judgments were much more powerfully influenced by immoral

rather than moral actions, in all cases except when the moral action was *far* better individually than the immoral action was bad. Again, the negative stimuli had an influence in moral judgments similar to that found in many other domains.

Many of these results parallel findings in non-moral domains, and appear to be incompatible with the idea that moral cognition is resistant to the heuristic effects that are so prevalent in non-moral cognition. The evidence presented here therefore provides some support for theories that treat moral cognition as domain-general cognition (e.g. Sunstein, 2005; Baron, 2010; Bucciarelli et al., 2008), rather than domain-specific (e.g. Mikhail, 2009; Dwyer et al., 2010). As discussed in the introduction, the theory of a universal moral grammar is a prominent domain-specific theory of moral judgment. This theory suggests that many moral judgments resemble legal rules, are made intuitively, and by a specific moral faculty (e.g. Mikhail, 2007, 2009; Dwyer et al., 2010). This last claim is the one addressed by the data in Experiments 1-3, and the data appear to be incompatible with the universal moral grammar theory. The empirically observed similarities between moral and non-moral judgments in these experiments call into question the necessity of hypothesising a specialised moral faculty.

Although the social intuitionist model (e.g. Haidt, 2001; Haidt & Bjorklund, 2008), and dual-process theories of moral judgment (e.g. Greene et al., 2008; Paxton & Greene, 2010) do not focus on heuristics in moral judgment, they do emphasise similarities between moral and non-moral cognition in terms of affective engagement and unconscious processing. The fact that both of these theories suggest that unconscious intuitions are likely to play a role in moral judgment indicates that they are compatible with the idea of moral heuristics. The role of emotions in conscious

and unconscious processing in moral judgment is tested and discussed in detail in Chapters 3 and 4.

Implications

It has been argued that, although people may have different philosophical moral beliefs, they are likely to make moral judgments that are errors even on their own system of moral beliefs (e.g. Sunstein, 2005). Philosophers have developed normative theories of how people should behave, such as utilitarianism (e.g. Mill, 1863/2007), according to which people should make moral decisions in terms of how to maximize utility for society as a whole, and deontology (e.g. Kant, 1788/2002), according to which people should act according to moral rules, and only make something a moral rule if it should be followed universally. The results reported in this chapter do not influence the normative content of such theories; such theories do not need to be psychologically realistic in order to describe what people *should* do. However, the reality that moral judgments are influenced by psychological factors outside of these systems may make adherence to such a set of rules very difficult, if not untenable.

Experiments 1 and 2 investigated whether people judged that two moral actions should be judged to be more moral than either individually, and two immoral actions should be judged to be more immoral than either individually. Although participants generally do so, they also commit errors, and these errors vary according to normatively irrelevant factors, such as whether the conjunction is judged before the conjuncts, and whether the scale the judgments are made on is coarse or fine.

Experiment 3 tested whether participants' judgments of conjunctions of moral and immoral actions should be equal to the sum of their judgments of the individual

actions. People persistently judged conjunctions of moral and immoral actions to more immoral than the sum of their judgments of the actions individually. It appears that, in combined judgments of moral and immoral actions, immoral actions are so influential that people's judgments are often inconsistent.

Certain economists have begun to include psychological knowledge about economic behaviour in their models, in order to develop a realistic model of how people behave economically, and to develop interventions in order to improve financial outcomes (e.g. Thaler & Benartzi, 2004). Knowledge of some of the heuristic influences on moral behaviour discussed in this chapter might similarly point the way for psychologists to develop interventions that might help people to make more consistent moral judgments.

The type of conflict explored in Experiment 3 (i.e. between the influence of moral and immoral actions in judgments) is not the only type of moral conflict. As discussed in Chapter 1, moral dilemmas, in which two opposing options represent morally desirable goals, and only one can be met, have been an important source of data for moral psychologists. The experiments reported in Chapter 3 investigate how people make judgments of others' decisions as they resolve these dilemmas, and what factors influence such judgments.

Chapter 3 Information about Emotions Modulates Reasoning about Moral Dilemmas

The experiments in Chapter 2 illustrated that unconscious processes, including heuristics, are often influential in moral judgments. As discussed in Chapter 1, a prominent strand of research in recent years has suggested that emotions may play an important role in prompting people to make intuitive, rather than deliberative moral judgments. It is these emotions that are addressed in Chapter 3, and specifically how information about the emotions experienced by people in moral dilemmas may influence how people judge the rightness and wrongness of people's actions.

The emotions experienced, or not experienced, by those who have made difficult, morally charged decisions are sometimes reported as factors that influence people's judgments of their choices, or that mitigate their guilt or responsibility. For instance, when the Costa Concordia cruise ship sank off the west coast of Italy in 2012, attention focused on the captain, not least because he abandoned ship before the passengers in his care had succeeded in doing so. The captain of the Concordia was reported as saying, "In the moment the floor started to become steeper, you have no other option: To die or to swim. I regret nothing" (Kosinski, 2013).

A statement like this, which minimises the emotion experienced in such cases, may seem callous, and prevent people from identifying with the person in the dilemma. Justification without emotion behind it may sound cold and uncaring, especially in cases involving death. Judges and media members have often imbued their judgments of criminals with emotions, and implicitly condemned the criminals by stating that they 'show no remorse' for their crimes in court (e.g. Peachey, 2013; McLaughlin & Brown, 2013). Statements such as these indicate that the failure to

experience, or at any rate to communicate emotions when carrying out actions that are widely considered to be morally wrong can lead people to view a moral actor more negatively. On the other hand, experiencing very strong emotions can sometimes be used as a defence for illegal acts of violence (e.g. Whitehead & Haugh, 2012), and expressing remorse can be used as a reason to give a criminal a lesser sentence than might be legally justifiable (e.g. Aaron, 2013), suggesting that people understand that emotions can play a powerful role in moral decisions and are more inclined to forgive emotionally charged immoral behaviour.

Cognitive and social psychology has begun to systematically investigate the role of emotions in moral judgments and moral dilemmas. The experiments reported in this chapter aimed to examine the role of emotions in providing a framework for assessing the appropriateness of decisions made in moral dilemmas. They tested whether appeals to the emotions experienced by the protagonist when faced with a moral dilemma were considered persuasive justifications for the decision the protagonist made, and whether information about the emotion a protagonist experienced as they were faced with a moral dilemma affected how people thought about their subsequent decision.

Emotion and Moral Reasoning

In this chapter, emotion is considered from the point of view of its role as additional information, as a source of extra knowledge and a further input into the reasoning process. This is distinct from the suggestion that many moral judgments are made via emotional processes without any conscious reflection, which considers moral reasoning to be, at least some of the time, the post-hoc justification of intuitions rather than a deliberative consideration of different options (e.g., Haidt 2001;

Cushman et al., 2006). However, understanding how emotions may influence moral judgments is a key step in understanding their potential role as an input in the reasoning process.

Unconscious and Emotional Processes in Moral Judgment

Although in the past, psychological theories of moral judgment had focused on conscious reasoning (e.g. Kohlberg, 1973), contemporary evidence has suggested that much of moral judgment occurs through unconscious, often emotional judgments (e.g. Haidt, 2001; Haidt, Bjorklund & Murphy, 2000), or through dual processes of unconscious emotional judgments and controlled cognitive processes (e.g. Greene et al., 2001, 2004, 2008). As discussed briefly in Chapter 1, diverse evidence has been gathered in the past decade to support the contention that emotions and unconscious processing can influence moral judgments. People have nearly instant implicit reactions to stories of moral violations (e.g., Luo, Nakic, Wheatley, Richell, Martin, & Blair, 2006). Furthermore, greater activation of brain areas associated with specific negative emotional states such as anger and disgust, compared with activation in areas associated with executive control and goal maintenance, is a good predictor of people rejecting unfair offers to share a sum of money (e.g. Sanfey, Rilling, Aronson, Nystrom & Cohen, 2003), suggesting that those emotions are an important cause of such rejection. The same experiment found evidence that greater activation of areas associated with executive control and goal maintenance is a good predictor of people accepting unfair offers, suggesting that conscious processing is required to overcome negative affect (e.g., Sanfey et al., 2003).

Not only do people apparently sometimes make moral judgments based on unconscious and emotional reactions, but as discussed in Chapter 1, the unconscious

origins of such judgments apparently sometimes leads people to be ‘dumbfounded’, meaning that they cannot always articulate the principle underlying their moral judgment (e.g., Haidt et al., 2000; Haidt 2001; Cushman et al., 2006). The example given in Chapter 1 of siblings deciding to have incestuous sexual relations (e.g. Haidt, 2001) is one scenario that can prompt people to make strong judgments of moral wrongness without being able to rationally defend their judgment. This result suggests that their moral judgment was reached unconsciously, and was perhaps influenced by a strong negative emotional reaction to reading about incest. It also suggests that the reasons people give for their moral judgments may not be the ones that led them to those judgments.

In cases where moral judgments have unconscious origins, people do not have conscious access to the principles on which their judgments are based, and so their justifications seem not to be blueprints for moral reasoning, but rather post hoc guesses at what might justify their judgments. However, justifications, whether post hoc or not, have been taken by some theorists to be likely to be a key component of how people persuade others of their moral views (e.g. Haidt, 2001; Paxton & Greene, 2010). These theorists suggest that social moral reasoning, such as discussing with friends the reasons why certain behaviours, for example eating meat, are either right or wrong, can help to change people’s moral intuitions, which may be a major factor in changing future moral judgments.

An implication of the evidence that some moral judgments are based on unconscious processes is that factors such as emotional priming that unconsciously alter such processes should also influence moral judgments. Researchers have demonstrated that this is indeed the case. Moral judgments are affected by, for example, changes in people’s mood after watching an amusing video (e.g., Valdesolo

& DeSteno, 2006), and after anger or disgust has been induced using either disgusting videos or odours (e.g., Ugazio, Lamm & Singer, 2012; Schnall et al., 2008). As well as judgment, moral behaviour can also be affected by emotions, as demonstrated by the moral emotion of elevation, the feeling of moral uplift that accompanies witnessing someone else do a good deed for someone else. As discussed in Chapter 1, this emotion has been noted to increase nurturing behaviour in nursing mothers (Silvers & Haidt, 2008), and to prompt an inclination to emulate morally good behaviour (Algoe & Haidt, 2009; Schnall et al., 2010).

Dual-Processes in Moral Judgment

Although moral judgments are often influenced by emotions and unconscious judgments, conscious reasoning and cognitive control processes also inform moral judgments. Furthermore, differences between scenarios can differentially activate different processes. For example, indirect, impersonal harm such as pushing the switch to save the five men by turning the trolley in the trolley problem (discussed in Chapter 1) engages emotional processes to a lesser extent than personal, direct harm such as pushing the man off the footbridge to save the five men in the footbridge problem (e.g. Greene et al., 2001). Cognitive control processes are recruited to solve difficult personal moral dilemmas (e.g. Greene et al., 2004). People also consciously reason about relevant features of moral dilemmas and distinguish between features such as whether harm is inevitable or not, and whether an argument is strong or not (e.g. Moore et al., 2008; Paxton et al., 2012). Furthermore, the unconscious reactions that can be so influential in moral judgments can be suppressed when participants are explicitly instructed to provide rationally justifiable responses (e.g. Pizarro, Uhlmann & Bloom, 2003).

Hence, people may rely on immediate emotional responses to make unconscious, heuristic moral judgments, and they may deploy a cognitive mechanism to engage in deliberative reasoning to a considered moral choice. Which process is used in a given scenario may depend on the type of behaviour required by the scenario (e.g. whether or not the behaviour requires personal contact, or whether it involves an emotionally charged moral norm such as incest). As discussed in Chapter 1, it has been argued that the reason the cognitive developmental theory emphasises conscious reasoning as the measure of interest (e.g. Kohlberg, 1973), while the social intuitionist model emphasises emotional intuitions (e.g. Haidt, 2001) may be the difference in the tasks they employed experimentally (Monin, et al., 2007). Experiments testing the cognitive developmental theory tended to employ moral dilemmas balanced between different obligations, such as the Heinz dilemma discussed in Chapter 1 (e.g. Colby et al., 1983), which may have encouraged participants to engage in reasoning. Experiments testing the social intuitionist model, on the other hand, have tended to ask participants for judgments about highly emotional moral violations such as the sibling incest example discussed earlier (e.g. Haidt, 2001), and may therefore be more likely to result in emotional reactions. Therefore, the content of the scenarios researchers use to test moral judgment is likely to be influential in determining the extent to which different processes are involved in those judgments.

In this chapter, the focus on emotion is at one remove from these processes: it is on the folk psychology of the role of emotion in moral judgment. It is suggested that people *expect* other people to experience emotions when they are confronted with moral dilemmas, and so their understanding of the moral choices a protagonist made is affected by information about the emotions the protagonist experienced. This

suggestion is tested in two ways: by examining whether people create explanations that rely on emotions to justify different sorts of moral choices, and by examining whether information about different emotions primes people to read about some moral choices more quickly than others.

Personal and Impersonal Moral Dilemmas

As discussed above, different moral dilemmas seem to engage emotional processes to different extents. The personal-impersonal distinction is a special case of this difference. Personal dilemmas, (such as whether or not it is acceptable to push a man off a bridge in order to stop five men from being hit by a train) result in greater activation of areas of the brain associated with emotion than impersonal dilemmas (such as whether or not it is acceptable to push a switch to turn a train away from five people and towards one; e.g. Greene et al., 2001). This difference in emotional engagement made personal and impersonal dilemmas useful for investigating the novel research question of whether people consider the emotions of others to be relevant when making moral judgments. It allowed for testing the hypothesis that emotions would be considered a source of information to a different extent for different dilemma types.

However, the proposal that people expect other people to experience emotions when they are confronted with moral dilemmas implicates not only personal dilemmas, but also impersonal dilemmas. Although impersonal dilemmas may result in less emotional activation than personal dilemmas, aspects of both types of scenario should still be interpreted as emotional. For instance, considering the five men who would be killed in the trolley problem if the protagonist does not act should prompt an

emotional reaction. It was therefore anticipated that information about the protagonist's emotions would affect even impersonal dilemmas.

In the first of the experiments reported in this chapter, this proposal is investigated in terms of the types of justifications people provide for their decisions in moral dilemmas. As discussed above, researchers have provided evidence that justifications for moral judgments may sometimes be post-hoc explanations of unconscious judgments, and may not identify any argumentative steps considered by a moral agent. However, there also seem to be cases in which conscious moral reasoning occurs (e.g. Paxton et al., 2012). Furthermore, different theorists have argued that one of the primary ways in which moral judgments are likely to be changed is through social contact and discussion of moral principles (e.g. Haidt, 2001; Haidt & Bjorklund, 2008; Paxton & Greene, 2010). According to this perspective, when conscious moral reasoning changes minds, it is likely to be less of an internal weighting of the different arguments, and more of a social process in which people persuade one another by appeal to either argument or intuition. Thus, regardless of whether moral justifications reveal the processes behind a particular moral judgment, or are merely post hoc attempts to justify a moral judgment, they are nonetheless likely to play an important role in influencing future moral judgments. Considering the evidence that emotions influence moral judgments at an unconscious level, they are also likely to be involved in justifications.

Experiment 4

The aim of this experiment was to examine appeals to emotion in post-hoc justifications for decisions in moral dilemmas. Participants' justifications were expected to appeal to emotive factors more for the personal footbridge dilemma, in

which an individual who is on a footbridge over a runaway trolley about to run into a group of five railway workers must decide whether to push a stranger off the bridge to cause the trolley to stop, compared to the corresponding impersonal trolley version, in which an individual who is at the wheel of a runaway trolley about to run into a group of five railway workers must decide whether to hit a switch that will cause the trolley to proceed to the right, where it will run into a single worker. However, as discussed above, the requirements of moral decision tasks may be crucial to whether or not emotions are expressed, and so this experiment also tested whether participants could provide justifications based on emotions if prompted to do so for impersonal dilemmas, and conversely, whether they could provide justifications based on reason if prompted to do so for personal dilemmas.

Participants were asked to imagine themselves in the situation depicted by the trolley³ or footbridge dilemmas, and to make the decision of whether or not they would push the switch (in the trolley problem), or push the man (in the footbridge problem). They were then given the following context for their justification task:

‘Later that evening at home, you go over and over the situation in your mind. Your friends and family rally around and do everything they can to help you. When your closest friend arrives, you talk things over in private. You recollect as best you can how the situation arose and how it unfolded. But you are distressed to see that your friend appears genuinely shocked at the decision you made, although they try to hide it.’

The participants’ task was to complete the following sentence:

³ The word ‘trolley’ was changed to ‘train’ for these dilemmas, as ‘train’ is used more commonly by the Irish population.

‘You decide to try your best to describe in detail to your friend your personal experience in the moments before you made your decision. You say, “I knew I had to make a decision fast. This is what I experienced in those seconds...”’

Justifications given by participants in this ‘non-directed’ baseline condition were compared to those given by participants who were given an ‘emotion-directed’ sentence completion task:

You decide to try your best to describe in detail to your friend *what feelings were in your heart* in the moments before you made your decision. You say, “I knew I had to make a decision fast. This is what I experienced in those seconds, *the feelings and emotions I had...*”

and to those given by participants who were given a ‘reason-directed’ task:

You decide to try your best to describe in detail to your friend *what thoughts were in your mind* in the moments before you made your decision. You say, “I knew I had to make a decision fast. This is what I experienced in those seconds, *the reasons and thoughts I had...*”

It was expected that participants could suppress emotional reactions and access reasoned justifications even in emotionally charged personal dilemmas when they

were directed to provide a reasoned justification. It was also expected that they could access emotional justifications even in impersonal dilemmas.

Method

Participants

180 volunteers, 86 women and 93 men (and one person did not indicate their gender), were recruited from the campus of Trinity College Dublin. Their ages ranged from 18 to 55 years, with an average age of 25 years. They were assigned at random to one of three groups: non-directed ($n = 60$), reason-directed ($n = 60$), and emotion-directed ($n = 60$).

Design and Materials

Participants in the three groups (non-directed, emotion-directed, or reason-directed) each received two dilemmas: a personal and impersonal dilemma, the footbridge and trolley problems. Half were given the personal problem first, and the other half the impersonal problem first, and no order effects were observed on the decision to act or not act⁴. They carried out a decision task first: ‘Would you hit the switch/push the man?’ They were asked to circle their answer: ‘(a) I would hit the switch/push the man. (b) I would not hit the switch/push the man.’ Then they carried out the justification task: they were asked to provide a justification for their decision as if to a friend, elicited by the prompts described earlier, directed towards either a reason-based justification or an emotion-based justification, or non-directed.

⁴ For the train problem, 84% of those who completed the train problem first and 82% of those who completed the footbridge problem first judged that they would push the switch, $\chi^2(1, n = 360) = .160, p = .689$. For the footbridge problem, 10% of those who completed the train problem first and 17% of those who completed the footbridge problem first judged that they would push the man, $\chi^2(1, n = 360) = 1.731, p = .188$.

Procedure

Participants were recruited individually and were provided with a booklet containing, on a separate page each, instructions, each problem, its decision task and justification task, and a debriefing paragraph (see Appendix 3.1 for the problems and their tasks; see Appendix 3.2 for the instructions and debriefing). The instructions informed them that they should answer the problems in the order they were given and that they should not change any answers.

Results

Decision Task

Participants judged that they would act rather than not act in the impersonal dilemma, binomial, $z = 8.87, p < .001, r = .66$, and not act rather than act in the personal one, binomial, $z = 9.76, p < .001, r = .73$. They were more likely to endorse pushing the switch in the impersonal dilemma than pushing the man in the personal dilemmas, with 83% vs. 13% overall 'yes' responses respectively, $\chi^2 (1, n = 360) = 176.596, p < .001, \phi = -.700$. The result is consistent with previous research and it occurred in each group: non-directed, $\chi^2 (1, n = 120) = 67.962, p < .001, \phi = -0.753$, emotion-directed, $\chi^2 (1, n = 120) = 50.714, p < .001, \phi = -0.650$, and reason-directed, $\chi^2 (1, n = 120) = 58.865, p < .001, \phi = -0.700$, as Figure 3.1 shows.

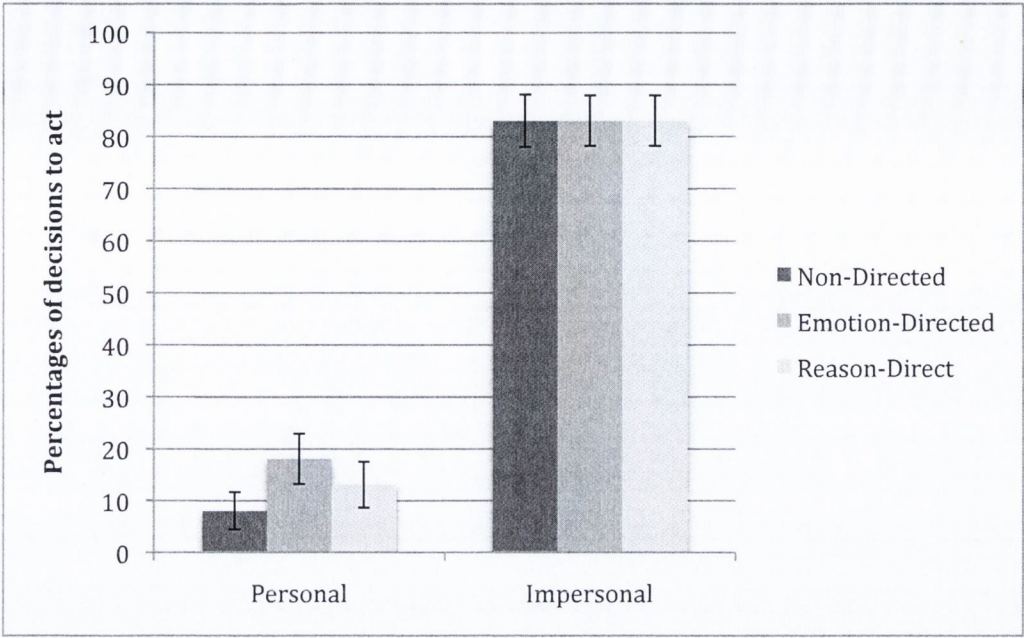


Figure 3.1 Percentages of decisions to act in Experiment 4 for personal and impersonal dilemmas in the non-directed, emotion-directed and reason-directed conditions. Error bars are standard error of the mean.

Justifications with Emotive Content

Justifications were categorized as emotive if they contained emotive content, that is, (a) a direct reference to emotions, for example, ‘shock’, (b) an indirect reference to emotions, for example, ‘the choice was horrid’, or (c) a reference to other people’s emotions, for example, the grieving of the workmen’s families⁵. Emotive content of the justification was coded regardless of the type of reason it contained; that is to say, an emotive justification could make reference to any reason, including a

⁵ All justifications were coded by the first author and an independent rater and agreement was 100%. A second independent rater coded a random selection of 20% of responses, and agreement between the independent raters was 83% for the trolley and 86% for the footbridge problem, Cohen’s Kappa = .557, $p < .005$, and Kappa = .706, $p < .001$ respectively. Disagreements were resolved by discussion.

‘utilitarian’ reason, such as, *“I experienced complete terror and confusion but I reasoned that one life lost is better than 5 lives lost”*, or a ‘direct killing’ reason, such as, *“Shock, anxiety and stress. Being the one to actually cause a death like that I couldn’t do. Physically pushing him would be too much, I wouldn’t be able to live with myself”*. Conversely, a non-emotive justification could make reference to a utilitarian reason, for example, *“I had to choose between 2 bad outcomes, one involved less loss of life. So I chose it”*, or to a ‘direct killing’ reason, for example, *“I couldn’t be directly responsible i.e. push someone on the tracks, it would be like murdering someone.”* More information on the types of reasons is provided below.

Justifications with emotive content comprised about one third of all justifications and they were provided as often for the personal and impersonal dilemmas overall, 34% vs. 32%, $\chi^2(1, n = 360) = 0.113, p = .737$. Three comparisons tested the prediction that emotive justifications would vary according to the instructions given, and these comparisons are illustrated in Figure 3.2. Emotive justifications were provided more often for the personal than the impersonal dilemma in the baseline non-directed condition, $\chi^2(1, n = 120) = 6.604, p < .05, \phi = 0.235$. This difference was eliminated in the emotion-directed condition, $\chi^2(1, n = 120) = 0.034, p = .855$. In the reason-directed condition, not only was the non-directed tendency for more justifications to be emotive in the personal dilemma eliminated, it was in fact reversed, with somewhat more justifications with emotive content being provided for the impersonal dilemma than for the personal dilemma, $\chi^2(1, n = 120) = 3.871, p < .05, \phi = -0.180$.

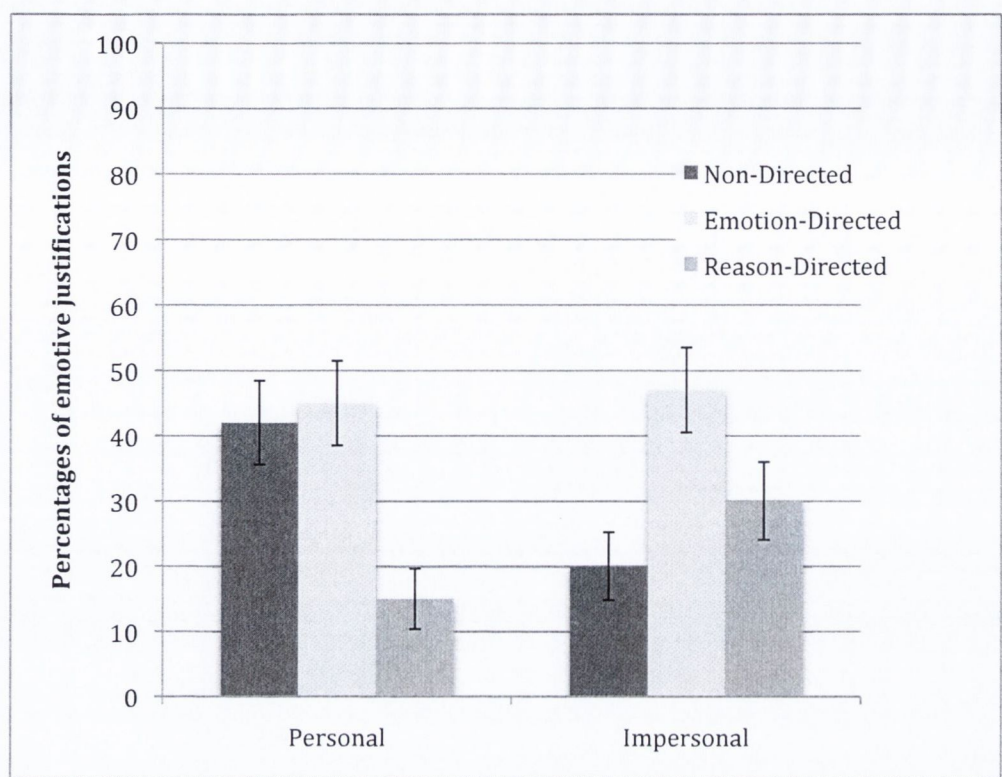


Figure 3.2 Percentages of justifications featuring emotive content for personal and impersonal dilemmas in the non-directed, emotion-directed and reason-directed conditions in Experiment 4. Error bars are standard error of the mean.

Types of Reasons

Participants produced several different types of reasons, which were categorized into 8 different types⁶. The categories are summarised here, and the percentage of responses in each of the main categories are displayed for each dilemma in Figure 3.3: (i) Utilitarian justifications that referred to saving multiple lives rather

⁶ All reasons were coded by the first author and an independent rater and agreement was 69% overall, 79% for the trolley and 59% for the footbridge problem, Cohen's Kappa =.620, $p < .001$ and .501, $p < .001$ respectively. A second independent rater coded a random selection of 20% of responses, and agreement between the independent raters was 72% for the trolley, and 42% for the footbridge problem, Cohen's Kappa =.559 and .311 respectively. Disagreements were resolved by discussion.

than just one life, were the most common type of justification overall (37%). An example is: “*I had to choose between 2 bad outcomes, one involved less loss of life. So I chose it*”. They were generated more often for the impersonal than the personal dilemma overall, 64% vs. 11%, $\chi^2(1, n = 360) = 109.555, p < .001, \phi = -0.552$; and within each group: non-directed, 57% vs. 7%; $\chi^2(1, n = 120) = 34.660, p < .001, \phi = -0.537$; reason-directed, 70% vs. 12%; $\chi^2(1, n = 120) = 42.254, p < .001, \phi = -0.593$; and emotion-directed, 65% vs. 13%; $\chi^2(1, n = 120) = 33.611, p < .001, \phi = -0.529$.

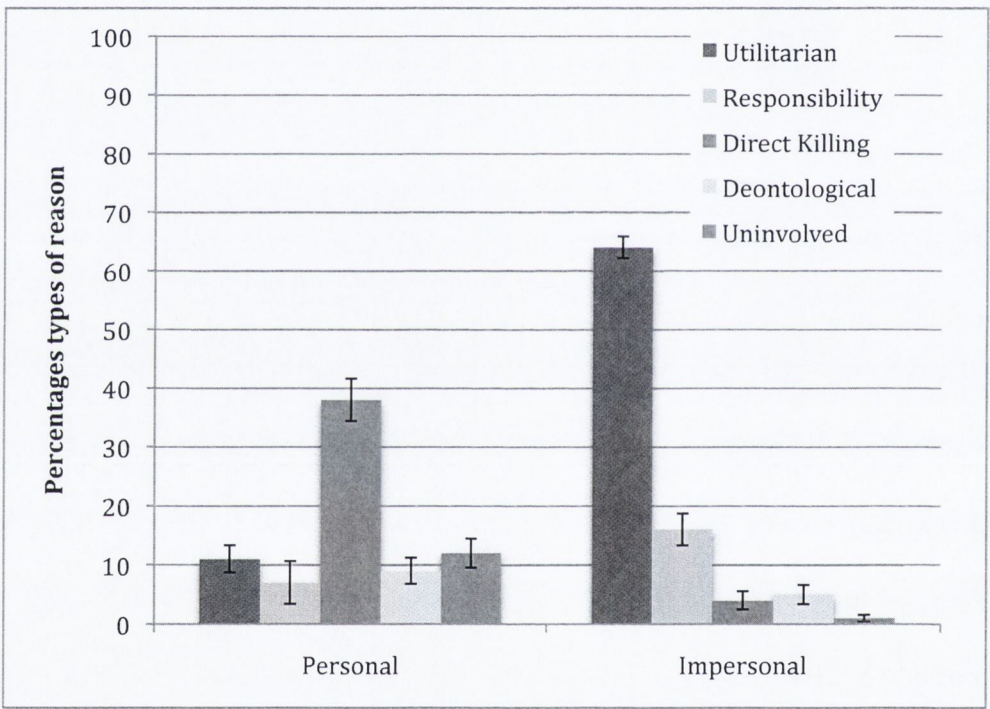


Figure 3.3 Percentages of main types of reasons for personal and impersonal dilemmas in Experiment 4. Error bars are standard error of the mean.

Key: Utilitarian: Saving multiple lives rather than one; Direct killing: Directness of action; Responsibility: Own personal responsibility; Uninvolved: Not involved in situation; Deontological: Action/ inaction is right or wrong

‘Direct killing’ justifications referring to the death being caused directly by the action, were the next most common justification overall (21%). An example is, “*If I hit the switch I would be intentionally killing the man. If I did nothing, it would be horrible but I could shout at the men to try and get them to move but I would not be killing anyone*”. They were generated more often for the personal than the impersonal dilemmas overall, 38% vs. 4%, $\chi^2(1, 360) = 60.044, p < .001, \phi = .408$, and in each condition: non-directed, 35% vs. 7%; $\chi^2(1, n = 120) = 14.602, p < .001, \phi = 0.349$; reason-directed, 47% vs. 5%; $\chi^2(1, n = 120) = 27.184, p < .001, \phi = 0.476$; and emotion-directed, 32% vs. 2%; $\chi^2(1, n = 120) = 19.440, p < .001, \phi = 0.402$.

Other categories of justifications, which did not appear in more than 11% of justifications overall and so are not analysed further, are summarised here. Some participants referred to (iii) the individual’s personal responsibility in the situation (11%), e.g., “*I don’t want to be the cause of the death of another individual. It’s not my fault that those five people died but it would have been my fault if that one person died*”. (iv) Some participants referred to whether individuals considered themselves involved in the situation (6%), e.g., “*I would do nothing, so that the event would unfold without my participation and so I would not feel guilty for pushing the man to his death as I decided not to take part in events*”. (v) Some participants provided deontological justifications about whether the action was right or wrong, or the impossibility of putting a value on someone’s life (7%), e.g., “*I couldn’t value somebody’s life more than others. It’s not right to say one life is less important than five*”. The remainder of responses comprised (vi) ‘dumbfounded’ responses or ‘non-explanations’ (4%), e.g., “*I felt sorry for the 5 workmen, but couldn’t do anything to save them*”, (vii) expressions of doubt about the situation (5%), e.g., “*I doubted my ability to push the man over the bridge. I doubted that this would stop the train in any*

case”, or (viii) the expression of an emotion only rather than a justification (7%), e.g., “*Uncertainty, doubt, scary, frightened, empowerment, apprehension*”.

Discussion

Participants naturally provided justifications with emotive content more often for personal dilemmas than they did for impersonal dilemmas, that is, in the baseline non-directed condition. This result is consistent with previous research indicating that personal dilemmas engage emotional processes to a greater extent than impersonal dilemmas (e.g. Greene et al., 2001; Valdesolo & De Steno, 2006).

However, this difference can be eliminated: when participants were prompted to tell a friend about their feelings, “This is what I experienced in those seconds, *the feelings and emotions I had...*” they produced as many emotive justifications for the impersonal dilemma as for the personal one. When they were prompted to tell a friend about their thoughts, “This is what I experienced in those seconds, *the reasons and thoughts I had...*”, they produced somewhat *more* emotive justifications for the impersonal dilemma than for the personal one, suggesting that suppression of emotive thought is more pronounced for the personal than the impersonal dilemma. The finding is consistent with the idea that participants can access emotion-based justifications for impersonal dilemmas just as they can for personal dilemmas, and likewise, they can access reason-based justifications for personal dilemmas just as they can for impersonal dilemmas.

These results are congruent with certain dual-process accounts of moral reasoning, according to which emotions and reasons are likely to be drawn out by differences between moral tasks (e.g. Greene et al., 2001; Paxton & Greene, 2012; Bucciarelli et al., 2008). These accounts suggest that certain moral scenarios, such as

the footbridge problem, are more likely than other scenarios to engage emotional processing due to a difference in the directness of contact involved in the scenario. They view emotions and conscious, controlled processing as independent processes, which are each capable of influencing moral judgments. The results of Experiment 4 cohere with this perspective in that participants appear to expect that emotions and reasons will both be considered relevant explanations for their behaviour in these moral dilemmas.

Various researchers have demonstrated that moral judgments vary depending on whether participants are instructed to judge rationally or intuitively (e.g. Pizarro et al., 2003), whether they have been primed by hypnosis, or odour (e.g. Wheatley & Haidt, 2005; Schnall et al., 2008), and whether they have been primed by different positive emotions (e.g. Valdesolo & DeSteno, 2006; Strohminger et al., 2011), among other factors. Those sceptical of the influence or frequency of reasoning in moral scenarios (e.g. Haidt, 2001; Haidt & Bjorklund, 2008) might interpret the fact that justifications change depending on the instructions participants are given to mean that the real reasons for people's moral decisions are inaccessible to consciousness. The results of Experiment 4 do not rule out this interpretation; it is possible that participants do not consider both emotional and reasoned factors when they make moral judgments, and only notice both when they are asked to after the judgment is made. On the other hand, these results may highlight a fundamental feature of judgment in a dilemmatic context: dilemmas feature multiple compelling, and sometimes opposing, reasons for different behaviours which participants can access, and instructions like those used in the justification task may differentially highlight these different reasons.

While this experiment does suggest that participants' reactions to moral dilemmas are somewhat malleable given explicit instruction, the baseline non-directed condition supported previous research in indicating that personal moral dilemmas seem to engage emotional processes more than impersonal dilemmas (e.g. Greene et al., 2001; Valdesolo & DeSteno, 2006). This result appears to support the distinction between such dilemmas in terms of emotional engagement.

,The results of this experiment provided evidence that participants understand both personal and impersonal dilemmas as emotional, and that they can use both emotional and reasoned language in their justifications, which may be a key factor in moral persuasion (e.g. Haidt, 2001; Paxton & Greene, 2010). The next experiment asked participants to make judgments of moral actions and inactions while considering information about the emotions felt by the person in the scenario. This experiment therefore provided a test of whether the kind of information participants provided in Experiment 4 when they were asked to direct their justifications towards emotions would be effective in influencing a third-party's thought processes and moral judgments.

Experiment 5

The second experiment reported in this chapter explored the role of emotional information in moral judgment from a different perspective to that taken in Experiment 4. Instead of being asked to provide emotive justifications for their *own* moral decisions, participants were asked to make judgments of the decisions of actors in moral dilemmas after reading information about the emotions the actors experienced, and the decisions they made. The experiments therefore tested the role of

emotional information in moral judgment from converging perspectives: from that of the person in the dilemma, and that of someone judging the person in the dilemma.

The aim of the experiment was to examine whether descriptions of a protagonist's emotions would affect participants' reading of the protagonist's decision and their judgments about its morality, for both personal and impersonal dilemmas. The dilemmas used were adapted from past literature on the personal-impersonal distinction (Greene et al., 2004), and moved beyond the prototypical trolley and footbridge dilemmas to a diverse range of other impersonal and personal scenarios. Participants' judgments of the actors' decisions were gathered, as well as latency data for the amount of time taken to read about the emotions experienced, the decision made, and the appropriateness of the decision.

As discussed previously, personal dilemmas such as the footbridge problem have been identified as more emotive than impersonal ones, and so it was expected that participants would read sentences about the emotions experienced by a protagonist in these dilemmas more quickly for personal than for impersonal dilemmas. However, the strength of emotion experienced by the protagonists of the dilemmas was varied, e.g., 'Facing this decision, X feels some strong/mild emotions'. Given the evidence that reading personal rather than impersonal dilemmas leads to greater activation of areas of the brain associated with emotion, it was expected that participants would read that a protagonist experienced strong emotions faster than mild emotions for personal dilemmas, and mild emotions faster than strong emotions for impersonal dilemmas.

The primary novel proposal investigated in this experiment concerned the effect of such emotion descriptions on the subsequent reading of the protagonist's decision. The length of time participants took to read a sentence describing the

protagonist's decision after they had been 'primed' by a sentence describing the protagonist's emotion as mild or strong was recorded. The decision made by the protagonist was also varied (i.e. the person in the dilemma decided to act or not to act). The logic of priming studies is that the provision of some information, in this case, a description of an emotion experienced by the actor, enables a participant to make an inference. For example, they may infer that the protagonist will act, and so when they subsequently read a sentence describing that the protagonist did in fact act, they should be able to read this sentence quickly (i.e., they are 'primed' to read it) because the information has already been mentally represented (e.g., Frosch & Byrne 2012).

It was expected that participants would construct varying representations of what the actors would do based on information about the type of dilemma described, and the type of emotion described, and that this would be revealed in the latencies to read the decision the actor made. A simple view of how this could occur is that strong emotion would inhibit the decision to act, whereas the absence of strong emotion, or the overriding of it, would allow a reasoned decision to act. It is suggested here instead that, since people do understand impersonal dilemmas as somewhat emotional (as evidenced by Experiment 4), it is not the absence of emotions in an impersonal dilemma that allows a reasoned decision to act, for example, to hit the switch, and that even in the presence of strong emotions, participants should expect a protagonist to act in an impersonal dilemma. Emotions contribute to the decision about whether to act or not act, but it is suggested that they do not necessarily inhibit action, and that in some cases they may even facilitate it. The variation of the dilemma (personal and impersonal) and the emotion felt (mild and strong) led to the decision being evaluated in four situations by participants:

- (a) Impersonal dilemma and strong emotion. If people expect others to experience only some emotion in impersonal dilemmas, e.g., hitting the switch in the trolley problem, then being told that someone experienced *strong* emotions is *incongruent* with their expectations. It was predicted that the information should affect the time it took participants to read about the protagonist's decision. For impersonal dilemmas the typical decision is to act. On a simple view, if strong emotions inhibit the decision to act, then people should expect the protagonist to decide not to act, and so they should read more quickly that he or she did not act than that he or she did act. Alternatively, on the view proposed here, it is not the absence of emotions in an impersonal dilemma that allows a reasoned decision to act, and participants should continue to expect the protagonist to decide to act based on additional factors such as the utilitarian cognitive calculation of greater good. The protagonist's experience of strong emotion may even be interpreted in such a way (e.g., strong emotion about the possible deaths of the five workmen) as to enhance the expectation that the protagonist will act.
- (b) Impersonal dilemma and mild emotion. If people expect others to experience at least a small amount of emotion in impersonal dilemmas, then being told that someone experienced mild emotions is *congruent* with their expectations. Hence, the information about a congruent emotion is redundant and so it should have no effects on subsequent reading times of the decision to act or not to act.
- (c) Personal dilemma and strong emotion. If people expect others to experience strong emotions in direct dilemmas, then being told that

someone experienced strong emotions is *congruent* with their expectations in this situation. Information about a congruent emotion is redundant information and so it should have no effects on subsequent reading times of the decision to act or not to act.

- (d) Personal dilemma and mild emotion. If people expect others to experience strong emotions in personal dilemmas, then being told that someone experienced *mild* emotions is *incongruent* with their expectations in this situation. It was predicted that the information should affect the time it takes participants to read about the protagonist's decision. For personal dilemmas the typical decision is not to act, and previous research has implied that strong emotions restrain people from acting in this situation. In fact, hearing that someone experiences only mild emotion at the prospect of, for example, pushing the man onto the tracks, may flag to listeners that there is something not quite normal about the protagonist and hence they may expect that he or she will behave in a way that is not quite normal. Hence it is suggested that the information that the protagonist experienced only mild emotions will lead participants to infer that the protagonist will act, and thus it will prime them to read more quickly that he acted than that he did not act.

Latencies to make the appropriateness judgment should not be primed by the variation of emotions, since these are decisions the participants make rather than information they integrate after having been primed. Emotional information may be integrated into people's judgments of the appropriateness of the actions of the protagonists; however, these judgments are not necessarily related to the emotion

prime, and may also be based on the behaviour in the dilemma, rather than the emotion.

Method

Participants

The participants were 23 students from Trinity College Dublin, 18 women and 5 men, whose ages ranged from 18 to 46 years, with an average age of 23 years, who participated in return for course credits.

Design and Materials

Participants acted as their own controls in a 2 (dilemma: personal versus impersonal) by 2 (emotion: strong versus mild) by 2 (decision: act versus not act) within-participants design. They received 8 moral dilemmas and 2 non-moral filler stories (see Appendix 3.3). The content of the dilemmas concerned harm violations such as deaths and injuries, and fairness violations such as stealing and lying. The dilemmas were adapted from Greene et al. (2004) to ensure that the primary difference between matched personal and impersonal pairs was the directness of contact of the protagonist's action in relation to another person. The syntactic structure was also modified to ensure the dilemmas were matched for features such as the number of sentences, and the length of the target sentences (see Appendix 3.3). The 8 contents were assigned to the 8 conditions in a Latin-square design, and presented in a different randomised order for each participant.

The participants' task was to judge the moral appropriateness of the protagonist's decision, in response to the question 'Was X's decision appropriate?' Four responses were recorded: (a) time to read the emotion sentence, e.g., 'Facing this

choice, X feels some strong/mild emotions’, (b) time to read the decision sentence, ‘X decides he will/won’t do it’, (c) their ‘yes’ or ‘no’ response to the question, ‘Was X’s decision appropriate?’ and (d) time to respond to the question, ‘Was X’s decision appropriate?’

Procedure

Participants were tested individually. The materials were presented using SuperLab software running on an Apple MacBook laptop. Participants were instructed that the task was not a test of intelligence and that the aim was to examine the responses made by most people (see Appendix 3.4 for full consent and debriefing materials). They were asked to read each story carefully, to take as much time as they required, and to answer the question before moving on to the next story.

Each dilemma was presented one sentence at a time on screen; participants progressed to the next sentence by pressing the spacebar key, labelled ‘continue’. The time they took to read the target sentences was recorded from key press to key press. They responded to the appropriateness question by pressing the keys labelled ‘yes’ and ‘no’ (the ‘y’ and ‘h’ keys respectively). Participants first completed 3 practice trials using materials distinct from those in the experiment, to familiarize themselves with the set-up. They involved all of the same steps as the dilemmas that would actually count as part of the experiment; after completing the practice trials, it was explained to participants that they could press the spacebar to continue to the main experiment. Finally, participants were debriefed and thanked (consent and debriefing forms can be seen in Appendix 3.4). Participants took on average about 12 minutes to complete the experiment.

Results

Outliers

Based on previous studies (e.g., Santamaría, Espino & Byrne, 2005; Frosch & Byrne, 2012) outliers were identified as any latency that was greater than the individual's mean latency plus 2 standard deviations, or less than their mean latency divided by 3. These criteria were applied to all three latency measures: time taken to read the emotion sentence, time taken to read the decision sentence, and time taken to make the judgment of appropriateness. The criteria were chosen after considering other choices. The chosen upper bound of the mean plus 2 standard deviations was the most conservative of those considered (the mean plus 2.5 and 3 standard deviations were also considered), as the less conservative upper bounds did not result in any of the longer latencies for emotion, decision, or appropriateness latencies being replaced. The chosen lower bound of the mean latency divided by 3 was also the most conservative considered (in terms of being the closest to the mean), and although it did not result in any of the quicker latencies to read the emotion sentence being replaced, the other criteria tested also failed to do so (these were the mean divided by 2.5 and by 2).

Outliers were replaced by the individual's mean latency. These boundaries led to the replacement of less than 5% of latencies for both emotion and decision latencies; for appropriateness latencies, 17% of latencies were removed. Due to the majority of the conditions for each latency analysis failing the Kolmogorov-Smirnov test of normality, the data was transformed to natural logs for the following analyses of the latencies to read the emotion and decision sentences, and to make the appropriateness decision.

Latencies to Read Emotion Sentences

A 2 (dilemma: personal vs. impersonal) x 2 (emotion: strong vs. mild) ANOVA, on the log transformed latencies⁷ to read the emotion sentence, e.g. ‘Facing this decision, X experienced some strong/mild emotions’ with outliers (4.3%) replaced, showed a main effect of dilemma, $F(1, 22) = 11.760, p < .005, \eta_p^2 = .348$, due to participants reading the emotion sentence faster for personal than impersonal dilemmas, as Figure 3.4 shows. There was no main effect of emotion, and the two variables did not interact, $F < 1$ in both cases.

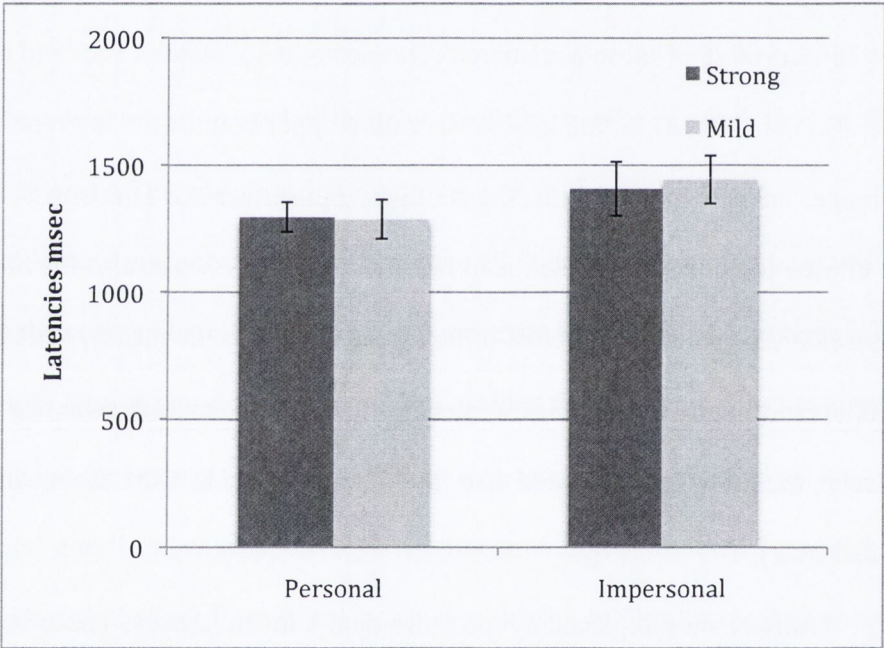


Figure 3.4 Mean latencies for the emotion sentences in Experiment 5 for personal and impersonal dilemmas. Error bars are standard error of the mean.

⁷ After the natural log transformation, emotion latencies were normal according to Kolmogorov-Smirnov tests for three of the four conditions, but not for impersonal dilemmas with mild emotions. However, a non-parametric analysis demonstrated the same results for main effects and the two-way interaction using Wilcoxon Signed Ranks tests.

Latencies to Read Decision Sentences

A 2 (dilemma: personal vs. impersonal) x 2 (emotion: strong vs. mild) by 2 (decision: act vs. not act) repeated measures ANOVA, on the log transformed latencies to read the decision sentence⁸, e.g. 'He decides he will/won't do it', with outliers (4.9%) replaced, showed no main effects of dilemma, $F < 1$, or emotion, $F(1, 22) = 1.475, p = .237$, but a significant main effect of decision, $F(1, 22) = 6.264, p < .05$) due to participants reading the decision to act faster than the decision not to act. Emotion did not interact with dilemma, or decision, and dilemma and decision did not interact, $F < 1$ in each case. However, the three variables interacted, $F(1, 22) = 7.966, p < .05, \eta_p^2 = .266$.

This three-way interaction was decomposed using t-tests. Four t-tests were carried out to test the effects of acting and not acting on reading times, for each combination of emotion strength and dilemma type, with a Bonferroni alpha of $p = .0125$. For impersonal dilemmas such as pushing the switch to move the train to hit the one man, rather than the five, strong emotions primed participants to read quickly that the protagonist decided to act, e.g., to push the switch, more quickly compared to the decision not to act, $t(22) = 3.456, p = .002, d = -0.73$; there was no difference between the decision to act and not to act for mild emotions, $t(22) = 0.186, p = .854$. For personal dilemmas, such as pushing the man from the footbridge, there was a marginal effect in which mild emotions primed participants to read quickly that the protagonist decided to act, e.g., to push the man, more quickly compared to the

⁸ After natural log transformations, 6 of the 8 conditions were normal according to Kolmogorov-Smirnov tests, but not the decision latencies for personal dilemmas with mild emotions when the actor decided not to act, and latencies for personal dilemmas with strong emotions when the actor decided not to act. However, again a non-parametric analysis using Wilcoxon Signed Ranks tests revealed the same results for both main effects and two-way interactions.

decision not to act, $t(22) = 2.636, p = .015, d = -0.58$; strong emotions had no effect, $t(22) = .04, p = .969$. These results are illustrated in Figure 3.5.

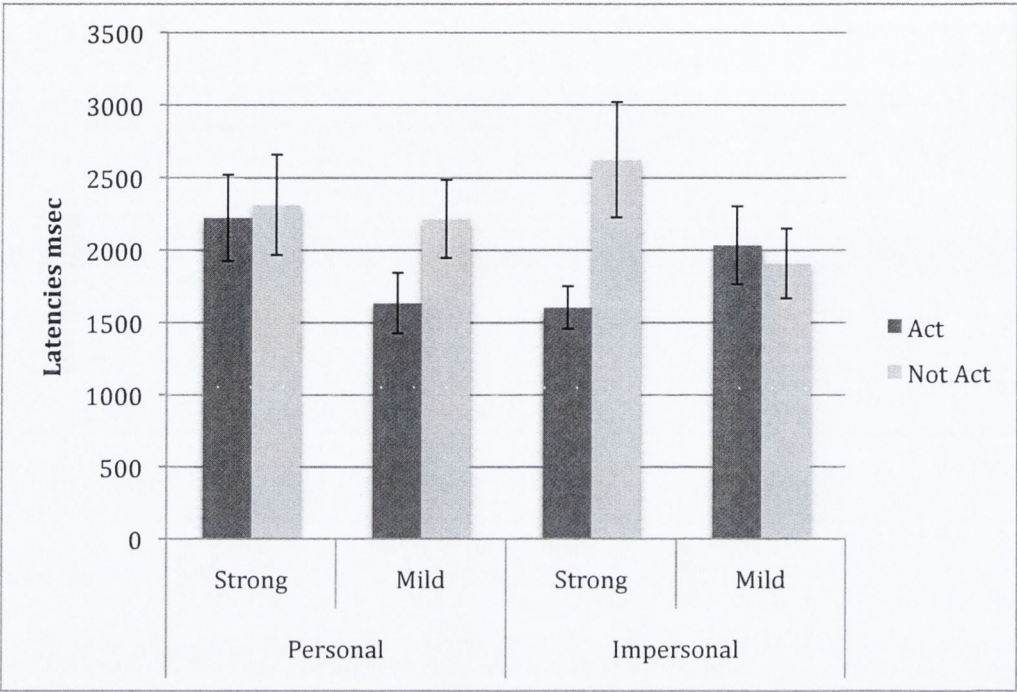


Figure 3.5 Mean latencies for the decision sentences in Experiment 5 for personal and impersonal dilemmas for the decision to act and not to act, prefaced by strong and mild emotions. Error bars are standard error of the mean.

Judgments of Appropriateness

The percentages of judgments that the protagonist’s decision was appropriate in each category of dilemma, emotion and decision are displayed in Figure 3.6. Wilcoxon’s tests on participants’ yes/no judgments about whether the decision was appropriate showed there was no overall effect of dilemma, $z = -.465, p = .642$, emotion, $z = -1.156, p = .248$, or decision, $z = -.464, p = .642$. Friedman’s tests showed emotion did not interact with dilemma, $\chi^2(3, n = 23) = 2.137, p = .544$, or

decision, $\chi^2(3, n = 23) = 2.255, p = .521$, and the interaction of dilemma with decision was also not significant, $\chi^2(3, n = 23) = 4.660, p = .198$.

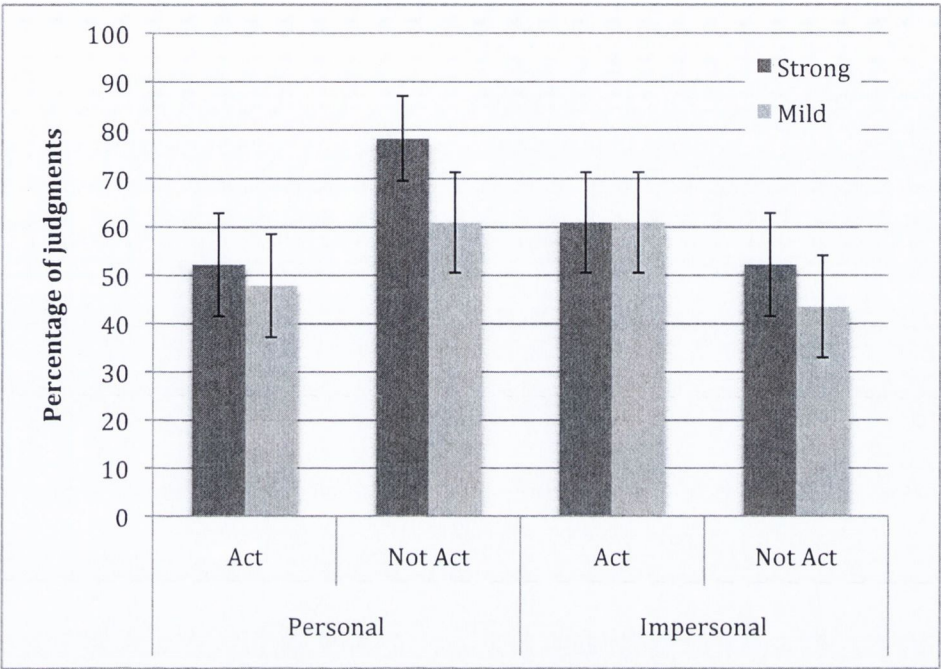


Figure 3.6 Percentages of judgments that the decision was appropriate in Experiment 5. Error bars are standard error of the mean.

Latencies to judge appropriateness

An ANOVA of the same design as that used to analyse latencies to read the decision sentences on the latencies for judgments⁹ (yes and no responses) and with outliers (17%) replaced, showed the same results: no main effects of dilemma, $F(1, 22) = 1.160, p = .293$, emotion, $F(1, 22) = 1.367, p = .255$, or decision, $F < 1$. Emotion did not interact with dilemma, $F < 1$, or decision, $F(1, 22) = 1.303, p = .266$,

⁹ After natural log transformations, appropriateness latencies were normal according to Kolmogorov-Smirnov tests for 5 of the 8 comparisons but not for impersonal dilemmas with mild emotions when the actor decided to act, personal dilemmas with mild emotions when the actor decided to act, and personal dilemmas with mild emotions when the actor decided not to act. Again however, a non-parametric analysis of the data using Wilcoxon Signed Rank tests revealed the same results for main effects and two-way interactions as the parametric analysis.

decision and dilemma did not interact, $F(1, 22) = 3.432, p = .077$, and the three variables did not interact, $F < 1$. These results are illustrated in Figure 3.7. As more than 5% of outliers were replaced in this analysis, a second analysis was carried out on the dataset, without any outliers being replaced. Again, there were no main effects of dilemma, $F < 1$, emotion, $F(1, 22) = 1.304, p = .266$, or decision, $F < 1$. Emotion did not interact with dilemma, $F(1, 22) = 1.694, p = .207$, or decision, $F(1, 22) = 1.019, p = .324$, decision and dilemma did not interact, $F < 1$, and the three variables did not interact, $F < 1$.

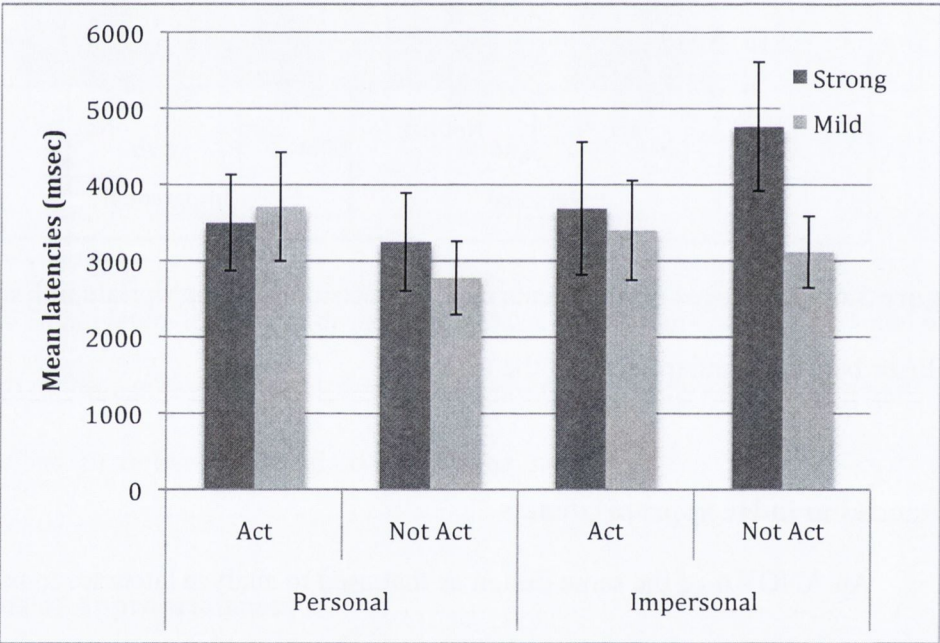


Figure 3.7 Mean latencies for judgments about the appropriateness of the decision in Experiment 5. Error bars are standard error of the mean.

Discussion

Experiment 5 showed that participants read sentences about the emotions experienced by a protagonist faster for personal dilemmas, than for impersonal dilemmas. This result is congruent with previous research indicating that personal

dilemmas are more emotive than impersonal ones, and provided validation for the adapted materials of this experiment by indicating that they follow the usual pattern in terms of emotional engagement (e.g. Greene et al., 2001; Valdesolo & DeSteno, 2006). The experiment also provided the novel result that strong emotions prime people to read quickly that the protagonist decided to act for impersonal dilemmas, such as hitting the switch on the runaway train: the result suggests that people expect others who experience strong emotions to act in such circumstances. The result also suggests that even impersonal moral dilemmas are affected by emotions, and that personal and impersonal dilemmas share some emotional factors (e.g., Nakamura 2013). It lends some support to the suggestion that it is not the absence of strong emotions in an impersonal dilemma that allows a reasoned decision to act, and that even in the presence of strong emotions, participants expect a protagonist to act in an impersonal dilemma. The experiment also shows that mild emotions prime people to read quickly that the protagonist decided to act for personal dilemmas, such as pushing the man onto the tracks in the footbridge dilemma, and the result provides some support for the idea that people may consider others who do *not* experience strong emotions in such circumstances to be responding atypically, and thus expect them to act atypically. This is consistent with research indicating that people with damage in areas of the brain associated with social emotion are more likely to approve of reasoned decisions in personal moral dilemmas, such as pushing the man onto the tracks in the footbridge problem (e.g. Koenigs et al., 2007; Ciaramelli et al., 2007).

The results suggest that in impersonal dilemmas, people reason to a decision to act even in the face of strong emotion. They provide some support for the idea that emotions can be an additional input to a process of reason, rather than a replacement

of reason. Strong emotions in an impersonal dilemma – for example, repugnance at the deaths of the five railway workers – may even amplify the decision to act.

These results also suggest that people understand emotions to be an important component of moral action, for both personal and impersonal dilemmas, but that they distinguish what emotion types are appropriate for different decision types. It is therefore suggested that the emotions experienced by the protagonist may modulate the alternatives that people consider when they understand their moral decisions. For example, reading about someone experiencing only mild emotions in personal dilemmas may strike people as unusual or callous, which may lead them to think about the person carrying out an action.

Information about strong versus mild emotions did not differentially affect participants' judgments of the appropriateness of the decision or their latency to make their judgment of its appropriateness. As can be seen from Figure 3.6, there was quite high acceptance of the appropriateness of the decision in every case, even for action in personal dilemmas, which may indicate that overall, information about the protagonist's emotions increased the acceptability of any decisions in these dilemmas.

This experiment provides some support for theories that emphasise a role for emotions in moral judgment (e.g. Haidt, 2001; Greene et al., 2004; Bucciarelli et al., 2008). Participants in the experiment appeared to understand emotions to have some role in moral judgment, and more so for certain dilemmas, that is, personal dilemmas, that are known to be more emotional than others. However, the experiment suggests a more complex view than has thus far been available, and indicates that information about emotions experienced by protagonists may prompt people to generate inferences about what actions the protagonists are likely to carry out. Emotional

information is implicated in both personal and impersonal dilemmas, indicating, as in Experiment 4, that aspects of both types of dilemma are understood to be emotional.

General Discussion

People expect others to experience emotions when they are confronted with difficult moral dilemmas. It is well known that some sorts of moral dilemmas evoke emotions more than others (e.g. Greene et al., 2001; Valdesolo & DeSteno, 2006; Bucciarelli et al 2008). The findings of Experiments 4 and 5 are consistent with this observation: participants naturally provide more emotive justifications for personal than impersonal dilemmas, and they read sentences about the emotions experienced by a protagonist faster for personal dilemmas than impersonal dilemmas.

However, the experiments also provided novel results suggesting a more complex integration of emotion and reasoning. People understand that even impersonal dilemmas contain emotional content. When people were asked to explain the feelings in their heart when in a moral dilemma, there was no difference between the dilemma types in the emotive justifications provided. They also produced fewer emotive justifications for the personal dilemma than for the impersonal one when they were prompted to justify their decision by telling a friend about their thoughts. The result suggests that not only can participants override the tendency to be emotive about personal dilemmas and access reason-based justifications for them, but they can also access emotion-based justifications for impersonal dilemmas. This experiment corroborates the idea that even impersonal moral dilemmas are affected by emotions.

Furthermore, when a protagonist was described as experiencing strong emotions in Experiment 5, participants were primed to read quickly that the protagonist decided to act for impersonal dilemmas, such as hitting the switch in the

trolley problem, compared to the decision not to act. This result suggests that when people hear that a protagonist experienced a strong emotion in the impersonal dilemma, they expect the protagonist to act. It is consistent with the idea that people understand emotion to be a component of the typical reaction that leads to action, even in impersonal dilemmas. The results of Experiment 5 suggest that people expect emotions experienced in a dilemmatic context to be linked to moral action in other ways.

When a protagonist does *not* experience strong emotions in situations in which people expect them to do so, for example, in deciding whether to push a man off a footbridge, their relative lack of emotion seems to lead people to expect them to act, as Experiment 5 also showed. The result is consistent with the idea that emotion is a component of the typical reaction that leads to decisions *not* to act in personal dilemmas; for example, people may usually experience a sort of moral repugnance when considering decisions such as whether to push a man off a bridge in order to save others. Other research has demonstrated that people with damage to areas of the brain associated with the generation of emotions make atypical moral judgments, particularly with regard to dilemmas that contrast utilitarian considerations (i.e. favouring the welfare of the many over that of the few), with emotionally aversive behaviour. For example, in personal dilemmas like the footbridge problem, participants with damage to the ventromedial prefrontal cortex (which results in deficits in social emotion) are more likely to endorse the personal moral harm than are normal controls or brain-damaged controls (Koenigs et al., 2007). This pattern repeated in another study that compared patients with lesions in the ventromedial prefrontal cortex with normal controls on personal and impersonal moral dilemmas (Ciaramelli et al., 2007). Again, patients with lesions were more likely to endorse the

utilitarian course of action when it involved an emotionally aversive action than were controls, and they also did so more quickly. It is possible to speculate that Experiment 5 indicates that people expect those who experience mild rather than strong emotions to act in personal moral dilemmas.

The results of Experiments 4 and 5 provide some support for theories that suggest dilemmas differentially engage emotions (e.g. Bucciarelli et al., 2008; Greene et al., 2001, 2004; Moore et al., 2011). However, they also emphasise that dilemmas that may be *less* intrinsically emotional than others may still have emotional content, and that this content is readily available to many participants. Another prominent theory of moral judgment, the moral grammar theory, suggests that emotions can importantly interface with moral cognition (e.g. Dwyer et al., 2010), and so it to seems congruent with the findings here. However, as discussed in Chapter 2, the role of domain-general heuristic processes in moral judgments, such as those observed in Experiments 1-3 are problematic for this theory.

It is important to note that, in Experiment 4, post-hoc justifications are treated as data in themselves, rather than as indicating how people reached their judgments. As previously discussed, there are cases in which the reasons for moral decisions are not consciously accessible to participants (e.g. Haidt, 2001; Cushman et al., 2006). The experiment illustrates that people can access both emotional and reasoned justifications; the data supports the idea that different instructions can lead people to focus on different aspects of scenarios, but does not lead to conclusions as to what processes underlie moral judgments.

The interpretation of Experiment 5 follows the logic of priming studies, that the provision of information about the emotion enables a participant to make an inference, for example, that the protagonist will act, and they are then primed to read

quickly a subsequent sentence describing that the protagonist did act because they have already mentally represented the information. An alternative interpretation is that the variation in reading times is not related to inferences participants made about what they expected the protagonist to do, but instead, that reading emotional words may induce an affective state or increase arousal which in turn might speed up processing, particularly of emotion-congruent information. However, this view is ruled out by the data: the observed differences in reading times across conditions were observed only for emotion-incongruent information, and only for decisions to act.

Emotions and Counterexamples

The results of Experiments 4 and 5 lead to a novel suggestion about the role of emotional information in moral judgments: that is, that emotions may provide additional information that modulates the counterexamples that participants consider to the decisions that protagonists take in moral dilemmas (e.g. Gubbins & Byrne, 2013a). People retrieve and inhibit counterexamples when they reason about non-moral content. Content of different sorts makes some counterexamples readily available and eliminates others from consideration (e.g. Byrne & Johnson-Laird, 2009; Markovits, Lortie Forgues, & Brunet, 2010). Moral content can also modulate the counterexamples that people consider. For example, when presented with a scenario in which a man who arrived home late to find his wife dead from a heart attack, having been delayed by a morally appropriate controllable action (stopping to visit his parents), or a controllable action that may be considered morally inappropriate in the context of the delay (stopping for a burger), and an uncontrollable situation (being stuck in traffic), participants mentally undid the morally inappropriate and controllable action more than the other two (McCloy & Byrne, 2000).

Information about the emotions experienced by the protagonist may also affect the availability of counterexamples that people consider when they understand others' moral decisions. When people think about what to do when faced with a moral situation, they may attempt to think through the alternative options, for example, in the trolley or footbridge scenarios, do something and one person will be harmed but five others will be saved, do nothing and one person will not be harmed but five others will not be saved.

The evidence from the literature indicates that the salience of different alternatives varies for different dilemma types. For example, the alternative to do something and allow one person to be harmed while saving five others is more salient when the something to be done is impersonal and does not require direct contact, such as hitting the switch. The alternative to do nothing and one person will not be harmed but five others will not be saved is more salient when the something to be done is personal and requires direct contact, for example, pushing the man.

Experiment 5 indicates that sometimes the salience of these alternative possibilities may be affected by the emotions that people believe the protagonist is experiencing. That is to say, in some cases it appears that unusual emotional reactions that participants might not experience themselves (e.g. mild emotions in personal dilemmas) prime people to expect unusual behavioural reactions that participants themselves might not do (e.g. acting in personal dilemmas). People may read decisions quickly in these inappropriate combinations in a similar fashion to how they focus on morally inappropriate behaviours when they undo events (e.g. McCloy & Byrne, 2000).

These findings suggest that a greater understanding of the emotions people experience when they make moral judgments might influence other people's

judgments of their actions. Public discussion of people who have faced moral dilemmas is often damning in tone; the captain of the Costa Concordia faced a great deal of criticism for his actions (e.g. Weathers & Evans, 2012). Experiment 4 suggests that people sometimes expect explanations of their emotions to influence judgments of their actions, and Experiment 5 suggests that they seem to have expectations of the combinations of emotions and actions that people will have in moral dilemmas. These findings may indicate that knowledge about emotions experienced while facing moral dilemmas may play a role in the kinds of judgments people make of others who have faced such dilemmas.

Implications

Although Experiment 5 did not demonstrate differential effects of emotions on judgments of the acceptability of protagonists choices in moral dilemmas, judgments of acceptability were quite high in all conditions. This provides some reason to think that the provision of emotional information may have increased the acceptability of action in all dilemmas, especially in personal moral dilemmas, where approval of action is normally quite low (e.g. Greene et al., 2001). An obvious place where this sort of influence might be important is in criminal trials, where the presentation of evidence of emotional hardship already takes place in some jurisdictions. In these cases the accused may argue that certain crimes were acts of ‘passion’, or ‘temporary insanity’ – in other words, of an overwhelming emotional reaction that left them out of control of their behaviour. Victims are also sometimes given an opportunity to appeal to the jury or judge in Victim Impact Statements, where they may explain the emotional traumas they have been exposed to as a consequence of the crimes perpetrated against them. It is possible to speculate, based on the results of

Experiments 4 and 5, that statements from the perpetrators of morally questionable acts may influence moral judgments.

Experiment 4 suggests that people understand the potential utility of emotive explanations in convincing people of the morality of their behaviour. This influence may not simply take place in legal situations. Many moral judgments take place outside of the legal context, in people's everyday lives, perhaps in contexts similar to that suggested by the directions participants were given in Experiment 4, where they attempt to convince their friends of the justifiability of their actions.

The evidence discussed throughout this chapter indicates that emotions may be integrated in thinking and reasoning about moral judgments. It also seems to play a role in influencing moral action. This idea of moral emotions being an important part of the decision to act is congruent with what psychological research has discovered about the moral emotion of elevation, as discussed in Chapter 1. This emotion, which often arises when people see someone doing something morally good, seems to provide a component of the transition from witnessing morally good behaviour, to behaving morally oneself (e.g. Silvers & Haidt, 2008; Algoe & Haidt, 2009; Cox, 2010). The next chapter focuses on this emotion, and particularly its role in leading people to behave morally. Specifically, the experiments reported in the next chapter investigated the cognitive processes that may be involved in the transition from hearing about someone else doing pro-social deeds, to thinking about carrying out pro-social deeds oneself.

Chapter 4 Cognitive Processes in Moral Elevation: Counterfactual ‘if only’ thoughts and intentions to emulate

Moral psychologists have often focused on judgments and reactions in immoral rather than morally good scenarios. Chapter 2 demonstrated the pervasive influence immoral rather than moral events have on people’s moral judgments, and often the moral dilemmas used by psychologists in experiments, such as the trolley and footbridge problems discussed in Chapter 3, contain choices between two actions that violate a moral principle, such as harming someone or allowing more people to be harmed. However, in recent years researchers have begun to investigate the effects of witnessing morally good behaviour. When people think of times they have seen people helping those less fortunate than them, or going out of their way to help make someone else’s day better, or protecting someone from danger, they may feel uplifted, and perhaps think about doing something good themselves (e.g. Algoe & Haidt, 2009; Schnall & Roper, 2012). This feeling of being morally uplifted, which has been referred to as moral elevation (e.g. Haidt, 2000), can inspire and motivate people to emulate the good behaviour they have witnessed. This emulation can be either specific, meaning that people carry out a similar behaviour to the one they witnessed (e.g. Cox, 2010), or general, meaning people are inspired to attempt to become a better person in some other way (e.g. Schnall et al., 2010).

In the experiments reported in this chapter, participants were asked to think about a time when they saw someone demonstrating humanity’s better nature. Consider for example, one participant, who recollected: *“When my elderly mother was locked out my neighbor brought her into their house and looked after her until I*

got home". The participant reported that she thought about changing her own behaviour after the event by "*Being less busy to have time and be thoughtful*". In the experiments reported here, the question of how the emotional experience of elevation leads people to emulate the behaviour of moral excellence is explored. It is suggested that the link may be provided by a set of cognitive processes, based on the creation of counterfactual alternatives to reality, that provide blueprints for the formulation of intentions for the future.

Moral elevation can be described as an emotional response to moral exemplars. For example, people may be elevated after witnessing someone perform an act of moral beauty such as charity, fidelity, kindness, heroism or any other display of virtue that improves the welfare of others while incurring a cost to the person performing it (Schnall et al., 2010; Schnall & Roper, 2012). It is linked to inspiration, and to feelings of empathy and benevolence toward others (Diessner, Iyer, Smith & Haidt, 2013; Algoe & Haidt, 2009). Its focus is on others, unlike general positive emotions such as joy, that focus people on themselves (e.g. Algoe & Haidt, 2009). For this reason, it is termed an 'other-praising' emotion (e.g. Algoe & Haidt, 2009). These emotions, including elevation, but also gratitude and admiration, are elicited when witnessing someone else engaged in a particular kind of behaviour. Gratitude is evoked when someone does something generous or kind that benefits oneself, and admiration is evoked when someone demonstrates great ability. Elevation, on the other hand, is evoked by acts of moral excellence, such as charity, that do not benefit the elevated person directly (Algoe & Haidt, 2009). The research reported here investigates how people think and reason about the type of events that elicit elevation, and particularly how these events lead people to consider changing their own behaviour to be better. In this chapter, the literature relating to elevation is reviewed,

and a new theory of how elevation may lead to emulation of good behaviour is presented. Three experiments testing this theory are reported, followed by a discussion of the results of how the experiments relate to the theory, and how the findings relate to theories of moral judgment generally.

Moral Exemplars and Moral Elevation

The three experiments reported in this chapter asked participants to think about an experience in which they saw or heard about someone demonstrating humanity's higher or better nature, in which they were not themselves the beneficiary: that is, someone did something good, honourable, or charitable for someone else (adapted from Algoe & Haidt, 2009). Participants provided various sorts of good memories, including memories from their family life, for example, "*Mam washing Gran's feet. My mother cared for her mother – allowing her to stay in her own home (as she wished) for much longer than anyone predicted*". Alternatively, they were asked to think about an experience in which they saw or heard about someone demonstrating humanity's lower or worse nature, in which they themselves were not the target: that is, someone did something bad, dishonourable, or uncharitable to someone else. Participants provided various sorts of bad memories, for example, "*I witnessed a mother, closed-fisted, punching her toddler in the head*". The experiments compared the way people thought about these morally good and bad events, and how such thought could lead to emulation.

As discussed in Chapter 1, experiments have shown that when people reflected on experiences of elevation, triggered either by watching a film clip in the laboratory, or by considering an event in their own daily life recorded in a diary, they were motivated to engage in pro-social and affiliative behaviour (Algoe & Haidt, 2009).

Participants who had watched an elevating film clip were more likely than those who had not to volunteer for an unpaid study, and to spend longer helping with a tedious task, compared to participants who had watched an amusing film clip (Schnall et al., 2010). The induction of elevation elicited greater donations from white US participants to black-oriented charities, suggesting that elevation makes people feel generous and closer to others (Freeman, Aquino, & McFerran, 2009). As discussed in Chapter 1, elevating experiences even affected nurturing behaviour in nursing mothers, including increased hugging and nursing, and may be linked to oxytocin release (e.g. Silvers & Haidt, 2008), a process associated with empathy for strangers (e.g. Barraza & Zak, 2009). Elevation may also be an important mediator of attitudes towards companies in the public eye - elevation helped participants to form stronger positive attitudes of companies that decided not to move offshore (e.g. Grappi, Romani & Bagozzi, 2013).

The impact of elevation on behaviour can be general: sometimes, elevation increased helping with a different kind of task from the one with which people were elevated, indicating that elevation inspired altruism and good deeds, rather than simple imitation (e.g. Schnall et al., 2010). In other cases, its impact was specific. Field studies of self-reports of elevation at the conclusion of a spring-break service trip (i.e. a group trip taken by university students to help people less fortunate than themselves during their spring break from university), and reports of volunteerism three months later, showed that participants were still motivated to volunteer after three months, but only with the kind of work which had originally elevated them (Cox 2010).

The primary question explored in this chapter is, how does moral elevation lead to emulation of good behaviour, or an inspiration to change one's behaviour for

the better? The experiments in this chapter tested a new theory of the cognitive processes that underlie the transformation of morally elevating experiences to the formation of intentions to emulate. This theory is based on the creation of counterfactual ‘if only’ thoughts about how things could have been different (Gubbins & Byrne 2013b). For example, the participant who described her mother being looked after by neighbours imagined how the event could have turned out differently if, “*she had wandered off and been lost*”. The theory is outlined below, and followed by the three experiments testing the theory.

Elevation Inspires Emulation

Previous work on elevation (e.g. Haidt, 2000) has suggested that the phenomena involved in and resulting from elevation are best described by the broaden-and-build theory of positive emotions (e.g. Fredrickson 2001). According to this theory, positive emotions turn people’s focus onto other people, broadening “momentary thought-action repertoires”, meaning they have a wider range of thoughts and actions available to them, which subsequently builds their physical, intellectual, social and psychological resources (Fredrickson 2001, p.1). On this account, elevation triggers positive and affectionate emotions for the altruist, and these emotions in turn increase the likelihood of pro-social behaviour (e.g., Haidt, 2000). However, the cognitive mechanisms by which this process might occur are underspecified, and it is hypothesised here that creating alternatives to reality may play an important role in this process. Specifically, it is suggested that the observation of moral excellence evokes an emotional experience of elevation, and elicits the spontaneous creation of counterfactual ‘if only’ thoughts about how reality could have turned out differently,

which in turn provide a blueprint for the formation of intentions for emulation (Gubbins & Byrne 2013b).

Consider a computer program designed to simulate a theory of the cognitive processes underlying elevation-inspired emulation. The program would take as input an elevating experience; for example, a participant in the experiments reported below described an elevating event: *“During heavy snow, I saw people getting out of their cars to spend 20 minutes helping push a car uphill”*. It would provide as output a blueprint for how to emulate the elevating event, or to engage in morally good activity. For example, the participant reported they thought about changing their own behaviour: *“Being more helpful to people during bad weather”*. Since different positive, other-praising emotions, such as gratitude, do not prompt emulation or pro-social behaviour to the same degree as elevation (Algoe & Haidt, 2009), it can be said that the translation of elevation into emulation is likely to depend on more than just an emotional reaction. Specifically, it is likely to involve a complex set of mental processes based on social, personal, motivational and emotional components. The theory proposed by Gubbins & Byrne (2013b) addresses what may be a key component: a set of cognitive processes.

The proposed algorithm to simulate the component of cognitive processes that lead from elevation to emulation consists of two primary components: (a) the creation of counterfactual alternatives to reality, and (b) the formulation of intentions to emulate. The observed event evokes an emotional response, but the theory proposes that further processes are required to translate the elevating experience to emulation. The theory proposes that emulation requires more than the emotionally uplifting experience of elevation to inspire the immediate desire to be better, and that eventual change is achieved through a set of cognitive processes that comprise the creation of

counterfactual alternatives that provide the blueprint for the formulation of future intentions. The relative contributions of emotional and cognitive processes in moral judgment are discussed in relation to the experiments.

Rationale for the Role of Counterfactual Thoughts in Elevation-Inspired Emulation

There is good reason to expect that elevating experiences spontaneously elicit counterfactual ‘if only’ thoughts, because elevating experiences are exceptional and it is well established that exceptional experiences do so (e.g. Kahneman & Tversky, 1982; Dixon & Byrne, 2011). People often think about how things turned out, and they imagine how they could have turned out differently, ‘if only...’. They create counterfactual alternatives to reality. For example, a participant in the experiments reported here said, after thinking of a friend who bought a meal for a homeless person, that things could have been different “*if my friend had turned a blind eye to the homeless person then he may have gone hungry*”.

People exhibit remarkable regularities in their creation of counterfactual alternatives to reality. Regret is an emotion linked to counterfactual thinking (i.e. people sometimes think about what might have been if only..., and regret the reality), and if elevation also involves counterfactual thinking, elevating events may share certain features with the type of events that prompt regret. People focus on actions rather than inactions, and so they regret bad outcomes that result from their actions more than their inactions (e.g. Roese, 1997; Ritov & Baron, 1990). People regret actions from a recent perspective (that is, they think about actions they wish they had not done), and inactions from a remote perspective (such as a failure to spend time with family and friends, or to pursue educational opportunities; e.g. Gilovich &

Medvec, 1995). When people regret general behaviour (that is, behaviour that did not take place at a particular time, but over a non-specific period), such regrets are more likely to be due to inactions, while when they regret specific behaviour (that is, behaviour that took place at a particular time), their regrets are no more likely to be due to action or inaction (Davison & Feeney, 2008). People also create counterfactuals to intentional events, that is, events within their control, rather than events outside their control (although, for events with bad outcomes, this occurs more for morally inappropriate behaviour, rather than morally appropriate behaviour, e.g. McCloy & Byrne, 2000).

These observations lead to the prediction that if, as the theory claims, morally good events trigger counterfactual thinking, then it should be observed that they share certain characteristics of counterfactual precursors: they should tend to be unexpected, since they are exceptional; they should be controllable, that is, intentional events rather than unintentional; they should tend to be actions rather than inactions, and, since they are likely to be actions rather than inactions, they should not be general events (i.e. events that take place over a period of time) more than they are specific (i.e. events that take place at a particular period of time).

One aim of the experiments reported in this chapter was to test these predictions about characteristics of morally good events as counterfactual precursors. Another aim was to test the sorts of counterfactuals people produce when they think about how a morally good event could have turned out differently. These counterfactual alternatives are discussed in the following sections.

Mental Representations and Counterfactuals

People create a counterfactual alternative to reality by modifying an aspect of

their mental representation of reality (e.g. Byrne, 2005; Kahneman & Tversky, 1982). People may construct a representation of reality by thinking about some of the true possibilities (e.g. Byrne, 2007; Johnson-Laird & Byrne 2002). Two distinctions between the types of counterfactuals that people generate are explored in the experiments reported below.

Additive and subtractive counterfactuals. People rely on their representation of reality as a basis for their construction of alternatives to reality, deleting aspects of reality or adding aspects to reality. ‘Subtractive’ counterfactuals delete something from reality; that is, people think about how the situation could have been different if something that happened had not happened (e.g. Markman, Lindberg, Kray & Galinsky, 2007; Kray, Galinsky & Markman, 2009). An example of such a counterfactual can be taken from one of the experiments reported in this chapter. One participant, thinking of a good memory, remembered his father helping a woman who had fallen at the side of the road, and imagined how the situation could have turned out differently if “*we had not stopped the car*”. Similarly, a participant thinking of a bad memory remembered people acting selfishly during a sale at a shop and thought about how the situation could have been different if “*greed was not hardwired into society/human DNA*”. In contrast, additive counterfactuals add new features to reality; for example, the participant, thinking of a good memory, who recollected neighbours looking after her mother imagined how the situation could have turned out differently if “*she had wandered away and been lost*”. Another participant, thinking of a bad memory, remembered the 2001 attacks on the World Trade Center in New York City, and imagined how things could have been different if “*my grandmother had decided to go on her holiday she would have been on the first plane that hit the tower*”.

Additive counterfactuals help people to prepare for the future more than subtractive counterfactuals because they provide a blueprint plan for what the future could be like, (e.g. Roese, 1994; Kray et al., 2009). Additive counterfactual mind-sets inspire creative thought compared to subtractive counterfactual mind-sets, whereas subtractive counterfactual mind-sets aid analytic problem solving compared to additive counterfactual mind-sets (Markman et al., 2007).

Upward and downward counterfactuals. People construct counterfactuals that make comparisons to a better outcome or a worse outcome. ‘Upward’ counterfactuals compare reality to a better alternative and help people to identify the causes of bad outcomes (Roese, 1994). For example, thinking of a good memory, one participant remembered an acquaintance paying off an unemployed student’s college loan, and imagined how things could have turned out better if “*the recession hadn’t happened*”. Another participant, thinking of a bad memory, remembered a time when their friend found out that the friend’s boyfriend had cheated on her by catching the other woman in his apartment, and thought about how things could have been better if “*my friend had left before the doorbell rang - she never needed to know*”. In contrast, ‘downward’ counterfactuals compare reality to a worse alternative (Roese, 1994). For example, thinking of a good memory, a participant remembered when his father stopped to help a motorcyclist who had crashed, and imagined how things could have turned out worse if “*we had just driven on past the accident to our destination*”. Another participant, thinking of a bad memory, remembered when a teenager was rudely snapping a balloon at a bus stop and hit a woman in the head with it, and imagined how things could have been worse if “*he had hit the woman in the eye*”. Upward counterfactuals are generated spontaneously more than downward ones (e.g.

Roese 1997).

Upward counterfactuals can lead people to experience negative emotions such as regret or guilt, but serve to improve performance on cognitive tasks, whereas downward counterfactuals can make people feel better, but possibly at the expense of enabling them to identify the causes of bad outcomes (Roese, 1994). Counterfactual thoughts amplify the experience of emotions such as regret, relief, guilt, satisfaction, and social ascriptions such as blame (e.g., Roese, 1997). People think about the way a situation turned out, and they think about how it could have turned out differently, and in the comparison of the two the emotion emerges or increases (e.g., Kahneman & Tversky, 1982). The experiments reported in this chapter test whether people tend to create upward or downward counterfactual ‘if only’ thoughts about morally good events and morally bad events, and whether their counterfactuals are created by additive or subtractive changes.

People do not always do morally good things, in spite of being able to think of someone doing something very morally good (as the experiments reported here illustrate). This suggests that people do not always succeed in emulating morally good behaviour. The theory proposed here suggests that part of the difficulty arises because the sorts of counterfactuals people naturally construct for morally good events may not facilitate the sorts of intentions that they can readily implement. These intentions are discussed in the next section.

Rationale for the Role of the Formation of Intentions from Counterfactuals

The function of counterfactual ‘if only’ thoughts is in part preparatory. These kinds of thoughts help people to identify causes of outcomes and form intentions for the future (e.g., Roese, 1994; 1997). The process of thinking about how things might

have been different provides a map for future behaviour: if people want things to be different, and have thought about how things might have been different, their intentions may be formed from this knowledge. Intentions, in turn, play an important role in how people think about moral behaviour. People distinguish between whether moral behaviour was intentional or accidental, and their interpretations of how intentional a behaviour was vary according to whether it had morally good or bad outcomes. For instance, as discussed in Chapter 1, in adulthood, and even at a young age, people judged side effects of decisions with bad consequences to be more intentional than those with good consequences (e.g. Knobe, 2003; Leslie et al., 2006).

People can also distinguish between specific and general intentions. Intentions to change may be specific. For example, with reference to a memory of a friend who had paid a toll for the car behind, a participant noted their intention to change their own behaviour, “*maybe try it again someday*”. Such intentions may also be general. For example, with reference to a memory of an individual who donated a kidney to a relative, a participant noted their intention to change their own behaviour, “*to be more giving*”. The question of whether elevating memories lead people to generate specific or general intentions to change is important because general intentions to change may be more difficult to implement. The theory proposed here suggests that the type of counterfactuals people generate in response to elevating events may not facilitate specific intentions, which would provide a clear path toward emulation, but general intentions, which may be more difficult to implement. Specifically it is proposed that thinking of people doing good things for others may result in people thinking about how the situation could have been worse (i.e. downward counterfactuals), and that they may do so by removing a feature from reality (i.e. subtractive counterfactuals). As discussed above, such counterfactuals are less effective than upward and additive

counterfactuals in identifying how to improve behaviour, and preparing people for the future. Thus, people's thoughts about emulating behaviour may be general rather than specific, and it is suggested that this may be part of the reason that people do not always emulate good deeds. The experiments reported below therefore tested not only how people thought about changing their behaviour after seeing or hearing about a moral exemplar, but also whether their intention to change was general or specific.

Experimental Task

In the experiments reported here, participants recalled good or bad events that they witnessed; they imagined how they could have turned out differently; and they indicated whether they had thought about changing their own behaviour. The tendencies to create downward or upward counterfactuals, to create additive or subtractive counterfactuals, and to create specific or general intentions to emulate were tested. The experiments also tested whether the memories participants recalled shared characteristics of memories that would produce lead people to spontaneously produce counterfactual alternatives: that is, to be unexpected rather than expected, intentional rather than unintentional, actions rather than inactions, and not more general than specific. The experiments also examined the effect of thinking about elevating experiences on mood. In some experiments, mood improved immediately after recalling an elevating event (Algoe & Haidt, 2009). However, in the experiments reported in this chapter, mood prior to recalling the event was compared to mood, not immediately after recalling the elevating experience, but after a period of time spent thinking about the event in detail, to test the impact of thinking or reflecting about elevating experiences on mood.

In order to obtain measures from a wide and diverse population, over 400 volunteers were recruited to take the experiments from the general public attending an exhibition to celebrate the 50th anniversary of the School of Psychology over a six-week period, the ‘Happy?’ Lab in the Gallery exhibition at the Science Gallery in Trinity College Dublin, one of the top 5 free visitor attractions in Dublin, <https://sciencegallery.com/happy/>. The participants ranged in age from 18 to 87 years of age¹⁰; two-thirds were Irish and the remaining one third comprised different nationalities, including from other European countries (UK, Denmark, Poland, Finland, Slovenia, Slovakia, Spain, Germany, Hungary, Netherland, France, Portugal, Italy), and other continents, North America (US, Canada), South America (Brazil), Africa (South Africa), and Asia (Georgia, South Korea).

The experiments also asked participants whether the main person in the scenario was very close to them, known, or a stranger, and also whether they fit into one of several categories of friends, family, social acquaintances, and work colleagues, among others. These questions were included in order to provide the Science Gallery with options in terms of how to display the cards of participants who consented to having their memory displayed. However, the questions did not relate to the theory tested by the other questions in the experiments, and they are not analysed here for that reason.

¹⁰ Since the Science Gallery is open to the public, people of all ages were able to take part in the experiments. Therefore 146 participants were removed across the three experiments prior to any analysis because they (a) indicated that they were under 18 years of age, (b) they did not provide their present age, and provided an age under 18 for their age at the time of the memory, or (c) they provided a memory, but no other information for any other measures.

Experiment 6

The aim of the experiment was to compare counterfactual thoughts and emulation intentions about morally good and bad experiences, including examining the counterfactuals created as upward or downward and additive or subtractive, and the emulation intentions as general or specific, as well as assessing the counterfactual precursors of whether a memory was an action or an inaction, took place at a specific time or over a general period, was expected or unexpected, and intentional or unintentional. The experiment therefore tested the theory discussed above (Gubbins & Byrne, 2013b) in terms of whether participants produced memories that are of a kind that would lead people to spontaneously produce counterfactual alternatives to reality, and in terms of what type of counterfactuals people generated. It also tested the ways in which people think about changing their behaviour when they think of such memories.

Method

Design and Materials

There were two groups, who recalled either a good experience or a bad experience. Participants in the good condition were prompted to recall a memory of a good experience as follows (adapted from Algoe & Haidt 2009):

“Please think of an experience in which you saw someone, or heard about someone, demonstrating humanity’s higher or better nature. Please pick an example in which you were not the beneficiary, that is, someone did something good, honorable, or charitable *for someone else*. Please think of a situation that resulted in little or no negative feeling, or at least in which the

positive feelings were much stronger than the negative. Please describe the experience briefly, in writing, here”.

Participants in the bad condition were prompted to recall a memory of a bad experience as follows:

“Please think of an experience in which you saw someone, or heard about someone, demonstrating humanity’s lower or worse nature. Please pick an example in which you were not the target, that is, someone did something bad, dishonorable, or uncharitable to someone else. Please think of a situation that resulted in little or no positive feeling, or at least in which the negative feelings were much stronger than the positive. Please describe the experience briefly, in writing, here”.

After they had recorded their memory, participants completed a series of tasks designed to measure their counterfactual thoughts about the past event, intentions to emulate, and characteristic precursors of counterfactual thoughts.

Counterfactual thoughts and intentions to emulate. Participants were asked to complete a standard counterfactual sentence stem task in response to the prompt: “People sometimes imagine how an event could have turned out differently ‘if only...’ Please think about the experience and about how it could have turned out differently. Complete the following sentence: ‘It could have turned out differently if...’ “. They were asked a question about their future emulation intentions: “Did you think about

changing your own behavior after the event? If so, describe what you thought about changing”.

Counterfactual precursors. Participants were asked, ‘Was the behaviour expected or unexpected?’ and they circled their answer on a scale from 0 (unexpected) to 10 (expected), and ‘Was the behavior intentional or unintentional?’ and they circled their answer on a scale from 0 (unintentional) to 10 (intentional). They were asked, “Was it an action (something someone does) or an inaction (something someone does not do)”, and they circled one, and “Was it a specific experience (an event that occurred at a specific time) or a general one (a non-specific experience)” and they circled one.

Emotional consequences. Participants were asked to ‘Please indicate your current mood’ on a scale from 0 (very sad) to 10 (very happy) at the outset of the experiment prior to recalling the experience or completing any tasks, and again after they had completed all of the tasks. They were also asked to provide “the single best word to describe your feelings at the time”. The full set of questions and their order of presentation are provided in Appendix 4.1.

Inter-rater reliability. The author and two independent raters (one coding the entire dataset, the other twenty percent of the dataset) scored the counterfactuals for two distinctions. They scored them as upward or downward. An upward counterfactual is a representation of a better world, such as that imagined by the participant who thought about “*if the recession hadn’t happened*”. A downward counterfactual is a representation of a worse world, such as that imagined by the

participant who thought about their father helping a motorcyclist and then considered how things might have been different had their family “*just driven on past the accident to our destination*”. They also scored the counterfactuals as additive or subtractive. An additive counterfactual adds a feature to reality. An example came from a participant who remembered a time when they and another pedestrian came close to being hit by a reckless driver, and thought about how things could have been different if “*the pedestrian crossed the road too quickly*”. A subtractive counterfactual removes some feature of reality. An example came from the participant who remembered seeing two men fighting outside of a pub, and thought about how things could have been different if “*the two men involved had not been drunk*”.

The raters also scored whether the participant expressed having thought about changing their behaviour. Although participants were asked to write whether they thought about changing their behaviour (yes or no), they were also given room to state *how* they thought about changing their behaviour. Many participants used this space to explain their thinking in some way, rather than giving a yes or no answer. For instance, one participant gave the response, “*I will drive more carefully*”, a response that did not include a yes or a no. Therefore the coders judged whether the participants’ responses indicated that they thought about changing their behaviour, or not.

For judgments of whether their memories were actions or inactions, and judgments of whether their memories were specific or general, participants were found to have frequently provided responses that did not appear to fit with their memories, such as judging an action memory as an inaction, or a memory that took place at a specific time as general. For instance, a participant recalled a memory of someone reversing into her husband’s car, and categorised it as an inaction. Similarly,

a participant described how their mother received no compensation from the government after her husband died as a result of experiments conducted on him, and categorised it as an action, though it seemed to be the government's inaction in not giving her compensation that was the bad behaviour involved. Another participant described a time that he came close to being run over by a car as a general memory, in spite of the fact that it took place at a specific time. Similarly, a participant described the story of a woman whose husband left her broke, only for her to later raise his children, born to another woman, as her own, as taking place over a period of many years, yet they categorised it as specific. Similar difficulties were encountered by a past study that asked participants to categorise their regrets as specific or general (Davison & Feeney, 2008). These authors had specific regrets independently re-categorised, since the only problems in their data concerned general regrets that had been categorised as specific. In the experiments reported here, however, instances of incorrect categorisation occurred for memories involving actions, inactions, specific memories and general memories. Therefore responses for all of these categorisations were re-coded.

Finally, whether people's thoughts about changing their behaviour were general or specific was coded by the author, the supervisor, and an independent rater (the same rater who had coded 20% of responses for the other variables). The author and the same independent rater also coded whether the single word summation of the feeling the event provoked in participants was positive, negative, or ambiguous between positive and negative (e.g. words such as 'shock' or 'surprise'). All disagreements were resolved by discussion. Information on inter-rater agreement for Experiment 6 can be found in Appendix 4.3.

Participants and Procedure

The participants were 87 members of the general public attending the ‘Happy’ Lab in the Gallery exhibition at the Science Gallery in Trinity College Dublin (see <https://sciencegallery.com/happy/>). The participants were assigned at random to the good condition ($n = 32$) or the bad condition ($n = 55$). There were 35 men and 51 women (and one person who did not record their gender), whose ages ranged from 19 to 73 years with an average age of 33 years (and 10 people did not record their age). Two thirds of the participants were Irish and the remaining one-third were one of 18 different nationalities. A further 25 participants were excluded prior to any analysis for a variety of reasons; the reasons for these exclusions, including examples of each reason for exclusion, can be seen in Appendix 4.2. Furthermore, some responses were ambiguous between two categories being compared, and were coded as missing in the analysis. For this experiment, three responses were coded as ambiguous for additive/subtractive counterfactuals; no other variables of whether the counterfactual was upward or downward, the memory was specific or general, an action or an inaction, whether the participant thought about changing their behaviour or not, or whether those thoughts were specific or general had a response coded as ambiguous. In rare instances in which participants failed to provide a response for a given variable, their response for that variable was also coded as missing. The materials were administered to participants by trained mediators in a quiet writing area of the Science Gallery.

Results

Participants recalled events from on average 3.69 years ago (calculated from information about their current age and their age at the time of the experience), and

there was no difference between the good ($M = 5.23$, $SD = 7.55$) and bad conditions ($M = 2.88$, $SD = 4.73$) in terms of how long ago they recalled memories from, Mann Whitney $U = 540.5$, $z = -1.11$, $p = .267$.

Counterfactuals and Emulation Intentions

Participants generated more downward than upward counterfactuals in the good condition (81% downward; binomial, $z = 3.24$, $p < .001$, $r = .58$) and more upward than downward in the bad condition (88% upward; binomial, $z = 5.4$, $p < .001$, $r = .73$); the difference between the two conditions was significant, $\chi^2(1, n = 83) = 39.639$, $p < .001$, $\phi = -0.691$, as Figure 4.1 shows. They generated more subtractive than additive counterfactuals in the good condition, although this result was not significant, (61% subtractive; binomial, $z = 1.28$, $p = .2$), and significantly more additive than subtractive counterfactuals in the bad condition (14% subtractive; binomial, $z = 5.04$, $p < .001$, $r = .71$); the difference between the two conditions was also significant, $\chi^2(1, n = 81) = 21.328$, $p < .001$, $\phi = -0.513$.

Participants indicated that they thought about changing their own behaviour in the future in the good condition (74%; binomial, $z = 2.52$, $p < .05$, $r = .45$), but not significantly so in the bad condition (63%; binomial, $z = 1.59$, $p = .11$). The difference between the two conditions was not significant, $\chi^2(1, n = 79) = 1.166$, $p = .28$, as can be seen in Figure 4.1.

There was no significant difference of whether intentions to change were general or specific in the good condition (71% general, binomial, $z = 1.84$, $p = .0658$, $r = .38$). There was no significant difference in the bad condition (43% general, binomial, $z = .55$, $p = .5823$). The difference between the two conditions was significant, $\chi^2(1, 54) = 4.084$, $p < .05$, $\phi = .275$, as can be seen from Figure 4.1.

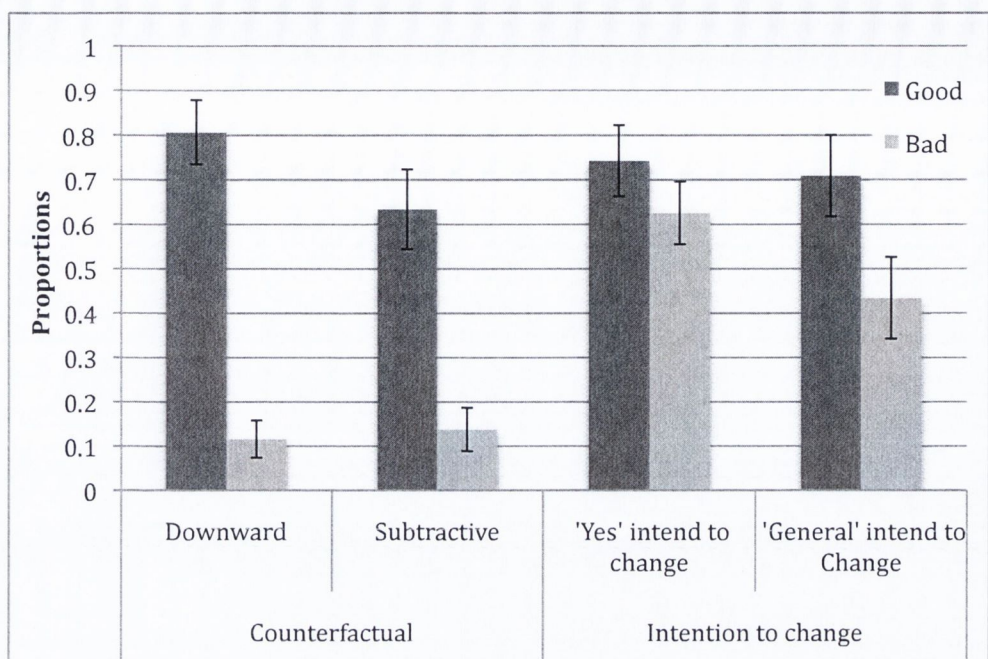


Figure 4.1. Proportions of counterfactuals as downward, subtractive, and intentions to change, as well as general rather than specific intentions to change, for good and bad memories in Experiment 6. Error bars are standard error of the mean.

Counterfactual Precursors.

Participants overall tended to judge the behaviour in their memories to be unexpected ($M = 3.69$, $SD = 2.96$), and more participants judged the behaviour to be unexpected (0-4 on the 11-point scale) than expected (6-10 on the scale) overall, binomial, $z = 3$, $p < .005$, $r = .35$ (with 12 people judging their memory to be neutral, or 5 on the scale). They did so for bad memories, binomial, $z = 3.53$, $p < .001$, $r = .5$ (5 neutral), but not for good memories, binomial, $z = 0.00$, $p = 1.00$ (7 neutral). Ratings of expectedness on the scale also indicated that memories in the bad condition were significantly more unexpected than those in the good condition, Mann Whitney $U = 508$, $z = -3.3$, $p < .005$, $r = -.35$, as Figure 4.2 shows.

Participants judged the behaviour in their memories to be more intentional than unintentional ($M = 8.72$, $SD = 2.3$), and more people judged their memory to be intentional (6-10 on the 11-point scale) than unintentional (0-4 on the scale), binomial, $z = 8.06$, $p < .001$, $r = .89$ (5 neutral, meaning a judgment of 5 on the scale). They did so for good memories, binomial, $z = 4.93$, $p < .001$, $r = .9$ (2 neutral), and bad memories, binomial, $z = 6.23$, $p < .001$, $r = .86$ (3 neutral). Ratings indicated that participants judged the behaviour to be equally intentional in the good and bad conditions, Mann Whitney $U = 830.5$, $z = .492$, $p = .622$, as can be seen from Figure 4.2.

Memories were judged to be actions rather than inactions overall, binomial, $z = 8.37$, $p < .001$, $r = .9$. This was the case in the good condition, binomial, $z = 5.48$, $p < .001$, $r = .97$, and in the bad condition, binomial, $z = 6.2$, $p < .001$, $r = .84$. There was no significant difference between the good and bad conditions, $\chi^2(1, n = 87) = 2.439$, $p = .118$. Memories were judged to be specific rather than general overall, binomial, $z = 7.08$, $p < .001$, $r = .76$. This was the case in the good condition, binomial, $z = 3.36$, $p < .001$, $r = .59$, and in the bad condition, binomial, $z = 6.2$, $p < .001$, $r = .84$. There was no significant difference between the conditions, $\chi^2(1, n = 87) = 2.620$, $p = .106$, as shown in Figure 4.2.

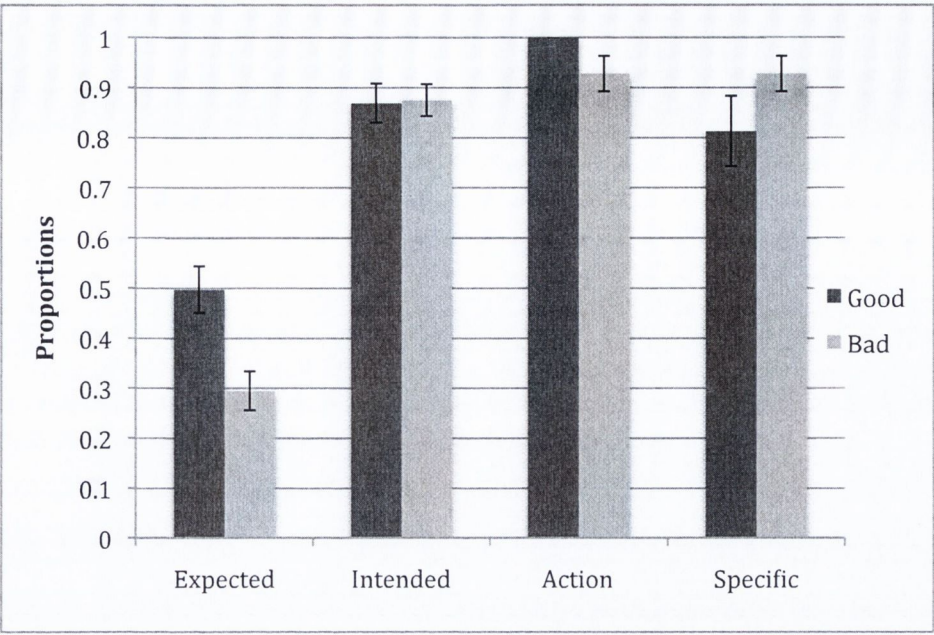


Figure 4.2. Proportions of counterfactual precursors (participants' ratings on the 0-10 scale for how expected and intended the behaviour in their memory was, and proportions of participants whose responses were categorized as actions, or as specific) for good and bad conditions in Experiment 6. Error bars are standard error of the mean.

Emotional consequences. Participants' ratings of their mood did not change before and after a period of thinking about the event in the good condition, Wilcoxon, $z = .497, p = .619$; however, it deteriorated in the bad condition, Wilcoxon, $z = 4.715, p < .001, r = .47$. Although mood deteriorated in the bad condition, in all conditions, both before and after, mood was more positive than negative, as Figure 4.3 shows; that is to say, the task did not put participants in a negative mood overall. The figure also shows that participants described their feelings using emotion words corresponding to positive emotions rather than words corresponding to negative emotions or ambiguous words in the good condition more than in the bad condition, $\chi^2(1, n = 86) = 66.721, p < .001, \phi = -0.881$.

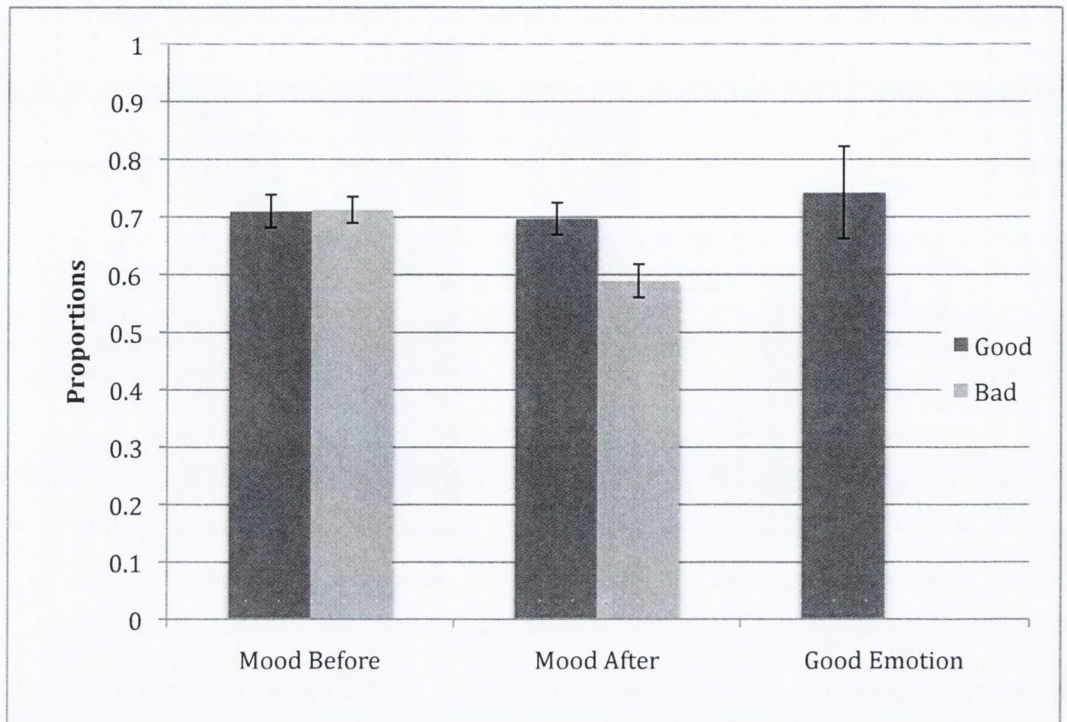


Figure 4.3. Proportions of results for mood before recording the memory and mood after (these are based on the ratings on the 0-10 scale participants provided), and of positive single words to describe the participants’ feelings, for good and bad conditions for Experiment 6. Error bars are standard error of the mean.

Discussion

The results showed that people created different counterfactual alternatives to reality when they imagined how morally good and bad experiences could have turned out differently. They imagined how things could have been worse when someone did something good, and they tended to do so by creating an alternative to reality from which something had been subtracted. In contrast, when someone did something bad, they tended to imagine how things could have been better, and they tended to do so by creating an alternative to reality to which something new had been added. As the examples at the beginning of the chapter illustrate, participants were capable of

producing either an upward or downward counterfactual to memories of someone doing something good, or memories of someone doing something bad.

Participants formed intentions to change their behaviour more than not after witnessing good actions, and somewhat, though not significantly more than after bad actions. Their descriptions of what they decided to change were more general for the good condition than for the bad condition. These results suggest that participants were more likely form the intention to become a better person generally after being elevated, rather than copying an act exactly, compared to the bad condition. This provides some support for the hypothesis that thinking about good events prompts intentions to change that are not necessarily straightforward to implement.

For the most part, the results supported the suggestion that memories of people doing morally good and bad things correspond to the sorts of events that spontaneously elicit counterfactuals: they were controllable, or intentional rather than unintentional; they were actions rather than inactions; and they were not more general than specific (and in fact were more specific than general). Contrary to prediction, highly honourable or charitable behaviour, although it was not expected, was also not unexpected, and morally bad behaviour was more unexpected than morally good behaviour.

Thinking about bad experiences affected participants' mood more than thinking about good experiences, which did not influence mood. Although elevation improves mood immediately after recall (e.g. Algoe & Haidt, 2009), this effect did not emerge after subsequent reflection on components of the memory.

Most recollected experiences in Experiment 6 came from the relatively recent past. Regret, an emotion in which counterfactual thought appears to operate (i.e. regret is the feeling that arises from thinking that things could have been better),

focuses more on actions for recent memories, but more on inactions for remote memories (e.g. Gilovich & Medvec, 1995). Experiment 7 examined whether more distant memories or more recent ones affect how people think and reason about elevation, an emotion that may also involve counterfactual elements.

Experiment 7

The aim of the second experiment in this series was to examine the effects of remote and recent perspectives on how people think and reason about morally good and bad experiences. Memories from longer ago and more recent memories are likely to differ for two reasons. Selecting from across their entire lives should allow participants to choose from a greater range of memories. Memories from longer ago are also ones that have remained memorable in spite of the passing of time. Both of these differences suggest that remote, rather than recent memories should be more likely to involve exceptional and unexpected behaviour, which in turn should make them more likely to prompt the generation of counterfactuals (e.g. Dixon & Byrne, 2011). Furthermore, due to their exceptional nature, memories recalled from longer ago might be more likely to be the kind of memories that would prompt participants to think about changing their behaviour.

Participants were asked to look back on their experiences *in their whole lives* and think about an experience in which they saw someone, or heard about someone, demonstrating humanity's higher or better nature. Alternatively they were asked to think about an experience *in the past month or so* in which they saw someone, or heard about someone, demonstrating humanity's higher or better nature. Once again, the experiment focused on how counterfactual thoughts were generated for both good and bad memories.

Method

Design, Materials and Procedure

There were four groups, who received the same instructions as in the previous experiment except that the two *recent* good and bad groups were instructed: “Please look back on your experiences *in the past month or so* and think about an experience ...” whereas the two *remote* good and bad groups were instructed: “Please look back on your experiences *in your whole life* and think about an experience ...”. The materials and procedure were otherwise the same as in Experiment 6.

Participants

The participants were 173 members of the general public attending the ‘Happy’ Lab in the Science Gallery exhibition¹¹. The participants were assigned at random to one of four conditions: recent good ($n = 37$), recent bad ($n = 44$), remote good ($n = 47$), and remote bad ($n = 45$). There were 75 men and 93 women (and 5 people did not record their gender), whose ages ranged from 19 to 70 years with an average age of 36 years (and 16 people did not record their age). Two thirds of the participants were Irish and the remaining one-third were one of 18 different nationalities. A further 64 participants were excluded prior to any data analysis for a variety of reasons. Information about these exclusion criteria can be found in Appendix 4.4. As in Experiment 6, some responses were ambiguous between two categories being compared, and were coded as missing in the analysis. For this experiment, one response for upward/downward counterfactuals was not a counterfactual (i.e. the

¹¹ Six participants were gathered after the run in the Science Gallery, but prior to any analysis, from the campus of Trinity College Dublin, in the remote good memory condition, in order to ensure close to equal numbers of participants in each condition.

participant did not consider how things could have been different; for additive/subtractive counterfactuals, one response was ambiguous and another was not a counterfactual; for responses to the questions of whether participants thought about changing their behaviour, and whether that change was general or specific, two responses were ambiguous. No responses for whether the memory was an action or an inaction, or specific or general were ambiguous. Again, in rare instances in which participants failed to provide a response for a given variable, their response for that variable was also coded as missing.

Results

Participants recalled events from on average 4.47 years ago (calculated from information about their current age and their age at the time of the experience), and there was no difference between the good ($M = 3.77$, $SD = 7.28$) and the bad conditions ($M = 5.08$, $SD = 9.31$), Mann Whitney $U = 2962.5$, $z = .404$, $p = .686$, as in Experiment 6. However, participants recalled events from about 8 years ago in the remote condition ($M = 7.9$, $SD = 10.16$), and from within the past few months in the recent condition ($M = .31$, $SD = .73$), Mann Whitney $U = 902.5$, $z = 8.028$, $p < .001$, $r = .64$, in line with their instructions. Information on inter-rater reliability for the same variables as those coded in Experiment 6 can be found in Appendix 4.5 for Experiment 7.

Counterfactual Thoughts and Intentions to Emulate

As in Experiment 6, participants generated more downward than upward counterfactuals in the two good conditions, (90% downward recent, binomial, $z = 4.32$, $p < .001$, $r = .78$, and 84% downward remote, binomial, $z = 4.48$, $p < .001$, r

= .67) and more upward than downward in the two bad conditions (3% downward recent, binomial, $z = 5.77, p < .001, r = .92$, and 5% downward remote, binomial, $z = 5.71, p < .001, r = .88$). The difference between the good and bad conditions was significant, $\chi^2(1, n = 157) = 110.02, p < .001, \phi = -0.837$, as in Experiment 6, and as Figure 4.4 shows. There was no difference between the two recent conditions and the remote ones overall, $\chi^2(1, n = 157) = .326, p = .568$.

As in Experiment 6, participants generated somewhat more subtractive than additive counterfactuals in the good conditions but not significantly so (67% subtractive recent, binomial, $z = 1.64, p = .101$, and 53% subtractive remote, binomial, $z = .3, p = .7642$); they generated significantly more additive than subtractive counterfactuals in the bad conditions (10% subtractive recent, binomial, $z = 4.81, p < .001, r = .77$, and 19% subtractive remote, binomial, $z = 3.86, p < .001, r = .6$). The difference between the good and bad conditions was significant, $\chi^2(1, n = 156) = 32.543, p < .001, \phi = -0.457$, replicating Experiment 6, as shown in Figure 4.4. There was no difference between the recent conditions and the remote conditions overall, $\chi^2(1, n = 156) = .067, p = .796$.

Participants tended to think about changing their behaviour more than not for the good conditions, significantly so for remote memories (85%, binomial, $z = 4.57, p < .001, r = .67$), but not significantly so for recent memories (66%, binomial, $z = .169, p = .091$). There was no significant tendency for participants to think about changing their behaviour in the bad conditions, either for recent memories (41%, binomial, $z = .96, p = .3371$), or for remote memories (63%, binomial, $z = 1.56, p = .1188$).

Participants indicated that they thought about changing their own behaviour more often in the good conditions compared to the bad conditions overall, $\chi^2(1, n = 161) = 10.174, p < .005, \phi = .251$, as in Experiment 6, and as illustrated in Figure 4.4. They

also indicated they thought about changing their behaviour more often in the remote conditions than the recent conditions overall, $\chi^2(1, n = 161) = 8.47, p < .005, \phi = .229$. Pairwise comparisons showed that participants indicated they thought about changing their behaviour more often in the remote good condition compared to the recent good condition, $\chi^2(1, n = 81) = 4.025, p < .05, \phi = .223$, and in the remote bad condition compared to the recent bad condition, $\chi^2(1, n = 80) = 4.018, p < .05, \phi = .224$.

For good memories, these intentions to change were general rather than specific for remote memories (68% general, binomial, $z = 2.11, p < .05, r = .34$). There was no significant difference for recent memories (61% general, binomial, $z = .83, p = .4065$). For bad memories, there was no difference between general and specific intentions for recent memories (53% general, binomial, $z = 0.00, p = 1$), or remote memories (39% general, binomial, $z = .98, p = .3271$). Intentions to change were more general in good conditions than in bad conditions, $\chi^2(1, n = 104) = 4.7, p < .05, \phi = .213$. There was no significant difference between the recent and remote conditions, $\chi^2(1, n = 104) = .016, p = .9$. Pairwise comparisons showed that intentions to change were general more than specific in the remote good condition compared to the remote bad condition, $\chi^2(1, n = 64) = 5.63, p < .05, \phi = .297$; there was no difference between the recent good condition and the recent bad condition, $\chi^2(1, n = 40) = .251, p = .616$, as Figure 4.4 shows.

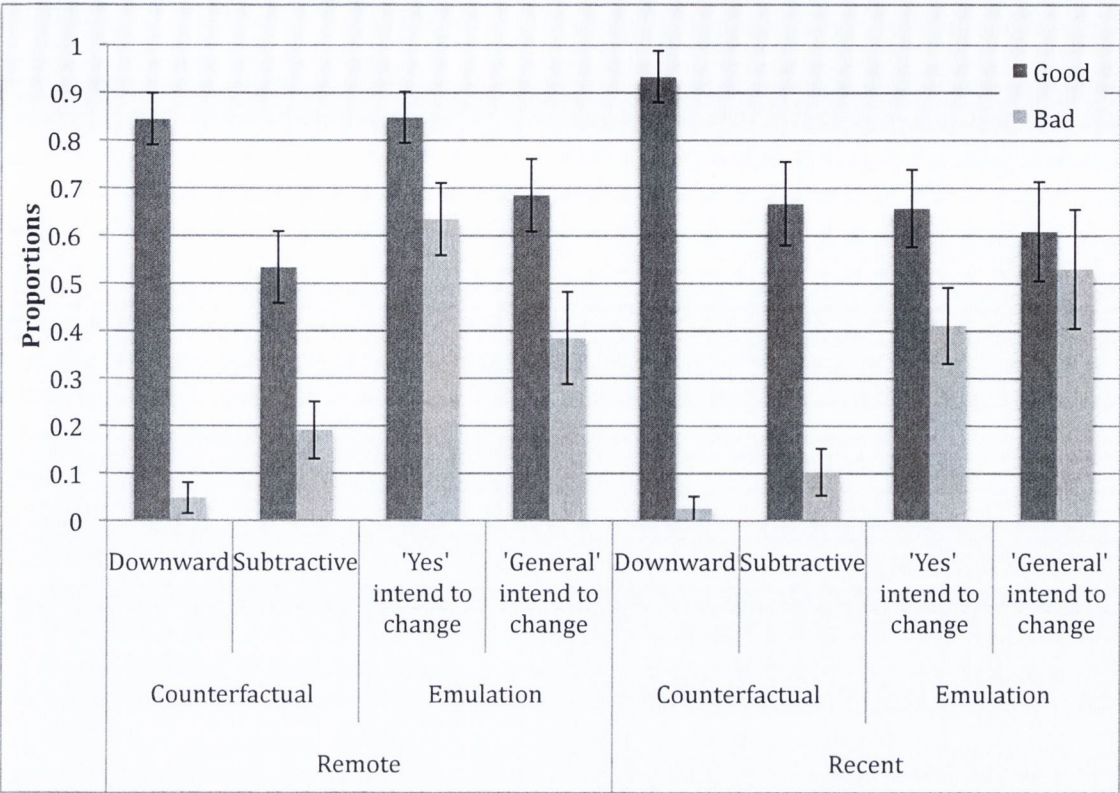


Figure 4.4. Proportions of counterfactuals as downward and as subtractive, and intentions to change, as well as general rather than specific intentions to change, for good and bad memories in remote and recent conditions in Experiment 7. Error bars are standard error of the mean.

Counterfactual Precursors

Participants judged the behaviour in their memory as unexpected (0-4 on the 11-point scale) more often than expected (6-10 on the scale) overall, binomial, $z = 2.51, p < .05, r = .2$ (9 neutral). Although people judged the behaviour in their memories as unexpected somewhat more often than expected for good conditions, as in Experiment 6, this difference was not significant, binomial, $z = 1.58, p = .1141$, (4 neutral); similarly, and as in Experiment 6 (although the trend was significant in Experiment 6), there was a non-significant trend for participants to judge the behaviour as unexpected for bad conditions, binomial, $z = 1.86, p = .0629, r = .2$ (5

neutral). There was no significant difference in unexpected and expected judgments in recent conditions, binomial, $z = 1.15, p = .2501$ (5 neutral); however, there were significantly more unexpected than expected judgments for remote conditions, binomial, $z = 2.24, p < .05, r = .24$ (4 neutral). Ratings of the expectedness of memories on the scale were not significantly different for memories in good and bad conditions, Mann Whitney $U = 3273, z = 1.297, p = .195$, or in recent and remote conditions, Mann Whitney $U = 3370, z = .958, p = .338$, as Figure 4.5 shows.

As in Experiment 6, participants judged events as intentional (6-10 on the 11-point scale) more often than unintentional (0-4 on the scale) overall, binomial, $z = 11.01, p < .001, r = .84$ (2 neutral). This occurred for good conditions, binomial, $z = 7.75, p < .001, r = .85$; bad conditions, binomial, $z = 7.73, p < .001, r = .83$ (2 neutral); recent conditions, binomial, $z = 6.98, p < .001, r = .79$ (2 neutral); and remote conditions, binomial, $z = 8.44, p < .001, r = .88$. As in Experiment 6, ratings indicated that there was no difference in how intentional the behaviour was between the good and bad conditions overall, Mann Whitney $U = 3631.5, z = .368, p = .713$, or between the recent and remote conditions overall, Mann Whitney $U = 3716, z = .035, p = .972$. These ratings are illustrated in Figure 4.5.

As in Experiment 6, and as illustrated in Figure 4.5, participants generated more memories that were actions than inactions overall, binomial, $z = 11.85, p < .001, r = .9$. The memories were actions in the good conditions (100% action, binomial, $z = 9.06, p < .001, r = .99$) and in the bad conditions (91% action, binomial, $z = 7.63, p < .001, r = .81$), and unlike in Experiment 6, more in the good conditions than the bad conditions overall, $\chi^2 (1, n = 173) = 7.917, p < .01, \phi = -0.214$. The memories generated were also actions rather than inactions in the recent (94% action, binomial, $z = 7.78, p < .001, r = .86$) and remote (97% action, binomial, $z = 8.85, p < .001, r$

= .92) conditions, and there was no significant difference between the recent and remote conditions overall, $\chi^2(1, n = 173) = .828, p = .363$.

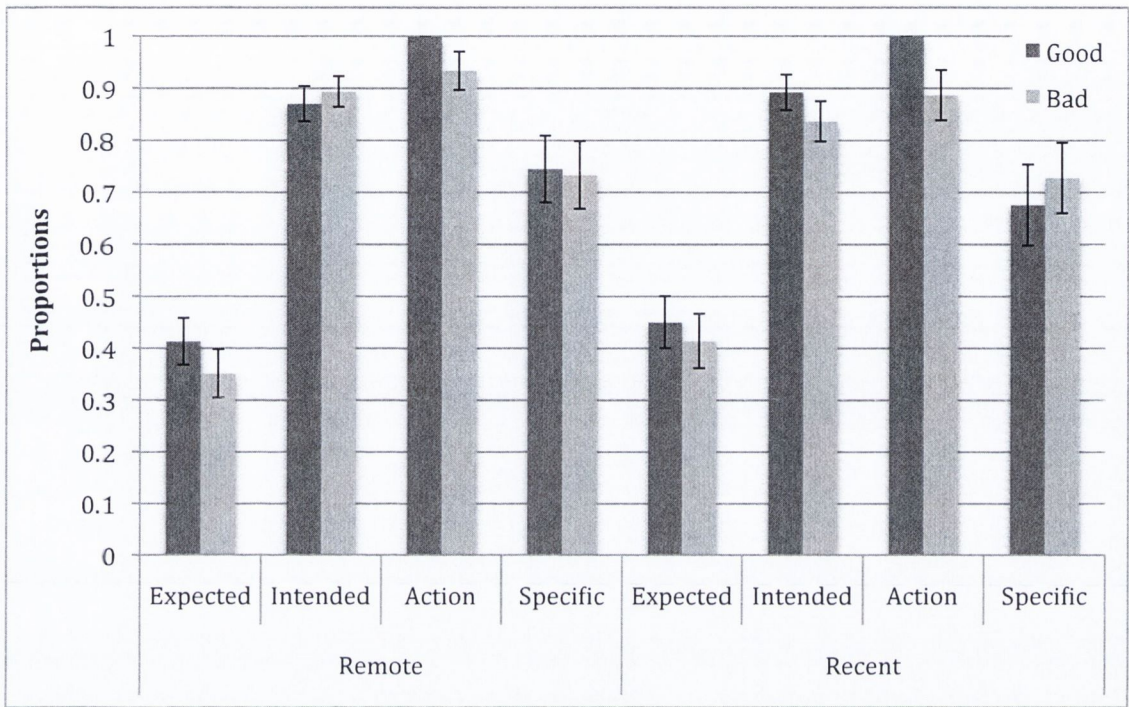


Figure 4.5 Proportions of counterfactual precursors (participants' ratings on 0-10 scale for how expected and intended the behaviour in their memory was, and proportions of participants whose responses were categorized as actions, or as specific) for good and bad memories, in remote and recent conditions, in Experiment 7. Error bars are standard error of the mean.

Again, as in Experiment 6, the memories people generated were specific more than general overall, binomial, $z = 6.61, p < .001, r = .5$. The memories were specific more than general for good conditions (71% specific, binomial, $z = 3.82, p < .001, r = .42$), and for bad conditions (73% specific, binomial, $z = 4.24, p < .001, r = .45$), and there was no difference between the good and the bad conditions overall, $\chi^2(1, n = 173) = .056, p = .814$, again replicating Experiment 6. Memories were also specific

more than general for recent conditions (72% specific, binomial, $z = 3.56, p < .001, r = .4$) and for remote conditions (74%, binomial, $z = 4.48, p < .001, r = .47$), and there was no difference between the recent and remote conditions overall, $\chi^2 (1, n = 173) = .270, p = .604$. These results are illustrated in Figure 4.5.

Emotional Consequences

As in Experiment 6, and as Figure 4.6 shows, participants' rating of their mood did not change between before and after the tasks in the good conditions overall, Wilcoxon, $z = 1.135, p = .256$, and it once again deteriorated for the bad condition, overall, Wilcoxon, $z = 4.985, p < .001, r = .4$. Again, although mood deteriorated in the bad condition, in all conditions, both before and after, mood was more positive than negative, as Figure 4.6 shows. Mood also deteriorated between before and after ratings in the recent conditions, Wilcoxon, $z = 3.088, p < .005, r = .25$, and although it deteriorated somewhat between before and after ratings in the remote conditions overall, the difference was not significant, Wilcoxon, $z = 1.743, p = .081, r = .14$.

Participants described their feelings using emotion words corresponding to positive emotions more than negative or ambiguous emotions more in the good conditions than in the bad conditions, $\chi^2 (1, n = 171) = 137.609, p < .001, \phi = .897$, as Figure 4.6 shows; there was no difference between the recent and remote conditions, $\chi^2 (1, n = 171) = 1.462, p = .481$.

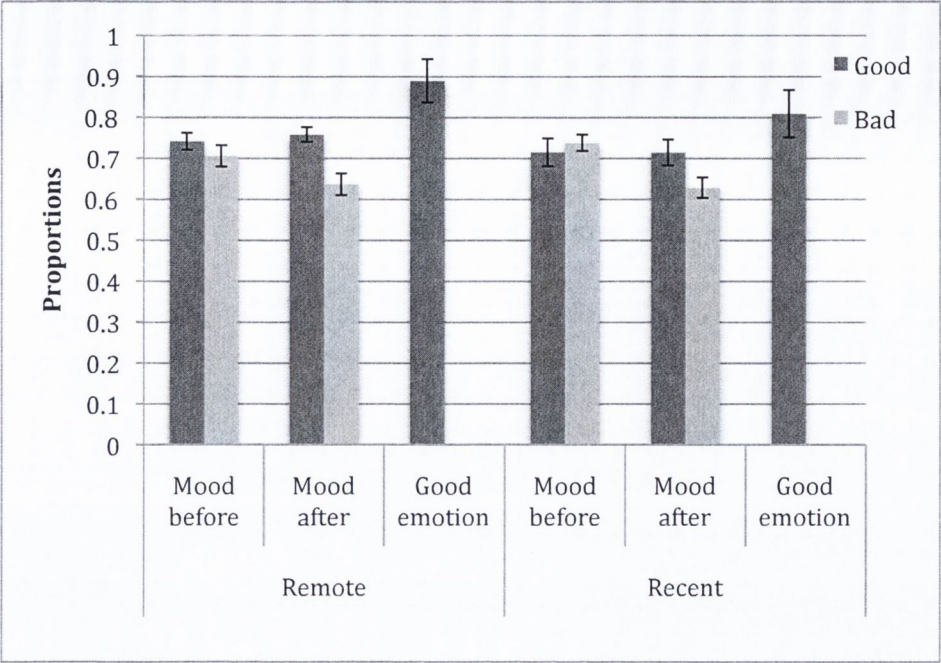


Figure 4.6. Proportions of results for mood before recording the memory and mood after (these are based on the ratings on the 0-10 scale participants provided), and of positive single words to describe the participants' feelings, for good and bad memories in recent and remote conditions for Experiment 7. Error bars are standard error of the mean.

Discussion

Experiment 7 demonstrated that recalling an event from one's entire lifetime in which one witnessed someone doing something morally good or morally bad generally results in the same counterfactual thoughts, characteristics, and emotions, as events recalled from the more recent past. One important difference was that participants were more inclined to think about changing their behaviour for good events recalled from over a longer period. The result suggests that morally good events recalled from throughout a lifetime provided greater impetus to develop intentions to emulate. This may be because being encouraged to recall good events from across a wide span of time allowed people to select the most memorably,

exceptionally honourable or charitable behaviours they had witnessed, which in turn may have resulted in a greater desire to emulate that behaviour.

However, thoughts about changing behaviour for remote good memories replicated Experiment 6 in that they were more likely to be general than specific, while thoughts about changing behaviour for recent good memories were not. As in Experiment 6, intentions to change were more general in good conditions than in bad conditions overall.

As in Experiment 6, the hypothesis that good events would be unexpected was not borne out. It appears that good behaviour, even the unusually good, honourable or charitable behaviours participants were asked to generate in these experiments, was not particularly unexpected (though it was not expected either).

Nonetheless, the memories did share several features with the types of memories that produce counterfactuals, as discussed above. The types of good and bad memories recalled were generally similar to those recalled in Experiment 6. That is to say, the memories were actions rather than inactions (more often for good than bad conditions), not more general than specific (and in fact again, they were more specific than general), and controllable (that is, intentional rather than unintentional).

Like elevating experiences, counterfactual thoughts about one's own behaviour can affect change in one's future behaviour by illustrating a plan (e.g. Roese, 1994). The third experiment in this series focused on how thoughts about people's own moral and immoral actions differ from thoughts about another's, and also, if there were differences in each type of memory's tendency to lead people to think of changing their behaviour. As Experiments 4 and 5 investigated the role of emotions as information in moral dilemmas by examining the same question from the perspective of the participant, and another person, Experiment 8 contrasted the effects

of thinking about another person doing something good or bad, and thinking about oneself doing similar things.

Experiment 8

The aim of the third experiment in this series was to examine how people think and reason about morally good and bad actions carried out by themselves, compared to those carried out by others. Elevation is generated when people see, hear about, or remember moral exemplars; this experiment tested whether people create the same kind of memories and counterfactuals, and whether they think in a similar way about the memories, when they think about times when they behaved very honourably or charitably themselves. In the good, self-focused condition, participants were asked to think about an experience in which *they themselves* demonstrated humanity's higher or better nature. Alternatively, in the good, other-focused they were asked to think of an experience in which they saw someone else demonstrating humanity's higher or better nature. Once again good and bad experiences were examined, and the experiment tested how people thought about the memories and how they could have turned out differently.

It was predicted that participants would not think about changing their own behaviour after thinking of a time when they did something good, since there would be no need to correct their behaviour in such cases, but that they would think about doing so after doing something bad. Participants who thought about someone else doing something good should again think of changing their behaviour, more so than participants thinking about someone else doing something bad. The counterfactuals generated in response to good memories should be downward and subtractive, and that in response to bad memories they should be upward and additive. Since

participants have privileged access to their own behaviour, it was predicted that their behaviour would be judged to be more expected than that of others, and also more intentional, but it was predicted that all memories produced would again be actions more than inactions, and that they would be no more general than specific.

Method

Design, Materials and Procedure

There were four groups, who received the same instructions as in Experiments 6 and 7 except that participants in the self-focused good condition were told: “Please think of an experience in which *you yourself* demonstrated humanity’s higher or better nature. Please pick an example in which you were not the beneficiary, that is, *you* did something good, honorable, or charitable for someone else”. Participants in the self-focused bad condition were told: “Please think of an experience in which *you yourself* demonstrated humanity’s lower or worse nature. Please pick an example in which you were not the target, that is, *you* did something bad, dishonorable, or uncharitable to someone else.” In this latter case, participants were not required to write down the memory, in case the requirement to share the details of their own bad behaviour inhibited them from reasoning openly about it. Participants in the other-focused good and bad conditions received the same instructions as in Experiment 6. The materials and procedure were otherwise the same as the previous experiments.

Participants

The participants were 141 members of the general public attending the ‘Happy’ Lab in the Gallery exhibition¹². The participants were assigned at random to one of four conditions: self-focused good (n = 56), self-focused bad (n = 29), other-focused good (n = 35), and other-focused bad (n = 21). There were 59 men and 80 women (and 2 people did not record their gender), whose ages ranged from 18 to 87 years with an average age of 34 years (11 people did not record their age). Approximately two thirds of the participants were Irish and the other third were one of 17 different nationalities and combinations of nationalities (2 participants neglected to provide nationalities, and 2 wrote ‘Caucasian’). A further 31 participants were excluded prior to any data analysis for a variety of reasons. Details on exclusion criteria can be found in Appendix 4.6. As in Experiments 6 and 7, some responses were ambiguous between two categories being compared, and were coded as missing in the analysis. For this experiment, three responses for upward/downward counterfactuals were ambiguous; one memory was ambiguous between being specific or general; and for responses to the questions of whether participants thought about changing their behaviour, and whether that change was general or specific, two responses were ambiguous. No responses for whether the memory was an action or an inaction, or whether the counterfactual was additive or subtractive, were ambiguous. Again, in rare instances in which participants failed to provide a response for a given variable, their response for that variable was also coded as missing.

¹² Twenty-five extra participants (15 in the other-focused good condition, 10 in the self-focused bad condition) were gathered from around the campus of Trinity College Dublin after the run in the Science Gallery but prior to any analysis, in order to ensure close to equal numbers of participants in each condition.

Results

Details on coding and inter-rater reliability for this experiment can be found in Appendix 4.7. Participants recalled events from longer ago in the bad conditions overall ($M = 7.9$, $SD = 10.81$) than the good conditions ($M = 3.22$, $SD = 5.55$), Mann Whitney $U = 1414.5$, $z = 2.643$, $p < .01$, $r = .23$; there was no significant difference in participants recalling events from longer ago in the self-focused conditions overall ($M = 5.58$, $SD = 9.76$) than in the other-focused conditions ($M = 4.04$, $SD = 5.05$), Mann Whitney $U = 1877$, $z = .615$, $p = .539$. Pairwise comparisons showed that the differences were due to participants recalling events from longer ago in the self-focused bad condition compared to the self-focused good condition, Mann Whitney $U = 493$, $z = 2.106$, $p < .05$, $r = .24$, but not in the other-focused bad condition compared to the other-focused good condition, Mann Whitney $U = 292.5$, $z = 1.288$, $p = .198$.

Counterfactuals and Future Intentions

As in Experiments 6 and 7, participants generated more downward than upward counterfactuals in the good conditions, (79% downward for other-focused memories, binomial, $z = 2.97$, $p < .005$, $r = .55$, and 98% downward for self-focused memories, binomial, $z = 6.41$, $p < .001$, $r = .94$), and more upward than downward counterfactuals in the bad conditions (5% downward for other-focused memories, binomial, $z = 3.79$, $p < .001$, $r = .85$, and 0% downward for self-focused memories, binomial, $z = 5$, $p < .001$, $r = .96$). There was also a significant difference between good and bad conditions, $\chi^2(1, n = 123) = 93.093$, $p < 0.001$, $\phi = -0.87$, as in Experiments 6 and 7. There was no difference between the self-focused and other-focused conditions overall, $\chi^2(1, n = 123) = 2.089$, $p = .148$, as Figure 4.7 shows.

As in Experiments 6 and 7, participants generated somewhat more subtractive than additive counterfactuals in the good conditions, although the differences were not significant (59% subtractive other-focused, binomial, $z = .74$, $p = .4593$, and 52% subtractive self-focused, binomial, $z = .14$, $p = .8887$). Participants generated somewhat more additive than subtractive counterfactuals in the bad conditions (28% subtractive, self-focused, binomial, $z = 2.23$, $p < .05$, $r = .41$, and 30% subtractive, other-focused, although this latter difference was not significant, binomial, $z = 1.56$, $p = .1188$). The difference between the good and bad conditions was significant, $\chi^2(1, n = 126) = 8.182$, $p < .005$, $\phi = -0.255$, as in Experiments 6 and 7. There was no significant difference between the self-focused conditions and the other-focused conditions, $\chi^2(1, n = 126) = .202$, $p = .653$. These results are also illustrated in Figure 4.7.

Participants tended to think about changing their behaviour more than not for the good conditions for other-focused memories (79% yes, binomial, $z = 3.25$, $p < .001$, $r = .56$), as in Experiment 6 and 7, but for self-focused good memories they tended not to think about changing their behaviour more than they thought about changing it (33% yes, binomial, $z = 2.24$, $p < .05$, $r = .31$). They thought about changing their behaviour in the bad conditions, as in Experiment 6 and 7. They did not significantly do so for other-focused memories (58%, binomial, $z = .46$, $p = .6455$), but significantly so for self-focused memories (86%, binomial, $z = 3.72$, $p < .001$, $r = .69$). Unlike in Experiments 6 and 7, they thought about changing their behaviour in the bad conditions more than the good conditions, $\chi^2(1, n = 133) = 6.909$, $p < .01$, $\phi = -0.228$, but this result was due to an effect of the focus of the memory. Pairwise comparisons showed that, similarly to Experiment 6 and 7, participants in the other-focused conditions thought about changing their behaviour somewhat more

often in the good condition than the bad condition, although the difference was not significant, $\chi^2(1, n = 53) = 2.781, p = .095$, but participants in the self-focused conditions thought about changing their behaviour more often in the bad condition than the good condition, $\chi^2(1, n = 80) = 20.725, p < .001, \phi = -.509$. These results are also illustrated in Figure 4.7.

For good memories these intentions to change tended to be general more than specific, but not significantly so either for other-focused memories, (63% general, binomial, $z = 1.15, p = .2501$), or for self-focused memories (61% general, binomial, $z = .71, p = .4777$). For bad memories, intentions to change were specific more than general for other-focused memories, but not significantly so (64% specific, binomial, $z = .6, p = .5485$), and general more than specific for self-focused memories, but again, not significantly so (54% general, binomial, $z = .2, p = .8415$). There was no significant difference between good and bad conditions in terms of general or specific intentions to change their behaviour, $\chi^2(1, n = 80) = 1.491, p = .222$; or between other- and self-focused conditions, $\chi^2(1, n = 80) = .731, p = .393$. Pairwise comparisons also showed that intentions to change were not significantly more general than specific in the other-focused good condition compared to the other-focused bad condition, 63% general vs. 36% general, $\chi^2(1, n = 38) = 2.237, p = .135$. This trend may have been significant, but only 11 people in the other-focused bad condition thought about changing their behaviour. There was also no difference between the self-focused good condition and the self-focused bad condition, 61% general vs 54% general, $\chi^2(1, n = 42) = .203, p = .653$, as Figure 4.7 shows.

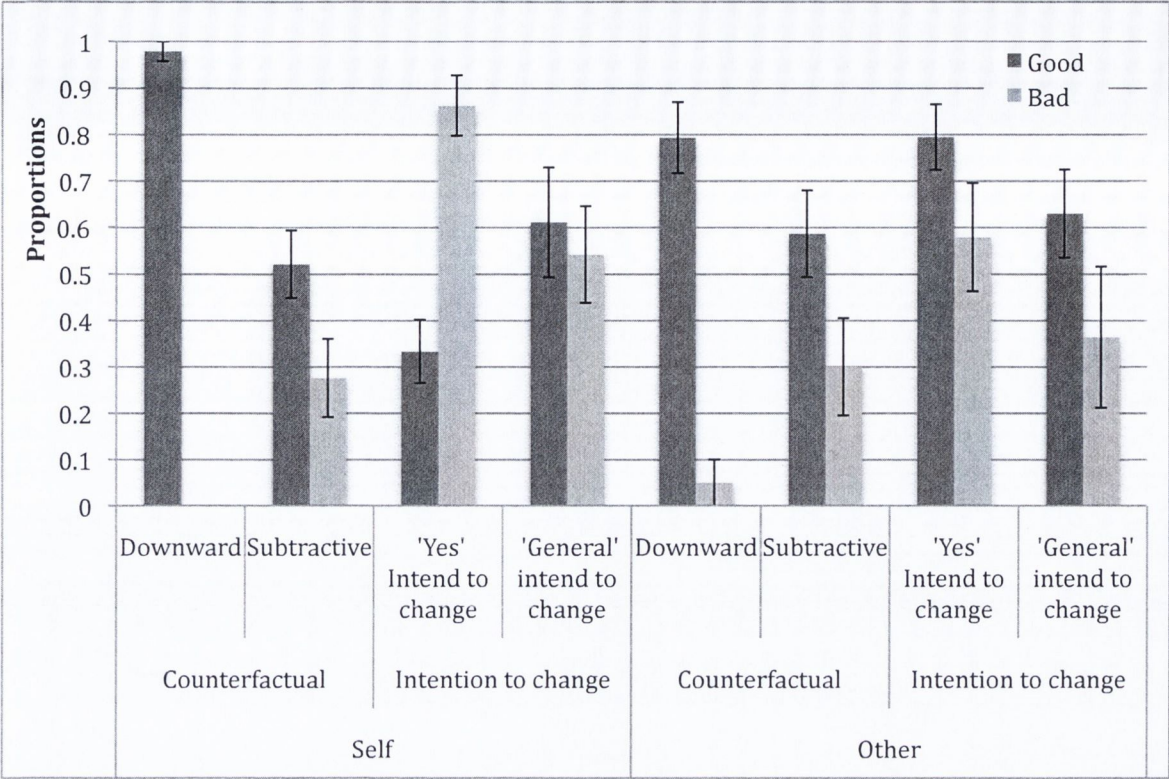


Figure 4.7. Proportions of counterfactuals as downward and as subtractive, and intentions to change, as well as general rather than specific intentions to change, for good and bad memories in self-focused and other-focused conditions in Experiment 8. Error bars are standard error of the mean.

Counterfactual Precursors

The manipulation of the person initiating the behaviour (i.e. self or other) played an important role in how expected the behaviour was. There was no overall difference in how many participants judged the behaviour in their memory to be unexpected (0-4 on the 11-point scale) or expected (6-10 on the scale), binomial, $z = 0.00, p = 1.00$ (6 neutral). Participants judged the behaviour in their memories as expected more often than unexpected for good conditions, binomial, $z = 2.05, p < .05, r = .22$ (4 neutral); in contrast, behaviour in bad memories was judged to be unexpected more often than expected, binomial, $z = 2.62, p < .01, r = .38$ (2 neutral).

More people in self-focused conditions judged the behaviour in their memories to be expected than unexpected, binomial, $z = 2.35$, $p < .05$, $r = .26$ (4 neutral). In contrast, there were significantly more unexpected than expected judgments for other-focused conditions, binomial, $z = 2.75$, $p < .01$, $r = .38$ (2 neutral). Ratings indicated that the behaviour recalled was more expected in the good conditions overall than the bad conditions, Mann Whitney $U = 1431.5$, $z = 3.429$, $p < .005$, $r = .29$, and more in the self-focused conditions overall than the other-focused conditions, Mann Whitney $U = 1396$, $z = 3.958$, $p < .001$, $r = .34$, as Figure 4.8 shows. Participants also rated events as expected in the self-focused good condition more than the self-focused bad condition, Mann Whitney $U = 331.5$, $z = 4.483$, $p < .001$, $r = .49$; however, there was no difference between the expectedness of another person's good and bad behaviour, Mann-Whitney $U = 353.000$, $z = .248$, $p = .804$.

Participants judged events as intentional (6-10 on the 11-point scale) more often than unintentional (0-4 on the scale) overall, binomial, $z = 8.09$, $p < .001$, $r = .7$ (4 neutral). This occurred for good conditions, binomial, $z = 7.78$, $p < .001$, $r = .83$ (1 neutral); bad conditions, binomial, $z = 2.92$, $p < .005$, $r = .43$ (3 neutral); self-focused conditions, binomial, $z = 5.41$, $p < .001$, $r = .61$ (2 neutral); and other-focused conditions, binomial, $z = 6.02$, $p < .001$, $r = .8$. Ratings indicated that, unlike Experiments 6 and 7, behaviour was judged to be more intentional in the good conditions overall than the bad conditions, Mann Whitney $U = 1317$, $z = 4.25$, $p < .001$, $r = .36$, and more in the other-focused conditions than the self-focused conditions overall, Mann Whitney $U = 1592$, $z = 3.352$, $p < .001$, $r = .28$, as Figure 4.8 also shows. Participants judged good events as more intentional both for other-focused conditions, Mann Whitney $U = 264.5$, $z = 2.038$, $p < .05$, $r = 0.27$, and for self-focused conditions, Mann Whitney $U = 344.5$, $z = 4.326$, $p < .001$, $r = .48$.

As in Experiments 6 and 7, memories were categorised as actions more than inactions overall, binomial, $z = 10.44, p < .001, r = .88$, both in the good conditions (100% actions, other-focused, binomial, $z = 5.74, p < .001, r = .97$; 98% actions, self-focused, binomial, $z = 7.09, p < .001, r = .95$) and the bad conditions (95% actions, other-focused, binomial, $z = 3.93, p < .001, r = .86$; 79% actions, self-focused, binomial, $z = 2.97, p < .005, r = .55$). They were categorized as actions more often in the good than bad conditions, $\chi^2(1, n = 141) = 10.036, p < .005, \phi = 0.267$. Events were categorised as actions equally in the self-focused and other-focused experiences overall, $\chi^2(1, n = 141) = 2.624, p = .105$. These results are also illustrated by Figure 4.8.

Events were categorised as specific more than general overall, binomial, $z = 5.32, p < .001, r = .45$. The same pattern occurred in the good conditions (other-focused 80%, binomial, $z = 3.38, p < .001, r = .57$; self-focused 71%, binomial, $z = 2.96, p < .005, r = .4$), and in the bad conditions, although the differences in the bad conditions were not significant (other-focused, 71%, binomial, $z = 1.75, p = .08$; self-focused, binomial, $z = 1.86, p = .0629$). They were not categorised as more specific in either the good conditions or the bad conditions, $\chi^2(1, n = 140) = .321, p = .571$, as in Experiments 6 and 7. They were also categorised as specific equally in the self-focused conditions and the other-focused ones overall, $\chi^2(1, n = 140) = .728, p = .393$. These results are also illustrated in Figure 4.8.

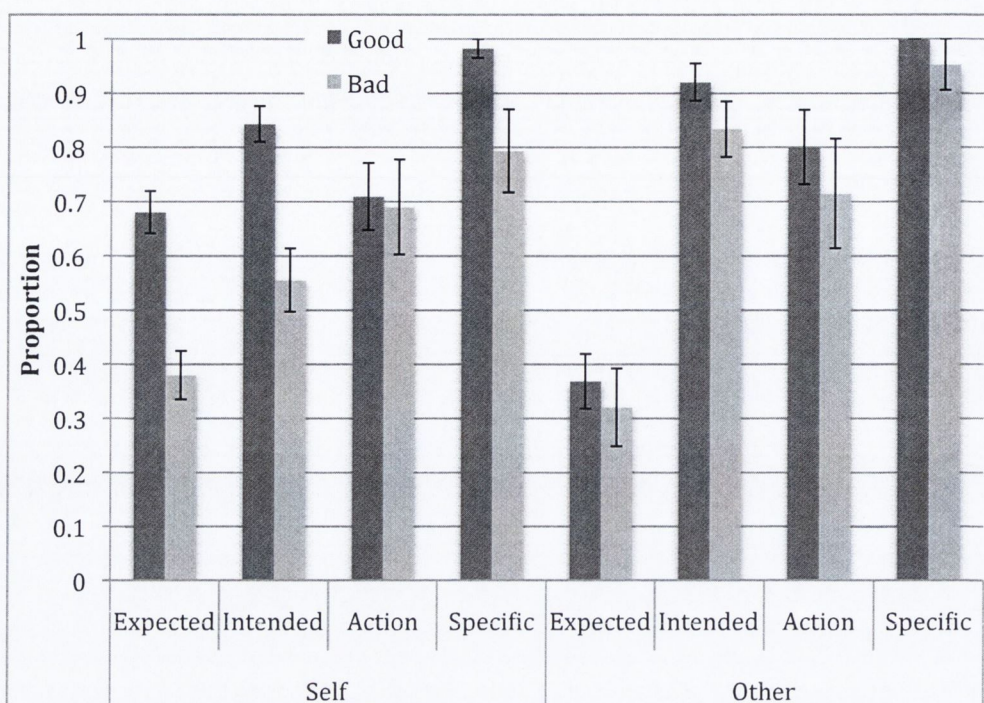


Figure 4.8. Proportions of counterfactual precursors (participants' ratings on 0-10 scale for how expected and intended the behaviour in their memory was, and proportions of participants whose responses were categorized as actions, or as specific) for good and bad memories, in self-focused and other-focused conditions, in Experiment 8. Error bars are standard error of the mean.

Emotional Consequences

Unlike in Experiments 6 and 7, participants' ratings of their mood improved from before to after thinking about the events in the good conditions overall, Wilcoxon, $z = 4.413$, $p < .001$, $r = .34$, and once again it deteriorated for the bad conditions overall, Wilcoxon, $z = 3.081$, $p < .005$, $r = .32$, as illustrated by Figure 4.9. There were no differences between before and after ratings of mood for the self-focused conditions overall, Wilcoxon, $z = .349$, $p = .727$, or the other-focused conditions, Wilcoxon, $z = .03$, $p = .976$, again as shown by Figure 4.9. The figure also shows that participants described their feelings using emotion words corresponding to

positive emotions more than negative or ambiguous words in the good condition more than the bad condition, $\chi^2(1, n = 137) = 99.9, p < .001, \phi = .854$; there was no significant difference between other- and self-focused conditions, $\chi^2(1, n = 137) = 1.852, p = .396$.

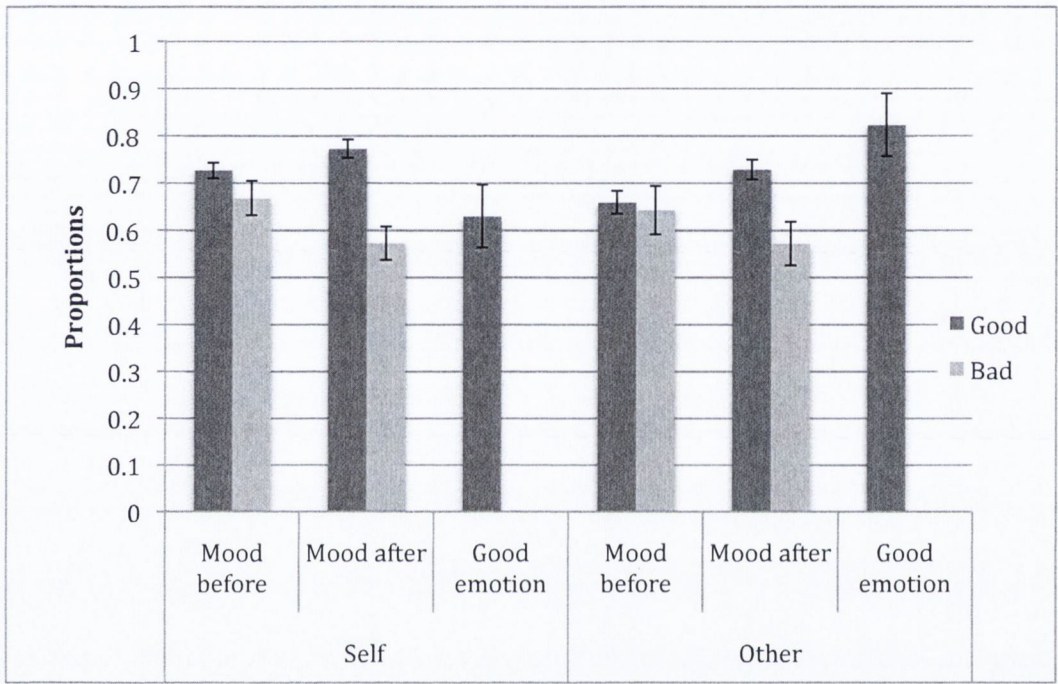


Figure 4.9. Proportions of results for mood before recording the memory and mood after (these are based on the ratings on the 0-10 scale participants provided), and of positive single words to describe the participants' feelings, in good and bad memories, in self-focused and other-focused conditions for Experiment 8. Error bars are standard error of the mean.

Discussion

Experiment 8 generally replicated the results of Experiments 6 and 7. Participants again tended to produce downward and subtractive counterfactuals in good conditions, compared to upward and additive counterfactuals in the bad

conditions. The most notable new result in this experiment was that participants were more inclined to form the intention to change their own behaviour in the self-focused conditions after doing something bad rather than good. Participants thought about changing their behaviour more when seeing or hearing about someone else doing something good rather than something bad, as in Experiments 6 and 7; however, the opposite effect occurred when participants focused on their own past behaviours. The tendency for intentions to change to be more general for other-focused good memories and more specific for other-focused bad memories that was observed in the previous experiments emerged in a trend, but was not significant for this experiment; this may have occurred due to the relatively few participants included in this test in the other-focused bad condition (i.e., 11 of 21).

As in Experiments 6 and 7, the memories participants recalled were actions rather than inactions (more often in the good than the bad conditions), not more general than specific (and in fact, again more specific than generally, equally in the good and the bad conditions), and intentional rather than unintentional (and more so for good conditions than bad, both when these memories focused on other people's behaviour, and when focused on participants' own behaviour). Good memories were also more expected than unexpected, but this difference was mediated by participants expecting their own good behaviour more than their own bad behaviour; there was no difference in expectation for the good and bad behaviour of others, as in Experiment 7, but differently from Experiment 6.

Participants also recalled events where they did something bad from longer ago than when they did something good, a difference that did not occur for other-focused events. This indicates that such events, which may be tied to regrets, have a special hold on people's memories compared to good things they have done.

Participants' memories of themselves doing good and bad things resembled, for the most part, their memories of other people doing good and bad things. They shared the features of events that are likely to produce spontaneous counterfactual thoughts, and when they were prompted to produce such thoughts about the memories, participants' responses followed the same pattern regardless of the agent involved. The main difference occurred in their thoughts about changing their behaviour. Naturally, participants were inclined to think about changing their own behaviour after doing something bad, but not something good; however, as in previous experiments, thinking about other people doing something good prompted them to think about changing their behaviour more than thinking about other people doing something bad. These results suggest that, while an emotional reaction to witnessing someone doing something good or bad may be involved in leading people to think about changing their behaviour or not, salient features of the memory such as the person involved are also taken into account.

General Discussion

The experiments showed that people create different counterfactual alternatives to reality when they imagine how morally good and bad experiences could have turned out differently. Participants tended to imagine how things could have been worse when someone did something good, and they did so by creating an alternative to reality from which something had been subtracted. In contrast they tended to imagine how things could have been better when someone did something bad, and they did so by creating an alternative to reality to which something had been added. These differences held in Experiments 6, 7 and 8, regardless of the length of time since the memory, or the person involved in the memory.

They thought about changing their behaviour after witnessing good actions more than bad actions, except when they were the ones carrying out the good and bad behaviours, in which cases they thought about changing their behaviour more in the light of bad actions. Thoughts about changing their behaviour were also influenced by the length of time since the memory. Thoughts about changing behaviour were prompted by memories from longer ago more often than by memories from more recently. Memories from longer ago may have involved more exceptional, memorable behaviour, which in turn may have prompted people to think more about changing their own behaviour. This result may suggest that differences in how exceptional moral behaviour is can affect how likely such behaviour is to prompt people to think about emulating it.

Although memories of people doing good things for others were generally more likely than memories of people doing bad things to others to result in thoughts about changing one's behaviour, participants' descriptions of what they decided to change were general rather than specific for the good conditions more so than the bad conditions. This difference occurred in Experiments 6 and 7, and a trend, though not a significant one, also occurred in the other-focused conditions in Experiment 8. These results provide some support for the hypothesis presented in this chapter that the type of intentions to emulate good behaviour that arise from hearing about other people doing good things may be general rather than specific, and therefore may be difficult for people to implement. As discussed at the beginning of this chapter, these results may help to explain the fact that people do not always follow through on intentions to emulate moral exemplars. Furthermore, downward counterfactuals, such as those generated in response to memories of people doing morally good things for others, are known to be less helpful in planning for the future than upward counterfactuals (e.g.

Roese, 1994), and subtractive counterfactuals, again generated in response to good memories, are less helpful in generating creative thinking (e.g. Markman et al., 2007). It may be conjectured that the generation of such counterfactuals may lead people not to think in detail about how they could emulate the behaviour of moral exemplars.

The results also provided some support for the idea that morally good and bad actions correspond to the sorts of events that spontaneously elicit counterfactuals. They were intentional rather than unintentional; they were actions rather than inactions, and they were specific rather than general (supporting the hypothesis that they would not be more general than specific). These memories did not match all predicted features of counterfactual precursors, however; memories of people doing honourable or charitable things were not judged to be unexpected to the extent that was predicted, whereas memories of dishonourable or uncharitable things were. On the other hand, only self-focused good behaviour was judged to be more expected than unexpected, indicating that such behaviour was only *expected* when participants had privileged access to the thought processes behind the behaviour. These studies do therefore provide some evidence that such morally good and bad memories are the kind that prompt people to generate counterfactual alternatives.

Although past studies indicated that elevating memories improved mood immediately after recall (e.g. Algoe & Haidt, 2009), thinking about bad experiences affected participants' mood more than thinking about good experiences, and mood only improved for good conditions in Experiment 8, and not in Experiments 6 and 7. These results suggest that thinking about memories of other people doing good things in more detail sometimes does not result in elevated mood. It may be the case that any immediate improvement in mood dissipated as participants considered the variety of questions they were asked in these experiments.

Alternative Views

It is suggested that the translation of elevation into emulation involves a set of cognitive processes as a key element, and that these processes consist of two primary components: (a) the creation of counterfactual alternatives to reality, and (b) the formulation of intentions to emulate. The observed event evokes an emotional response, but on this account further processes are required to translate the elevating experience to emulation. Emulation requires not only the emotionally uplifting experience to inspire the immediate desire to be better, but eventual change through a set of cognitive processes that comprise the creation of counterfactual alternatives that provide the blueprint for the formulation of future intentions.

An alternative view is that an emotional response of elevation leads to emulation by turning people's perspectives to others and making a broader range of thoughts and actions available to them (e.g. Haidt, 2000). This process may well play a role in encouraging people to behave more pro-socially after witnessing or hearing about a moral exemplar. Furthermore, emotions and unconscious processes certainly can influence moral judgments and behaviour; as discussed in Chapters 1-3, such processes are often very influential in moral judgments, whether they are activated due to features of the scenario being considered, or through affective priming (e.g. Haidt, 2001; Greene et al., 2001; Wheatley & Haidt, 2005; Luo et al., 2006; Valdesolo & DeSteno, 2006; Monin et al., 2007).

However, cognitive processes also interact with automatic emotional processes, and are differentially relied upon to solve different moral problems (e.g. Greene et al., 2004, 2008; Moore et al., 2011; Paxton et al., 2012). Strikingly, as discussed at the beginning of this chapter, elevation leads people to think about acting

pro-socially or emulating good actions more than even another other-praising emotion such as gratitude (e.g. Algoe & Haidt, 2009). This suggests that positive emotion is not the only process required in order to transition from feeling uplifted to wanting to change one's own behaviour. Although emotions seem to play a key role in moral judgments, and emotions such as moral elevation may help to influence the transition from moral judgment to moral action, the evidence suggests that cognitive, non-emotional processes also play a part in both judgments and this transition. The experiments presented in this chapter indicate that the generation of counterfactual thoughts may be a key component in this transition.

Implications

Although this chapter supported previous work in indicating that witnessing or hearing about moral exemplars leads people to think about changing their behaviour, it also suggested that more is needed than an immediate emotional reaction in order to trigger such emulation. While the influence of an emotional reaction such as that described by previous accounts of elevation (e.g. Haidt, 2000; Algoe & Haidt, 2009) was supported, the experiments further identified how domain-general cognitive processes such as the generation of counterfactual alternatives, and the identification of the person acting morally, may be an integral part of the tendency to behave morally after being exposed to moral exemplars. As discussed in Chapter 2, the observation of domain-general cognitive processes in moral judgment provides support for the idea that moral reasoning is constituted by the same processes as non-moral reasoning (e.g. Sunstein, 2005; Bucciarelli et al., 2008), rather than by domain-specific processes (e.g. Dwyer et al., 2010). The experiments reported here also

provide some support for the idea that unconscious emotional and controlled cognitive processes interact in moral judgment, as discussed in Chapter 3.

The results of these experiments also have implications for everyday reasoning. The theory proposed in this chapter has suggested that formulating specific, rather than general intentions to emulate good behaviour may be key in determining whether or not people can easily implement their intentions. However, the experiments found evidence that people often generate general, rather than specific, thoughts about changing in response to moral exemplars. It may be speculated that people could possibly improve their moral behaviour by making a conscious effort to formulate specific rather than general intentions when morally inspired. This is a question that may be answered by future research.

An important limitation of the experiments reported here is that they did not establish a causal role for the generation of counterfactual alternatives in the transition from elevation to emulation. Instead, the studies establish certain regularities in the types of counterfactuals that are generated, as well as establishing that the kind of memories that prompt elevation share features with the kinds of scenarios that prompt people to consider counterfactual alternatives. Future studies could usefully attempt to test whether such a causal relationship exists. One possible method for testing this relationship would be think-aloud studies; these can provide access to some of the intermediate steps in thinking, particularly those that deliver thoughts to conscious awareness.

In the next chapter the results of this chapter, and the preceding chapters, are discussed in detail. The discussion focuses especially on how these results relate to the psychological debates of the roles of emotional, unconscious and conscious

processing in moral judgment, and whether moral reasoning is domain-general or domain-specific. The relevance of the results for everyday reasoning is also discussed.

Chapter 5 Discussion

The primary goal of this thesis was to investigate some of the cognitive processes underlying moral reasoning. The experiments reported in the chapter did this by investigating consistency in moral judgment (Chapter 2), the role of emotions as information in moral judgment (Chapter 3), and the role cognitive processes such as counterfactual generation may play in motivating people to emulate moral exemplars (Chapter 4). By investigating these issues, the thesis has touched on a series of psychological debates. The results of Chapters 2 and 4 relate to the debate over whether moral reasoning is similar to other types of reasoning, or different. The results of Chapters 3 and 4 relate to the question of how emotional and cognitive processes interact in moral reasoning. This information is useful not only academically, but also practically, in identifying the processes that are influential in everyday moral reasoning, and suggesting areas of moral thought that may be appropriate for psychological intervention.

In this discussion, the implications of the experimental results reported in the preceding chapters for psychological theories of moral reasoning, for practical reasoning, and for the applicability of normative theories will be discussed. In what follows, the findings of the three sets of experiments reported throughout this thesis are summarised, along with the important conclusions that can be taken from them. The following three sections will summarise the results of the three experimental series (i.e. the first section will focus on the experiments on moral consistency, the second on the experiments on the role of emotion as information in moral reasoning, and the third on the experiments on elevation). Each section will summarise the

primary discoveries of each series, then consider the implications of these discoveries for psychological theories, and then discuss their practical and normative implications.

Moral Consistency

It seems likely that people believe that their moral reasoning is consistent; certainly there is evidence that people often view their moral judgments as objectively true (e.g. Goodwin & Darley, 2008), a criterion that is much more stringent than that of consistency within one's own judgments. In Chapter 2, the role of heuristic processes in moral judgment was investigated, specifically with regard to the question of how people make individual and conjoined numeric moral judgments, and whether such judgments are consistent. Three experiments (Experiments 1-3) were reported that demonstrate that numeric moral judgments sometimes violate consistency, and that they tend to do so in similar ways to judgments in other domains.

The basic task in these experiments investigated how participants judged individual moral and immoral behaviours, compared to pairs of such actions. In probability judgments, participants often violate the principle of conjunctions: that is, that the probability of two events occurring together cannot be greater than the probability of either occurring individually (e.g. Tversky & Kahneman, 1983; Khemlani et al., 2012). The analogous principle in morality is that two moral actions should be judged to be more moral than either individually, and two immoral actions should be judged to be more immoral than either individually. Experiments 1 and 2 tested whether people employed this principle when making moral judgments; that is, whether their judgments matched this principle. Thus, using numeric scales from 'Most immoral' to 'Most moral', participants were asked to judge how moral or immoral actions were individually, and in conjunction. For example, in Experiments 1

and 2, participants made judgments of individual behaviours such as, “*A babysitter beat the infant in her care*”, and “*A babysitter stole her employer’s jewellery*”, and of the conjunction, “*A babysitter beat the infant in her care, and stole her employer’s jewellery*”. In these experiments, participants generally succeeded in keeping their judgments congruent with the above principle, but they also reliably committed violations.

More importantly, in both experiments consistency was improved when the conjunction was judged after the individual conjuncts as opposed to before, and in Experiment 2, when judgments were made on a fine scale instead of a coarse one. These experiments showed for the first time that moral consistency could be improved simply by allowing people to think about the morality of each action separately before thinking about them together. This result is also consistent with results from research on probability, in which consistency was also improved by judging the two individual items first. This suggests that the same processes may be involved in judgments from both domains.

In Experiment 3, participants made judgments of mixed-valence pairs of moral and immoral actions, and their judgments were compared to a similar principle: that pairs of mixed moral and immoral actions should not be disproportionately influenced by either moral or immoral actions, and so ratings of a conjunction should be equal to the sum of the ratings of the items individually. Participants made judgments of how moral or immoral actions were individually, such as, “*A person forced the foreign girl who was working for him to accept a salary below minimum wage*”, and “*A person organized Christmas gifts for children at an orphanage*”, and in conjunction, “*A person forced the foreign girl who was working for him to accept a salary below*

minimum wage, and the next week organized Christmas gifts for children at an orphanage”.

Participants reliably gave greater influence to the immoral in their judgments of mixed-valence conjunctions than their judgments of the individual items indicated they should. This result is consistent with results from across many different domains in which negative stimuli were given greater influence than positive stimuli (e.g. Baumeister et al., 2001). In contrast to the results for same-valence moral judgments, consistency was not improved by considering the conjunction last compared to first.

These data show that when people judge the morality of a conjunction of a good and a bad action, they tend to be inconsistent, in that their judgments of the conjunction exaggerate the immorality of a pair of actions. These data show for the first time that they do so not merely compared to how the individual actions are judged by other people (e.g., Birnbaum, 1973; Risky & Birnbaum, 1974) but more strikingly, even compared to how they themselves judge the individual actions.

Psychological Implications

These results have implications for theories of moral judgment, particularly with regard to the differences and similarities between moral and non-moral judgments. Some past evidence has suggested that moral judgments are likely to be relatively immune to the kind of heuristic effects that influence judgments in other domains. For instance, the order in which dilemmas are presented generally does not affect responses to the dilemmas (e.g. Dwyer et al., 2010), and some moral distinctions, such as the difference in permissibility between personal and impersonal dilemmas, generalise cross-culturally (e.g. Mikhail, 2007; Moore et al., 2011). This evidence has led to the suggestion that moral judgments are the product of a domain-

specific, innate moral grammar (e.g. Mikhail, 2007, 2009; Hauser, 2006; Dwyer et al., 2010).

However, the evidence reported in Chapter 2 conflicts with this idea. The results of these experiments suggest that domain-general processes such as heuristics influence moral judgments. The primary results discussed in Chapter 2 have direct parallels in other domains, and appear to bias judgments in a similar fashion to heuristics in other domains.

Some researchers have objected to the concept of moral heuristics on the basis that previous discussions of such heuristics (e.g. Sunstein, 2005; Baron, 2010) have taken different forms of utilitarianism as the correct view of moral judgment against which bias should be measured, even though it is not necessarily the case that everyone would agree with such theories (e.g. Waldmann et al., 2012). However, the experiments reported in Chapter 2 do not require such a strong moral commitment; indeed, the form of internal consistency tested in these experiments is something any moral theory would require in order for people to maintain coherence between past and future judgments. The fact that a simple manipulation such as a change in the position of conjunction influenced moral judgments, as it did judgments of probability, is a strong objection to the idea that moral judgments are not influenced by the same kind of heuristics that affect judgments of non-moral items (e.g. Dwyer et al., 2010), and strengthens the idea that the same processes influence moral and non-moral cognition. The result that mixed-valence moral judgments tend to be disproportionately influenced by an immoral item (Experiment 3), as occurs in many other domains of judgment (e.g. Baumeister et al., 2001), further suggests that moral judgments are similar to non-moral judgments.

Normative and Practical Implications

These results also have implications for the application of normative philosophical theories of moral judgment, which describe what people *should* do morally. The results reported in Chapter 2 suggest that people will have difficulty remaining consistent with the type of all-encompassing systems suggested by utilitarianism (e.g. Mill, 1863/2007), or deontological systems (e.g. Kant, 1788/2002). Since people cannot always maintain consistency within triples of judgments about a single set of moral actions, it seems unlikely that they would be capable of, for example, consistently acting to increase utility for everyone. As discussed in Chapter 2, this issue does not suggest that any particular philosophical systems are incorrect in identifying what people *should* do, but such systems are likely to remain impractical guides for behaviour, given that heuristics sometimes appear to limit consistency.

The results of Chapter 2 also have practical implications for moral reasoners. The presence of inconsistency in judgments of same-valence moral actions suggests that sometimes when people commit multiple immoral actions, those actions may not be judged as immoral as individual judgments of those same items indicate they should be. A similar effect may occur when people carry out multiple moral actions. This effect occurred more often when the conjunction was read before the individual items. The effect of the position of conjunction in Experiments 1 and 2 demonstrates that, when people make judgments of pairs of moral actions, they sometimes judge subsequent individual actions to be just as good (and a similar effect occurs for bad actions).

These latter results suggest that, if people have made multiple judgments of immoral actions, they may sometimes judge subsequent, similar immoral actions to be just as bad as the ones they originally judged. A similar effect may occur for

judgments of morally good actions. This may result in moral judgments sometimes being influenced by what order people hear about them in. For example, if someone hears about one person who attacked a man and stole money from him, and subsequently hears about another person who just stole money from a man, they may judge the individual action to be just as immoral as the pair of actions. On the other hand, if they heard about the person who just stole money first, they might judge that action to be less morally wrong than the action of attacking and stealing money from a man.

Alternative Views

A possible explanation of the results of Experiments 1 and 2 is that, when people made judgments of pairs of moral actions, or pairs of immoral actions, before judging the subsequent actions, they may have made judgments as though all actions were committed by the same person, rather than judging the actions separately. Thus, when they read about the pair of actions first, they may have formed an impression of the person's character as a good person (for a pair of good actions), or as a bad person (for a pair of bad actions). These impressions, if formed, may have led people to judge subsequent individual actions to be as good, or as bad, as the pair because they viewed the person as good or bad. This interpretation makes sense in terms of explaining why people sometimes judge, for example, individual good actions to be as good as a pair of good actions when they judge the pair first.

However, such an interpretation does not render such judgments rational, nor does it undermine the idea that the results of Experiments 1 and 2 provide evidence of heuristic processes operating in moral judgments. Regardless of the judgments people

make of the moral character of others, two good actions should still be judged to be better than one in order for one's judgments to remain rational and consistent.

Although these experiments did not measure the blame, praise or responsibility of people, and focused instead on individual actions, the findings of Experiments 1 and 2 may be tentatively extrapolated to how people make judgments of other people (rather than of their actions). It may be the case that people who do many bad things, such as criminals who offend repeatedly, could be judged more harshly for subsequent bad actions than they would be for their first bad action. Similarly, people who often do good things may be judged to have done something better for subsequent individual good actions than they would be without their reputation. Although it may seem rational that someone who often does bad things should be more of a cause for concern than a one-time offender, or that someone who often does good things should have their reputation burnished, it is possible that people may be blamed or praised disproportionately due to their past immoral and moral deeds.

Similarly, Experiment 3 provided evidence that, when judging both good and bad actions, the bad actions persistently outweigh the good in people's judgments. Only when someone has done an extremely good thing and a negligibly bad thing is this effect eliminated. As with Experiments 1 and 2, although this experiment focused on the morality of actions rather than judgments of praise, blame or responsibility, it may be tentatively suggested that once someone has done something immoral, they may face an uphill battle in attempting to reform their reputation. Although a case can be made that people who act immorally are likely to do so again, it may also be true that people who act morally are likely to do so again. It is tentatively suggested therefore, that judgments that overemphasise the immoral, such as those in

Experiment 3, may make the road to redemption more difficult for those who have made a mistake, and perhaps even make redemptive behaviour seem less appealing than continuing to act immorally.

The experiments reported in Chapter 2 identified three factors that seem to unconsciously influence consistency in moral judgments: whether people judged a pair of moral judgments before or after its individual components; what kind of scale such judgments were made on; and a disproportionate influence of immoral rather than moral actions in mixed-valence judgments. The next section of this discussion focuses on the experiments reported in Chapter 3, which investigated the role of emotions as information, and how they may influence conscious and unconscious processing in moral judgment.

Emotion as Information in Moral Reasoning

The experiments reported in Chapter 3 investigated the role of information about other people's emotions in moral reasoning. It was hypothesised that participants would sometimes provide information about their own emotions when justifying their own moral decisions, and that the provision of such information would be influential when reasoning about the rightness and wrongness of other people's actions.

It was also hypothesised that these influences would vary depending on the type of scenario involved. To test the role of information about emotions in moral judgments, the experiments reported utilised a common distinction in moral psychology: that is, the distinction between the permissibility of personal and impersonal harm. As discussed in Chapter 3, personal harm, involving direct, physical contact with the person to be harmed, is usually considered to be less permissible than

impersonal harm, and some studies support the idea that the cause of this difference may be the fact that personal dilemmas result in greater emotional engagement than impersonal dilemmas (e.g. Greene et al., 2001; Valdesolo & DeSteno, 2006; Koenigs et al., 2007; Ciaramelli et al., 2007).

As discussed in Chapter 3, the justifications people give for their moral judgments are sometimes insufficient to explain said judgments, yet people often do not abandon these moral decisions in the face of contravening evidence (e.g. Cushman et al., 2006; Haidt, Murphy & Bjorklund, 2000). However, people sometimes *can* successfully justify their moral judgments, and the discussion of moral justifications (whether post-hoc or not) is likely to be a key factor in conscious moral reasoning (e.g. Haidt, 2001; Paxton & Greene, 2010). Due to the important role that moral justifications may play in social reasoning, Experiment 4 asked participants to make a decision in a moral dilemma, and subsequently to justify that decision to a friend.

Participants provided justifications involving direct or indirect references to emotions more often for the personal footbridge problem than for the impersonal trolley problem when they were given no direction. However, when asked to refer to their feelings in their justifications, there was no difference in emotive justifications between the trolley and footbridge dilemmas, and when they were directed to provide reasoned justifications, participants actually provided significantly more emotive justifications for the trolley than the footbridge problems. These results suggested that participants understood the act of pushing the switch in the trolley problem to be emotive, even if not usually to the same degree as the act of pushing the man in the footbridge problem. The experiment provided the novel finding that when people judged whether to act or not when faced with a moral dilemma that pitted the violation of a moral principle against a potentially more beneficial consequence, they

could access emotive or non-emotive justifications as readily for impersonal moral dilemmas as for personal ones.

Experiment 5 also examined the role of emotions as information in moral dilemmas, but from a different perspective: participants were asked to make judgments of actors, in a range of diverse moral dilemmas, who experienced emotions before making their decision. These dilemmas included some involving violations of harm and fairness, and both personal (like the footbridge problem) and impersonal (like the trolley problem) versions. Participants were provided with information about the strength of the emotions the actors felt, and about the decisions they made. The experiment investigated whether the time taken to read about the emotion experienced by the actor in the dilemma varied according to the dilemma type (personal or impersonal) and the emotion type (strong or mild). It also investigated whether participants were primed to form expectations by reading information about the protagonists' emotions, and thus to read certain decisions more quickly than others. This was tested by recording the time taken to read about the decision made by the actor in the dilemma. The time taken to make a judgment of the appropriateness of the decision, and the judgment itself, were also recorded.

Participants read about the emotions experienced by actors faster for personal than for impersonal dilemmas, regardless of the strength of the emotions, indicating that participants considered emotional responses to be more appropriate for personal than impersonal dilemmas. However, when participants were subsequently given information about the decision the actor made, the three variables (dilemma type, emotion strength, and decision type) interacted in reading times for the decision sentence. For impersonal dilemmas, participants were primed by reading about strong emotions to read quickly that the protagonist decided to act compared to the decision

not to act. This is a novel result, indicating that when people know someone experienced a strong emotion when faced with an impersonal moral dilemma, such as harm caused indirectly by hitting a switch, they expect them to act. The result suggests that strong emotion, rather than simply prohibiting action, can enhance the decision to act in a moral dilemma.

In contrast, for *personal* dilemmas, participants were primed by reading about *mild* emotions to read quickly that the protagonist decided to act, as opposed to not act. This result showed for the first time that when people know a person experiences only a mild emotion when faced with a moral dilemma involving personal harm, such as directly by pushing a man, they expect them to act. The result suggests that mild emotions were not perceived as inhibiting acting in moral dilemmas.

The results of Experiment 5 indicate that expectations of the decision to act rather than not act are influenced differentially by different emotion-types, and particularly when people read about protagonists experiencing incongruent emotions. The results suggest that being primed by the knowledge of what kind of emotions actors experienced led people to form inferences of what actions they would take. In impersonal dilemmas, reading that the actor felt strong emotions prompted people to think they will act, whereas in personal dilemmas reading that the actor felt *mild* emotions prompted participants to think they will act. The first of these effects suggests that people think of strong emotions as a key component of action in impersonal dilemmas, even though they may experience less emotion when considering such dilemmas than they would for other scenarios, such as personal moral dilemmas. The second effect suggests that people may understand someone who only experiences mild emotions in personal dilemmas, such as when they are considering pushing a man off a bridge to save five others, to be lacking in a key

component of what usually prohibits people from acting in such a situation. If emotions are indeed part of the reason people tend to judge action to be impermissible in personal moral dilemmas, it may be the case that people consider someone who does not experience strong emotions in such cases to be particularly likely to act. As discussed in Chapter 3, people with lesions in areas of the brain associated with emotion tend to choose to carry out the harm in personal dilemmas in order to save the most possible lives, more than normal controls (e.g. Koenigs et al., 2007; Ciaramelli et al., 2007). The participants in Experiment 5 seemed to believe that people lacking in the usual emotional response to such dilemmas would also be likely to fail to produce the usual behaviour response.

Psychological Implications

The results of Experiments 4 and 5 have implications for theories of moral reasoning. Sometimes the features of a moral problem (e.g. Haidt et al., 2000; Greene et al., 2001; Pizarro et al., 2003; Monin et al., 2007), or irrelevant contextual details such as a disgusting odour, or having been shown a video designed to elicit a particular emotion (e.g. Schnall et al., 2008; Valdesolo & DeSteno, 2006; Strohminger et al., 2011) lead people to respond to moral scenarios either emotionally, or non-emotionally. It is suggested that information about emotions experienced by actors in moral dilemmas may make certain possibilities more and less salient. Participants may form expectations of what an actor will do when faced with a difficult choice based on the type of dilemma the actor is in, the type of emotions the actor experiences, and the decision the participants themselves would make. For example, because strong emotions are the usual response to personal dilemmas, when participants read about someone experiencing mild emotions in such a dilemma, it

may make salient the possibility that they will carry out the unusual behavioural response of acting rather than not acting.

These experiments provide some support for theories such as the social intuitionist model that emphasise the role of emotion in moral judgment (e.g. Haidt, 2001; Haidt & Bjorklund, 2008). However, they also provide support for a more integrated view of emotions, unconscious and conscious processing, such as that suggested by the idea that the features of different moral scenarios influence the roles of emotions or cognitive deliberation (e.g. Greene et al., 2004; 2008; Monin et al, 2007; Bucciarelli et al., 2008). The results demonstrate that people can provide emotive or non-emotive justifications for their judgments, and that they may form expectations of actions based on emotional primes. These results suggest that emotions operate not only as reactions that guide moral judgment one way or another, such as the emotional reactions that appear to lead people to judge incest to be impermissible (e.g. Haidt, 2001), or to forbid pushing the man over the footbridge (e.g. Greene et al., 2001), but also as informational inputs into moral reasoning. Information about the type of emotions actors experienced in Experiment 5 affected reading speeds differently for different dilemma and decision types, suggesting that decisions were made following reasoning that involved information about the actors' dilemmas, their emotions, and their actions.

Participants also used emotional information in their justifications of their own actions in Experiment 4, suggesting they consider such information to be a relevant input into moral reasoning, and perhaps a factor that mitigates any blame that might be attributed to their actions. Along with the evidence of Experiment 5, this result suggests that although emotions may often play the instant, intuitive role indicated by the social intuitionist model (e.g. Haidt, 2001), they might also be considered as a

component of moral reasoning. This interpretation suggests a view of emotions being more integrated with slower cognitive processing about morality than indicated by the social intuitionist model, and perhaps something closer to the dual-process model (e.g. Greene et al., 2004, 2008). It is also consistent with the view that moral evaluations may precede emotions and that emotions may precede moral evaluations in moral judgments, and that they may interact (e.g. Bucciarelli et al., 2008).

Practical and Normative Implications

The results of Experiments 4 and 5 also have practical implications. As discussed in Chapter 3, emotions are often considered as inputs into judgments of guilt; for instance, judges may reduce sentences by reason of temporary insanity, or condemn a criminal for expressing no remorse for their crimes. Such judgments cohere with the evidence presented in Chapter 3, and indicate that people consider emotional experiences to be legitimate inputs into reasoning about moral rightness and wrongness.

Some important philosophical theories of morality have suggested that moral judgment is nothing but emotion (e.g. Hume, 1739-1740/2004). While the evidence of conscious moral reasoning discussed throughout this thesis (e.g. Moore et al., 2008; Paxton et al., 2012) contradicts this strong sentimentalist position, the results of Experiments 4 and 5 indicate that people seem to believe that emotions do have a role in terms of mitigating judgments of appropriateness. Most modern psychological theories of moral judgment also suggest that people sometimes experience emotions that may influence their judgments. It may not therefore be practical or plausible for people to make all their judgments to cohere with a broad deontological or utilitarian system. As discussed with regard to Experiments 1-3, normative systems need not

cohere with psychological realities. However, the reality that emotions often influence moral judgments may make it difficult for most people to cohere with such systems, even if they desired to do so.

Alternative Views

Due to the preponderance of evidence that emotions often appear to play a role in moral decisions (as discussed in Chapters 1 and 3), most, if not all, modern psychological theories of moral reasoning hypothesise some sort of causal role for emotions, integrated in some way with cognition. Theories that emphasise dual-processes of emotional and controlled cognition (e.g. Haidt, 2001; Greene et al., 2008; Bucciarelli et al., 2008), as well as the moral grammar theory (e.g. Mikhail, 2011; Dwyer et al., 2010) acknowledge that emotions interact in some way with cognitive processes about moral behaviour (the latter theory suggesting that emotions may play a particular role in motivating moral behaviour). The results of Experiments 4 and 5 therefore do not appear to definitively rule out any prominent theory of moral reasoning, although it is unclear if the moral grammar theory could predict that participants would read emotion-incongruent decisions faster than emotion-congruent decisions, as was observed in Experiment 5. This latter theory is sceptical as to the causal role of emotion in moral judgment (e.g. Dwyer et al., 2010), a view that seems inconsistent with the interpretation put forward here that emotional information causes people to form inferences as to the moral decisions people will make (i.e. Experiment 5).

The most important contribution of these experiments is to help to identify how emotions may be integrated into justifications and expectations of how people may act in moral dilemmas. The next section discusses the experiments reported in

Chapter 4. These experiments also examined the interaction of emotional and cognitive processes in moral judgment. They focused on the moral emotion of elevation, and how this emotion, along with a set of cognitive processes, may encourage people to behave morally.

Elevation, emulation and counterfactual thoughts

The experiments reported in Chapter 3 discussed how information about emotions might be used as inputs into moral reasoning. Other recent research has begun to investigate specifically moral emotions; that is, positive emotions that arise when someone does something good. These emotions may influence behaviour in different ways; for instance, gratitude, an emotional response to people doing good things for oneself, can inspire people to thank and praise the person who has done the good thing (e.g. Algoe & Haidt, 2009). Chapter 4 examined the moral emotion of elevation, which is elicited when people see or hear about moral exemplars (that is, people doing good things for other people). It is characterized by a sense of uplift, and people are sometimes inspired to emulate the behaviour of the exemplar after being elevated (e.g. Haidt, 2000; Schnall et al., 2010). The experiments reported in Chapter 4 tested how the emotion of elevation, and the generation of counterfactual alternatives to reality, may both be involved in the generation of intentions to emulate good behaviour.

Participants were asked to think of a situation where someone either did something good *for* someone else, or bad *to* someone else. They were then asked to generate a counterfactual thought about their memory, followed by a series of questions about the memory, such as whether it involved an action or an inaction, whether it took place at a specific time or over a general period, how expected it was,

how intentional it was, and whether they thought about changing their behaviour. In Experiment 6, participants were asked for either good or bad memories; in Experiment 7, they were asked for either a good or bad memory from the past month, or from across their entire lives; and in Experiment 8, they were asked for either a memory where they themselves did something good or bad, or another person did something good or bad.

All three experiments showed, for the first time, that participants who recalled memories of someone doing something honourable or charitable for someone else tended to generate downward and subtractive counterfactuals. That is, when they thought of how the situation might have been different ‘if only’, they thought about how the world might have been worse, and they removed a feature of the scenario to do so. In contrast, they also showed for the first time that participants who recalled someone doing something *dishonourable* or *uncharitable* generated upward and additive counterfactuals. That is, they thought about how the world might have been better, and they did so by adding something to the scenario that had not previously been present.

As well as demonstrating that people tend to generate different kinds of counterfactuals for memories of people doing good and bad things, these experiments also provided evidence that the memories they produce share features with scenarios that tend to prompt people to spontaneously generate counterfactual thoughts. They tended to be actions, to be specific rather than general, and to be controllable (i.e. intentional), which are features of the kind of scenarios that tend to produce counterfactuals (e.g. Roese, 1997; Davison & Feeney, 2008; McCloy & Byrne, 2000), although only the bad events were judged to be unexpected.

Good more than bad memories tended to cause people to think about changing their behaviour. They did so for recent and remote memories, as in Experiment 7, and for other-focused memories, as in Experiments 6 and 8. The only exception was self-focused memories. In this case, participants thought about changing their bad behaviour more than their good behaviour.

The experiments also led to the discovery that people formulate general intentions to change in the future (that is, intentions not related to a specific behaviour) when they witness morally good events, whereas they formulate specific intentions to change when they witness morally bad events, as Experiment 6 showed. They did so for remote memories, although not for recent memories, as Experiment 7 showed. They did not do so for other-focused memories or for self-focused memories in Experiment 8, although there was a non-significant trend suggesting that intentions to change were more general in good other-focused conditions than in bad other-focused conditions (and as discussed in Chapter 4, this result may have arisen because there were relatively fewer participants involved in this comparison than in the other experiments). It is suggested that the tendency of intentions to change following good events to be general may explain why it can be difficult to emulate good behaviour: general intentions may not provide a clear enough path for emulation. This suggestion is consistent with the observation that people remembering elevating memories tended to generate downward rather than upward, and subtractive rather than additive counterfactuals; downward counterfactuals are less helpful than upward counterfactuals when people are attempting to solve cognitive tasks (e.g. Roese, 1994), and subtractive counterfactuals are less helpful in inspiring creative thought than additive counterfactuals (although they do help with analytical problem-solving, e.g. Markman et al., 2009). It may be the case that the generation of downward and

subtractive counterfactuals, and general intentions to change behaviour, are linked, and that their occurrence following elevating memories may help to explain why people sometimes fail to emulate moral exemplars.

Psychological Implications

The findings have implications for theories of moral reasoning. The fact that other-focused good memories are more likely than other-focused bad memories to prompt people to think about changing their behaviour accords with the broaden-and-build theory of positive emotions, according to which negative emotions narrow the focus of people's thoughts, but positive emotions broaden the possibilities people think about, offering them an increased range of potential behavioural choices (e.g. Fredrickson 2001; Haidt, 2000). However, they also provide some support for the hypothesis that the generation of counterfactual thoughts may contribute to the transformation of elevation into intentions to emulate morally good behaviour. As discussed in Chapter 4, the process of deciding to behave morally because of witnessing a moral exemplar would appear to require cognitive processes to move from emotional inspiration to thoughts about acting.

Thinking about how things might have been may be a part of this process. Memories of people doing something honourable or charitable appear to match several of the features of events that prompt counterfactual thought, and counterfactual thoughts about these memories tend to consider how much worse the situation might have been without some aspect of the good behaviour. These findings suggest that a complex interaction of emotional reactions, and domain-general cognitive processes, rather than a simple emotional reaction, is key in generating intentions to emulate good behaviour. This suggestion is congruent with the idea of

dual-processes of emotions and controlled cognitive processes influencing moral reasoning (e.g. Greene et al., 2004, 2008). These experiments support the idea that emotions play a role in generating moral action, but also suggest a more complex interaction between emotional and cognitive processes than would be predicted by the social intuitionist model, according to which most moral judgments are driven by emotional intuitions (e.g. Haidt, 2001; Haidt & Bjorklund, 2008).

The suggestion that the generation of counterfactuals may be involved in the transition from elevation to emulation also has implications for the question, discussed in detail in the section on moral consistency, of whether moral reasoning is similar to, or different from non-moral reasoning. Although the experiments in Chapter 4 have not established a causal role for counterfactual thinking in emulation of good behaviour, they have provided evidence that memories of the sort that lead to elevation are also the sort that tend to generate counterfactual thinking. If it is the case that counterfactual thoughts are involved in generating thoughts about emulating good behaviour, as indicated by Experiments 6-8, this again suggests that domain-general cognitive processes are a key component of moral cognition.

There is some previous evidence for the operation of counterfactual thought in moral reasoning: when people were asked to alter moral dilemmas in which they had already judged action to be impermissible, in order that the action would be permissible, they did so in ways that resembled the generation of counterfactual alternatives (Bucciarelli et al., 2008). Furthermore, although people tend to imagine alternatives to controllable more than uncontrollable events, they focus more on morally inappropriate controllable than appropriate controllable events (e.g. McCloy & Byrne, 2000). These results suggest that future research focused on the role of counterfactuals in moral judgment might reveal other interesting distinctions.

Practical and Normative Implications

The results also have practical implications for everyday morality. The moral emotion of elevation may be a key factor in increasing morally good behaviour in the world; witnessing moral exemplars appears to inspire people to think about changing their behaviour. However, the evidence from Experiments 6 and 7 suggests that often, their intentions to change are general rather than specific, and may therefore be difficult for them to implement. There may be potential for psychologists to help to design interventions to encourage people to formulate specific rather than general intentions when they feel morally uplifted by witnessing morally exemplary behaviour. The results also have normative implications, in that prescriptive theories are usually designed to produce the best possible world, morally speaking. Such theories could take into account the observation that witnessing and thinking about moral exemplars appears to increase people's intentions to behave well themselves. The experiments here focused only on the link between elevation and intention to behave morally. However, it is possible to speculate that, if this link extends to actual moral behaviour, emotional and cognitive processes that lead to intentions to emulate good behaviour may provide a natural, psychological mechanism to increase pro-social behaviour throughout society.

The experiments reported in this thesis show for the first time that when people judge the morality of actions, their moral consistency can be improved, by simply allowing people to think about the morality of each action separately before thinking about them together. However, they also show that people remain inconsistent when they judge the combined morality of a moral and an immoral action,

in that their judgments place undue emphasis on the immoral action compared to how they themselves judge the individual actions.

The experiments reported here also provide the novel finding that when people decide whether to act or not when faced with a moral dilemma, they can construct emotive justifications for their decisions as readily for impersonal moral dilemmas as for personal ones. Furthermore, when people know a person has experienced a strong emotion when faced with an impersonal moral dilemma, such as harm caused indirectly by hitting a switch, they expect them to act. When they know a person experienced only a mild emotion when faced with a personal moral dilemma, such as harm caused directly by pushing a man, they also expect them to act.

Finally, the experiments also provide the novel discovery that when people recall witnessing someone doing something honourable or charitable for someone else, they think about how the situation might have been different by imagining how the world might have been worse, and they create a subtractive counterfactual that removes something from the event. Their intentions to change for the future are usually general, rather than specific. In contrast, when they recall witnessing someone doing something *dishonourable* or *uncharitable*, they think about how the world might have been better, and they create additive counterfactuals that add something to the event. Their intentions to change for the future are neither specific nor general.

Implications for understanding moral psychology

This thesis has discussed two major themes with implications for how psychologists understand moral reasoning. The first has been the idea that moral and non-moral cognition often seem to be constituted by similar mental processes. The second has been the interaction of emotional and controlled cognitive processes,

including reasoning and the generation of counterfactual alternatives, in moral judgment. Each of these themes has important implications for the understanding of how people think about moral matters.

The experiments discussed in this thesis, and many results reported by other researchers on moral psychology, seem to indicate that many features of cognition about moral matters are similar to features of cognition about non-moral matters. As discussed in Chapter 2, heuristic processes seem to operate in moral and non-moral cognition. Specifically, Experiments 1 and 2 demonstrated results similar to the conjunction fallacy in probability judgments. Experiment 3 also demonstrated a heuristic tendency to weigh immoral actions more heavily than moral actions in mixed-valence moral judgments, and importantly, more heavily than their own judgments of the actions individually would warrant. These results parallel results from a variety of domains in which negative stimuli are given greater emphasis than positive stimuli (e.g. Baumeister et al., 2001).

Chapter 4 also provides evidence that the generation of counterfactual alternatives, which are important in non-moral reasoning (e.g. Roese, 1994; Byrne, 2005), may play an important role in prompting moral action. As discussed in Chapter 2, the moral grammar theory, which posits a specifically moral mechanism for moral judgment, and suggests that moral judgments should not be susceptible to the kinds of heuristics that non-moral judgments are, has difficulty with this evidence (e.g. Dwyer et al., 2010; Mikhail, 2007, 2009). Instead, the results support theories that emphasise the role of domain-general cognitive processes in moral judgments, such as the moral heuristics approach (e.g. Sunstein, 2005; Baron, 2010), and dual-process theories of moral judgment (e.g. Greene et al., 2004, 2008; Paxton & Greene, 2010; Bucciarelli et al., 2008).

The evidence suggests that reasoning about moral matters involves the same processes as reasoning about non-moral matters. Importantly, it has been observed that there is no simple way to distinguish between moral and non-moral matters – that is to say, while people can identify what qualifies as moral and what qualifies as immoral, there are likely to be disagreements between individuals within cultures (e.g. Bucciarelli et al., 2008; Graham, Haidt & Nosek, 2009), and especially between different cultures (e.g. Henrich et al., 2005). Essentially, moral propositions are deontic propositions – that is, propositions about what is permissible and obligatory. It has been hypothesised, therefore, that moral reasoning is just deontic reasoning about moral matters (e.g. Bucciarelli et al., 2008).

Most contemporary theories of reasoning depend on a distinction between fast, intuitive processes, and slower, deliberative processes (e.g. Evans, 2008; Byrne & Johnson-Laird, 2009; Kahneman, 2011). Most reasoning appears to take place unconsciously using fast, intuitive processes, but slower, deliberative processes are required in order to correctly make more difficult inferences. Although intuitive judgments are generally accurate, they sometimes lead people into error. As demonstrated in Chapter 2, these errors affect not only non-moral, but also moral judgments.

One important difference between moral and non-moral cognition is that emotion often appears to play a special role in moral cognition, as discussed in Chapter 3. However, judgments of right and wrong can be made independently of emotion. For instance, it has been suggested that matters of petty theft can be judged to be morally wrong, but may inspire no emotion if the loss is inconsequential (e.g. the theft of a paper clip is wrong, but would be unlikely to cause an emotional reaction; Bucciarelli et al., 2008). As discussed previously, emotions also affect other

kinds of judgment, such as judgments of the likability of neutral stimuli (e.g. Winkielman et al., 1997). Therefore the presence of emotion in many moral judgments is not a strong objection to the idea that domain-general processes operate in moral judgment.

The debate over the relative roles of emotional and cognitive processes in moral judgment has been another important theme in this thesis. Emotion may not be a necessary component of moral judgment, but it does often influence such judgments, and this aspect of moral judgment has garnered a great deal of attention in modern theories of moral reasoning (e.g. Haidt, 2001; Greene et al., 2004, 2008; Bucciarelli et al., 2008). The experiments in Chapters 3 and 4 focused on how emotional and cognitive processes interact when people reason about the emotions and decisions of people who are in moral dilemmas, and when people are moved by witnessing good and bad moral actions to change their own behaviour. People can produce emotive or reasoned justifications for their own moral judgments in moral dilemmas that are either highly emotional, or somewhat emotional (Experiment 4). When they read about what other people felt and did in moral dilemmas, their reading times vary depending on the strength of feelings people have, the type of decision people make, and whether dilemma the people are in is highly emotional or not (Experiment 5). Lastly, the emulation of good actions may be influenced not only by positive emotion turning people's perspectives outward, but also by the generation of counterfactual alternatives (Experiments 6-8). These results suggest a complex interaction of emotions and domain-general cognitive processes influencing moral judgments and moral actions. They suggest that people incorporate information about the emotions experienced by others in their moral judgments, and also that they form expectations about the type of moral actions people will carry out based on the type of emotions

those people experience. They also suggest that people understand that others may experience emotions even in impersonal moral dilemmas, and that they consider such emotions to be components of legitimate explanations of moral decision-making.

Limitations and Directions for Future Research

Future research could develop the findings of the experiments presented in this thesis in a number of ways. Experiments 1 and 2 established that numeric moral judgments are vulnerable to certain heuristic effects similar to those demonstrated in judgments of probability (e.g. a version of the conjunction fallacy, which, in probability judgments, leads participants to sometimes judge the occurrence of two events together to be no less probable than the occurrence of either individually, and which in morality sometimes leads people to judge two moral actions to be no more moral than one or both individually). Experiment 3 demonstrated that, when participants make judgments of mixed-valence moral and immoral actions, they tend to weigh the immoral action more heavily than the moral action, and more heavily than would be expected given their own judgments of the actions individually. This result directly parallels results from a variety of domains in which negative stimuli are given greater weight than positive stimuli. A limitation of Experiments 1-3 is that they focused only on judgments of moral actions, rather than on judgments praise and blame, or of the morality of people. Future research could focus on whether the heuristic effects identified in judgments of moral actions in these experiments might also be identified in these kinds of judgments.

Furthermore, with regard to future research, cognitive shortcuts that sometimes lead people into error have been identified in a much wider variety of non-moral judgments than have thus far been studied with regard to morality (e.g. Tversky,

1969; Kahneman & Tversky, 1984; Thaler & Sunstein, 2008; Kahneman, 2011).

More of these research paradigms could be adapted to study moral judgments, and such experiments could be useful in identifying if there are areas in which moral judgments are resistant to heuristics that non-moral judgments are not.

Experiments 4 and 5 indicated that people consider information about emotions experienced during moral dilemmas to be an important factor in moral judgments. However, a limitation of these experiments was that they did not test the influence of information about particular emotions on reading times or on moral justifications, but only the influence of different strengths of feelings. Since priming studies indicate different influences on moral judgments for emotions such as disgust (e.g. Schnall et al., 2008) and different positive emotions such as mirth and elevation (e.g. Valdesolo & DeSteno, 2006; Strohminger et al., 2011), there is good reason to believe that reading that people in different moral dilemmas experienced different specific emotions, and even different strengths of different specific emotions, might result in varying influences on moral reasoning. Future studies might focus on how people process information about participants experiencing different emotions, in different dilemmas, and making different decisions.

A limitation of the studies of the moral emotion of elevation (Experiments 6-8) was that many participants had to be eliminated. However, this concern is balanced by the diversity of the samples that were accessed for these experiments in terms of nationality, gender and age. Another limitation of Experiments 6-8 was that although they provided supporting evidence that the generation of counterfactuals may play a role in transferring people's emotional feeling of elevation into an intention to behave well themselves, and indeed in what kind of intentions people generated, the experiments did not establish a causal link between these processes. They established

regularities in the types of counterfactuals generated in response to memories of people doing both good and bad things, and found some evidence indicating that such memories are of a kind that are likely to lead people to generate counterfactual alternatives. Future studies might focus on establishing a causal link. As discussed in Chapter 4, think-aloud studies might be useful in this regard, as they can provide information as to certain intermediate steps in reasoning – specifically the steps that bring thoughts to conscious awareness. Such studies might possibly help to identify whether people spontaneously generate counterfactual alternatives in response to elevating experiences.

In general, future research could consider turning towards testing for the influence of domain-general processes in moral judgment. This thesis supports a great deal of other research indicating that moral reasoning is similar to non-moral reasoning: judgments are sometimes made unconsciously, and are sometimes influenced by emotions, but people can also reason when they are presented with relevant information. Furthermore, there is reason to believe that thought about moral problems is often influenced by similar mechanisms to thought about non-moral problems, such as counterfactual alternatives to reality. This research suggests that investigating the roles of domain-general processes in moral reasoning may improve psychological explanations of moral phenomena.

As discussed in Chapter 2, economists have utilised psychological knowledge of financial decision-making to design programs to improve such decision-making, including saving behaviour (e.g. Thaler & Benartzi, 2004). The research presented in this thesis, along with other, similar research, lends itself to the possibility of improving moral behaviour. Experiments 1-3 demonstrate that people are capable of reasoning rationally about moral issues, but also that their judgments may sometimes

be biased by intuitive heuristics. Experiments 4 and 5 demonstrate that people are not inevitably swayed by emotions and unconscious intuitions, but suggest that such processes may influence judgments through social reasoning. It has previously been observed that witnessing moral exemplars appears to lead people to behave morally themselves. It may be conjectured, on the basis of the results of Experiments 6-8, that this improvement may be increased if people consider how, specifically, they might emulate that behaviour, rather than generally resolving to behave more morally. Future research may clarify whether this speculation is borne out.

Psychologists may be able to use the information provided by these experiments to develop interventions that may help people to behave more rationally and, perhaps, more morally. Such interventions may help people to maintain internal consistency of moral judgments, so that they do not allow normatively irrelevant factors, such as in what order they make their judgments, to influence those judgments. They could make people aware that their reasoning may sometimes be influenced by emotion (but that such judgments are not necessarily biased). Finally, interventions focused on how cognitive and emotional processes interact in moral judgments might make it possible for people to formulate clearer plans for how to meet their intentions to behave more morally. Taken together, the results presented in this thesis may help psychologists to develop methods that could allow people to become more moral individuals, and thus increase the amount of morally good behaviour in the world.

References

- Aaron, K. (2013) Judge: Remorse reason for no jail time. Retrieved September 19th 2013: <http://www.news4jax.com/news/judge-remorse-reason-for-no-jail-time/-/475880/20634864/-/73rdf1/-/index.html>
- Algoe, S. B., & Haidt, J. (2009). Witnessing excellence in action: The 'other-praising' emotions of elevation, gratitude, and admiration. *The journal of positive psychology*, 4 (2), 105-127.
- Baron, J. (2010). Cognitive biases in moral judgments that affect political behavior. *Synthese*, 172 (1), 7-35.
- Barraza, J. A., & Zak, P. J. (2009). Empathy toward strangers triggers oxytocin release and subsequent generosity. *Annals of the New York Academy of Sciences*, 1167(1), 182-189.
- Baumeister, R. F., Bratslavsky, E., Finkenauer, C., & Vohs, K. D. (2001). Bad is stronger than good. *Review of general psychology*, 5 (4), 323-370.
- Belsky, G. (2012). Charity on Steroids. Retrieved September 19th 2013 at: <http://nymag.com/news/intelligencer/lance-armstrong-charity-2012-10/>
- Bielenberg, K. (2013) The secret, softer side of Paul O'Connell. Retrieved September 19th 2013 at: <http://www.independent.ie/lifestyle/the-secret-softer-side-of-paul-oconnell-29193472.html>
- Birnbaum, M. H. (1972). Morality judgments: Tests of an averaging model. *Journal of Experimental Psychology*, 93 (1), 35-42.
- Birnbaum, M. H. (1973). Morality judgments: Tests of an averaging model with differential weights. *Journal of Experimental Psychology*, 99, 395-399.

- Bucciarelli, M., & Johnson-Laird, P. N. (2005). Naïve deontics: A theory of meaning, representation, and reasoning. *Cognitive Psychology*, 50 (2), 159-193.
- Bucciarelli, M., Khemlani, S., & Johnson-Laird, P.L. (2008). The Psychology of Moral Reasoning. *Judgment and Decision Making*, 3, (2), 121–139.
- Byrne, R.M.J. (2005). *The rational imagination: How people create alternatives to reality*. Cambridge: MIT Press.
- Byrne, R.M.J., (2007). Precis of The Rational Imagination: How People Create Alternatives to Reality. *Behavioral and Brain Sciences*, 30, 2007, 439-453
- Byrne, R. M. J., Espino, O., & Santamaria, C. (1999). Counterexamples and the suppression of inferences. *Journal of Memory and Language*, 40, 347-373.
- Byrne, R.M.J. & Johnson-Laird, P.N. (2009). 'If' and the problems of conditional reasoning. *Trends in Cognitive Sciences*. 13, 282-287
- Ciaramelli, E., Muccioli, M., Làdavas, E., & di Pellegrino, G. (2007). Selective deficit in personal moral judgment following damage to ventromedial prefrontal cortex. *SCAN*, 2, 84-92.
- Colby, A., Kohlberg, L., Gibbs, J., Lieberman, M., Fischer, K., & Saltzstein, H.D. (1983). A longitudinal study of moral judgment. *Monographs of the society for research in child development*, 48 (1/2), 1-124.
- Cox, K. S. (2010). Elevation predicts domain-specific volunteerism 3 months later. *The Journal of Positive Psychology*, 5 (5), 333-341.
- Cushman, F., Young, L., & Hauser, M. (2006). The Role of Conscious Reasoning and Intuition in Moral Judgment: Testing Three Principles of Harm. *Psychological Science*, 17 (12) 1082-1089.

- Davison, I. M., & Feeney, A. (2008). Regret as autobiographical memory. *Cognitive Psychology*, 57(4), 385-403.
- De Neys, W. (2006). Dual processing in reasoning two systems but one reasoner. *Psychological Science*, 17 (5), 428-433.
- Diessner, R., Iyer, R., Smith, M. M., & Haidt, J. (2013). Who engages with moral beauty? *Journal of Moral Education*, 42 (2), 139-163.
- Dixon, J.E., & Byrne, R.M. (2011). "If only" counterfactual thoughts about exceptional actions. *Memory & cognition*, 39 (7), 1317-1331.
- Dwyer, S., Huebner, B., & Hauser, M. D. (2010). The linguistic analogy: Motivations, results, and speculations. *Topics in cognitive science*, 2 (3), 486-510.
- Evans, J. S. B. (2008). Dual-processing accounts of reasoning, judgment, and social cognition. *Annual Review of Psychology*, 59, 255-278.
- Frederick, S. (2005). Cognitive reflection and decision making. *The Journal of Economic Perspectives*, 19 (4), 25-42.
- Fredrickson, B. L. (2001). The role of positive emotions in positive psychology: The broaden-and-build theory of positive emotions. *American psychologist*, 56 (3), 218-226.
- Freeman, D., Aquino, K., & McFerran, B. (2009). Overcoming beneficiary race as an impediment to charitable donations: Social dominance orientation, the experience of moral elevation, and donation behavior. *Personality and Social Psychology Bulletin*, 35(1), 72-84.
- Frosch, C. A., & Byrne, R. M. (2012). Causal conditionals and counterfactuals. *Acta Psychologica*, 141 (1), 54-66.
- Gilovich, T., & Medvec, V. H. (1995). The experience of regret: what, when, and why. *Psychological review*, 102 (2), 379.

- Goodwin, G. P., & Darley, J. M. (2008). The psychology of meta-ethics: Exploring objectivism. *Cognition*, 106 (3), 1339-1366.
- Graham, J., Haidt, J., & Nosek, B. A. (2009). Liberals and conservatives rely on different sets of moral foundations. *Journal of personality and social psychology*, 96 (5), 1029-1046.
- Grappi, S., Romani, S., & Bagozzi, R. P. (2013). The effects of company offshoring strategies on consumer responses. *Journal of the Academy of Marketing Science*, 1-22.
- Greene, J.D., Nystrom, L.E., Engell, A.D., Darley, J.M., and Cohen, J.D. (2004). The Neural Bases of Cognitive Conflict and Control in Moral Judgment. *Neuron*, 44, 389-400.
- Greene, J.D., Morelli, S.A., Lowenberg, K., Nystrom, L.E., & Cohen, J.D. (2008). Cognitive load selectively interferes with utilitarian moral judgment. *Cognition*, 107, 1144-1154.
- Greene, J.D., Sommerville, R.B., Nystrom, L.E., Darley, J.M., and Cohen, J.D. (2001). An fMRI Investigation of Emotional Engagement in Moral Judgment. *Science*, 293, 2105-2108.
- Greenwald, G., MacAskill, E., & Poitras, L. (2013). Edward Snowden: the whistleblower behind the NSA surveillance revelations. Retrieved September 19th 2013 from: <http://www.guardian.co.uk/world/2013/jun/09/edward-snowden-nsa-whistleblower-surveillance>.
- Gubbins, E. & Byrne, R.M.J. (2013a). Dual processes of emotion and reason in direct and indirect moral dilemmas. Paper revised and resubmitted to *Thinking and Reasoning*.

- Gubbins, E. & Byrne, R.M.J. (2013b). Cognitive Processes in Moral Elevation: Counterfactual 'if only' thoughts and intentions to emulate. Paper in preparation.
- Gubbins, E., Byrne, R.M.J. & Johnson-Laird (2013). Moral inconsistency. Paper in preparation.
- Haidt, J. (2000). The positive emotion of elevation. *Prevention and Treatment*, 3, (3) 1-5.
- Haidt, J. (2001). The Emotional Dog and its Rational Tail. *Psychological Review*, 108 (4) 814-834.
- Haidt, J., & Bjorklund, F. (2008). Social intuitionists answer six questions about morality. In W. Sinnott-Armstrong (Ed.), *Moral psychology: Vol. 2. The cognitive science of morality* (pp. 181– 217). Cambridge, MA: MIT Press.
- Haidt, J., Bjorklund, F., & Murphy, S. (2000). Moral dumbfounding: When intuition finds no reason. Unpublished manuscript, University of Virginia.
- Hauser, M. D. (2006). The liver and the moral organ. *Social cognitive and affective neuroscience*, 1 (3), 214-220.
- Henrich, J., Boyd, R., Bowles, S., Camerer, C., Fehr, E., Gintis, H., McElreath, R., Alvard, M., Barr, A., Ensminger, J., Henrich, N.S., Hill, K., Gil-White, F., Gurven, M., Marlowe, F.W., Patton, J.Q., & Tracer, D. (2005). "Economic man" in cross-cultural perspective: Behavioral experiments in 15 small-scale societies. *Behavioral and brain sciences*, 28 (6), 795-814.
- Hume, D. (1739-1740/2004) *A Treatise of Human Nature*. Mineola, N.Y.: Dover.
- Johnson-Laird, P. N., & Byrne, R.M.J. (2002). Conditionals: a theory of meaning, pragmatics, and inference. *Psychological review*, 109 (4), 646-678.
- Kahneman, D. (2011). *Thinking, fast and slow*. Macmillan.

- Kahneman, D., & Tversky, A. (1982). The simulation heuristic. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases* (pp. 201–208).
- Kahneman, D., & Tversky, A. (1984). Choices, values and frames. *American Psychologist*, 39 (4), 341-350.
- Kant, I. (1788/2002). *Critique of Practical Reason*. Indianapolis; Cambridge: Hackett Publishing Company.
- Khemlani, S. S., Lotstein, M., & Johnson-Laird, P. (2012). The probabilities of unique events. *PloS one*, 7 (10), e45975.
- Kinsella, M. (2013). J'accuse! The French media on the Paul O'Connell 'conspiracy'. Retrieved September 19th at: <http://www.thescore.ie/paul-oconnell-french-media-883005-Apr2013/>
- Knobe, J. (2003). Intentional action in folk psychology: an experimental investigation. *Philosophical Psychology*, 16, 309-323.
- Koenigs, M., Young, L., Adolphs, R., Tranel, D., Cushman, F., Hauser, M., & Damasio, A. (2007). Damage to the prefrontal cortex increases utilitarian moral judgements. *Nature*, 446 (7138), 908-911.
- Kohlberg, L. (1973). The claim to moral adequacy of a highest stage of moral judgment. *The journal of philosophy*, 70 (18), 630-646.
- Kosinski, M. (2013). Former Costa Concordia captain: I regret nothing. World News on NBC News.com. Retrieved March 6th 2013 at: http://worldnews.nbcnews.com/_news/2013/01/16/16531308-former-costa-concordia-captain-i-regret-nothing?lite

- Kray, L. J., Galinsky, A. D., & Markman, K. D. (2009). Counterfactual structure and learning from experience in negotiations. *Journal of Experimental Social Psychology*, 45 (4), 979-982.
- Leslie, A. M., Knobe, J., & Cohen, A. (2006). Acting Intentionally and the Side-Effect Effect Theory of Mind and Moral Judgment. *Psychological Science*, 17 (5), 421-427.
- Luo, Q., Nakic, M., Wheatley, T., Richell, R., Martin, A., & Blair, R. J. R. (2006). The neural basis of implicit moral attitude? An IAT study using event-related fMRI. *Neuroimage*, 30 (4), 1449-1457.
- Markman, K.D., Lindberg, M.J., Kray, L.J., and Galinsky, A.D. (2007). Implications of Counterfactual Structure for Creative Generation and Analytical Problem Solving. *Personality and Social Psychology Bulletin* 33, 312-324.
- Markovits, H., Lortie Forgues, H., and Brunet, M-L (2010). Conditional reasoning, frequency of counterexamples, and the effect of response modality. *Memory & Cognition*, 38, 485-492
- McCloy, R. & Byrne, R.M.J. (2000) Counterfactual thinking about controllable actions. *Memory & Cognition*, 28, 1071-1078.
- McLaughlin, E.C., & Brown, P. (2013). Judge sentences Cleveland kidnapper Ariel Castro to life, plus 1,000 years. Retrieved September 19th 2013 at: <http://edition.cnn.com/2013/08/01/justice/ohio-castro/index.html>
- Mikhail, J. (2007). Universal moral grammar: theory, evidence and the future. *Trends in Cognitive Sciences*, 11 (4), 143-152.

- Mikhail, J. (2009). Moral Grammar and Intuitive Jurisprudence: A Formal Model of Unconscious Moral and Legal Knowledge. *Psychology of Learning and Motivation*, 50, 27-100.
- Mikhail, J. (2011). Emotion, neuroscience, and law: a comment on Darwin and Greene. *Emotion Review*, 3 (3), 293-295.
- Mill, J.S. (1863/2007). *Utilitarianism*. Mineola, N.Y.: Dover.
- Minihan, M., & Cullen, P. (2013). Government indicates it will not be influenced by bishops' intervention on abortion law. Retrieved September 19th 2013 at: <http://www.irishtimes.com/news/politics/government-indicates-it-will-not-be-influenced-by-bishops-intervention-on-abortion-law-1.1382405>.
- Monin, B., Pizarro, D.A., & Beer, J.S. (2007). Reason and Emotion in Moral Judgment: Different Prototypes Lead to Different Theories. In Vohs, K.D., Baumeister, R.F., & Loewenstein, G. *Do emotions help or hurt decision making?: A hedgefoxian perspective*. New York: Russell Sage Foundation.
- Moore, A.B., Clark, B.A., & Kane, M.J. (2008). Who Shalt Not Kill? Individual Differences in Working Memory Capacity, Executive Control, and Moral Judgment. *Psychological Science*, 19 (6) 549-557.
- Moore, A. B., Lee, N. L., Clark, B. A., & Conway, A. R. (2011). In defense of the personal/impersonal distinction in moral psychology research: Cross-cultural validation of the dual process model of moral judgment. *Judgment and Decision Making*, 6 (1), 186-195.
- Nakamura, K. (2013). A closer look at moral dilemmas: latent dimensions of morality and the difference between trolley and footbridge dilemmas. *Thinking and Reasoning*. In press.

- Paxton, J.M., & Greene, J.D. (2010). Moral Reasoning: Hints and Allegations. *Topics in Cognitive Science*, 2, 511-527.
- Paxton, J. M., Ungar, L., & Greene, J. D. (2012). Reflection and reasoning in moral judgment. *Cognitive science*, 36 (1), 163-177.
- Peachey, P. (2013). Mother and partner who showed no remorse for murder of Daniel Pelka to serve at least 30 years in jail. Retrieved September 19th 2013 at: <http://www.independent.co.uk/news/uk/crime/mother-and-partner-who-showed-no-remorse-for-murder-of-daniel-pelka-to-serve-at-least-30-years-in-jail-8743576.html>
- Peavoy, T. (2013). Paul O'Connell will not be cited over kick to Dave Kearney's head at Thomond Park. Retrieved September 19th 2013 at: <http://www.rte.ie/sport/rugby/rabo-direct-pro-12/2013/0415/381486-oconnell-avoids-citing-over-kearney-kick/>
- Peele, S. (2011). Blaming Mike McQueary. Retrieved September 19th 2013 at: http://www.huffingtonpost.com/stanton-peelee/mike-mcqueary-penn-state_b_1090428.html
- Pizarro, D. A., & Bloom, P. (2003). The intelligence of the moral intuitions: A comment on Haidt (2001). *Psychological Review*, 110 (1), 193-196.
- Pizarro, D.A., Uhlmann, E., & Bloom, P. (2003). Causal deviance and the attribution of moral responsibility. *Journal of Experimental Social Psychology*, 39, 653-660.
- Riskey, D. R., & Birnbaum, M. H. (1974). Compensatory effects in moral judgment: Two rights don't make up for a wrong. *Journal of Experimental Psychology*, 103 (1), 171-173.
- Ritov, I., & Baron, J. (1990). Reluctance to vaccinate: Omission bias and

- ambiguity. *Journal of Behavioral Decision Making*, 3 (4), 263-277.
- Roese, N.J. (1994). The functional basis of counterfactual thinking. *Journal of Personality and Social Psychology*, 66, 805–818.
- Roese, N. J. (1997). Counterfactual thinking. *Psychological Bulletin*, 121 (1), 133-148.
- Sanfey, A. G., Rilling, J. K., Aronson, J. A., Nystrom, L. E., & Cohen, J. D. (2003). The neural basis of economic decision-making in the ultimatum game. *Science*, 300 (5626), 1755-1758.
- Santamaría, C., Espino, O., & Byrne, R. M. (2005). Counterfactual and semifactual conditionals prime alternative possibilities. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 31 (5), 1149-1154.
- Schnall, S., Haidt, J., Clore, G.L., & Jordan, A.H. (2008). Disgust as Embodied Moral Judgment. *Personality and Social Psychology Bulletin*, 34, 1096-1109.
- Schnall, S., & Roper, J. (2012). Elevation puts moral values into action. *Social Psychological and Personality Science*, 3 (3), 373-378.
- Schnall, S., Roper, J., & Fessler, D. M. (2010). Elevation leads to altruistic behavior. *Psychological Science*, 21 (3), 315-320.
- Shanahan, C., & Gallagher, C. (2012). Outrage as sex attacker avoids jail term. Retrieved September 20th 2013 at:
<http://www.irisht Examiner.com/archives/2012/1005/world/outrage-as-sex-attacker-avoids-jail-term-209976.html>
- Silvers, J. A., & Haidt, J. (2008). Moral elevation can induce nursing. *Emotion*, 8 (2), 291-295.

- Strohminger, N., Lewis, R.L., & Meyer, D.E. (2011). Divergent effects of different positive emotions on moral judgment. *Cognition*, 119, 295- 300.
- Sunstein, C.R. (2005). Moral Heuristics. *Behavioral and Brain Sciences*. 28, 531-573.
- Thaler, R. H., & Benartzi, S. (2004). Save more tomorrow™: Using behavioral economics to increase employee saving. *Journal of political Economy*, 112 (S1), S164-S187.
- Thaler, R. H., & Sunstein, C. R. (2008). *Nudge: Improving decisions about health, wealth, and happiness*. Yale University Press.
- Tversky, A. (1969). Intransitivity of preferences. *Psychological review*, 76 (1), 31-48.
- Tversky, A., & Kahneman, D. (1983). Extensional versus intuitive reasoning: The conjunction fallacy in probability judgment. *Psychological review*, 90 (4), 293-315.
- Ugazio, G., Lamm, C., & Singer, T. (2012). The Role of Emotions for Moral Judgments Depends on the Type of Emotion and Moral Scenario. *Emotion*, 12 (3) 579-590.
- Valdesolo, P., & DeSteno, D. (2006). Manipulations of Emotional Context Shape Moral Judgment. *Psychological Science*, 17 (6) 476-477.
- Waldmann, M. R., Nagel, J., & Wiegmann, A. (2012). Moral judgment. *The Oxford handbook of thinking and reasoning*, 364-389.
- Weathers, H., & Evans, R. (2012). Captain Coward had his eye on English dancer: After claims he was distracted by a woman on the bridge, passengers and crew reveal Casanova antics of crash skipper. Retrieved April 19th 2013 at: <http://www.dailymail.co.uk/news/article-2089680/Costa-Concordia-Captain-Coward-Francesco-Schettino-eye-dancer-Domnica-Cemortan.html>

- Wetzel, D. (2012). Mike McQueary, broken down and estranged, files \$4 million whistleblower lawsuit. Retrieved September 23rd 2013 at:
<http://sports.yahoo.com/news/ncaaf--mike-mcqueary-estranged--files-4-million-whistleblower-lawsuit.html>
- Wheatley, T., & Haidt, J. (2005). Hypnotic disgust makes moral judgments more severe. *Psychological science*, 16 (10), 780-784.
- Whitehead, A., & Haugh, A. (2012). Murder can be 'crime of passion', says top judge. Retrieved September 19th 2013.
<http://www.telegraph.co.uk/news/uknews/law-and-order/9020905/Murder-can-be-crime-of-passion-says-top-judge.html>
- Winkielman, P., Zajonc, R.B., & Schwarz, N. (1997). Subliminal affective priming resists attributional interventions. *Cognition & Emotion*, 11 (4), 433-465.

Appendix 2.1 – Experimental Materials Experiments 1 and 2

Triples of Morally Good Actions

1. A woman took her sister's child into her home after her sister died, and scrimped and saved to pay for the child to go to University.
2. A man gave blood so that more would be available to help people who had been in accidents, and donated money to a charity that helped to fund cancer research.
3. A man intervened to help stop a fight, and tended to the wounds of a stranger injured in the fight.
4. A husband rejected an opportunity to cheat on his wife, and took his wife for a romantic walk.
5. The staff of a hospital banded together to subsidize the construction of a badly needed burns unit in their hospital, and worked together to save the victims of a terrorist bomb.
6. The director of a factory hired an equal number of able-bodied and disabled people, and paid them equally.
7. A woman worked for many years to raise money for research into kidney disease, and donated one of her kidneys to a friend of hers, saving her friend's life.
8. A young man helped an old woman to cross the street, and he volunteered at a home for the elderly.

Triples of Morally Bad Actions

1. A man attacked a security guard and stole money from him.
2. A husband slept with his wife's friend, and left his wife, stealing her car.
3. A violent bully terrorized the playground and beat up a younger girl with 4.
A man insulted a colleague for no reason, and made it look as though she had committed a crime so that she lost her job.
5. A babysitter beat the infant in her care, and starved her so that the child became malnourished.
6. The headmaster of a school physically and sexually abused one of his pupils.
7. The owner of a bar forced the foreign girl who was working for him to accept a salary below minimum wage, and to have sex with him.
8. The financial director of a company embezzled funds for many years, and covered his embezzlements by framing his colleague.

Triples of Morally Neutral Actions

1. A man bought a newspaper in a corner shop beside the train station, and posted an important document to a business associate in another city.
2. A man sat on the couch in his apartment on a wet, rainy evening in the city, and read a collection of detective stories quickly because he was very interested in what would happen next.
3. A woman finished her dinner at eight o'clock on a Tuesday evening, and, having watched a television program, prepared to go to bed at eleven in the evening.
4. A woman wrote an email to a colleague to arrange to meet for lunch in order to discuss a new sales strategy, and made a phone call to a hotel in a city she would be visiting to enquire about the availability of their rooms.

Appendix 2.2 – Consent and Debriefing Experiment 1

Instructions

Thank you for agreeing to participate in this study on reasoning. This study is being carried out by Eoin Gubbins (Visiting Student Research Collaborator) with the supervision of Prof. Phil Johnson-Laird (Princeton University) and Prof. Ruth Byrne (Trinity College Dublin). This study examines everyday reasoning; it is not a test of intelligence. The aim of the study is to examine the sorts of answers most people provide. The data will be stored anonymously.

Participation in this study will take about fifteen minutes. There is no time limit on your responses, but please try to finish all of the problems before leaving the study. You are free to withdraw at any time.

You will read a series of sentences. Each sentence describes one or more actions. Some of them are morally good actions, and some of them are morally bad actions. Your task is to rate each sentence in turn, on a scale from -100 through 0 to 100, with -100 meaning morally the worst possible action, and 100 meaning morally the best possible action. 0 is a midpoint meaning neither good nor bad.

If you have any questions about this study, or would like to access a summary of its findings, please do not hesitate to contact one of us at the following contact details:

Eoin Gubbins
Room 3.01,
Lloyd Building,
Trinity College Dublin,
Dublin 2,
Ireland.
[*egubbins@tcd.ie*](mailto:egubbins@tcd.ie)

Prof. Phil-Johnson-Laird
3-C-3 Green Hall,
Psychology Department,
Princeton University,
Princeton, NJ 08540,
USA.
[*phil@princeton.edu*](mailto:phil@princeton.edu)

By clicking to continue, you indicate that you understand the instructions given, and give your consent for your data to be used as part of this study.

This research is funded by a Government of Ireland PhD Scholarship from the Irish Research Council for the Humanities and Social Sciences.

Debriefing

Thank you for participating in this study. The aim of the study is to examine how people think about different kinds of moral issues. The study investigates how people judge the rightness and wrongness of moral actions both individually, and as pairs.

Thank you again for your participation. If you have any questions or concerns about the study, please do not hesitate to ask, or contact one of us at the contact details below.

Eoin Gubbins
Room 3.01,
Lloyd Building,
Trinity College Dublin,
Dublin 2,
Ireland.
egubbins@tcd.ie

Prof. Phil-Johnson-Laird
3-C-3 Green Hall,
Psychology Department,
Princeton University,
Princeton, NJ 08540,
USA.
phil@princeton.edu

Appendix 2.3 – Complete criteria for consistency for Experiments 1 and 2

Triples were judged to be consistent if:

- (i) The conjunction was larger than both conjuncts, including when one of the conjuncts was equal to 0 (but not when both were equal to 0).
- (ii) Both conjuncts and the conjunction were equal to 0.
- (iii) The conjunction was equal to 100.
- (iv) The conjunction was equal to one conjunct, and the other conjunct was equal to 0.

Triples were judged to have made one violation of consistency if:

- (i) The conjunction was larger than one conjunct, but smaller than the other.
- (ii) The conjunction and both conjuncts were equal to one another, but larger than 0 and less than 100.
- (iii) The conjunction was equal to one conjunct and larger than the other (but the conjunction was less than 100, and the smaller conjunct was greater than 0).

Triples were judged to have made two violations of consistency if:

- (i) The conjunction was smaller than both conjuncts
- (ii) The conjunction was smaller than one conjunct, and equal to the other.
- (iii) The conjunction was greater than both conjuncts, but both conjuncts were equal to 0.

Appendix 2.4 – Experiment 2 Consent and Debriefing:

Instructions

Thank you for agreeing to participate in this experiment on reasoning. This study is being carried out by Eoin Gubbins (PhD. student) with the supervision of Prof. Ruth Byrne (Trinity College Dublin) and in collaboration with Prof. Phil Johnson-Laird (New York University), funded by a Government of Ireland PhD Scholarship from the Irish Research Council for the Humanities and Social Sciences. This experiment examines everyday reasoning; it is not a test of intelligence, and the aim of the study is to examine the sorts of answers most people provide. The study is anonymous and confidential. The data will be stored anonymously, and in accordance with the Freedom of Information and the Data Protection Acts. You are entitled to access data stored about you, in this case the data will be scored as group scores and participants will not be identified individually. You may request access to a copy of the final report when it is available if you wish. Participation in the experiment will take about 20 minutes and you are free to withdraw at any time. If you would like to participate, you can start by clicking on the 'next' button. There is no time limit on your responses, but please try to complete all the problems in a single session before exiting the program. Your task is to read a series of sentences. Each sentence describes one or more actions. Some of them are morally good actions, some of them are morally bad actions, and some of them are morally neutral actions. Your task is to rate each sentence in turn, on a scale that runs from the worst possible action, to the best possible action, with a midpoint meaning morally neither good nor bad. For some questions you will be given a scale from -7 (Most immoral) to +7 (Most moral), with a midpoint of 0 (Neither moral nor immoral). You should click on the button that represents your judgement. For other questions you will be given a scale from -100 (Most immoral) to +100 (Most moral), with a midpoint of 0 (Neither moral nor immoral). You should slide the cursor from 0 to the point on the scale that represents your judgement. If you want to judge an item as neutral (0), it is necessary to move the slider away from 0, and then back to settle on 0 for it to record 0 as your judgement. Thank you again for agreeing to take part in the study. If you have any questions about it, or would like to access a summary of its findings, please do not hesitate to contact one of us at the following contact details: Eoin Gubbins, Room 3.01, Lloyd Building, Trinity College Dublin, Dublin 2, Ireland. Email:

egubbins@tcd.ie OR Prof. Ruth Byrne, Lloyd Building, Trinity College Dublin, Dublin 2, Ireland. Email: rmbyrne@tcd.ie. By clicking to continue, you indicate that you understand the instructions given, and give your consent for your data to be used as part of this study.

Debriefing

Thank you for participating in this study. The aim of the study is to examine how people think about different kinds of moral issues. The study investigates how people judge the rightness and wrongness of moral actions when they hear about actions singly, compared to when they hear about them together. Specifically, we are investigating if people make consistent judgments (i.e. that two good actions should be morally better than either action individually, and the same for bad actions). This study also investigates if people are more consistent when they make moral judgements on a scale with fewer intermediate points, or more. Thank you again for your participation. If you have any questions or concerns about the study, please do not hesitate to ask, or contact one of us at the following details: Eoin Gubbins, Room 3.01, Lloyd Building, Trinity College Dublin, Dublin 2, Ireland. egubbins@tcd.ie
Ruth Byrne: Lloyd Building, Trinity College Dublin, Dublin 2, Ireland.
rmbyrne@tcd.ie

Appendix 2.5 – Criteria for Consistency and Inconsistency in Experiment 3

Triples were judged to be consistent if they were additive (e.g. $A = 30$, $B = -20$, $A \& B = 10$).

Four categories of possible moral inconsistency were identified:

- (i) The conjunction was better than either conjunct (e.g. $A = 30$, $B = -10$, $A \& B = 50$).
- (ii) The conjunction was as good as the good conjunct (e.g. $A = 50$, $B = -20$, $A \& B = 50$).
- (iii) The conjunction was good, but not as good as the good conjunct (e.g. $A = 50$, $B = -30$, $A \& B = 40$).
- (iv) The conjunction was 0, eliminating a tendency to rate the immoral item higher in the conjuncts (e.g. $A = 50$, $B = -70$, $A \& B = 0$).

These categories were combined into a single category of moral inconsistent triples – these were triples in which the conjunction was more moral than the sum of the conjuncts would suggest it should have been.

Four categories of possible immoral inconsistency were identified:

- (i) The conjunction was worse than either conjunct (e.g. $A = 30$, $B = -10$, $A \& B = -30$).
- (ii) The conjunction was as bad as the bad conjunct (e.g. $A = 50$, $B = -20$, $A \& B = -20$).
- (iii) The conjunction was bad, but not as bad as the bad conjunct (e.g. $A = 50$, $B = -30$, $A \& B = -10$).
- (iv) The conjunction was 0, eliminating a tendency to rate the moral item higher in the conjuncts (e.g. $A = 50$, $B = -20$, $A \& B = 0$).

These categories were combined into a single category of moral inconsistent triples – these were triples in which the conjunction was more moral than the sum of the conjuncts would suggest it should have been.

Appendix 2.6 – Experiment 3 Materials

These materials are presented here with the immoral conjunct before the moral conjunct, although both orders were balanced in this experiment. The rating of the item in Experiment 2 is in parentheses.

Matched Pairs

1. A person beat the infant in her care (-82), and the next week donated one of her kidneys to a friend of hers (+84).
2. A person hit his son (-70), and the next week took his sister's child into his home after his sister died (+65).
3. A person forced the foreign girl who was working for him to accept a salary below minimum wage (-63), and the next week organized Christmas gifts for children at an orphanage (+62).
4. A person slept with his wife's friend (-67), and the next week organized a charity event to raise money for research into kidney disease (+62).
5. A person gave a teaching job to his nephew over better-qualified candidates (-53), and the next week donated money to a charity that helped to fund cancer research (+49).
6. A person embezzled funds (-49), and the next week intervened to help stop a fight (+51).
7. A person attacked a security guard (-49), and the next week paid for his dead sister's child to go to University (+48).
8. A person stole money from his employer (-49), and the next week helped an old woman to cross the street (+46).

Higher Moral

1. A person humiliated her assistant (-45), and the next week gave her neighbours money after their house burned down (+55).
2. A person cheated in an exam (-19), and the next week gave blood so that more would be available to help people who had been in accidents (+54).
3. A person parked in the space reserved for disabled employees (-35), and the next week helped save the victims of a terrorist bomb (+55).

4. A person terrorized their workplace canteen (-32)¹³, and the next week subsidized the construction of a badly needed burns unit (+54).

Higher Immoral

1. A person knowingly sold alcohol to a twelve-year-old (-55), and the next week took his wife for a romantic walk (+26).

2. A person stole her employer's jewellery (-55), and the next week stopped to help a motorist whose car had broken down by the side of the road (+47).

3. A person sexually abused one of his pupils (-92), and the next week rejected an opportunity to cheat on his wife (+49).

4. A person made it look as though a colleague had committed a crime so that he lost his job (-70), and the next week hired an equal number of able-bodied and disabled people (+57).

Neutral

1. A person finished her dinner at eight o'clock on a Tuesday evening, and the next week prepared to go to bed at eleven in the evening.

2. A person wrote an email to a colleague to arrange to meet for lunch, and the next week made a phone call to a hotel to enquire about the availability of their rooms.

3. A person bought a newspaper in a corner shop beside the train station, and the next week posted an important document to a business associate in another city.

4. A person sat on the couch in his apartment on a wet, rainy evening in the city, and the next week read a collection of detective stories quickly because he was very interested in what would happen next.

¹³ Score refers to "A school child terrorized the playground".

Appendix 2.7 – Experiment 3 Consent and Debriefing:

Instructions: Thank you for agreeing to participate in this experiment on reasoning. This study is being carried out by Eoin Gubbins (PhD. student) with the supervision of Prof. Ruth Byrne (Trinity College Dublin) and in collaboration with Prof. Phil Johnson-Laird (New York University), funded by a Government of Ireland PhD Scholarship from the Irish Research Council for the Humanities and Social Sciences. This experiment examines everyday reasoning; it is not a test of intelligence, and the aim of the study is to examine the sorts of answers most people provide. The study is anonymous and confidential. The data will be stored anonymously, and in accordance with the Freedom of Information and the Data Protection Acts. You are entitled to access data stored about you, in this case the data will be scored as group scores and participants will not be identified individually. You may request access to a copy of the final report when it is available if you wish. Participation in the experiment will take about 20 minutes and you are free to withdraw at any time without any penalty. If you choose to withdraw, your data will be destroyed. If you would like to participate, you can start by clicking on the 'next' button. There is no time limit on your responses, but please try to complete all the problems in a single session before exiting the program. Your task is to read a series of sentences. Each sentence describes one or more actions. Some of them are morally good actions, some of them are morally bad actions, and some of them are morally neutral actions. Your task is to rate each sentence in turn, on a scale that runs from -100 (Most immoral) to +100 (Most moral), with a midpoint of 0 (Neither moral nor immoral). You should click on the button on the scale that represents your judgement, and then click Next. Thank you again for agreeing to take part in the study. If you have any questions about it, or would like to access a summary of its findings, please do not hesitate to contact one of us at the following contact details: Eoin Gubbins, Room 3.01, Lloyd Building, Trinity College Dublin, Dublin 2, Ireland. Email: egubbins@tcd.ie OR Prof. Ruth Byrne, Lloyd Building, Trinity College Dublin, Dublin 2, Ireland. Email: rmbyrne@tcd.ie. By clicking to continue, you indicate that you understand the instructions given, and give your consent for your data to be used as part of this study.

Debriefing

Thank you for participating in this study. The aim of the study is to examine how people think about different kinds of moral issues. The study investigates how people judge the rightness and wrongness of moral actions when they hear about actions singly, compared to when they hear about them together. Specifically, we are investigating how people judge conjunctions of moral and immoral actions, and if this varies when they read the moral conjunct first or the immoral conjunct first. Thank you again for your participation. If you have any questions or concerns about the study, please do not hesitate to ask, or contact one of us at the following contact details: Eoin Gubbins: Room 3.01, Lloyd Building, Trinity College Dublin, Dublin 2, Ireland. egubbins@tcd.ie Ruth Byrne: Lloyd Building, Trinity College Dublin, Dublin 2, Ireland. rmbyrne@tcd.ie

Appendix 3.1 – Experiment 4 Materials

Train problem

You are at the wheel of a runaway train quickly approaching a fork in the tracks. On the tracks extending to the left is a group of five railway workmen. On the tracks extending to the right is a single railway workman. If you do nothing the train will proceed to the left, causing the deaths of the five workmen. The only way to avoid the deaths of these workmen is to hit a switch on your dashboard that will cause the train to proceed to the right, causing the death of the single workman

1. Would you hit the switch? (Please circle your answer)

- A. I would hit the switch
- B. I would not hit the switch

Footbridge problem

You are on a footbridge over the railway tracks towards which a runaway train is quickly approaching. On the tracks beyond the footbridge is a group of five railway workmen. If you do nothing the train will proceed on the tracks, causing the deaths of the five workmen. The only way to avoid the deaths of these workmen is to push a nearby stranger off the bridge so that his large body will stop the train, causing the death of the stranger.

1. Would you push the man? (Please circle your answer)

- A. I would push the man
- B. I would not push the man

2. Later that evening at home, you go over and over the situation in your mind. Your friends and family rally around and do everything they can to help you. When your closest friend arrives, you talk things over in private. You recollect as best you can how the situation arose and how it unfolded. But you are distressed to see that your friend appears genuinely shocked at the decision you made, although they try to hide it.

Non-directed

You decide to try your best to describe in detail to your friend your personal experience in those seconds the moments before you made your decision. You say “I knew I had to make a decision fast. This is what I experienced in those seconds...”

Reason-directed

You decide to try your best to describe in detail to your friend what thoughts were in your mind in the moments before you made your decision. You say, “I knew I had to make a decision fast. This is what I experienced in those seconds, the reasons and thoughts I had...”

Emotion directed

You decide to try your best to describe in detail to your friend what feelings were in your heart in the moments before you made your decision. You say, “I knew I had to make a decision fast. This is what I experienced in those seconds, the emotions and feelings I had...”

Appendix 3.2 – Experiment 4 Consent and Debriefing Forms

Consent Form

Thank you for agreeing to participate in this study on reasoning. The study is being carried out by Eoin Gubbins (PhD student), with the supervision of Prof. Ruth Byrne. The study examines everyday reasoning; is not a test of intelligence. The aim of the study is to examine the sorts of answers that most people provide.

Participation in this study will take about 5 minutes. You may withdraw at any time if you wish. The study is anonymous and confidential. The data will be scored as group scores and participants will not be identified individually. The data will be stored in accordance with the Freedom of Information Act (1997).

You will be given 2 brief stories to read, and your task is to answer a series of questions about it. Please read each one carefully in the order it appears in the booklet and respond to the questions that follow the scenario. Please answer the questions in the order that you are given them and please do not change any of your answers.

If you have any further questions about the study or would like to access a summary of the findings of the study, please do not hesitate to contact one of us at the following contact details:

You may detach this debriefing sheet and take it away with you if you wish. Please return the booklet with your answers to the experimenter.

Contact details:

PhD. Student:

Eoin Gubbins,

School of Psychology,

Trinity College Dublin.

Tel: 01 8968417

Supervisor:

Prof. Ruth Byrne,

School of Psychology,

Trinity College Dublin

Tel: 01 896 4054

This research is funded by a Government of Ireland PhD Scholarship from the Irish Research Council for the Humanities and Social Sciences.

Please first sign below to indicate that you have read and understand the information above and that you give your consent to participate.

Name (please print): _____

Signature: _____

Please also provide the following information:

Age: _____

Gender: _____

Debriefing

Thank you for participating in this study. The aim of the study is to examine how people think about different kinds of moral issues. The study investigates under what circumstances people would carry out certain actions, as well as the reasons and emotions involved in their decisions.

Thank you again for your participation. If you have any questions or concerns about the study please do not hesitate to ask, or contact one of us at the contact details below.

We do not anticipate that anything in the scenario presented will cause any distress to participants. However, if you find that the story does cause you distress, Trinity College Dublin operates a counselling service for students. They can be contacted by telephone at 01-8961407, or by email at student-counselling@tcd.ie. Their address is Student Counselling Service, 3rd Floor, Trinity College, 7-9 Sth. Leinster St., Dublin 2.

You may detach this debriefing sheet and take it away with you if you wish. Please return the booklet with your answers to the experimenter.

Contact details:

PhD. Student:

*Eoin Gubbins,
School of Psychology,
Trinity College Dublin.
Tel: 01 8968417*

Supervisor:

*Prof. Ruth Byrne,
School of Psychology,
Trinity College Dublin
Tel: 01 896 4054*

This research is funded by a Government of Ireland PhD Scholarship from the Irish Research Council for the Humanities and Social Sciences.

Appendix 3.3 – Experiment 5 Materials

The modification for impersonal and personal versions is given in italics. The modifications for the emotion (strong, mild) and decision (will, won't) conditions are presented in parentheses.

1. Train

Luke is *at the wheel of/on a bridge above* a runaway train quickly approaching a fork in the tracks. On the tracks extending to the left is a group of five railway workmen. *On the tracks extending to the right/On the bridge standing near to him* is a single railway workman/ *stranger*. The train is on the point of proceeding to the left, causing the deaths of the five workmen. The only way to avoid the deaths of these workmen is for him to *hit a switch on his dashboard/ push the stranger off the bridge* that will cause the train to *proceed to the right/ come to a stop*, leading to the death of the *single workman/stranger*. Facing this choice, Luke feels some strong (mild) emotions. He decides he will (won't) do it. Was Luke's decision appropriate?

2. Taxes

John is the owner of a small business completing his end of year tax return. His business has not done well this year and he is now trying desperately to make ends meet. He has just received a large tax bill that could lead to his business having to close down. It occurs to him that he could pretend that some of his personal expenses are business expenses when he *fills in the online form on his computer/ speaks to the revenue officer on the telephone*. Facing this choice, John feels some strong (mild) emotions. He decides he will (won't) do it. Was John's decision appropriate?

3. Lifeboat

Jill is on a cruise ship when there is a fire on board and the ship has to be abandoned. The lifeboat Jill is in charge of is overcrowded and begins to fill with water as the seas get rougher. The lifeboat is about to sink, forcing everyone in it into the water. The only way to stop the lifeboat from sinking is for Jill *to ask some of the other occupants to go overboard/ to push some of the other occupants overboard* into the freezing seas to float alongside it. Facing this choice, Jill feels some strong (mild) emotions. She decides she will (won't) do it. Was Jill's decision appropriate?

4. CV

Jen has been trying to find jobs without much success, when she is called for an interview for a great job for which she is very well qualified. While completing detailed questionnaires before her first interview, she discovers from the other candidates that they all have the experience that is required in the particular setting specified. Jen has the best qualifications for the job and plenty of experience but not in the particular setting specified. The decisions are about to be made about which candidates will be called for a second interview. The only way for her to get a second interview is *to tick the box on the questionnaire/ to tell lies to the interview panel during her interview* that she has the required experience in the setting specified. Facing this choice, Jen feels some strong (mild) emotions. She decides she will (won't) do it. Was Jen's decision appropriate?

5. Wallet

Mike is walking down the street when he comes across a wallet lying on the ground and close-by several 20 euro notes that have most likely fallen out of the wallet. From the credit cards and other items in the wallet it's clear that the wallet's owner is wealthy. Mike, on the other hand, has been hit by hard times recently and could really use some extra money. Mike sees that he will shortly pass by *the police station where he could leave the wallet at the desk/ the owner in her garden where he could hand the wallet to her himself*. As he is walking, he considers *leaving the wallet at the police desk/ handing the wallet to the owner and keeping the 20 euro notes for himself/ hiding the 20 euro notes in his pocket*. Facing this choice, Mike feels some strong (mild) emotions. He decides he will (won't) do it. Was Mike's decision appropriate?

6. Baby

Late one night while Pat's family is sleeping burglars break into her house. There has been a spate of very violent robberies in the area where families have been attacked and beaten badly. Pat and her mother are hiding in a wardrobe in the bedroom and her mother is on the mobile phone to the police whispering her address to them. Pat's baby brother in her arms begins to cry and she *holds his soother in his mouth firmly / presses her hand over his mouth* to block the sound. His crying will soon attract the

attention of the burglars who will harm the family. The only way for Pat to keep the family safe from the burglars is to *hold his soother in his mouth firmly/keep her hand pressed over the baby's mouth* even though he is struggling in her arms, clearly panicking and distressed. Facing this choice, Pat feels some strong (mild) emotions. She decides she will (won't) do it. Was Pat's decision appropriate?

7. Vaccine

Ann is a senior officer in charge of a government agency attempting to contain a serious viral epidemic. Many people who catch the virus suffer long-term illnesses from it. A minority of people who are vaccinated against the virus are allergic to the vaccine and suffer long-term illnesses from the side effects. An epidemic is predicted in which many people will catch the virus and suffer illnesses from it. The only way to avoid the illnesses from the virus is to *launch the vaccine for the whole population/to inject allergic volunteers to refine the vaccine*, which will cause the illnesses of allergic people. Facing this choice, Ann feels some strong (mild) emotions. She decides she will (won't) do it. Was Ann's decision appropriate?

8. Fumes

Paul is the late-night watchman in a hospital in which mildly toxic fumes are rising up through the hospital's ventilation system. There are 30 patients in wards on the first floor of the hospital. There are 10 patients in wards on the ground floor. The fumes are about to rise up into the wards on the first floor containing the 30 patients and cause them severe breathing discomfort all night. The only way to avoid the discomfort to these patients is for *Paul to close a vent to the first floor so that the closed vent cover/to lie across a vent to the first floor so that his body* will cause the fumes to remain on the ground floor leading to severe breathing discomfort all night for the 10 patients there. Facing this choice, Paul feels some strong (mild) emotions. He decides he will (won't) do it. Was Paul's decision appropriate?

Filler 1: Turnips

Tom is a farm worker driving a turnip-harvesting machine. He is approaching two diverging paths. By choosing the path on the left, Tom will harvest ten bushels of turnips. By choosing the path on the right, Tom will harvest twenty bushels of turnips. Tom's turnip-harvesting machine is on the point of turning to the left. The only way

for Tom to harvest twenty bushels of turnips instead of ten is to turn his machine to the right. Facing this choice, Tom feels some strong (mild) emotions. He decides he will (won't) do it. Was Tom's decision appropriate?

Filler 2: Brownies

Jess has decided to make a batch of brownies for herself. She opens her recipe book and finds a recipe for brownies. The recipe calls for a cup of chopped walnuts. Jess doesn't like walnuts, but she does like macadamia nuts. As it happens, she has both kinds of nuts available to her. The only way for Jess to really enjoy the brownies is to replace the walnuts with macadamia nuts. Facing this choice, Jess feels some strong (mild) emotions. She decides she will (won't) do it. Was Jess' decision appropriate?

Appendix 3.4 – Experiment 5 Consent and Debriefing Forms

Thank you for agreeing to participate in this study on reasoning. The study is being carried out by Eoin Gubbins (PhD student), with the supervision of Prof. Ruth Byrne. The study examines everyday reasoning, it is not a test of intelligence. The aim of the study is to examine the sorts of answers that most people provide.

Participation in this study will take about 20 minutes. You may withdraw at any time if you wish, without penalty. The study is anonymous and confidential. The data will be stored in accordance with the Freedom of Information and the Data Protection Acts. You are entitled to access data stored about you, in this case the data will be scored as group scores and participants will not be identified individually. You may request access to a copy of the final report when it is available if you wish.

In this study, first you will read and respond to 3 practice stories presented on the computer screen; then you will proceed to the main experiment, where you will be given 8 brief stories to read presented on the computer screen. These stories will appear sentence by sentence – to move on to read each sentence, please press the ‘Spacebar’ key, which is marked ‘Continue’. Please read each sentence carefully. Your task is to answer the question at the end of each story. You can provide your answer by pressing the keys marked ‘yes’ and ‘no’. You may take as much time as you wish. Please read each story carefully and answer the question that follows the story before moving on to the next one.

If you have any further questions about the study or would like to access a summary of the findings of the study, please do not hesitate to contact one of us at the following contact details:

PhD. Student:

*Eoin Gubbins,
School of Psychology,
Trinity College Dublin.
egubbins@tcd.ie
Tel: 01 896 4234.*

Supervisor:

*Prof. Ruth Byrne,
School of Psychology,
Trinity College Dublin
rmbyrne@tcd.ie
Tel: 01 896 4054*

This research is funded by a Government of Ireland PhD Scholarship from the Irish Research Council for the Humanities and Social Sciences.

Consent Form

Please first sign below to indicate that you have read and understand the information above and that you give your consent to participate. You may keep the information section of this page if you like; please give the consent section to the experimenter.

Name (please print): _____

Signature: _____

Please also provide the following information:

Age: _____ Gender: _____

What is your first language? _____

Debriefing

Thank you for participating in this study by Eoin Gubbins (PhD student), supervised by Professor Ruth Byrne. The aim of the study is to examine how people think about different kinds of moral issues. The stories varied whether the emotion experienced by the actor (strong or mild) is in line with what most people expect, and whether the decision the actor made (to do the action or not) is in line with what most people expect. We are examining whether these expectations 'prime' people to make moral judgments of certain sorts, and whether people make moral judgments more quickly when the emotions and decisions match expectations.

Thank you again for your participation. If you have any questions or concerns about the study please do not hesitate to ask, or contact one of us at the contact details below.

We do not anticipate that anything in the stories presented will cause any distress to participants. However, if you find that any of the stories or your participation in the study does cause you some distress, you may wish to contact a support service, e.g. Samaritans Ireland at Tel: 01 872 7700 (address 151 Marlborough Street, Dublin 1, Website: www.samaritans.org). You may detach this debriefing sheet and take it away with you if you wish. Please return the consent form to the experimenter.

PhD. Student:

Eoin Gubbins,
School of Psychology,
Trinity College Dublin.
egubbins@tcd.ie
Tel: 01 896 4234.

Supervisor:

Prof. Ruth Byrne,
School of Psychology,
Trinity College Dublin
rmbyrne@tcd.ie
Tel: 01 896 4054

Appendix 4.1 – Materials for the Good condition in Experiment 6

Materials for the other conditions and other experiments are a variation of these as described in the text).

Please indicate your current mood (circle one):

0-1-2-3-4-5-6-7-8-9-10

very sad

very happy

Please think of an experience in which you saw someone, or heard about someone, demonstrating humanity's higher or better nature. Please pick an example in which you were not the beneficiary, that is, someone did something good, honorable, or charitable *for someone else*. Please think of a situation that resulted in little or no negative feeling, or at least in which the positive feelings were much stronger than the negative. Please describe the experience briefly, in writing, here:

I consent to my memory being displayed on the Science Gallery noticeboard:

Yes / No

Please now answer the questions on the other side of the page.

PLEASE DO NOT READ THE QUESTIONS UNTIL YOU HAVE COMPLETED THE TASK ON THE OTHER SIDE OF THE PAGE! PLEASE ANSWER THE QUESTIONS IN THE ORDER IN WHICH THEY APPEAR.

ID No.:

Age:

Gender: M / F

Nationality:

Country of residence:

1. People sometimes imagine how an event could have turned out differently 'if only...' Please think about the experience and about how it could have turned out differently. Complete the following sentence: 'It could have turned out differently if...'

2. Was the behaviour expected or unexpected (circle one):

0-1-2-3-4-5-6-7-8-9-10

Unexpected

Expected

3. Was the behaviour intentional or unintentional (circle one):

0-1-2-3-4-5-6-7-8-9-10

Unintentional

Intentional

4. Was it an action (something someone does) or an inaction (something someone does not do) (circle one):

Action / Inaction

5. Was it a specific experience (an event that occurred at a specific time) or a general one (a non-specific experience) (circle one):

Specific / General

6. The single best word to describe your feelings at the time?

7. What was your age at the time? ____ years

8. Did you think about changing your own behaviour after the event? If so, describe what you thought about changing:

9. Was the main person in the event someone who was at the time (circle one):

Very close to you / Known to you / A stranger

10. Did the main person in the event come from (circle one):

Family/ Friends/ Social acquaintances/ Work Colleagues/ Sports/Politics/

Religion/ Stranger/

Other (please specify) _____

11. Please indicate your current mood (circle one):

0-1-2-3-4-5-6-7-8-9-10

very sad

very happy

Appendix 4.2. – Experiment 6 Exclusion Criteria

Twenty-five participants were excluded prior to any data analysis for a variety of reasons. Twenty-two participants were excluded because their answers to one of the questions, ‘Did you think about changing your own behaviour after the event?’ (9 participants) and ‘the single best word to describe your feelings at the time’ (4 participants), or their answers to both of those questions (9 participants) did not appear to relate to the memory they produced. For example, a participant who remembered their friend being punched by a security guard responded to the question about thinking about changing their behaviour, “*Yey, learn more*”. Another participant who recalled hearing about a man who mistreated his dog summed up his feelings about the event in the single word, “*Relief*”. Some participants also failed to follow the instructions of the condition they were given, i.e., they recorded a memory of something involving themselves either as actor, beneficiary or victim (2 participants; for instance, one participants remembered a time when they were injured falling off their bike and another cyclist remained with them until help arrived), or, they recorded a memory that did not involve someone else doing something (1 participant who provided a memory of his brother dying as a bad memory, but did not provide any details of someone doing something bad to someone else). Furthermore, the analysis of the counterfactual thoughts excluded 3 participants because their counterfactuals did not relate to their memory (e.g. one participant remembered a situation where a group of people were aggressive and yelled at a homeless man, and recorded the counterfactual, if only “*I had understood what was being offered and pursued it*”, suggesting the counterfactual related to a different memory). However, these 3 participants were included in all other analyses, which were not based on the counterfactual but on the memory and the participants’ feelings towards them.

Appendix 4.3 – Experiment 6 Inter-Rater Reliability

The author and one independent rater scored the entire dataset. For the upward/downward categorisation, inter-rater agreement was 94%, Cohen's Kappa = .955, $p < .001$; for the additive/subtractive categorisation, agreement was 90%, Cohen's Kappa = .875, $p < .001$; and for intentions to change behaviour, agreement was 93%, Cohen's Kappa = .887, $p < .001$. Another independent rater scored 20% of responses, and agreement between the independent raters was 77% for upward/downward, Cohen's Kappa = .700, $p = .001$, 91% for additive/subtractive, Cohen's Kappa = .897, $p < .001$, and 100% for intentions to change behaviour, Cohen's Kappa = 1.00, $p < .001$. Disagreements were resolved by discussion. In cases in which a participant either did not provide a response, or their response was ambiguous between two categories, their response was coded as missing in the analysis (4% of responses for upward/downward; 7% for additive/subtractive; 4.7% for thinking about changing one's own behaviour).

The author, the supervisor and an independent rater coded all intentions to change on whether the intention was specific (i.e. related to a specific behaviour, such as one participant who said, "*I thought about giving blood more often*"), or general (i.e. a desire to change not focused on a specific behaviour, such as one participant who said, "*developing more patience*"). Agreement between the author and supervisor was 81%, Cohen's Kappa = .651, $p < .001$. Agreement between the author and the independent rater was 81%, Cohen's Kappa = .669, $p < .001$. Agreement between the supervisor and the independent rater was 85%, Cohen's Kappa = .728, $p < .001$. Disagreements were resolved by discussion.

All responses of whether a memory was an action or an inaction, and specific or general were coded by the author and an independent rater, and agreement was 92% for action-inaction, Cohen's Kappa = .169, $p = .051$, and 88% for specific-general, Cohen's Kappa = .569, $p < .001$. Twenty percent of responses were coded by another independent rater, and agreement between the independent raters for action-inaction was 91%, Cohen's Kappa = .450, $p = .035$, and 82% for specific-general, Cohen's Kappa = .425, $p = .008$. Disagreements were resolved by discussion. (The latter relatively low Cohen's Kappas, compared to the high percentage of inter-rater agreement, are a result of the asymmetrical distribution of action/inaction and

specific/general responses, wherein most memories involved specific actions, leaving relatively few responses in other categories).

The author and an independent rater coded whether the single words to sum up the participants' feelings were positive, negative or ambiguous between the two. Agreement was 94%, Cohen's Kappa = .903, $p < .001$. Disagreements were resolved by discussion.

Appendix 4.4. – Experiment 7 Exclusion Criteria

Sixty-four participants were excluded prior to any data analysis for failure to follow the instructions. The most common reason for participants to be excluded in this experiment was that their statement about their intention to change, and also their single word to sum up their feelings about the memory did not match their memory (22 participants). Participants were also excluded if only one of those issues arose – that is, if their statement about their intention to change did not match their memory (12 participants) or if their single word to sum up their feelings towards the event did not match their memory (9 participants). Participants were excluded if their memory involved themselves as either actor, beneficiary or victim (5 participants), or if nobody did anything good or bad for or to someone else in their memory (depending on condition; 5 participants). Two participants were excluded because they provided a memory with no details (e.g. a participant responded to the memory prompt by saying, *“I really can't think of anything, so I'm going to power through and fill in the stuff on the back, sorry!”*). Four participants were excluded because their memory involved themselves, and also their statement about intentions to change did not match their memory. Two participants did not provide a memory. One participant was excluded because their memory involved themselves and their single word to sum up their feelings did not match their memory. One participant was excluded because their memory involved themselves and contained no details as to the event involved. One participant was excluded because their memory was in French, and an attempt at translating it revealed the participant indicated they did not understand a detail of the question. As in Experiment 6, the analysis of counterfactuals excluded some participants (9) because their counterfactual did not relate to their memory. However, these participants were included in all other analyses, which were not based on the counterfactual but on the memory and the participants' feelings towards them.

Appendix 4.5. – Experiment 7 Inter-Rater Reliability

As in Experiment 6, two independent raters scored the counterfactuals as upward/downward and as additive/subtractive; and whether the participant indicated that they thought about changing their behaviour. The author and one independent rater scored the entire dataset, and inter-rater agreement was 91% for upward/downward, Cohen's Kappa = .969, $p < .001$, 90% for additive/subtractive, Cohen's Kappa = .868, $p < .001$, and 91% for intentions to change their behaviour, Cohen's Kappa = .821, $p < .001$. Another independent rater scored 20% of responses, and inter-rater agreement between the independent raters was 94% for upward/downward, Cohen's Kappa = 1.00, $p < 0.001$, 89% for additive/subtractive, Cohen's Kappa = .748, $p < .001$, and 100% for intention to change their behaviour, Cohen's Kappa = 1.00, $p < .001$. All disagreements were resolved by discussion. In cases where a participant either did not provide a response, or their response was ambiguous between two categories, their response was coded as missing in the analysis (9% of responses for upward/downward; 10% for additive/subtractive; 7% for thinking about changing one's own behaviour).

The author, the supervisor and an independent rater coded whether the intentions to change behaviour provided by participants were specific or general. Agreement between the author and supervisor was 74%, Cohen's Kappa = .557, $p < .001$. Agreement between the author and the independent rater was 78%, Cohen's Kappa = .612, $p < .001$. Agreement between the independent rater and supervisor was 71%, Cohen's Kappa = .486, $p < .001$. Disagreements were resolved by discussion.

The author and an independent rater scored all memories for action/inaction, and specific/general. Inter-rater agreement for action/inaction was 92%, Cohen's Kappa = .600, $p < .001$, and for specific/general, agreement was 82%, Cohen's Kappa = .584, $p < .001$. A second independent rater scored 20% of all memories, and agreement between the two independent raters was 94%; Cohen's Kappa could not be calculated because one rater coded all items as actions, meaning one variable was a constant. For specific/general, agreement between the independent raters was 79%, Cohen's Kappa = .669, $p < .001$. Again, disagreements were resolved by discussion.

The author and an independent rater coded whether the single words to sum up the participants' feelings were positive, negative or ambiguous between the two.

Agreement was 92%, Cohen's Kappa = .868, $p < .001$. Disagreements were resolved by discussion.

Appendix 4.6. – Experiment 8 Exclusion Criteria

Thirty-one participants were excluded prior to any data analysis for a variety of reasons. The most common reason people were excluded was because both their single word to sum up their feelings, and their thinking about changing responses did not relate to their memory (14 participants). Another 8 participants' single word to sum up their feelings did not relate to their memory. Three participants did not produce a memory. Another 2 participants' thinking about changing response did not relate to their memory. Two participants produced memories that involved themselves in an other-focused condition. Two participants produced memories in which it was not clear anybody did anything either good or bad. Overall 15 of the participants were excluded only from the analysis of the counterfactuals because their counterfactual thought did not relate to the memory they recorded or because they did not provide a counterfactual. These participants were included in all other analyses, which were not based on the counterfactual but on the memory and the participants' feelings towards them.

Appendix 4.7. – Experiment 8 Inter-Rater Reliability

The author and an independent rater scored all of the counterfactuals as upward/downward, and as additive/subtractive, and whether they intended to change their behaviour. Agreement was 95% for upward/downward, Cohen's Kappa = .899, $p < .001$; for additive/subtractive counterfactuals agreement was 85%, Cohen's Kappa = .816, $p < .001$; and for intention to change behaviour, agreement was 90%, Cohen's Kappa = .905, $p < .001$. Another independent rater coded 20% of counterfactuals, and agreement between the independent raters was 97% for upward/downward counterfactuals, Cohen's Kappa = .934, $p < .001$, 91% for additive/subtractive counterfactuals, Cohen's Kappa = .814, $p < .001$, and 94% for intentions to change their behaviour, Cohen's Kappa = .835, $p < .001$. Disagreements were resolved by discussion.

The author, the supervisor, and an independent rater coded all intentions to change on whether the intention was specific or general. Agreement between the author and the supervisor was 68%, Cohen's Kappa = .454, $p < .001$. Agreement between the author and the independent rater was 72%, Cohen's Kappa = .477, $p < .001$. Agreement between the supervisor and the independent rater was 58%, Cohen's Kappa = .278, $p < .005$. Disagreements were resolved by discussion.

The author and an independent rater coded all memories as actions or inactions, and as specific or general. For action/inaction, agreement was 95%, Cohen's Kappa = .205, $p = .003$; for specific/general, agreement was 83%, Cohen's Kappa = .549, $p < .001$. A second independent rater coded 20% of memories, and agreement between the independent raters was 93% for action/inaction, Cohen's Kappa = .460, $p = .017$, and 89% for specific/general, Cohen's Kappa = .779, $p < .001$. Disagreements were resolved by discussion. Since memories were not provided for the self-focused bad memory condition, participants' own judgments of whether the memory was an action or an inaction, and specific or general, were used for this condition.

The author and an independent rater coded all single words to sum up the participants feelings as to whether they were positive, negative or ambiguous. Agreement was 89%, Cohen's Kappa = .819, $p < .001$. Disagreements were resolved by discussion.