"I don’t know why it is wrong, but it is just wrong!": An empirical investigation of moral dumbfounding and what (if anything) it reveals about the nature of moral judgment.

Tammy Orreal
Bachelor of Psychology (Hons) Class 1

This dissertation is presented for the degree of Doctor of Philosophy (Psychology) at Charles Sturt University
Faculty of Business, Justice, and Behavioural Sciences
School of Psychology
25 August 2019
# Table of Contents

Table of Contents ............................................................................................................ i
Certificate of Authorship .................................................................................................... v
Acknowledgement of Assistance ....................................................................................... vi
List of Tables ...................................................................................................................... viii
List of Figures .................................................................................................................... xii
List of Appendices ............................................................................................................. xiii
Glossary ............................................................................................................................. xiv
Abstract ........................................................................................................................... xv

Chapter 1: Moral Dumbfounding: A Brief History ....................................................... 1
  Haidt's Social Intuitionist Model ................................................................................... 1
  Evidence for the Social Intuitionist Model ................................................................. 3

Chapter 2: Empirical Investigations of Moral Dumbfounding .................................... 7
  Extant Empirical Research ......................................................................................... 7

Chapter Summaries ......................................................................................................... 12
  Chapter 3: Study 1 – Replicating and extending the original moral dumbfounding study ........................................................................................................... 12
  Chapter 4: Study 2 – Defining moral dumbfounding .................................................. 13
  Chapter 5: Study 3 – Investigating the moral dumbfounding measures ....................... 14
  Chapter 6: Study 4 – Exploring individual differences in moral dumbfounding ....... 14
  Chapter 7: Does moral dumbfounding reveal the nature of moral judgment? .......... 15

Chapter 3: Study 1 – Replicating and Extending the Original Moral Dumbfounding Study ........................................................................................................ 17

Method ............................................................................................................................. 18
  Participants .................................................................................................................... 19
  Research design ......................................................................................................... 20
  Apparatus and materials ........................................................................................... 20
  Measures ..................................................................................................................... 23
  Procedure ................................................................................................................... 30
  Data analysis .............................................................................................................. 32

Results .............................................................................................................................. 36
  Moral dumbfounding .................................................................................................. 36
Judgments of wrongness of an action ..............................................37
Change of judgment ........................................................................40
Response timing .............................................................................40
Basis underpinning judgments .........................................................44
Confidence in judgment ...................................................................45
Confusion and irritation ...................................................................47
Psychic discomfort ..........................................................................48
Speech dysfluency ...........................................................................49
Linguistic responses .........................................................................50
Novel measures ................................................................................54

Discussion ..........................................................................................60
Moral dumbfounding .........................................................................60
Claim 1: Differences between the reasoning stimulus and the intuition stimuli ...........................................................................61
Claim 2: Moral judgment is an intuitive response ................................65
Claim 3: Moral judgment is emotionally based ...................................65
Claim 4: Reasoning is post-hoc in the formation of moral judgment .........................................................................................66
Claim 5: Moral dumbfounding is evidence of judgment without reason ..........................................................67
Limitations ..........................................................................................70
Conclusion ..........................................................................................71

Chapter 4: Study 2 – Defining Moral Dumbfounding ..........................73
Definitions of moral dumbfounding .....................................................73
Common definition ...........................................................................73
Original definition ............................................................................74
Judgment without justification definition ...........................................74
Confidence definition ........................................................................75
Credulity definition ...........................................................................75
Puzzlement definition ........................................................................76
Evidence-based definition ...................................................................76
The relative contribution of key indicators of moral dumbfounding ...77
Belief that an action is harm-free .......................................................77
Puzzlement at an inability to justify ...................................................78
Methods ...............................................................................................79
Participants .......................................................................................79
Research design ...............................................................................79
Measures and procedure ....................................................................80
Certificate of Authorship

I, Tammy Orreal, hereby declare that this submission is my own work and to the best of my knowledge and belief, understand that it contains no material previously published or written by another person, nor material which to a substantial extent has been accepted for the award of any other degree or diploma at Charles Sturt University or any other educational institution, except where due acknowledgement is made in the dissertation. Any contribution made to the research by colleagues with whom I have worked at Charles Sturt University or elsewhere during my candidature is fully acknowledged. I agree that this dissertation be accessible for the purpose of study and research in accordance with normal conditions established by the Executive Director, Library Services, Charles Sturt University, or nominee, for the care, loan and reproduction of dissertation, subject to confidentiality provisions as approved by the University.

Signature: [Signature]

Date: 25 August 2019
Acknowledgement of Assistance

This research was supported financially by a Charles Sturt University Postgraduate Research Scholarship.

Dr Andrew McGrath and Dr Lauren Saling provided supervisory support. The role of Dr McGrath was primarily as a ‘methodological’ specialist and Dr Saling’s role was primarily as a ‘theoretical knowledge/development’ specialist; however, both provided guidance on all aspects of the research process as required, and constructive criticism on the content and format of this dissertation. Special acknowledgement and gratitude are extended to Dr Saling whose dedication to this research was such that she continued supervision in an adjunct capacity after resigning from Charles Sturt University for employment with RMIT University.

Sincere appreciation is offered to Dr Daniel Cohen for providing guidance in navigating the intersection of moral psychology and philosophy.

I am indebted to Jason Scarborough and Kaitlyn Scarborough without whom the fine-grained empirical research within this dissertation would have been unattainable. Jason provided technical support such as setting up computer and recording equipment and designing online processes for the purpose of data collection. He also provided research assistance by reviewing and editing interview transcripts, peer-reviewing investigator coding, and checking the integrity of each row of merged data-sets, to ensure the accuracy of data. Kaitlyn provided research assistance by independently coding all interviews for verbal and behavioural variables for reliability of results. She also reviewed and edited interview transcripts, and checked the integrity of each row of merged data-sets, to ensure the accuracy of data.

Assistance was received from Stewart Street Veterinary Hospital in Bathurst, NSW, who provided autoclaving services for the purpose of sterilising the cockroaches used in the Roach task.

Corrective Services NSW provided employment relief in the form of four hours paid study leave per week.

Third party editorial assistance was provided in the preparation of the dissertation by professional editor, Rosemary Purcell, who provided copyediting and proofreading services according to the guidelines laid out in the university-endorsed national Guidelines for Editing Research Theses.

And finally, last but by no means least, very special acknowledgment is given to my family. Thank you for your interest in this research and the many hours of
stimulating discussion. Thank you for picking up the slack at home when I needed to concentrate on my work. Thank you for your encouragement at every step along the way and for holding my head above water when I didn’t think I could make it. Because of your unfailing support, any attainment deriving from this research is owed to you, and I will be forever grateful.
List of Tables

Table 1 Frequency of Participants who were Classified as Morally Dumbfounded at Least Once in Response to the Dumbfounding Stimuli ................................................. 36

Table 2 Frequency and Mean Ranking of Initial and Final Judgments of ‘Wrong’ in Response to the Dumbfounding Stimuli ................................................................. 39

Table 3 Frequency of Participants Who Changed Their Judgment of the Dumbfounding Stimuli, from Initial to Final Judgment ......................................................... 40

Table 4 Range and Mean of Total Response Time to the Dumbfounding Stimuli ......... 41

Table 5 Mean Response Times and Mean Ranking of Time to Initial Judgment, Time to Initial Justification, and Time Difference Between Judgment and Justification, Across the Dumbfounding Stimuli .................................... 42

Table 6 Mean and Mean Ranking of the Basis (i.e., Reasoning or Gut Feeling) Underpinning Initial and Final Judgments of the Dumbfounding Stimuli ................. 44

Table 7 Mean and Mean Ranking of Confidence in Judgment of the Dumbfounding Stimuli ........................................................................................................... 46

Table 8 Mean and Mean Ranking of Confused by Task, and Irritated by Task, in Response to the Dumbfounding Stimuli ............................................................. 48

Table 9 Mean and Mean Ranking of Psychic Discomfort in Response to the Dumbfounding Stimuli ............................................................................................ 49

Table 10 Mean and Mean Ranking of Speech Dysfluency in Response to the Dumbfounding Stimuli ........................................................................................... 50

Table 11 Mean and Mean Ranking of Dead-End Arguments in Response to the Dumbfounding Stimuli ..................................................................................... 51

Table 12 Mean and Mean Ranking of “I Don’t Know” Statements in Response to the Dumbfounding Stimuli ............................................................................. 52

Table 13 Mean and Mean Ranking of Unsupported Declarations in Response to the Dumbfounding Stimuli ............................................................................ 53
Table 14 Mean and Mean Ranking of Dumbfounding Statements in Response to the Dumbfounding Stimuli ................................................................. 54

Table 15 Mean and Mean Ranking of the Composite Variable of Failure to Provide a Reason for Judgment, and the Component Variables of Tautological Reasons and Illogical Arguments, in Response to the Dumbfounding Stimuli ........................................ 55

Table 16 Mean and Mean Ranking of the Composite Variable of Struggle to Produce a Reason, and the Component Variable of Statements of Unwillingness to Support, in Response to the Dumbfounding Stimuli ............................................................................. 57

Table 17 Mean and Mean Ranking of Unique Justifications for a Judgment of the Dumbfounding Stimuli ........................................................................ 58

Table 18 Mean and Mean Ranking of Statements of Puzzlement at an Inability to Justify a Judgment of the Dumbfounding Stimuli ........................................ 59

Table 19 Frequency of Participants Classified as Morally Dumbfounded (at Least Once) in Response to the Dumbfounding Stimuli, According to Each Dumbfounding Definition .............................................................................. 86

Table 20 Mean Ranking, z Scores, Corrections for Ties, and Effect Sizes of Pairwise Comparisons Between the Common Definition and Other Definitions of Moral Dumbfounding ................................................................. 87

Table 21 z Scores, Corrections for Ties, and Effect Sizes of Pairwise Comparisons Between the Judgment Without Justification Definition and Other Definitions of Moral Dumbfounding ........................................................................ 88

Table 22 z Scores, Corrections for Ties, and Effect Sizes of Pairwise Comparisons Between the Confidence Definition and Other Definitions of Moral Dumbfounding ................................................................. 89

Table 23 Unstandardised and Standardised Beta Regression Coefficients, and Squared Semi-Partial Correlations, of Each Predictor Variable, on Each Step of a Hierarchical Multiple Regression Model Predicting Moral Dumbfounding ......................................................... 90

Table 24 Frequency of Participants who were Classified as Morally Dumbfounded at Least Once in Response to the Alternative Dumbfounding Stimuli ....................... 112
<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>25</td>
<td>Mean and Mean Ranking of Belief that the Action Depicted in the Dumbfounding Stimuli, by Presentation Medium, is Realistic</td>
</tr>
<tr>
<td>26</td>
<td>Frequency and Mean Ranking of Initial and Final Judgments of ‘Wrong’ in Response to the Alternative Dumbfounding Stimuli</td>
</tr>
<tr>
<td>27</td>
<td>Frequency of Participants Who Changed Their Judgment of ‘Wrong’ to the Alternative Dumbfounding Stimuli</td>
</tr>
<tr>
<td>28</td>
<td>Range and Mean of Total Response Time to the Alternative Dumbfounding Stimuli</td>
</tr>
<tr>
<td>29</td>
<td>Mean Response Times and Mean Ranking of Time to Initial Judgment, Time to Initial Justification, and Time Difference Between Judgment and Justification, Across the Alternative Dumbfounding Stimuli</td>
</tr>
<tr>
<td>30</td>
<td>Mean and Mean Ranking, of the Basis (i.e., Reasoning or Gut Feeling) Underpinning Initial and Final Judgments of the Alternative Dumbfounding Stimuli</td>
</tr>
<tr>
<td>31</td>
<td>Mean and Mean Ranking of Psychic Discomfort in Response to the Alternative Dumbfounding Stimuli</td>
</tr>
<tr>
<td>32</td>
<td>Mean and Mean Ranking of “I Don’t Know” Statements in Response to the Alternative Dumbfounding Stimuli</td>
</tr>
<tr>
<td>33</td>
<td>Mean and Mean Ranking of the Composite Variable of Failure to Provide a Reason for Judgment, and the Component Variables of Unsupported Declarations, Tautological Reasons, and Illogical Arguments, in Response to the Alternative Dumbfounding Stimuli</td>
</tr>
<tr>
<td>34</td>
<td>Mean and Mean Ranking of the Composite Variable of Struggle to Produce a Reason, and the Component Variables of Dumbfounding Statements and Statements of Unwillingness to Support, in Response to the Alternative Dumbfounding Stimuli</td>
</tr>
<tr>
<td>35</td>
<td>Mean and Mean Ranking of Unique Justifications for a Judgment of the Alternative Dumbfounding Stimuli</td>
</tr>
<tr>
<td>36</td>
<td>Mean and Mean Ranking of Statements of Puzzlement at an Inability to Justify a Judgment of the Alternative Dumbfounding Stimuli</td>
</tr>
</tbody>
</table>
Table 37 Mean and Mean Ranking of Belief that the Action Depicted in the Dumbfounding Stimulus is Realistic .............................................................. 135

Table 38 Frequency of Participants who were Classified as Morally Dumbfounded at Least Once in Response to the Original and Alternative Dumbfounding Stimuli for Both the Common and Evidence-based Definitions ........................................ 136

Table 39 Descriptive Demographic Statistics of the Sample for Study 4 (and Studies 1 and 3 that Comprise this Sample) ............................................................. 155

Table 40 Sample Items and Internal Consistency Reliabilities for The Big Five Aspects Scales (DeYoung et al., 2007) .......................................................................... 165

Table 41 Items and Internal Consistency Reliabilities for the Susceptibility to Experimenter Pressure Scale ............................................................................. 167

Table 42 Frequency of Participants who were Classified as Morally Dumbfounded at Least Once in Response to the Stimuli Across Socioeconomic Proxy Measures of Parental Class and Income ........................................................................ 170

Table 43 Frequency of Participants who were Classified as Morally Dumbfounded at Least Once in Response to the Stimuli Across Non-Religious Belief ...................... 171

Table 44 Unstandardised Beta Regression Coefficients, Wald Statistics, and Odds Ratios of Each Predictor Variable in a Model Predicting Moral Dumbfounding ........ 173
List of Figures

Figure 1. Diagrammatic representation of standardised interview protocol .................. 31

Figure 2. Clustered bar chart illustrating the difference between the Heinz vignette and the three intuition stimuli for which there was a significant effect in the prevalence of moral dumbfounding ................................................................. 37

Figure 3. Clustered column chart illustrating the frequency of participant initial and final judgments to the dumbfounding stimuli ................................................................. 38

Figure 4. Clustered column chart illustrating the significant effect between puzzlement and moral dumbfounding elicited using the dumbfounding stimuli .......................... 60

Figure 5. Clustered column chart illustrating the frequency of participant initial and final judgments to the alternative dumbfounding stimuli ........................................ 115

Figure 6. Clustered column chart illustrating the significant effect between puzzlement and moral dumbfounding elicited using the alternative dumbfounding stimuli ......... 133

Figure 7. Clustered column chart illustrating the significant effect between moral dumbfounding when defined according to the Common definition compared to the Evidence-based definition .................................................... 137
# List of Appendices

<table>
<thead>
<tr>
<th>Appendix</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Ethics Documents</td>
<td>197</td>
</tr>
<tr>
<td>A.1</td>
<td>CSU Human Research Ethics Committee Approval</td>
<td>197</td>
</tr>
<tr>
<td>A.2</td>
<td>Information Statement</td>
<td>199</td>
</tr>
<tr>
<td>A.3</td>
<td>Consent Form</td>
<td>201</td>
</tr>
<tr>
<td>B</td>
<td>Photographs of the Data Collection Laboratory</td>
<td>203</td>
</tr>
<tr>
<td>B.1</td>
<td>Equipment View from Participant Seat</td>
<td>203</td>
</tr>
<tr>
<td>B.2</td>
<td>Equipment View from Experimenter Seat</td>
<td>203</td>
</tr>
<tr>
<td>B.3</td>
<td>Data Collection Laboratory Set-Up</td>
<td>204</td>
</tr>
<tr>
<td>B.4</td>
<td>Participant and Experimenter Positioning</td>
<td>204</td>
</tr>
<tr>
<td>C</td>
<td>Stimuli Employed in Study 1</td>
<td>205</td>
</tr>
<tr>
<td>C.1</td>
<td>The Heinz Dilemma</td>
<td>205</td>
</tr>
<tr>
<td>C.2</td>
<td>The Incest Story</td>
<td>206</td>
</tr>
<tr>
<td>C.3</td>
<td>The Cannibal Story</td>
<td>207</td>
</tr>
<tr>
<td>C.4</td>
<td>The Roach Task</td>
<td>208</td>
</tr>
<tr>
<td>C.5</td>
<td>The Soul Task</td>
<td>209</td>
</tr>
<tr>
<td>D</td>
<td>Questionnaires Employed in the Current Study</td>
<td>210</td>
</tr>
<tr>
<td>D.1</td>
<td>After Each Task Questionnaire</td>
<td>210</td>
</tr>
<tr>
<td>D.2</td>
<td>Demographic Questionnaire</td>
<td>212</td>
</tr>
<tr>
<td>E</td>
<td>Hierarchical Coding Scheme for Variables Derived from Data</td>
<td>216</td>
</tr>
<tr>
<td>F</td>
<td>Stimuli Employed in Study 3</td>
<td>219</td>
</tr>
<tr>
<td>F.1</td>
<td>The John Dilemma</td>
<td>219</td>
</tr>
<tr>
<td>F.2</td>
<td>The Masturbation Story</td>
<td>220</td>
</tr>
<tr>
<td>F.3</td>
<td>The Liver Story</td>
<td>221</td>
</tr>
<tr>
<td>F.4</td>
<td>The Lolly Task</td>
<td>222</td>
</tr>
<tr>
<td>F.5</td>
<td>The Extended Soul Task</td>
<td>223</td>
</tr>
<tr>
<td>F.6</td>
<td>The Atlas Video</td>
<td>224</td>
</tr>
<tr>
<td>F.7</td>
<td>The Spot Video</td>
<td>225</td>
</tr>
<tr>
<td>G</td>
<td>The Big Five Aspect Scales Personality Questionnaire</td>
<td>226</td>
</tr>
<tr>
<td>H</td>
<td>Susceptibility to Experimenter Pressure</td>
<td>229</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Description</td>
<td></td>
</tr>
<tr>
<td>--------------</td>
<td>-----------------------------------</td>
<td></td>
</tr>
<tr>
<td>ANOVA</td>
<td>analysis of variance</td>
<td></td>
</tr>
<tr>
<td>BFAS</td>
<td>Big Five Aspect Scales</td>
<td></td>
</tr>
<tr>
<td>CSU</td>
<td>Charles Sturt University</td>
<td></td>
</tr>
<tr>
<td>MANOVA</td>
<td>multivariate analysis of variance</td>
<td></td>
</tr>
<tr>
<td>MRA</td>
<td>multiple regression analysis</td>
<td></td>
</tr>
<tr>
<td>PBAC</td>
<td>Pharmaceutical Benefits Advisory Committee</td>
<td></td>
</tr>
<tr>
<td>PBS</td>
<td>Pharmaceutical Benefits Scheme</td>
<td></td>
</tr>
<tr>
<td>SES</td>
<td>socioeconomic status</td>
<td></td>
</tr>
<tr>
<td>SIM</td>
<td>Social Intuitionist Model</td>
<td></td>
</tr>
<tr>
<td>TGA</td>
<td>Therapeutic Goods Administration</td>
<td></td>
</tr>
</tbody>
</table>
Abstract

Think of any contentious political-social issue in which you resolutely judge the other side of the debate to be wrong. It is likely that you would remain steadfast in your judgment even if you had difficulty articulating a reason for your judgment or were made aware of the shortcomings of your justification. Sometimes we know that something is wrong despite an inability to provide coherent reasons. This phenomenon, known as moral dumbfounding, has been proposed as evidence that moral judgments are not caused by a process of reasoning; presumably, one would be able to provide reasons to justify a judgment if reasoning was involved in the formation of moral judgment. The possible implications of this assumption can be understood in everyday terms: if it is the case that moral judgments are merely primitive feelings without logical justification, then genuine moral discourse is not possible. Moral dumbfounding is clearly an important and interesting phenomenon but the empirical evidence for it, to date, is very limited. The broad aim of the present dissertation was to establish whether moral dumbfounding is a real phenomenon. Through a series of empirical studies, support for moral dumbfounding as a genuine, robust phenomenon was found. Moral dumbfounding can be elicited using a broad range of stimuli and does not vary as a function of individual differences. The phenomenon does not, however, offer support for models that claim that moral judgments are intuitive, and that reason plays no causal role in the formation of moral judgment. A number of alternative causal processes that are consistent with moral dumbfounding, but that accommodate a causal role for reasoning in the construction of moral judgments, are discussed here; thus, moral dumbfounding is broadly consistent with several theories of moral judgment, including rationalist models.
Chapter 1: Moral Dumbfounding: A Brief History

Consider the following story: A brother and sister have mutually-consensual sexual relations with one another. More than likely you would judge this action to be wrong. You may justify this judgment by appealing to possible harm that may result; for instance, the probability that a child resulting from this union will have genetic abnormalities, the siblings’ relationship could be harmed by their actions, or another person could be harmed if they found out. If, however, it is stipulated that no harm will result from the siblings’ actions given (a) multiple forms of birth control were used making pregnancy very unlikely, (b) the action enhanced the siblings’ relationship and was a once only occurrence, and (c) no one else will ever find out about the siblings’ actions, you may find yourself in a position where your justifications are undermined, but your judgment that the actions described in the scenario are wrong nevertheless persists. If pressed, you may assert something like, “I don’t know why it is wrong, but it is just wrong!” This dissociation between judgment and justification has been termed moral dumbfounding, a phenomenon first demonstrated empirically by Haidt, Björklund, and Murphy (2000). Understanding moral dumbfounding is the key aim of this dissertation.

Before engaging with moral dumbfounding, some discussion of the broader context in which this phenomenon was initially identified is required. In particular, it is important to understand Haidt’s (2001, 2012) theoretical goals because it illuminates the possible theoretical significance of the dumbfounding phenomenon. Moral dumbfounding is crucial for Haidt as he appeals to it as evidence for a grand theory of morality. In the next section, Haidt’s theory is described. Ultimately, understanding of the dumbfounding phenomenon ought to play a role in our evaluation of Haidt’s theory, and other theories about the nature of morality and moral judgment. These are the theoretical stakes associated with moral dumbfounding.

Haidt’s Social Intuitionist Model

Honour killing (where an individual is killed because they have brought dishonour to their family) is an acceptable practice in some cultural groups while in others this is considered immoral. To explain this kind of striking cultural variation in moral judgment and behaviour, Jonathan Haidt developed the Social Intuitionist Model (SIM; 2001). He later extended this model to explain differences in moral judgment as a function of political ideology (2012). The SIM is the idea that “moral judgment is
generally the result of quick, automatic evaluations (intuitions)” (2001, p. 814). Haidt, however, does not offer a consistent characterisation of moral intuitions; for instance, he describes moral intuitions as a kind of cognition that is not reasoning (2001, p. 814) but also claims that moral intuitions are fueled by moral emotions (2005, p. 73), and elsewhere they are described as “affective flashes” (Haidt & Joseph, 2004, p. 60). It is worth pausing here to explain how these various sources suggest quite different models.

The characterisation in Haidt (2001) seems to suggest that moral intuitions involve a kind of ‘System 1’ automatisation of moral judgments. This is not an especially radical view; for instance, it seems that reasoning, in general, must employ ‘quick and dirty’ reflexes to deal with familiar situations (Barth & Chartrand, 1999). Haidt (2005) is more radical because here he proposes that moral intuitions are not ultimately caused by cognitive processing but, rather, our moral intuitions are non-cognitive emotional responses. Haidt and Joseph (2004) seem to go further, stating that moral intuitions just are emotional responses. Although Haidt aligns himself closely with sentimentalism, affect is not explicitly a feature of SIM; rather, SIM incorporates only intuitions (which may or may not be affective in nature, see Clarke, 2008 for further discussion). Note also that Haidt (2001) does not distinguish between product and process so, for instance, it is not clear whether moral intuitions cause moral judgments or whether moral judgments simply are moral intuitions.

Moral reasoning, according to SIM, is a post-hoc process that occurs after a moral judgment is already made; that is, although reasoning and judgment may co-occur, correlation does not imply causation. Haidt (2001) argues that although we think of ourselves as scientists who generate hypotheses, seek relevant evidence, and evaluate this evidence prior to making moral judgments, we actually operate more like lawyers defending our quick, intuitive judgments. He uses the metaphor of the tail wagging the dog, where the dog is reason and the tail intuitive judgment. We think that the dog wags the tail (i.e., reason drives the process) but, in fact, intuition is primary, and reason plays only a post-hoc role.

According to SIM, our moral intuitions are shaped by social influences; private reasoning is de-emphasised in favour of interpersonal reasoning. We use reason after our judgments are made in order to “influence the intuitions (and hence judgments) of other people” (Haidt, 2001, p. 814). The purpose of our post-hoc reasoning, therefore, is not always to justify our moral judgments per se but to convince others of the veracity of our moral judgments. The mere fact, however, that “friends, allies, and acquaintances
have made a moral judgment exerts a direct influence on others, even if no reasoned persuasion is used" (2001, p. 819). Haidt, therefore, acknowledges that reasoning may play some role in moral judgments but, normally, reasoning is merely post-hoc (i.e., confabulated to justify moral judgments that have already been made).

Haidt (2001) argues that the notion of moral judgment being driven by intuitive processes coheres well with more general observations that human cognition is predominantly automatic (Bar gh & Chartrand, 1999) and that we are typically unable to accurately report our internal mental processes (Nisbett & Wilson, 1977). This, by itself, is weak evidence for the intuitive nature of moral judgment. Haidt does, however, seek independence evidence as outlined in the next section.

**Evidence for the Social Intuitionist Model**

To test a key prediction of SIM, that moral judgments are not sensitive to reason, Haidt and colleagues (e.g., Haidt et al., 2000; Haidt & Hersh, 2001) undertook a series of experiments. These studies had different stated aims, but each was designed to show the importance of intuition in moral judgment. Haidt developed a series of ‘moral intuition’ vignettes to evoke quick intuitive judgments that the actions described in the scenarios were wrong coupled with an inability to articulate reasons in support of those judgments. In 2000, Haidt et al. (unpublished manuscript) used these moral intuition vignettes to demonstrate that judgment is primarily intuitive (and not responsive to reason) by asking participants to judge whether the actions depicted in the stories were wrong and to justify their judgments.

The moral intuition stories were as follows: (a) *Incest*, a story describing a brother and sister engaging in sexual relations, and (b) *Cannibal*, a story in which a woman cooks and consumes human flesh from a disease-free cadaver. In addition, Haidt et al. (2000) used ‘non-moral intuition’ stimuli which, like the moral intuition stimuli, were designed to elicit a quick judgment that the action is wrong (which would presumably co-occur with a refusal to perform the action). These tasks were: (c) *Soul*, in which the participant is asked to sign over their soul to the experimenter, and (d) *Roach*, in which the participant is asked to drink a glass of juice into which a sterilised cockroach has been immersed and removed. The *Heinz* story (Kohlberg, 1969), a story in which a man steals a life-saving drug for his dying wife, was also used as a reasoning measure. In the *Heinz* story, one needs to weigh up competing interests, those of the pharmacist and the dying woman and, therefore, such a case is not readily amenable to intuitive judgments (Haidt et al., 2000). Indeed, Monin, Pizzaro, and Beer (2007)
propose that a complete account of moral judgment will invoke both reasoning and intuition partly because different situations (and stimuli) demand different mechanisms.

Haidt et al. (2000) found that approximately 80% of participants characterised the actions described in Cannibal and Incest as wrong. Roach and Soul tasks were performed by 37% and 23% of participants, respectively. Only 20% of participants characterised Heinz’s stealing of the drug as wrong. After the participant made an initial judgment and provided a justification for their judgment, the experimenter played ‘devil’s advocate’ reminding the participant that there is no harm in these cases given particular stipulations (Haidt et al., 2000, p. 6. Note, however, that Haidt et al. do not quantify the rate of harm-based justifications nor do they indicate whether participants typically offered harm-based justifications). In the Incest case, for instance, the stipulations were as follows: (a) multiple birth control measures were used making pregnancy extremely unlikely, (b) the siblings’ relationship was strengthened by the act, (c) the act would not be repeated, and (d) no one else would ever find out about the siblings’ actions. Following this process, the experimenter asked participants for their final judgment. When a participant’s harm-based reasons were rejected by the experimenter, they were left with three choices: (a) change their judgment, (b) change their justification, (c) maintain their judgment without generating further justifications. Participants who maintained their judgments in the absence of supporting reasons were classified as ‘morally dumbfounded’.

Although Haidt does not make an explicit argument from moral dumbfounding to SIM, a possible formulation of an argument is as follows. If we used reasoning to make moral judgments, then we would have access to the reasoning process and would never be left without reasons for our moral judgments. The fact that we are sometimes left without reasons for our moral judgments is evidence that moral judgments are not, typically, based on reason. Rather, moral judgments are based on intuition. Haidt (2001) has used moral dumbfounding as evidence for SIM and against the historically dominant paradigm (Kohlberg, 1969; Piaget, 1932; Turiel, 1983) that moral judgments are the product of reasoning. Moral dumbfounding is clearly a very interesting phenomenon but Haidt’s conclusion that it offers evidence that moral judgment is intuitive and typically insensitive to reason is precipitous; that is, does dumbfounding really show that reasoning plays no role in the formation of moral judgments? Instead, dumbfounding could, in some cases, just be evidence that moral judgments have
become automatised (that is, oft made judgements are a product of System 1 processes). This, and other alternative explanations, will be discussed in Chapter 7.

Moral dumbfounding has received substantial critical attention in both the psychological and philosophical literature with respect to dumbfounding itself as well as in terms of what this phenomenon commits us to with respect to the nature of moral judgment (for example, Clarke, 2008; Dwyer, 2009; Kennett & Fine, 2009; Sneddon, 2007). This is because the idea that moral judgments are the product of fast, intuitive (and perhaps affectively driven) processes poses a challenge to our conception of ourselves as rational agents “capable of grasping and responding to reason” (Kennett & Fine, 2009, p. 77). Although Kennett and Fine argue that moral dumbfounding does not ultimately present a challenge to our agency, the phenomenon is nevertheless taken very seriously.

Here we evaluate the challenge that moral dumbfounding poses to dominant paradigms of moral judgment. Before one can even begin to think about the higher order considerations associated with moral dumbfounding (such as what it commits us to with respect to the nature of moral judgment) it is important to first establish empirically the nature of the phenomenon itself. To this end, a series of empirical investigations are reported in Chapters 3, 4, 5, and 6. A total of 88 individuals provided the data for the investigations; participants were randomly assigned to one of two conditions: The first sample (n = 44) participated in a replication of the original dumbfounding study (Haidt et al., 2000) and the second sample (n = 44) participated in a modification of the original study using alternative stimuli. Note that, although the data for all studies were collected simultaneously, the studies are described separately for ease of exposition. Before describing the studies undertaken here, a detailed analysis of the original dumbfounding study as well as further studies of moral dumbfounding is presented in Chapter 2. Methodological issues (many of which are addressed in the empirical investigations undertaken in the present dissertation) are highlighted in this chapter.
Chapter 2: Empirical Investigations of Moral Dumbfounding

This chapter outlines research in the extant literature, before describing the research undertaken for this dissertation.

Extant Empirical Research

Empirical evidence of the prevalence and predictors of moral dumbfounding is very limited and the evidence to date is not compelling. Haidt et al. (2000; unpublished manuscript) produced the most famous paper on moral dumbfounding. In their paper, Haidt et al. describe a study into whether moral judgments are intuitive and, typically, unresponsive to reason. They claim to have demonstrated that moral dumbfounding is ubiquitous, particularly in response to scenarios in which harm-based justifications are not acceptable (or at least, not accepted).

Haidt et al. (2000) asked a small group of participants (17 female and 13 male undergraduates from the University of Virginia) to make judgments as to whether the actions described in three vignettes were wrong, and to provide a justification for their each judgment. The vignettes were as follows: (a) Heinz (Kohlberg, 1969), a story in which a man steals a life-saving drug for his dying wife, (b) Incest, a story describing sexual relations between a brother and sister, and (c) Cannibal, a story in which a woman cooks and consumes human flesh from a disease-free cadaver donated for medical research. There were also two behavioural tasks in which participants were asked to perform an action: (d) Roach, in which participants were asked to drink juice that has had a dead, sterilised cockroach submerged and then removed, and (e) Soul, in which participants were asked to sign a non-legally binding ‘contract’ entitling the experimenter to own their soul. The researchers expected that the intuition measures (Incest and Cannibal), which were designed to be harmless yet offensive (in essence, taboo violations), would produce a quick, intuitive judgment that the actions described are morally wrong. In contrast, the Heinz vignette is considered to be a reasoning task as it requires participants to weigh up competing interests and, therefore, relies on a process of deliberative reasoning.

Haidt et al. (2000) found that 87% of participants characterised the action described in Cannibal as wrong, and 80% did so for the Incest vignette. For the Roach task, 37% performed the task, while the Soul task was performed by only 23% of participants. For the Heinz vignette, only 20% of participants characterised Heinz’s
stealing of the drug as wrong. After participants made an initial judgment the experimenter played ‘devil’s advocate’ whereby the participant’s judgment, and the justification they offered for their judgment, were challenged (p. 6). With respect to the Incest scenario, for example, if a participant claimed that the action was wrong because harm could result from the siblings’ actions (such as genetic abnormalities in any resulting offspring, a threat to the siblings’ relationship, or to other people associated with the siblings), the experimenter reminded the participant that the siblings’ actions were actually harmless in light of the following stipulations: (a) multiple birth control measures were used making pregnancy extremely unlikely, (b) the siblings relationship was strengthened by the act, (c) the act would not be repeated, and (d) no one else would ever find out about the siblings’ actions. It is noteworthy that Haidt et al. (2000) do not quantify the rate of harm-based justifications nor did they indicate whether participants typically offered harm-based justifications. Following this process, the experimenter asked participants for their final judgment. When a participant’s harm-based reasons were rejected by the experimenter, they were left with three choices: (a) change their judgment, (b) change their justification, or (c) maintain their judgment without generating further justifications.

Those participants who, when challenged by the experimenter, maintained their initial judgment without offering further reasons were classified as ‘morally dumbfounded’. The researchers (Haidt et al., 2000) reported that some participants in this category stated that “they had no reasons to support their judgments” (p. 1), subsequently declaring “It’s just wrong to do that!” (p. 9). Often associated with such declarations were non-verbal behaviours such as face touching and laughter, which was taken as evidence of nervous-embarrassed puzzlement. Participants were, on average, significantly more likely to produce verbal statements consistent with moral dumbfounding when judging actions associated with the intuition scenarios, compared to the Heinz vignette, as wrong. The authors do not, however, provide an indication of the prevalence of dumbfounding. After discussion with the experimenter, typically following a rejection of the participants’ harm-based reasons, there was a decrease in the percentage of participants who judged the actions depicted as wrong for all scenarios, and the percentage of change did not differ significantly between the reasoning and intuition scenarios (see Haidt et al., 2000, Table 1, p. 13). Moral dumbfounding was operationalised in terms of verbal statements (such as unsupported declarations and dumbfounding statements) and non-verbal behaviours (including
laughing and fidgeting). More details regarding the specific variables that were used by Haidt et al. to measure dumbfounding are provided in Chapter 3 (method section, see p. 23). The authors defined moral dumbfounding as the “stubborn and puzzled maintenance of a judgment without supporting reasons” (p. 1).

Haidt et al. (2000) explain that participants experienced a flash of emotion when presented with the harmless but offensive intuition measures and intuitively judged the action described in the scenario as wrong. As the source of moral judgment in these cases is outside of conscious awareness (i.e., unconscious affective intuition), participants were not aware of how they reached their decision but, nevertheless, attempted to justify their judgment by offering reasons that were incompatible with the stipulations in each case. Having had their harm-based reasons rejected, some participants noted (with embarrassment) that they had no explanation for their judgment (with some admitting as much), and simply reaffirmed their initial, emotion-laden, intuitive judgment. Haidt et al. concluded that moral dumbfounding is typical of moral judgment (with little attention paid to participants who were not dumbfounded, and without quantifying the prevalence of dumbfounding). In 2001, Haidt (together with Hersh) followed up his inquiry into moral dumbfounding.

Again using an interview format, Haidt and Hersh (2001) presented a small group of participants (17 female and 19 male undergraduates from the University of Virginia) with six taboo sexual violations (two scenarios for each of three classes of sexual norms: homosexuality, unusual masturbation, and consensual sibling incest). The participants self-identified as holding conservative or liberal political preference (politically conservative individuals tend to hold attitudes of maintaining status-quo and tradition whereas politically liberal individuals typically hold progressive attitudes toward social change and equality; Jost, Nosek, & Goslin, 2008). Both groups of participants were asked for their judgment (notably couched in affective terms: “What are your feelings about this act/situation”, p. 202). Many participants condemned the taboo sexual violations and their emotional reactions better predicted their moral judgments than did their justifications. Compared to their liberal counterparts, conservative participants had significantly stronger affective condemnation and higher prevalence of dumbfounding in response to homosexuality. Moral dumbfounding was observed at a rate of 49% in politically conservative participants and 33% for liberal participants. Indicators of the phenomenon included an inability to justify a judgment accompanied by confusion. Dumbfounded participants often retorted “I don’t know”,

9
laughing, and taking long pauses while stumbling through their responses, which were punctuated with linguistic fillers such as “um” (pp. 209, 215).

In discussing their results, Haidt and Hersh (2001) argue that moral dumbfounding occurs when people have strong emotionally based intuitions (as is often the case in taboo violations) that conflict with conscious deliberation of the scenario at hand (for instance, considerations of harmlessness). Given that emotions are highly resistant to change (even in the context of attempted persuasion), participants stubbornly adhere to their emotionally based judgment. Haidt and Hersh conclude that, upon having the supporting reasons challenged, people can literally be left without reasons for their emotionally based moral judgments (like Haidt et al., 2000, Haidt and Hersh assume that dumbfounding is typical of moral judgment without acknowledging that the phenomenon may, in fact, be limited to so-called ‘intuition’ stories).

Royzman, Kim, and Leeman (2015) undertook a replication and extension of the research conducted by Haidt et al. (2000). Royzman et al. (2015, Study 2) employed the interview protocol of Haidt et al. and presented participants (19 female and nine male American university undergraduates) with the *Incest* vignette. Participants were asked to judge whether the action performed is wrong, confirm their decision, and then justify their view. Those who indicated the described action was wrong were given a response booklet containing a list of harm-based reasons (both actual and potential harm). If a participant did not wish to endorse any of the options provided, and could not offer any rationale of their own, they were able to nominate a declaration of moral dumbfounding: “I don’t have a strong reason at this point, but I just feel it’s wrong for them to do what they did” (Royzman et al., p. 302). The experimenter then reviewed the responses and challenged the participants in the same manner as Haidt et al.’s ‘devil’s advocate’ (p. 301). After providing their post-counterargument judgment, participants were required to nominate a reason from the booklet once again.

Royzman et al. (2015) found that, upon being challenged by the experimenter, the majority of participants who had indicated that the action was wrong withdrew their previous justifications and declared themselves to be without a reason for their judgment. Participants were asked whether they accepted the experimenter’s stipulations regarding a lack of harm associated with the described actions. Only one participant indicated that they found the specified action to be credibly harm-free and, therefore, only this participant was classified as morally dumbfounded. On this basis, Royzman et al. (p. 305) argue that Haidt et al. (2000) failed to “discriminate between
the cases in which the criterial features of the moral dumbfounding response (judgment without supporting reasons) are genuinely met from those in which they only appear to be met (supporting reasons are abundant but remain unexpressed)". They conclude that if moral dumbfounding exists at all (in which a person resolutely condemns an act without supporting reasons), it is extraordinarily rare.

In contrast to Royzman et al. (2015), a dumbfounding rate comparable to that of Haidt et al. (2000) was observed by McHugh, McGann, Igou, and Kinsella (2017) in their replication and extension of the research conducted by Haidt et al. McHugh et al. (2017; Study 1) presented a small group of participants (15 female and 16 male American undergraduates, postgraduates, and alumni) with the reasoning and intuition vignettes from the original dumbfounding study (Haidt et al., 2000; described above). The researchers also included a fourth reasoning vignette which describes a man stopping a runaway trolley by pushing an overweight bystander off a bridge onto its path, thus, saving five lives (commonly called the Trolley problem; Thomson, 1986). The method utilised by McHugh et al. was the same as that used in the original study (i.e., a face-to-face interview in which the experimenter challenged participants’ justifications and stipulated an absence of harm associated with the action). In operationalising moral dumbfounding for their study, the researchers argue that stubbornness and puzzlement might not be essential features of moral dumbfounding and, hence, define moral dumbfounding as the maintenance of a judgment without supporting reasons. McHugh et al. (2017; Study 1) report that 70.97% of their participants produced a dumbfounding response and the overall pattern of responses (and associated dumbfounding behaviour, for example, laughter and a lack of utterances such as “ums” and “uhhs” taken to indicate participants were "working towards a reason", pp. 8-9) resembled that observed by Haidt et al. (2000). Based on their results, the researchers conclude that moral dumbfounding is a robust phenomenon. There is, therefore, only one study with comparable findings concerning the manifestation of moral dumbfounding (note, though, that it is not possible to determine whether the rate of dumbfounding is similar to that obtained by Haidt et al. as this is not reported).

It is noteworthy that the studies of Royzman et al. (2015) and McHugh et al. (2017) were partial replications of the study by Haidt et al. (2000); neither researchers included the behavioural measures (first-person cases) employed by Haidt et al. Restricting measures to third-person vignettes (in which participants are presented with hypothetical situations and asked to judge the depicted action of an unknown third-
party) is potentially problematic as there is a distinction between judgments made in third-person and first-person scenarios (see Kennett, 2012). People typically make different decisions when they imagine themselves in a moral situation compared to imagining others in the same position (Nadelhoffer & Feltz, 2008). Different judgments are also made when people respond to a moral dilemma verbally compared with action (Bostyn, Sevenhant, & Roets, 2018). The measures in the original dumbfounding experiment (Haidt et al., 2000) comprised three moral dilemma vignettes and two behavioural tasks. By omitting tasks that call for a decision on a first-hand imminent action, both Royzman et al. and McHugh et al. neglected to investigate (and replicate) judgments relevant to personal decision-making. Furthermore, they did not incorporate measures that were comparable to Haidt et al., nor did they generate an empirically supported definition of moral dumbfounding, or carefully characterise rates of different features of the phenomenon.

The findings of Haidt et al. (2000) and subsequent studies leave a number of key questions unanswered, namely: (a) is moral dumbfounding a robust phenomenon and what is its prevalence; (b) what are the key defining features of moral dumbfounding; (c) when people maintain their judgments, do they do so despite giving up on the harm-based argument, or do they perhaps not accept the stipulations of no harm and ultimately maintain their judgments and the associated harm-based reasons; (d) would people be able to provide alternative justifications for their judgments if given enough time; (e) Would people abandon their initial judgments if given enough time to process their inability to justify them; and (f) when people assert that something is wrong, do they necessarily mean that it is morally wrong? For this dissertation, a careful replication and extension of the original dumbfounding study (Haidt et al., 2000) was undertaken. Chapter summaries are presented in the following section.

**Chapter Summaries**

This dissertation is presented in eight chapters. The following summarises Chapters 3 to 7, with Chapter 8 providing the conclusion.

**Chapter 3: Study 1 – Replicating and extending the original moral dumbfounding study**

While there is some empirical support for moral dumbfounding (Haidt & Hersh, 2001; McHugh et al., 2017), other researchers (Royzman et al., 2015) have concluded that if this phenomenon exists at all it is extraordinarily rare. These are the only replications of
the original dumbfounding study (Haidt et al., 2000) and they are not faithful replications; therefore, further empirical work exploring this phenomenon is clearly required. In Chapter 3 (Study 1), a replication and extension of the original moral dumbfounding study of Haidt et al. (2000) is presented. In addition to a faithful replication of the original study, an extension was undertaken to address the methodological shortcomings of the original research.

The way in which moral dumbfounding is operationalised in Study 1 corresponds with the most common definition of the phenomenon in the extant literature; that is, the maintenance of a judgment without provision of supporting reasons. Other definitions of this phenomenon do, however, exist and their specific details vary. Some definitions are more ‘inclusive’ (such as the definition used in Study 1), while others are more ‘exclusive’ meaning that behaviour representative of moral dumbfounding under this class of definition is more limited. Unsurprisingly, therefore, the reported rates of observed dumbfounding (Haidt & Hersh, 2001; McHugh et al., 2017; Royzman et al., 2015) depend upon how it is operationalised, which is partly responsible for the discrepancies in empirical findings.

Chapter 4: Study 2 – Defining moral dumbfounding

Interestingly, the original definition proposed by Haidt et al. (2000, p. 1) of moral dumbfounding being the “stubborn and puzzled maintenance of a judgment” only appears in the abstract of the paper and not in its body. When asked about the phenomenon of moral dumbfounding, Haidt indicated that many of the participants who were classified as morally dumbfounded in the original dumbfounding study (Haidt et al., 2000) were “comfortably dumbfounded”; that is, they demonstrated no discomfort with, and were not puzzled by, their inability to provide a justification for their moral judgments (Haidt, 2005, p. 72). Only a subset of individuals who were classified as morally dumbfounded were puzzled by their failure to offer a justification for their moral judgment; therefore, it is unclear as to why puzzlement was included in the original definition of moral dumbfounding. Since the original study, different definitions of moral dumbfounding have been used by different researchers and, therefore, an empirical examination of the key defining features of this phenomenon is clearly warranted and is described in Chapter 4.

Study 2 (presented in Chapter 4) is the first systematic quantification of (a) the difference in prevalence of moral dumbfounding when applying various definitions of the phenomenon, and (b) the proportion of variance in the phenomenon accounted for
by the different elements. The aim of this study was to generate an empirically supported definition of moral dumbfounding. The rate of moral dumbfounding was compared as a function of the various definitions of this phenomenon in the extant literature. While differences in definition clearly have the potential to contribute to discrepancies in the rates of moral dumbfounding reported in various studies, other factors such as the measures used to elicit the phenomenon may also impact on rates of dumbfounding. This was explored in Study 3 (reported on in Chapter 5) by using alternative measures to those employed by Haidt et al. (2000).

**Chapter 5: Study 3 – Investigating the moral dumbfounding measures**

To faithfully replicate Haidt et al.’s (2000) study, the moral judgments of participants in Study 1 (and utilised in Study 2) were elicited using their original ‘dumbfounding’ stimuli; however, some researchers (including Jacobson, 2012; Kennett, 2012) dismiss moral dumbfounding as an artefact of the stimuli used. They argue that the participants of Haidt et al. (2000) were presented with situations that they are unlikely to encounter in real life and were most likely dumbfounded by what they were asked to consider (rather than by their judgments). All empirical research conducted after the original study (Haidt & Hersh, 2001; McHugh et al., 2017; Rozyman et al., 2015) has utilised only these original tasks. Study 3 (reported in Chapter 5) is the first study to use a range of novel, more realistic scenarios (i.e., stimuli that one may encounter in real life). With greater ecological validity than extant studies, the aim of Study 3 was to determine whether moral dumbfounding can be elicited by ‘real-life’ scenarios.

**Chapter 6: Study 4 – Exploring individual differences in moral dumbfounding**

Although situational factors alone (such as the nature of the elicitation stimuli) may contribute to moral dumbfounding, it is also possible that individual differences contribute to the phenomenon; that is, some people may be more ‘dumbfounding’ than others. Some researchers (Haidt et al., 2000; Haidt & Hersh, 2001; McHugh et al., 2017) assume that moral dumbfounding is typical of moral judgment; personal and situational characteristics that may impact on dumbfoundability has been largely neglected. In the studies by Haidt et al. (2000), Haidt and Hersh (2001), and McHugh et al. (2017), participants were pressured by the experimenter to respond quickly and without deliberation. One possible mechanism underpinning dumbfoundability may be susceptibility to such pressure; however, these factors have not been previously
empirically tested. Study 4 explores the impact of individual differences on moral dumbfounding and is reported in Chapter 6.

**Chapter 7: Does moral dumbfounding reveal the nature of moral judgment?**

Haidt (2001) takes observations of moral dumbfounding to be evidence for his SIM, specifically the key claim of this model that reasoning does not play a causal role in moral judgment. Haidt has used moral dumbfounding to counter the historically dominant psychological paradigm (Kohlberg, 1969; Piaget, 1932; Turicel, 1983) in which moral judgments were understood to be a product of a process of conscious reflection. Moral dumbfounding is clearly a very interesting phenomenon, but Haidt’s conclusion that it offers evidence for SIM is precipitous. Even if moral dumbfounding is a real phenomenon, it remains an open question as to what this commits us to with respect to the nature of moral judgment. In Chapter 7, a discussion of the different processes that could explain moral dumbfounding is presented. All of these processes accommodate a causal role for reason and are plausible alternative explanations to Haidt’s account.
Chapter 3: Study 1 – Replicating and Extending the Original Moral Dumbfounding Study

The extensive critical engagement that the phenomenon of moral dumbfounding has received is premised on the idea that the existing evidence is sufficient. In other words, it is often taken at face value that moral dumbfounding is a genuine phenomenon, is (perhaps) typical of moral judgment, and may, therefore, have important implications for our understanding of the nature of moral judgment. Before we accept this, however, there are a number of open questions that need to be addressed, including (a) what exactly is moral dumbfounding; (b) is there empirical evidence for this phenomenon, and, if so; (c) is the existing evidence valid; and (d) what further empirical investigation is required? The answers to these questions are essential for confidence in theories that use moral dumbfounding in support of their respective accounts of moral judgment either directly (e.g., Cushman, Young, & Greene, 2010; Dwyer, 2009; Haidt, 2001; Hauser, Young, & Cushman, 2008; Hindriks, 2014; Prinz, 2005; Saunders, 2009; Sneddon, 2007) or indirectly (e.g., Crockett, 2013; Cushman, 2013; Greene, 2007).

Replication is the cornerstone of the scientific method; it is the means by which we can be assured that research results are valid and reliable, theoretical interpretation is verified, and alternative explanations ruled out (Open Science Collaboration, 2015). With only one partial replication in support of Haidt et al.’s (2000) seminal research (McHugh et al., 2017), it cannot yet be said that moral dumbfounding is a pervasive feature of moral judgment (despite the conclusion drawn by proponents of the dumbfounding paradigm, e.g., Cushman et al., 2010; Dwyer, 2009; Haidt, 2001; Hauser et al., 2008; Hindriks, 2014; McHugh et al., 2017; Prinz, 2005; Saunders, 2009; Sneddon, 2007). Here, the first complete replication (and extension) of the original moral dumbfounding study of Haidt et al. (2000) is presented. While replication is essential to the scientific enterprise, duplicating flawed research does not really accomplish much; that is, simply replicating Haidt et al.’s unpublished study of a poorly measured and ill-defined phenomenon is uninformative. Consequently, shortcomings of the original study (and other limited dumbfounding research) are addressed in the present study by conducting a methodologically rigorous empirical investigation into the phenomenon of moral dumbfounding (defined in the current study as ‘the maintenance of a judgment without provision of supporting reasons’; please refer to the Measures section on page 23 for the rationale behind the utilisation of this particular definition). In particular, analysis of the manifestations of moral dumbfounding is
undertaken here, and both first-person and third-person scenarios are included, with a larger sample size than that previously incorporated in this kind of research to improve generalisability. Although predicting exact prevalence rates is problematic, the following hypotheses were tentatively proposed based on results reported in the original moral dumbfounding study:

1. Moral dumbfounding will be elicited in response to the stimuli presented.
2. The intuition stimuli will elicit moral dumbfounding with significantly more frequency than the reasoning stimulus.
3. The intuition stimuli will be judged as wrong with significantly more frequency than the reasoning stimulus.
4. Participants will respond significantly faster when presented with the intuition stimuli than the reasoning stimulus.
5. The intuition stimuli will elicit judgment prior to justification, whereas the reasoning stimulus will elicit justification prior to judgment.
6. Participants will report basing their judgments significantly more on gut feeling in response to the intuition stimuli than to the reasoning stimulus.
7. Participants will report significantly less confidence in their judgment to the intuition stimuli than to the reasoning stimulus.
8. Participants will report significantly more confusion, and more irritation, in response to the intuition stimuli than to the reasoning stimulus.
9. There will be significantly more psychic discomfort in response to the intuition stimuli than to the reasoning stimulus.
10. There will be significantly less speech dysfluency in response to the intuition stimuli than to the reasoning stimulus.
11. There will be significantly more dead-end arguments, statements of “I don’t know”, unsupported declarations, and dumbfounding statements in response to the intuition stimuli than to the reasoning stimulus.

Method

The data for all four studies undertaken and reported in this dissertation were collected simultaneously. Participants (\(N = 88\)) were randomly allocated to the current study (to receive the original measures of Haidt et al., 2000; \(n = 44\)) or Study 3 (to receive the alternative measures; \(n = 44\)). There were no age or gender differences detected between groups (age: \(t(86) = 0.988, p = 0.326\); gender: \(\chi^2(1, N = 88) = 0.051, p = 0.821\)).
Participants
The sample for the current study consisted of Australian university students ranging in age between 18 and 63 years ($M = 28.18$, $SD = 13.25$). The male participants ($n = 15$) were aged between 18 and 63 years ($M = 25.2$, $SD = 13.05$) and the female participants ($n = 29$) were aged between 18 and 56 years ($M = 29.72$, $SD = 13.32$). As an inferential analysis between demographic variables and moral dumbfounding was not conducted in the current study, full descriptive demographic statistics (for participants of Studies 1 and 3) and inferential analyses (for the full sample; $N = 88$) are presented in Chapter 6 (Study 4).

Sampling procedure
A convenience sampling method was utilised to recruit participants. Participants in the study of Haidt et al. (2000) were 30 psychology undergraduates from the University of Virginia (17 females and 13 males) ranging in age between 18 and 20 years (with one participant 48 years old). Individuals eligible to participate in the present study were those over the age of 18 years and enrolled in any tertiary education course at any Australian tertiary institution (this was done to allow for greater variation in the sample, rather than restricting it to psychology students from a single university). While it was hoped that equal gender numbers from a range of disciplines and backgrounds would participate, more females volunteered for the study than did males.

Advertisements were placed on noticeboards in communal areas of Charles Sturt University (CSU), Bathurst campus, seeking volunteers. The first-year psychology student pool (in which students select research for participation as part of their course requirements) was also utilised to recruit participants. Advertisements were also posted via online university forums and public social networking sites. Participants were randomly allocated to Studies 1 or 3 using Random.org List Randomiser (https://www.random.org/lists/).

Sample size and power
Sample size and power were calculated for all statistical tests used in the current study using G*power (version 3.1):

- Cochran’s q - At a significance level of 0.05, odds ratio = 4, and 80% power, the required sample size was 29. As Cochran’s q is an extension of McNemar’s test, the sample size calculation for this test was used.
- McNemar test of change - At a significance level of 0.05, odds ratio = 4, and 80% power, the required sample size was 29.
• Friedman two-way ANOVA: At a significance level of 0.05, small effect size (0.4), and 80% power, the required sample size was 22 (this was calculated by adding 15% to the resulting sample size as this is a non-parametric test; Lehmann, 2006).

• Wilcoxon signed ranks test: At a significance level of 0.05, medium effect size = 0.5, and 80% power, the required sample size was 28.

• Kendall’s tau b: At a significance level of 0.05, correlation of .5, and 80% power, the required sample size was 35.

Thus, there was sufficient sample size and power for statistical analyses in the current study given \( n = 44 \).

**Ethical considerations**

The CSU Human Research Ethics Committee (protocol number 2016/45; annexed as Appendix A.1) provided ethical approval. An Information Statement and Consent Form (annexed as Appendix A.2 and Appendix A.3, respectively) was presented to participants and outlined the possibly confronting nature of the stimuli. They were also advised of their rights and expectations with respect to participation. Data were de-identified by assigning a numerical code to each participant with video recording and questionnaire data saved under this identifier.

**Research design**

Replicating the research design of the original moral dumbfounding study (Haidt et al., 2000), data were collected in a standardised open-ended interview. Qualitative data were obtained using open-ended interview questions and survey items, and behavioural observation; quantitative data were obtained using closed-ended survey items. Standardised interviewing was important to keep the interview protocol constant across participants, but open-ended interviewing allowed collection of in-depth information, with the researcher probing for clarity when needed. The open-ended survey items allowed participants to respond in their own words and presented an opportunity for evidence of moral dumbfounding to be revealed that may otherwise remain hidden. The behavioural observation was integral to understanding non-verbal indicators of moral dumbfounding. Comparison with the behavioural observations of the original moral dumbfounding study (Haidt et al., 2000) was undertaken.

**Apparatus and materials**

The following apparatus was used for all four studies undertaken here.
**Apparatus**

Sole-operator recording equipment (a Dell computer, teleprompter, JVC camera, Rode stereo microphone, a microphone boom, and lighting) and NewsBoss Software (version 6.2) were used to record the three vignettes (described in the next section). Each recording was embedded in a PowerPoint presentation to present the stimuli to participants. Each moral vignette was embedded on an individual slide within the presentation. A separate presentation was created for each order variant.

An Asus F555LA laptop computer was used to present the stimuli to participants (a Phillips 21.5” High Definition computer monitor was used, together with two Logitech Z200 multimedia speakers). The distance that participants sat from the computer screen was approximately 50 centimetres (a photograph of the view from the participant’s seat is annexed as Appendix B.1). The computer’s specifications could handle the processing loads of simultaneously presenting the scenarios and recording audio and video of participants’ responses without skipping or lags in the recordings.

A Logitech C920 High Definition webcam, Blue Yeti USB microphone (produces studio quality recordings), and Lenovo Tab 3.7” tablet computer were used to capture participants’ responses. The microphone was set to a bidirectional interview mode which enabled recording of the verbal output of experimenter and participant. Open Broadcaster Software Studio (version 20.0, https://obsproject.com) was used to record participant audio and video. This software enabled customisation of the recording such as picture-in-picture filming so that participant responses, and the presented scenario, could be recorded side by side to capture immediate reactions. An Asus 16” computer monitor faced the experimenter and was used to verify that the correct stimulus was presented to the participant, and the stimulus was playing without error (a photograph of the view from the experimenter’s seat is annexed as Appendix B.2). The tablet computer was used to capture the participants’ responses to each questionnaire item. The responses entered into the tablet were directly transferred to a cloud database which eliminated potential transcription errors that otherwise may have occurred when using paper-based platforms. Google Forms was used to create the interface to capture participant responses to each questionnaire item, and participant responses were transmitted directly to a file stored on Google Drive (with a two-stage authentication to access the stored data).
**Materials**

Tasks included in the current study were those employed by Haidt et al. (2000, pp. 18-19). The procedure Haidt et al. used to present the vignettes to participants is unclear; all stories “were read” (p. 7) but whether this was by the experimenter, or participants themselves, is not stated. The way in which vignettes are presented can influence participant responses and, therefore, the outcome of the research (see, Rosenthal, 1969/2009). If, for instance, the experimenter read the stories to participants, confounding may occur through different vocal intonations, subtle gestures, and slight changes in posture influencing participant perception of the presented scenario. On the other hand, if participants read the vignettes themselves, their judgments and justifications may be affected by inattention to pertinent information. To ensure that the three vignettes were presented identically to each participant, they were video recorded by the researcher who was seated in front of a black screen and read each of the stories verbatim from a teleprompter. Care was taken to ensure minimal facial expression and verbal emphasis during each of the recordings. Each recording began with the words, “Please let me know if the following story is familiar to you” as familiarity with the vignettes may influence the results.

**Reasoning vignette**

The Heinz vignette was developed initially by Kohlberg (1969). The Heinz vignette depicts a man stealing (for his terminally-ill wife) a life-saving but cost-prohibitive drug from a pharmacist who refuses to provide it at a reduced cost or for free (annexed as Appendix C.1). This moral dilemma is thought to elicit impassive consideration (i.e., reasoning) between potentially competing concerns of harm and rights.

**Intuition vignettes**

In contrast to the reasoning stimulus, all intuition stimuli (which represent taboo violations) were intended to elicit quick emotional condemnation but to be resistant to harm-based justifications. The first intuition vignette, referred to as Incest, describes consensual adult sex between a brother and sister (annexed as Appendix C.2). The second intuition vignette, Cannibal (annexed as Appendix C.3), portrays a medical school laboratory assistant who cuts off (and cooks) a piece of flesh from a cadaver that had been donated (and used) to teach medical students about anatomy (but that was now considered waste).
Behavioural tasks

Roach (developed initially by Rozin, Millman, & Nemeroff, 1986) involves asking participants to drink a cup of juice into which a sterilised cockroach had been immersed and then removed (annexed as Appendix C.4). Cockroaches were purchased from a pet shop to ensure they had been raised in a clean environment. The cockroaches were euthanised by placing them in a freezer and were then sterilised in an autoclave. Once sterilised, the cockroaches were individually placed in separate clean containers and stored for use. During the experiment (and in front of the participant), the researcher used tweezers to put the cockroach into the cup of apple juice or water (to prevent bacterial transfer from the experimenter’s hands to the cockroach or plastic cups).

Soul (annexed as Appendix C.5) involves asking participants to sign a non-legally binding ‘contract’ entitling the experimenter to own their soul after their death. Participants were informed that they could tear up the ‘contract’ immediately after signing. The following measures were derived for each vignette and behavioural task.

Measures

The following measures comprise those replicated from Haidt et al. (2000). 

Moral dumbfounding

Moral dumbfounding, as initially defined, consists of four components: (a) stubbornness, (b) puzzlement, (c) maintenance of a judgment, and (d) without supporting reasons (Haidt et al., 2000, p. 1). The most common definition of moral dumbfounding in the literature encompasses only components (c) and (d); that is, stubbornness and puzzlement are not frequently mentioned elements of dumbfounding (s.c., Astuti & Bloch, 2015; Dwyer, 2009; Gray, Schein, & Ward, 2014; Kasachkoff & Saltzstein, 2008; Kennett, 2012; Kennett & Fine, 2009; McHugh et al., 2017; and Railton, 2014). McHugh et al. (2017) correctly point out that this is likewise true in the original paper of Haidt et al. (2000) in which components (a) and (b) are not explicitly operationalised (only appearing in the abstract). Accordingly, stubbornness and puzzlement are omitted from the operationalisation of moral dumbfounding here and the phenomenon is defined as ‘the maintenance of a judgment without provision of supporting reasons’.

Inferential analyses of moral dumbfounding were undertaken in the current study rather than providing only a description of indicators of the phenomenon (as presented by Haidt et al., 2000). A moral dumbfounding variable was computed for each stimulus. The criteria for being categorised as morally dumbfounded was:

23
1. The participant needed to have made a judgment of ‘wrong’ for both the initial and the final judgment to satisfy the requirement of maintenance of a moral judgment; and

2. To satisfy the requirement of a judgment made without supporting reasons, the participant needed to have offered:
   i. an ‘unsupported declaration’, a ‘tautological reason’ or an ‘illogical argument’ (i.e., failed to provide a reason); or
   ii. a ‘dumbfounding statement’ or a statement of ‘unwillingness to support’ (i.e., struggled to produce a reason).

Explanation of the individual variables comprising the moral dumbfounding variable is provided below. A global measure of moral dumbfounding was computed for each participant by aggregating their frequency of moral dumbfounding across stimuli. Higher numbers indicate greater frequency of moral dumbfounding during the interview.

Judgments of wrongness of an action
In the current study, participants were asked to indicate whether they judged the action described in the vignettes to be ‘wrong’, ‘undecided’, or ‘OK’. The option of ‘undecided’ reflected a small digression from the methodology of Haidt et al. (2000) and was included here as it enabled participants to express ambivalence rather than forcing them to nominate a response that did not reflect their attitude. Also deviating slightly from the methodology of Haidt et al. (who only asked for initial and final judgments and justifications), participants in the current study were asked for an intermediate judgment and explanation (after the experimenter’s initial counterarguments). Although not relevant to the current study, the intermediate response was included as it enabled a measure of the stubbornness of judgment (consistent initial, intermediate, and final) as opposed to maintenance of judgment (consistent initial and final) to satisfy the definitional specificity examined in Study 2 (presented in Chapter 4).

Change of judgment
Whether or not there was a change of judgment was determined by recording variation in participants’ initial and final judgments. While Haidt et al. (2000) also asked participants to self-report whether their judgment had changed, this subjective measure was not considered a necessary addition to the objective measure of actual judgment
change. Intermediate judgment was not included in the calculation of the change of judgment variable for this study in keeping with the analyses of Haidt et al.

**Response timing**

The time-based variables reported in the original dumbfounding study (Haidt et al., 2000) were also measured here. The total response time for each task was measured from completion of the experimenter’s request for a first task initial judgment to the end of participant final justification for their last task judgment. According to Haidt et al. (2000), when presented with the intuition cases (i.e., the measures that elicit moral dumbfounding), participants typically provide a quick judgment followed by justification. The order of responses for participants (judgment or justification given first) was measured using the time difference between the commencement of participants’ initial judgment and that of their initial justification. Timings to the start of initial judgment and initial justification were used to test the assumption that moral dumbfounding occurs with quick judgment and slow justification (Haidt et al.).

**Basis underpinning judgments**

Participants reported whether they based each judgment on reasoning or gut feeling using two separate 5-point rating scales ranging from 1 (*strongly disagree*) to 5 (*strongly agree*). This procedure was comparable to that used by Haidt et al. (2000); however, in contrast to Haidt et al., participants in the current study were required to answer these two items concerning their initial judgment, and then again for their final judgment. Not only did this avoid the ambiguity for participants as to which judgment (i.e., initial or final) the question refers, but it enabled a measure of change in the self-reported basis of initial and final judgments.

The foundation underpinning judgments (specifically, a decision made on gut feeling) is central to moral dumbfounding according to Haidt (Haidt, 2001; Haidt et al., 2000; Haidt & Hersh, 2001). A foundation of judgment variable was computed using the difference in participant ratings of the degree to which they based their decision on reasoning versus gut feeling (with lower ratings indicating more gut feeling and less reasoning as the foundation underpinning the judgment). The basis of initial judgment, and the basis of final judgment, was computed for each stimulus. When investigating the basis underpinning judgment and moral dumbfounding, only the measure of the basis underpinning final judgment was used, as moral dumbfounding can only be demonstrated after a concluding decision. A global measure of the basis of final judgment was computed for each participant by averaging their basis of final judgment
across stimuli. Lower ratings were taken to indicate more gut feeling and less reasoning as the basis underpinning their decision.

Confidence in judgment
The degree to which participants felt confident in their judgments about whether the described action was right or wrong was assessed using a 5-point rating scale ranging from 1 (strongly disagree) to 5 (strongly agree). Again, in contrast to the original moral dumbfounding study (Haidt et al., 2000), this item was asked twice: once regarding participants’ initial judgment, and again for their final judgment. Once more, this avoided any ambiguity as to which judgment (i.e., initial or final) the question refers. Having had their judgment and justification questioned by the ‘devil’s advocate’, it was thought that a measure of change in confidence of judgment would yield interesting results. Variables of confidence in initial judgment, and confidence in final judgment, were computed for each stimulus. In addition, a global measure of confidence in judgment was computed for each participant by averaging their confidence in initial and final judgments across stimuli. Higher ratings were taken to indicate more confidence in the judgment.

Confusion and irritation
The final two items on the End-of-Task Questionnaire replicated that of Haidt et al. (2000) and were only asked once. Participants were asked to indicate on two separate 5-point rating scales the degree to which they were confused and irritated by the presented vignette or behavioural task. Both scales ranged from 1 (strongly disagree) to 5 (strongly agree). Measures of confusion and irritation were computed for each stimulus, with higher ratings indicating more confusion and irritation, respectively.

Psychic discomfort
The behaviour of participants was video recorded and various responses thought to be associated with moral dumbfounding were coded. Behavioural variables taken to indicate physical manifestations of psychic discomfort were (a) laughter, (b) face and body touching, (c) fiddling (i.e., manipulating hands or an object held in hand), and (d) posture shifting (i.e., moving around in the seat, leaning forward, leaning backwards). The range of behavioural phenomena coded in the present study differed somewhat from those included by Haidt et al. (2000). ‘Doubt faces’, for example, was a category used by Haidt et al. that was not used here because of the difficulty associated with classifying a facial expression in this way (i.e., the difficulty associated with attributing
a mental state to a facial expression with sufficient certainty). Also, ‘posture shifting’ was not a category used by Haidt et al. (2000).

The total response time differed between tasks ranging from an average of approximately 37 seconds for the Roach task (many participants drank the Roach juice immediately upon being asked without any discussion) through to an average of approximately four minutes for the Cannibal vignette. The variation in total task time meant that there was more or less time for participants to demonstrate behaviour (i.e., laughter, face touching, fiddling, and posture shifting). Accordingly, the behavioural variables were standardised by dividing the total number of each behavioural variable, per person per task, by the total task time (in seconds).

The four behavioural variables were combined as (a) the different manifestations of a psychological state of unease was not the focus of interest (so it was considered unnecessary to analyse individual behavioural variables), and (b) to prevent unnecessary inflation of error rates. Consequently, this measure deviates from Haidt et al. (2000) who analysed (some) individual behavioural variables. The combined four behavioural variables were collectively taken to indicate the physical manifestation of psychic discomfort.

**Speech dysfluency**

Paralinguistic responses coded in the present study included breaks, irregularity, or non-lexical expression that occurred within the flow of otherwise fluent speech. Coded speech dysfluencies were (a) ‘false starts’ (i.e., words and sentences that were cut off mid-utterance; phrases that were restarted or repeated, and repeated syllables), (b) ‘fillers’ (i.e., grunts or non-lexical utterances such as “uh”, “erm”, “um”, “well”), (c) ‘repaired utterances’ (i.e., instances of participants correcting their own slips of the tongue or mispronunciations), and (d) ‘long pauses’ (a lengthy pause before answering or a long break between words or sentences during a response). These responses were the same as those coded by Haidt et al. (2000) and, like McHugh et al. (2017), speech dysfluencies were taken to indicate working toward articulating a reason.

Unlike the other paralinguistic variables (i.e., fillers, false starts, and repaired utterances), the number of long pauses was divided by the total task time (seconds) as the opportunity for pausing was a function of total time taken to complete the task. It was considered unnecessary to analyse individual paralinguistic variables as the nuances of speech dysfluency was not the focus of interest; thus, to prevent unnecessary inflation of error rates, the four paralinguistic variables were combined and collectively
taken to indicate the participant working toward articulating a reason. Once again, the
current study measure deviates slightly from Haidt et al. (2000) who analysed (some)
individual paralinguistic variables.

**Linguistic responses**

Linguistic responses included those deductively coded using the coding scheme of the
original dumbfounding research (i.e., dead-end arguments, statements of “I don’t
know”, unsupported declarations, and dumbfounding statements; Haidt et al., 2000) and
those inductively coded from the current study data (discussed in the next section).
Deductively coded responses included (a) ‘dead-end arguments’ (i.e., an incomplete
justification for a judgment; e.g., participant stops in the middle of giving their
justification without prompting from the interviewer), (b) statements of “I don’t know”,
(c) ‘unsupported declarations’ (e.g., “It’s just wrong to do that”; “That’s terrible”), and
(d) ‘dumbfounding statements’ (e.g., “I believe X is wrong, but I can’t find a reason”
and “I can’t find the words to explain it”). These linguistic responses were also
measured by Haidt et al. (2000); however, there was some digression here from the
variables measured by Haidt et al. as some of the variables included in this category
were difficult to operationalise such as ‘arguments kept’ and ‘arguments dropped’ (and
no operationalisation was offered in the original study).

**Novel measures**

The following measures were inductively derived from the data collected and reported
in this dissertation.

**Failure to provide a reason**

During the inductive development of the codification scheme, several categories that
were not coded in the original dumbfounding study (Haidt et al., 2000) were identified
and operationalised. The first two categories appeared distinct from, but overlapping
with, the existing ‘unsupported declaration’ category (e.g., “It’s just wrong to do that”;
“That’s terrible”). The first category was that of ‘tautological reason’. Tautological
reasons (also coded in the study of McHugh et al., 2017) are details that are restated
from the presented vignette or task that are offered as justification for a judgment.
Examples of tautological reasons include, “It’s cannibalism” and, “They’re related”
when responding to the Cannibal and Incest vignettes, respectively. The second
category was that of ‘illogical argument’. Illogical arguments are claims used to support
broader justifications (i.e., a statement for which we can ascertain the truth value). An
eexample of an illogical argument is, “They might get pregnant” when responding to the
Incest vignette (this is considered to be an illogical statement here given the stipulation that two forms of birth control were used).

Both tautological reasons and illogical arguments were considered distinct from the existing category because unsupported declarations are descriptive judgments (i.e., an assertion that the participant claims to be true). As such, unsupported declarations are not merely restatements of detail, nor can they be falsified (i.e., they cannot be appraised as true or false because subjective claims cannot be right or wrong). While distinct, the three categories can, however, be mainly considered equivalent because (taking the Incest vignette as an example) statements such as "That's terrible" (unsupported declaration), "They're related" (tautological reason), or “They might get pregnant” (illogical argument), do not satisfy the criteria of having a logical basis for a judgment (see, Mallon & Nichols, 2011; McHugh et al., 2017). The three categories were analysed both separately and together (in the latter case the variables were combined to form a composite variable ‘failure to provide a reason’).

Struggle to produce a reason
Once again, during the inductive development of the codification scheme, a novel category was identified as distinct from, but overlapping with, the existing ‘dubbing statement’ category (e.g., “I can’t find the words to explain it”). This category was labelled ‘unwillingness to support’ and encompassed statements supporting acceptability of the scenario or task, but an unwillingness to commit to such a judgment. An example of unwillingness to support is, “I understand that there is no harm, and it should be OK, but I just still think it is wrong”. This category was considered distinct from admissions of dubbing as statements of unwillingness to support include an acknowledgement that the judgment should be acceptable, whereas dubbing statements omit any such declaration. The two categories were analysed both separately and together (in the latter case the variables were combined to form a composite variable ‘struggle to produce a reason’).

Unique justifications
Instead of ‘arguments kept’ and ‘arguments dropped’ (given no operationalisation was offered in the original study of Haidt et al., 2000), the number of ‘unique justifications’ offered by each participant per task was measured in the current study. The number of unique justifications offered during a task were taken as persistence in defence of a judgment; that is, the smaller the number of justifications offered, the greater the persistence in defence of that judgment. A global measure of unique justifications was
computed for each participant by averaging their number of unique justifications across stimuli. Lower numbers were taken to indicate persistence of justification.

**Puzzlement**

Despite puzzlement being a component of the original definition of moral dumbfounding (Haidt et al., 2000) and appearing in the descriptions of other researchers (such as Haidt & Björklund, 2008; Haidt & Hersh, 2001; Jacobson, 2012; and Royzman et al., 2015), puzzlement over an inability to justify a judgment is yet to be quantitatively measured. Explicit statements of puzzlement at a failure to provide a reason for a judgment were coded in the current study.

**Procedure**

The order of task presentation was randomised to control for order effects. Per the original experiment, the Soul task was always presented last because pilot testing had shown it to have the potential to alter the participants' mood which could then affect the performance of other tasks (Haidt et al., 2000). The stimuli were presented in the following two orders:

1. Incest, Roach, Cannibal, Heinz, Soul
2. Heinz, Cannibal, Roach, Incest, Soul

Participants completed the interview individually in a research laboratory room, (a photograph of the arrangement of the laboratory room is annexed as Appendix B.3) to replicate the data collection strategy of Haidt et al. (2000). Participants were seated opposite the experimenter (a photograph of the seating arrangement is annexed as Appendix B.4). Participants were informed that they would be presented with five possibly confronting tasks; they would be asked to make a judgment regarding whether a described action was wrong, and to do something (for two behavioural tasks), and then to justify their decision. They were advised that there are no right or wrong answers but that the experimenter would play 'devil's advocate' by questioning the justifications given for their decision. Standardised probe questions were used in keeping with the protocol of Haidt et al. (2000). If the participant responded that what the person in the story did was wrong, the main counterargument was that no harm was done; if the participant refused to engage in the action prescribed by the Roach or Soul tasks, the main counterargument was that no harm would be done. Figure 1 depicts the standardised protocol for the interview.
Figure 1. Diagrammatic representation of standardised interview protocol.

After each task, participants were asked to complete a short questionnaire indicating how they made their judgment and how they perceived the task (annexed as Appendix D.1). Once the final task was completed, participants were asked to respond to a final questionnaire incorporating demographic items (annexed as Appendix D.2). Upon completion, participants were thanked for their time and debriefed during which the theoretical underpinning of the study was explained. The participants were asked not to discuss the experiment (both theory and their experience) with anyone, as knowledge of the experiment could influence responses of potential volunteers.
The methodology utilised in the current study deviated slightly from that adopted in the original research of Haidt et al. (2000). First, Haidt et al. told their participants that interviews were not being recorded (going to the effort of pointing out a “large and conspicuous video camera on a tripod, visibly unplugged and pointed away from the participant”, p. 6). It was only upon completion of the interview that participants were informed that the session had been video and audio recorded (one participant refused permission for her recording to be analysed upon learning of the deception). While it may be thought that such short-term deception might avoid the confound of the observing process influencing participant behaviour (Hayes, 2000), regardless, an observer effect on participant behaviour is expected given that the interview is conducted in a research laboratory room with an experimenter present. Accordingly, in the current study, there was no deception, and participants were informed that the session was being video and audio recorded.

**Data analysis**

Both qualitative and quantitative data were collected and analysed.

**Qualitative analysis**

Audio recordings of each interview were sent to Rev.com for transcription. This service guarantees maximum precautions are taken to ensure data security and transcription accuracy. Once the transcriptions were returned, the researcher and two research assistants independently reviewed and edited the transcripts to ensure the accuracy of the data. Transcripts were then imported into NVivo for Windows (Version 11 Pro; QSR, 2015). NVivo is a commonly-used qualitative data-analysis program. The video files were converted from MP4 to an AVI format (NVivo does not recognise MP4 format). The AVI video files were then imported into NVivo for analysis.

A mixed deductive/inductive coding approach was applied to the data. As recommended by Bandara, Furtmueller, Gorbacheva, Miscon, and Beckhuizen (2015), guidelines were developed for coding. Per the guidelines, sentences (and even paragraphs) from transcripts and qualitative questionnaire responses, and large sections of recording from video files, were initially coded (rather than just fragments) to capture the context of the information. The content was coded under multiple codes (rather than only one code), which enabled the gathering of information in different combinations. A hierarchical codification scheme, derived from the research of Haidt et al. (2000) was adapted and extended during an iterative process as the researcher, and a research assistant, developed an understanding and appreciation of what was in the interviews.
All interviews (both transcripts and video footage) were coded and re-coded continually during the development of the codification scheme until no new categories emerged. The categories were labelled and operationalised. Interrelations between classes and their subcategories were identified, and groups were integrated and refined. Preliminary results from prior interviews were used to guide subsequent verbal analysis. Previous interviews were then revisited and coded according to the most recent hierarchical codification scheme. All were then compared, linked, and analysed until no new categories or links were evident (saturation occurred). The final hierarchical codification scheme is annexed as Appendix E.

Thoughts that emerged during the coding process were recorded systematically by the researcher and research assistant using NVivo’s memo function with annotations made on relevant pieces of text in a source. An Excel spreadsheet was designed to record the number of observations of participant behaviour once coded and quantified through NVivo analysis (NVivo produces a count of the number of instances coded under each node). The time variables (such as total task response period, and seconds that judgment preceded justification) were calculated for each participant using the formula function of Excel.

**Quantitative analysis**

SPSS Statistics for Windows (Version 25.0; IBM Corp, 2012) was used to undertake quantitative analyses. Questionnaire responses were downloaded from GoogleForms to Excel. Data were inspected for completeness and consistency, and no missing data were detected. The Excel file containing the questionnaire responses of participants and the Excel file containing quantification of behavioural observations were merged. The researcher and two research assistants reviewed the merged dataset to ensure the integrity of each row. The Excel file was then imported into SPSS Statistics for analysis.

**General data-analysis information**

Coding of verbal and behavioural variables was conducted independently by the researcher and a research assistant (in contrast to Haidt et al., 2000). An intraclass correlation was applied to demonstrate consistency among observational interval ratings (as opposed to Cohen’s kappa for nominal variables; Hallgren, 2012). The two-person coding procedure resulted in a very high degree of inter-rater reliability between 20 measurements. The average measure intraclass correlation coefficient was 0.987 with a 95% confidence interval from 0.968 to 0.995 ($F_{(19, 19)} = 77.96, p < 0.001$). Further
reliability was performed whereby the two coders together reviewed each of the instances recorded under each variable node in NVivo. The representativeness of the cases to each variable was discussed with differences in node assignments resolved. Finally, a subset of interviews \((n = 22)\), together with the coding allocations, was peer-reviewed by a second research assistant. There was no discrepancy in the coding.

The sample size of the original study (Haidt et al., 2000) was 30 participants (the smallest sample size that can be safely drawn from a non-normal distribution of observations if one wants to produce a normal sampling distribution of sample means). While the sample size of the current study was larger than the original research \((N = 44)\), a Monte Carlo simulation (MCS) was applied to compensate for a relatively small sample size and preserve statistical power. A resampling technique, MCS is utilised to estimate quantities of interest using the Law of Large Numbers. The process of MCS involves values being sampled at random from the obtained data. The resulting outcome from that sample is recorded. This is done repeatedly, each time using a different set of random values (called an iteration). In the analyses of this dissertation, iteration was set at 10,000 recalculations. The result is a probability distribution of possible outcome values which yields unbiased estimates about the characteristics of the population from which the sample at hand is drawn (sc., Field, 2009).

The current study also addresses several limitations present in the analyses undertaken by Haidt et al. (2000). First, the original researchers did not report whether their data met parametric assumptions so the appropriateness of these tests for these analyses is uncertain. In the current study, all variable distributions were checked for normality (i.e., skewness and kurtosis less than \(\pm 1\), and significance of the Shapiro-Wilk statistic), and sphericity (i.e., the significance of the Mauchly’s Test of Sphericity) using SPSS. As SPSS does not provide a test of homogeneity of variance for repeated measures variables, an \(F\)max test was computed by hand using the formula \(F_{\text{max}} = \frac{\text{Largest Sample Variance}}{\text{Smallest Sample Variance}}\), where the largest sample variance is simply the largest standard deviation squared, and the smallest sample variance is the smallest standard deviation squared (Allen & Bennett, 2008). Homogeneity of variance can be assumed when \(F_{\text{max}}\) is less than 10 (Tabachnick & Fidell, 2001). The assumptions of normality, sphericity, and homogeneity of variance were violated for many of the variables and, therefore, non-parametric tests were used (Allen & Bennett, 2008).

Second, Haidt et al. (2000) inflated the risk of Type 1 error as they conducted 180 separate \(t\)-tests and, when they performed multiple-contrast statistics, they failed to
adjust their significance level for pairwise comparisons. In the current study, a Bonferroni adjustment was used to maintain a family-wise alpha rate of 0.05 over multiple comparisons and, therefore, reduce the risk of false-positive results. Unless otherwise stated, a one-tailed significance level was set at $\alpha = 0.0125$ (i.e., a standard level of .05 divided by the four planned comparisons between the reasoning stimulus and the four intuition stimuli; see, Allen & Bennett). Effect sizes are reported where possible, with the effect size formula for significant pairwise comparisons being $r = \frac{z}{\sqrt{N}}$ (Allen & Bennett). Descriptions of effect size were guided by J. D. Cohen (1988) in which $r = 0.1$ can be considered small, $r = 0.3$ can be considered medium, and $r = 0.5$ can be considered large. Where the effect size was unable to be ascertained (e.g., there is no commonly-used effect size formula for the McNemar test of change; Allen & Bennett, 2008) a graphical overview is provided to give a sense of the size and direction of the effect of significant comparisons.

A note regarding the presentation of data: It is customary to present the median and interquartile range for the non-parametric tests applied in the current study given that the analysis of data is based on ranks (from lowest to highest, corrected for ties) and not actual scores (which would be the case in parametric tests). Instead, the mean and standard deviation are presented in the current study as measures of central tendency and variation. The nonconformity of presentation was required due to the narrow scale of self-response items (typically 0 to 5) and the small frequency of some variables (e.g., statements of puzzlement usually occurred once during an interview, if at all). The benefit of presenting the mean instead of the median can be appreciated when considering, for example, the descriptive results of the variable ‘basis of initial judgment’ to the presented stimuli. The median basis of initial judgment for Heinz, Incest, Cannibal, Roach, and Soul were all 0.0; however, the mean of each stimulus were $-0.66$, $0.11$, $0.16$, $-0.16$, and $-0.18$, respectively. While presenting the mean provided a nuanced gauge of difference between the measures (unable to be revealed in presenting the median), it is important to remember that the assumption of normality was violated for many of the variables in the current study (which is why non-parametric tests were used). As such, the mean should not be considered the most accurate representation of central tendency. Nonetheless, the median for each variable in the current study is available upon request, and the mean ranking is provided below for each variable (in the relevant sections), which is an accurate indicator of the performance of the presented stimuli. Differences between mean ranks of pairwise
comparisons that are equal to, or bigger than, the critical difference are interpreted as significant.

**Results**

Five participants indicated familiarity with the *Heinz* vignette. No familiarity with any of the other stimuli was reported. The experimenter ascertained the degree of familiarity with each of the five participants; all indicated that their familiarity was not extensive (e.g., “I think I heard a story like it in Developmental Psych”). Due to the small frequency found for one stimulus, and the superficial nature of the familiarity, it was not considered that exposure to, and previous contemplation of, the scenarios would influence the results.

**Moral dumbfounding**

It was hypothesised that moral dumbfounding would be elicited in response to the stimuli presented. Table 1 shows the frequency of participants who were classified as morally dumbfounded across stimuli. Only 4 of the 44 participants were not dumbfounded in response to any of the presented stimuli; thus, 40 participants were classified as morally dumbfounded at least once during the interview (90.9%). The global measure of moral dumbfounding ranged between 0 and 4 with a mean of 2.09 ($SD = 1.14$).

<table>
<thead>
<tr>
<th>Stimuli</th>
<th>Morally Dumbfounded</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Heinz</td>
<td>4</td>
<td>9.1</td>
</tr>
<tr>
<td>Incest</td>
<td>26*</td>
<td>59.1</td>
</tr>
<tr>
<td>Cannibal</td>
<td>28*</td>
<td>63.6</td>
</tr>
<tr>
<td>Roach</td>
<td>13</td>
<td>29.5</td>
</tr>
<tr>
<td>Soul</td>
<td>21*</td>
<td>47.7</td>
</tr>
</tbody>
</table>

*Note. $n = $ frequency of participants from $N = 44$.

*Significantly different to *Heinz* (i.e., the reasoning stimulus) at the Bonferroni adjusted $\alpha$ of 0.0125.

It was further hypothesised that the intuition stimuli would elicit moral dumbfounding with significantly more frequency than the reasoning stimulus. A Cochran’s $Q$ ($u =$
0.05) was used to test the prevalence of moral dumbfounding across the tasks. The prevalence of dumbfounding varied significantly across the stimuli, $Q (df = 4, N = 44) = 37.09, p < 0.001$. The McNemar test of change (Bonferroni adjusted $\alpha = 0.0125$) was used for post-hoc comparisons. The observed tendency for participants to become dumbfounded was significantly higher on both intuition vignettes (Incest, $p < 0.001$; Cannibal, $p < 0.001$) than the Heinz vignette. Participants also tended to become dumbfounded significantly more on the Soul task ($p < 0.001$) than the Heinz vignette. Despite expectations, the Roach task did not differ significantly from the reasoning stimulus in the prevalence of dumbfounding ($p = 0.025$). Figure 2 provides a graphical overview of the size and direction of the effect of significant comparisons.

![Bar chart](image)

**Figure 2.** Clustered bar chart illustrating the difference between the Heinz vignette and the three intuition stimuli for which there was a significant effect in the prevalence of moral dumbfounding.

**Judgments of wrongness of an action**

It was hypothesised that the intuition stimuli would be judged as wrong with significantly more frequency than the reasoning stimulus. Figure 3 presents a graphical overview depicting the frequency of participant initial and final judgments to the five presented stimuli. As the relevant hypothesis focuses on the wrongness of an action,
Table 2 presents only the frequency and mean ranking of initial and final judgments of ‘wrong’ to the presented stimuli. A Friedman two-way analysis of variance (ANOVA; $\alpha = 0.05$) was used to compare three or more related samples of data (this was used as the data were not amenable to a one-way repeated measures ANOVA due to violations of assumptions). Initial and final judgments were analysed separately. The Wilcoxon Signed Rank Test (Bonferroni adjusted $\alpha = 0.0125$) was used for post-hoc comparisons. The use of non-parametric analyses differs from the analysis strategy of Haidt et al. (2000) who used a parametric multivariate analysis of variance (MANOVA) for analysis of the binary measure of their participants’ judgments in response to the stimuli (i.e., ‘OK’, ‘wrong’).

![Clustered column chart](image)

*Figure 3.* Clustered column chart illustrating the frequency of participant initial and final judgments to the dumbfounding stimuli.
Table 2
Frequency and Mean Ranking of Initial and Final Judgments of ‘Wrong’ in Response to the Dumbfounding Stimuli

<table>
<thead>
<tr>
<th>Judgment</th>
<th>Stimuli</th>
<th>n</th>
<th>%</th>
<th>Mean Rank</th>
</tr>
</thead>
<tbody>
<tr>
<td>Initial</td>
<td>Heinz</td>
<td>19</td>
<td>43.2</td>
<td>2.65</td>
</tr>
<tr>
<td></td>
<td>Incest</td>
<td>32</td>
<td>72.7</td>
<td>3.34*</td>
</tr>
<tr>
<td></td>
<td>Cannibal</td>
<td>38</td>
<td>86.4</td>
<td>3.81*</td>
</tr>
<tr>
<td></td>
<td>Roach</td>
<td>14</td>
<td>31.8</td>
<td>2.23</td>
</tr>
<tr>
<td></td>
<td>Soul</td>
<td>26</td>
<td>59.1</td>
<td>2.98</td>
</tr>
<tr>
<td>Final</td>
<td>Heinz</td>
<td>22</td>
<td>50.0</td>
<td>2.93</td>
</tr>
<tr>
<td></td>
<td>Incest</td>
<td>28</td>
<td>63.6</td>
<td>3.17</td>
</tr>
<tr>
<td></td>
<td>Cannibal</td>
<td>36</td>
<td>81.8</td>
<td>3.65*</td>
</tr>
<tr>
<td></td>
<td>Roach</td>
<td>14</td>
<td>31.8</td>
<td>2.30</td>
</tr>
<tr>
<td></td>
<td>Soul</td>
<td>25</td>
<td>56.8</td>
<td>2.95</td>
</tr>
</tbody>
</table>

Note. n = frequency of participants from N = 44.
*Significantly different to Heinz (i.e., the reasoning stimulus) at Bonferroni adjusted α of 0.0125.

Initial judgments of wrongness of an action

Rankings of initial judgments varied significantly across the five presented stimuli, $\chi^2_F (df = 4, N = 44) = 40.60$ (corrected for ties), $p < 0.001$. There were significantly more initial judgments of ‘wrong’ to the Incest vignette than to the Heinz vignette, $z = -2.95$, $N - \text{Ties} = 22$, $p = 0.002$, $r = 0.44$ (large effect size). There were also significantly more initial judgments of ‘wrong’ to the Cannibal vignette than to the Heinz vignette, $z = -3.91$, $N - \text{Ties} = 25$, $p < 0.001$, $r = 0.59$ (large effect size). Neither of the behavioural tasks differed significantly to Heinz in initial judgments of wrongness (Roach, $p = 0.036$; Soul, $p = 0.134$).

Final judgments of wrongness of an action

Rankings of final judgments also varied significantly across the five presented stimuli, $\chi^2_F (df = 4, N = 44) = 25.50$ (corrected for ties), $p < 0.001$. There were significantly more final judgments of ‘wrong’ to the Cannibal vignette than to the Heinz vignette, $z = -3.22$, $N - \text{Ties} = 21$, $p < 0.001$, $r = 0.49$ (large effect size). None of the other intuition
tasks differed significantly to Heinz in final judgments of wrongness (Roach, $p = 0.018$; Incest, $p = 0.168$; Soul, $p = 0.474$).

**Change of judgment**

No prediction was made regarding differences between the measures with respect to change of judgment given that Haidt et al. (2000) report that the percentage of participants who changed their minds did not differ significantly across tasks. Table 3 presents the frequency, across stimuli, of participants who changed their judgment (from initial to final judgment). A Cochran’s $Q$ ($\alpha = 0.05$) was used to test the prevalence of changing judgment (from initial to final) across tasks. Again, this is in contrast to Haidt et al. (2000) who used a Friedman two-way ANOVA to compare participants who changed their mind from their initial to final judgment on Heinz compared to Incest, Cannibal, Roach, and Soul (and explicitly state on p. 15 that this variable was binary); however, the Friedman test should only be used on ordinal or interval data. The tendency for participants to change their judgment did not vary significantly across the five presented stimuli, $Q (df = 4, N = 44) = 5.25, p = 0.281$.

<table>
<thead>
<tr>
<th>Stimuli</th>
<th>Change of Judgment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Heinz</td>
<td>9</td>
</tr>
<tr>
<td>Incest</td>
<td>6</td>
</tr>
<tr>
<td>Cannibal</td>
<td>6</td>
</tr>
<tr>
<td>Roach</td>
<td>2</td>
</tr>
<tr>
<td>Soul</td>
<td>5</td>
</tr>
</tbody>
</table>

*Note. $n =$ frequency of participants from $N = 44$. No stimulus was significantly different to Heinz (i.e., the reasoning stimulus) at the Bonferroni adjusted $\alpha$ of 0.00625 (two-tailed).*

**Response timing**

The total response time for interviews ranged from 7 minutes and 50 seconds through to 30 minutes and 19 seconds. On average, the total response time for interviews was 17 minutes and 2 seconds ($SD = 5$ minutes and 32 seconds). Table 4 displays the range and mean response time (in seconds) to the presented stimuli.
Table 4
*Range and Mean of Total Response Time to the Dumbfounding Stimuli*

<table>
<thead>
<tr>
<th>Stimuli</th>
<th>Minimum</th>
<th>Maximum</th>
<th>M</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Heinz</td>
<td>37</td>
<td>354</td>
<td>141.34</td>
<td>80.25</td>
</tr>
<tr>
<td>Incest</td>
<td>86</td>
<td>491</td>
<td>177.14</td>
<td>73.35</td>
</tr>
<tr>
<td>Cannibal</td>
<td>67</td>
<td>965</td>
<td>243.77</td>
<td>142.95</td>
</tr>
<tr>
<td>Roach</td>
<td>2</td>
<td>129</td>
<td>37.5</td>
<td>37.28</td>
</tr>
<tr>
<td>Soul</td>
<td>2</td>
<td>322</td>
<td>89.32</td>
<td>72.66</td>
</tr>
</tbody>
</table>

*Note. N = 44. M = mean; SD = standard deviation.*

It was hypothesised that participants would respond significantly faster when presented with the intuition stimuli than the reasoning stimulus. Table 5 presents the mean (in seconds) and mean ranking of time to initial judgment, time to initial justification, and the difference in time between judgment and justification, to the presented stimuli. A series of Friedman two-way ANOVA ($\alpha = 0.05$) were used to compare three or more related samples (i.e., response times to *Heinz, Incest, Cannibal, Roach*, and *Soul*) for time to initial judgment, time to initial justification, and the order of responses, respectively. The Wilcoxon Signed Rank Test (Bonferroni adjusted $\alpha = 0.0125$) was used for post-hoc comparisons.
Table 5
Mean Response Times and Mean Ranking of Time to Initial Judgment, Time to Initial Justification, and Time Difference Between Judgment and Justification, Across the Dumbfounding Stimuli

<table>
<thead>
<tr>
<th>Stimuli</th>
<th>Initial Judgment</th>
<th>Initial Justification</th>
<th>*Judgment Precedes Justification</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M</td>
<td>SD</td>
<td>Mean Rank</td>
</tr>
<tr>
<td>Heinz</td>
<td>11.91</td>
<td>13.39</td>
<td>3.49</td>
</tr>
<tr>
<td>Incest</td>
<td>8.27</td>
<td>11.69</td>
<td>2.93</td>
</tr>
<tr>
<td>Cannibal</td>
<td>12.55</td>
<td>17.10</td>
<td>3.10</td>
</tr>
<tr>
<td>Roach</td>
<td>4.59</td>
<td>7.26</td>
<td>2.00*</td>
</tr>
<tr>
<td>Soul</td>
<td>14.86</td>
<td>22.59</td>
<td>3.48</td>
</tr>
<tr>
<td>Total</td>
<td>10.44</td>
<td>9.16</td>
<td>-</td>
</tr>
</tbody>
</table>

Note. N = 44, M = mean; SD = standard deviation.
Time is in seconds.
*Negative sign indicates that justification precedes judgment (of ‘wrong’).
*Significantly different to *Heinz* (i.e., the reasoning stimulus) at the Bonferroni adjusted a of 0.0125.
Time to initial judgment

The time to initial judgment varied significantly across the five presented stimuli, \( \chi^2_F (df = 4, N = 44) = 26.84 \) (corrected for ties), \( p < 0.001 \). Time to initial judgment for the Roach task was significantly faster than for the Heinz vignette, \( z = -3.41, N - \text{Ties} = 43, p = <0.001, r = 0.51 \) (large effect size). None of the other intuition stimuli differed significantly from the reasoning stimulus in time to initial judgment (Incest, \( p = 0.021 \); Cannibal, \( p = 0.221 \); Soul, \( p = 0.329 \)).

Time to initial justification

The time to initial justification also varied significantly across the five presented stimuli, \( \chi^2_F (df = 4, N = 44) = 30.56 \) (corrected for ties), \( p < 0.001 \). Time to initial justification for the Roach task was significantly faster than for the Heinz vignette, \( z = -4.14, N - \text{Ties} = 43, p = <0.001, r = 0.62 \) (very large effect size). Compared to the Heinz vignette, the slower time to initial justification for the Soul task approached, but did not reach, significance (\( p = 0.018 \)); however, the remaining two intuition stimuli were clearly non-significant (Incest, \( p = 0.403 \); Cannibal, \( p = 0.477 \)).

Order of responses

It was also hypothesised that the intuition stimuli would elicit judgment prior to justification, whereas the reasoning stimulus would elicit justification prior to judgment. The order of judgment and justifications varied significantly across the five presented stimuli, \( \chi^2_F (df = 4, N = 44) = 9.65 \) (corrected for ties), \( p = 0.047 \). The order of responses for the Incest vignette was significantly different than for the Heinz vignette, \( z = -2.58, N - \text{Ties} = 41, p = 0.004 \). There was a medium to large difference between the two stimuli (\( r = 0.39 \)) with judgment preceding justification for the Incest vignette and justification preceding judgment for the Heinz vignette. None of the other intuition tasks differed significantly to the Heinz vignette in order of responses (Cannibal, \( p = 0.315 \); Roach, \( p = 0.393 \); Soul, \( p = 0.092 \)).

Response timing and moral dumbfounding

Separate inferential analyses between moral dumbfounding and time to initial judgment, and time to initial justification, were undertaken. Kendall’s tau-b (\( \alpha = 0.05 \)) was used as an alternative to Pearson’s correlation as the assumptions of normality and linearity could not be met (Kendall’s tau-b is generally preferred over Spearman’s rho because it provides a better estimate of the true population correlation, and is not artificially inflated by multiple tied ranks; Allen & Bennett, 2008). There was no correlation
between moral dumbfounding ($M = 2.09; SD = 1.14$) and ranked time to initial judgment ($\tau = -0.18, p = 0.086, N = 44$) nor ranked time to initial justification ($\tau = 0.05, p = 0.341, N = 44$).

**Basis underpinning judgments**

It was hypothesised that participants would report basing their judgments significantly more on gut feeling in response to the intuition stimuli than to the reasoning stimulus. Given the assumed differences in the use of reason and gut feeling in making initial and final judgments, within-subject analyses were conducted using a series of Wilcoxon Signed Ranks Tests (a two-tailed $\alpha$ of 0.025 was set). Table 6 presents the mean and mean ranking of the basis underpinning initial and final judgments, across stimuli. Lower ratings indicate more gut feeling and less reasoning as the foundation underpinning judgment.

**Table 6**

*Mean and Mean Ranking of the Basis (i.e., Reasoning or Gut Feeling) Underpinning Initial and Final Judgments of the Dumbfounding Stimuli*

<table>
<thead>
<tr>
<th>Stimuli</th>
<th>Basis of Initial Judgment</th>
<th></th>
<th>Basis of Final Judgment</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$M$</td>
<td>$SD$</td>
<td>Mean Rank</td>
<td>$M$</td>
</tr>
<tr>
<td>Heinz</td>
<td>$-0.66$</td>
<td>1.80</td>
<td>2.83</td>
<td>$1.02$</td>
</tr>
<tr>
<td>Incest</td>
<td>$0.11$</td>
<td>1.85</td>
<td>3.17</td>
<td>$0.80$</td>
</tr>
<tr>
<td>Cannibal</td>
<td>$0.16$</td>
<td>1.99</td>
<td>3.06</td>
<td>$1.23$</td>
</tr>
<tr>
<td>Roach</td>
<td>$-0.16$</td>
<td>2.30</td>
<td>2.99</td>
<td>$0.80$</td>
</tr>
<tr>
<td>Soul</td>
<td>$-0.18$</td>
<td>2.34</td>
<td>2.95</td>
<td>$-0.34$</td>
</tr>
</tbody>
</table>

*Note. $N = 44$, $M =$ mean; $SD =$ standard deviation. Lower ratings indicate more gut feeling and less reasoning as the basis of judgment.

*Significantly different to Heinz (i.e., the reasoning stimulus) at the Bonferroni adjusted $\alpha$ of 0.0125.*

**Comparison of original and alternative dumbfounding stimuli**

The only stimulus to not differ significantly between the self-reported basis of initial and final judgments was the *Soul* task (mean rank$_{inital} = 11.07$, mean rank$_{final} = 13.44$; $p = 0.610$). For the remaining stimuli, participants were significantly more likely to base their final judgment on reasoning than they were their initial judgment. Reasoning was used significantly more for the final judgment of the *Heinz* vignette (mean rank = 20.10) than for the initial judgment (mean rank = 14.35), $z = -2.99$, $N - Ties = 36$, $p = 0.002$, $\nu = 0.45$ (large effect size). Final judgment for the *Incest* vignette (mean rank = 12.07).
involved significantly more reasoning than the initial judgment (mean rank = 8.33), 
\( z = -2.33, N - \text{Ties} = 21, p = 0.018, r = 0.35 \) (medium effect size). For the Cannibal 
vignette, the final judgment (mean rank = 17.02) was based significantly more on 
reasoning than the initial judgment (mean rank = 12.50), 
\( z = -3.18, N - \text{Ties} = 31, p = 0.002, r = 0.48 \) (large effect size). Finally, the basis of final judgment for the Roach task 
(mean rank = 10.15) was significantly more reasoning than the initial judgment (mean 
rank = 8.75), 
\( z = -3.14, N - \text{Ties} = 19, p = <0.001, r = 0.47 \) (large effect size). Given 
that almost all of the five within-subject analyses confirm that the self-reported 
foundation underpinning moral judgment varies across time, it was necessary to analyse 
initial and final judgments separately when comparing the basis underpinning judgment 
between measures.

**Comparison between alternative dumbfounding stimuli**

A pair of Friedman two-way ANOVAs (\( \alpha = 0.05 \)) were used to compare the stimuli in 
regard to the self-reported basis of initial judgment, and final judgment. The Wilcoxon 
Signed Rank Test (Bonferroni adjusted \( \alpha = 0.0125 \)) was used for post-hoc comparisons. 
For initial judgments, the self-reported basis of judgment did not vary across the five 
presented stimuli, \( \chi^2_P (df = 4, N = 44) = 1.29, p = 0.866 \). For final judgments, a 
significant difference was found between the stimuli in the self-reported basis 
underpinning the judgment, \( \chi^2_P (df = 4, N = 44) = 11.21 \) (corrected for ties), \( p = 0.023 \). 
Participants were significantly more likely to base their final judgment on gut feeling 
when responding to the Soul task than to the Heinz vignette, 
\( z = -3.25, N - \text{Ties} = 33, p = 0.001, r = 0.49 \) (large effect size). None of the other stimuli differed significantly from 
the reasoning stimulus in the basis underpinning final judgments (Incest, \( p = 0.075 \ ); 
Cannibal, \( p = 0.489 \ ); Roach, \( p = 0.299 \ ).

The mean of the basis for final judgment (\( M = 0.70, SD = 1.35 \)) was a positive 
integer indicating that, on average, this basis was more reasoning than gut feeling.
Kendall's tau-b (\( \alpha = 0.05 \)) was used to explore the relationship between moral 
dumbfounding (\( M = 2.09; SD = 1.14 \)) and the basis underpinning judgments. No 
relationship was found; \( \tau = -0.12, p = 0.151, N = 44 \).

**Confidence in judgment**

It was hypothesised that participants would report significantly less confidence in their 
judgment to the intuition stimuli than to the reasoning stimulus. Given possible 
differences in participants' confidence in initial and final judgments, within-subject
analyses were conducted using the Wilcoxon Signed Ranks Test. A two-tailed \( \alpha \) of 0.025 for each tail was set to test for the possibility of a significant relationship in both directions. There was no significant difference found in confidence of initial and final judgments for any of the stimuli (Heinz, \( p = 0.487 \); Incest, \( p = 0.861 \); Cannibal, \( p = 0.247 \); Roach, \( p = 0.573 \); Soul, \( p = 0.411 \)). As such, a confidence variable was computed for each stimulus by averaging each participants’ responses on the ‘confidence in initial judgment’ and ‘confidence in final judgment’ variables, with higher ratings taken to indicate more confidence in judgment. Table 7 presents the mean and mean ranking for confidence in judgment across stimuli. A Friedman two-way ANOVA (\( \alpha = 0.05 \)) was used to compare confidence in judgment of Heinz to that of Incest, Cannibal, Roach, and Soul. There was no significant difference in confidence in judgment across tasks \( \chi^2 \) (\( df = 4, N = 44 \)) = 2.77 (corrected for ties), \( p = 0.604 \).

Table 7
Mean and Mean Ranking of Confidence in Judgment of the Dumbfounding Stimuli

<table>
<thead>
<tr>
<th>Stimuli</th>
<th>M</th>
<th>SD</th>
<th>Mean Rank</th>
</tr>
</thead>
<tbody>
<tr>
<td>Heinz</td>
<td>4.18</td>
<td>0.99</td>
<td>3.17</td>
</tr>
<tr>
<td>Incest</td>
<td>4.11</td>
<td>0.77</td>
<td>2.97</td>
</tr>
<tr>
<td>Cannibal</td>
<td>4.15</td>
<td>0.73</td>
<td>2.73</td>
</tr>
<tr>
<td>Roach</td>
<td>4.30</td>
<td>0.61</td>
<td>3.14</td>
</tr>
<tr>
<td>Soul</td>
<td>4.09</td>
<td>0.96</td>
<td>3.00</td>
</tr>
</tbody>
</table>

*Note. \( N = 44 \). M = mean; SD = standard deviation. No stimulus was significantly different to Heinz (i.e., the reasoning stimulus) at the Bonferroni adjusted \( \alpha \) of 0.0125.*

A global measure of confidence in judgment was computed for each participant by averaging their confidence in judgment across stimuli. Higher ratings were taken to indicate more confidence in judgment. The mean of the confidence variable was 4.17 (\( SD = 0.52 \)). In investigating confidence in judgment and moral dumbfounding, it was necessary to test for the possibility of a significant relationship in both directions. This is because confidence in judgment was associated with the reasoning measure in the original dumbfounding study (the only measure not to elicit moral dumbfounding; Haidt et al., 2000); yet, confidence is the focus of some definitions of moral dumbfounding (such as those of Cushman et al., 2010; Pizarro & Bloom, 2003; and Sneddon, 2007).
As such, a two-tailed adjusted α of 0.025 for each tail was set. Kendall’s tau-b was used to explore the relationship between moral dumbfounding ($M = 2.09; SD = 1.14$) and confidence in judgment. No relationship was found; $τ = 0.012, p = 0.916, N = 44$.

**Confusion and irritation**

**Confused by task**

It was hypothesised that participants would report significantly more confusion in response to the intuition stimuli than to the reasoning stimulus. Table 8 presents the mean and mean ranking of the ‘Confused by Task’ variable, across stimuli. A Friedman two-way ANOVA ($α = 0.05$) was used to compare confusion elicited by *Heinz* to that of *Incest, Cannibal, Roach*, and *Soul*. Confusion varied significantly across the five presented stimuli, $χ^2 (df = 4, N = 44) = 9.77$ (corrected for ties), $p = 0.041$. The Wilcoxon Signed Rank Test (Bonferroni adjusted $α = 0.0125$) was used for post-hoc comparisons. Participants were significantly more confused during the *Soul* task than during the *Heinz* vignette, $z = -2.33, N - Ties = 18, p = 0.009, r = 0.35$ (medium effect size). While both the *Incest* and *Cannibal* vignettes approached, but did not reach, significance, (both $p = 0.017$), there was clearly no difference between the *Roach* task and the reasoning stimulus in self-reported confusion ($p = 0.224$). It may be noticed when viewing Table 8 that the mean of *Soul* is lower than *Heinz*, but the mean rank is higher. This is possible because the non-parametric test used here is based on ranks (from lowest to highest; corrected for ties) and not actual scores. In other words, *Soul* yielded a significantly higher mean rank compared with *Heinz*, which means that *Soul* had a greater number of high scores (corrected for ties), despite a lower mean. The same effect occurred with the variable ‘irritated by task’ with the measure of *Roach* (presented next).
Table 8

<table>
<thead>
<tr>
<th>Stimuli</th>
<th>Confused by Task</th>
<th>Irritated by Task</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$M$</td>
<td>$SD$</td>
</tr>
<tr>
<td>Heinz</td>
<td>1.59</td>
<td>0.73</td>
</tr>
<tr>
<td>Incest</td>
<td>1.93</td>
<td>0.93</td>
</tr>
<tr>
<td>Cannibal</td>
<td>1.93</td>
<td>0.95</td>
</tr>
<tr>
<td>Roach</td>
<td>1.70</td>
<td>0.80</td>
</tr>
<tr>
<td>Soul</td>
<td>1.15</td>
<td>1.11</td>
</tr>
</tbody>
</table>

*Note. $N = 44$. $M =$ mean; $SD =$ standard deviation.

*Significantly different to Heinz (i.e., the reasoning stimulus) at the Bonferroni adjusted $a$ of 0.0125.

Irritated by task

It was hypothesised that participants would report significantly more irritation in response to the intuition stimuli than to the reasoning stimulus. Table 8 (above) presents the mean and mean ranking of the ‘Irritated by Task’ variable, across stimuli. A Friedman two-way ANOVA ($\alpha = 0.05$) was used to compare irritation in response to Heinz to that of Incest, Cannibal, Roach, and Soul. While irritation varied significantly across the five presented stimuli, $\chi^2_F (df = 4, N = 44) = 23.95$ (corrected for ties), $p < 0.001$, post-hoc comparisons (using the Wilcoxon Signed Rank Test; Bonferroni adjusted $a = 0.0125$) indicated that participants were significantly less irritated during the Roach task than during the Heinz vignette, $z = -2.28$, $N - Ties = 17$, $p = 0.011$, $r = 0.34$ (medium effect size). There was no difference in irritation between the remaining three intuition stimuli and the reasoning stimulus (Incest, $p = 0.087$; Cannibal, $p = 0.019$; Soul, $p = 0.472$).

Psychic discomfort

It was hypothesised that there would be significantly more psychic discomfort in response to the intuition stimuli than to the reasoning stimulus. Table 9 presents the mean and mean ranking of observed manifestation of psychic discomfort across stimuli. A Friedman two-way ANOVA ($\alpha = 0.05$) was used to compare psychic discomfort in response to Heinz to that of Incest, Cannibal, Roach, and Soul. Psychic discomfort
varied significantly across the five presented stimuli, $\chi^2_p (df = 4, N = 44) = 89.94$ (corrected for ties), $p < 0.001$.

Table 9
*Mean and Mean Ranking of Psychic Discomfort in Response to the Dumbfounding Stimuli*

<table>
<thead>
<tr>
<th>Stimuli</th>
<th>$M$</th>
<th>$SD$</th>
<th>Mean Rank</th>
</tr>
</thead>
<tbody>
<tr>
<td>Heinz</td>
<td>2.78</td>
<td>2.06</td>
<td>2.03</td>
</tr>
<tr>
<td>Incest</td>
<td>2.90</td>
<td>1.92</td>
<td>2.20</td>
</tr>
<tr>
<td>Cannibal</td>
<td>3.14</td>
<td>2.12</td>
<td>3.32</td>
</tr>
<tr>
<td>Roach</td>
<td>26.98</td>
<td>33.65</td>
<td>4.39*</td>
</tr>
<tr>
<td>Soul</td>
<td>10.02</td>
<td>13.79</td>
<td>4.06*</td>
</tr>
</tbody>
</table>

Note. $N = 44$. $M$ = mean; $SD$ = standard deviation.
*Significantly different to Heinz (i.e., the reasoning stimulus) at the Bonferroni adjusted $\alpha$ of 0.0125.

The Wilcoxon Signed Rank Test (Bonferroni adjusted $\alpha = 0.0125$) was used for post-hoc comparisons. Psychic discomfort during the Roach task was significantly higher than during the Heinz vignette, $z = -5.43$, $N - Ties = 42, p = <0.001, r = 0.82$ (extremely large effect size). In addition, participants exhibited significantly more psychic discomfort during the Soul task than during the Heinz vignette, $z = -5.51$, $N - Ties = 43, p = <0.001, r = 0.83$ (extremely large effect size). Neither of the intuition vignettes differed significantly to the reasoning stimulus in the observed manifestation of psychic discomfort (Incest, $p = 0.335$; Cannibal, $p = 0.110$).

**Speech dysfluency**

It was hypothesised that there would be significantly less speech dysfluency in response to the intuition stimuli than to the reasoning stimulus. Table 10 presents the mean and mean ranking of speech dysfluency across stimuli. A Friedman two-way ANOVA ($\alpha = 0.05$) was used to compare speech dysfluency in response to Heinz to that of Incest, Cannibal, Roach, and Soul. Speech dysfluency varied significantly across responses to the five presented stimuli, $\chi^2_p (df = 4, N = 44) = 85.18$ (corrected for ties), $p < 0.001$.  

49
Table 10
Mean and Mean Ranking of Speech Dysfluency in Response to the Dumbfounding Stimuli

<table>
<thead>
<tr>
<th>Stimuli</th>
<th>M</th>
<th>SD</th>
<th>Mean Rank</th>
</tr>
</thead>
<tbody>
<tr>
<td>Heinz</td>
<td>161.00</td>
<td>30.84</td>
<td>3.13</td>
</tr>
<tr>
<td>Incest</td>
<td>122.00</td>
<td>34.40</td>
<td>3.84</td>
</tr>
<tr>
<td>Cannibal</td>
<td>158.00</td>
<td>43.54</td>
<td>4.24*</td>
</tr>
<tr>
<td>Roach</td>
<td>7.56</td>
<td>7.88</td>
<td>1.65*</td>
</tr>
<tr>
<td>Soul</td>
<td>15.90*</td>
<td>17.58</td>
<td>2.15*</td>
</tr>
</tbody>
</table>

*Significantly different to Heinz (i.e., the reasoning stimulus) at the Bonferroni adjusted α of 0.0125.

Note. N = 44, M = mean; SD = standard deviation.

The Wilcoxon Signed Rank Test (Bonferroni adjusted α = 0.0125) was used for post-hoc comparisons. There was significantly less speech dysfluency in responses to the Roach task than to the Heinz vignette, z = -4.66, N – Ties = 44, p < 0.001, r = 0.70 (very large effect size). There was also significantly less speech dysfluency in responses to the Soul task than to the Heinz vignette, z = -3.24, N – Ties = 43, p < 0.001, r = 0.49 (large effect size). There was, however, significantly more speech dysfluency in responses to the Cannibal vignette than in responses to the Heinz vignette, z = -4.37, N – Ties = 42, p < 0.001, r = 0.66 (very large effect size). The Incest vignette did not differ significantly to the Heinz vignette in speech dysfluency (p = 0.033).

Linguistic responses

Dead-end arguments

It was hypothesised that there would be significantly more dead-end arguments in response to the intuition stimuli than to the reasoning stimulus. Table 11 presents the mean and mean ranking of dead-end arguments across stimuli. A Friedman two-way ANOVA (α = 0.05) was used to compare dead-end arguments used in response to Heinz to that of Incest, Cannibal, Roach, and Soul. Dead-end arguments present in the responses of participants varied significantly across the five presented stimuli, χ²_r (df = 4, N = 44) = 20.05 (corrected for ties), p < 0.001.
Table 11
Mean and Mean Ranking of Dead-End Arguments in Response to the Dumbfounding Stimuli

<table>
<thead>
<tr>
<th>Stimuli</th>
<th>M</th>
<th>SD</th>
<th>Mean Rank</th>
</tr>
</thead>
<tbody>
<tr>
<td>Heinz</td>
<td>0.05</td>
<td>0.30</td>
<td>2.65</td>
</tr>
<tr>
<td>Incest</td>
<td>0.45</td>
<td>0.70</td>
<td>3.40*</td>
</tr>
<tr>
<td>Cannibal</td>
<td>0.39</td>
<td>0.78</td>
<td>3.19*</td>
</tr>
<tr>
<td>Roach</td>
<td>0.07</td>
<td>0.26</td>
<td>2.74</td>
</tr>
<tr>
<td>Soul</td>
<td>0.30</td>
<td>0.80</td>
<td>3.02</td>
</tr>
</tbody>
</table>

Note. N = 44. M = mean; SD = standard deviation.
*Significantly different to Heinz (i.e., the reasoning stimulus) at the Bonferroni adjusted $\alpha$ of 0.0125.

The Wilcoxon Signed Rank Test (Bonferroni adjusted $\alpha = 0.0125$) was used for post-hoc comparisons. Participants offered significantly more dead-end arguments in response to the intuition vignettes than to the Heinz vignette. The difference between the Incest vignette and Heinz, $z = -3.22$, N - Ties = 15, $p = <0.001$, had a large effect size ($r = 0.49$). The difference between the Cannibal vignette and Heinz, $z = -2.41$, N - Ties = 12, $p = .009$, had a medium effect size ($r = 0.36$). Neither of the behavioural tasks differed significantly from the reasoning stimulus with respect to dead-end arguments (Roach, $p = 0.497$; Soul, $p = 0.036$).

Statements of “I don’t know”
It was hypothesised that there would be significantly more statements of “I don’t know” in response to the intuition stimuli than to the reasoning stimulus. Table 12 presents the mean and mean ranking of “I don’t know” statements across stimuli. A Friedman two-way ANOVA ($\alpha = 0.05$) was used to compare statements of “I don’t know” in response to Heinz to that of Incest, Cannibal, Roach, and Soul. Statements of “I don’t know” present in the responses of participants varied significantly across the five presented stimuli, $\chi^2_f (df = 4, N = 44) = 13.76$ (corrected for ties), $p = 0.006$. 

51
Table 12
Mean and Mean Ranking of “I Don’t Know” Statements in Response to the Dumbfounding Stimuli

<table>
<thead>
<tr>
<th>Stimuli</th>
<th>“I Don’t Know” Statements</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M</td>
<td>SD</td>
<td>Mean Rank</td>
</tr>
<tr>
<td>Heinz</td>
<td>0.25</td>
<td>0.81</td>
<td>2.64</td>
</tr>
<tr>
<td>Incest</td>
<td>0.66</td>
<td>1.18</td>
<td>3.13</td>
</tr>
<tr>
<td>Cannibal</td>
<td>0.64</td>
<td>1.33</td>
<td>3.14</td>
</tr>
<tr>
<td>Roach</td>
<td>0.25</td>
<td>0.62</td>
<td>2.74</td>
</tr>
<tr>
<td>Soul</td>
<td>0.86</td>
<td>1.44</td>
<td>3.36*</td>
</tr>
</tbody>
</table>

Note. N = 44. M = mean; SD = standard deviation.
*Significantly different to Heinz (i.e., the reasoning stimulus) at the Bonferroni adjusted α of 0.0125.

The Wilcoxon Signed Rank Test (Bonferroni adjusted α = 0.0125) was used for post-hoc comparisons. Participants stated “I don’t know” significantly more in response to the Soul task than to the Heinz vignette, z = -2.76, N – Ties = 18, p = .003, r = 0.42 (medium to large effect size). None of the responses to the remaining intuition stimuli differed significantly from the reasoning stimulus in statements of “I don’t know” (Incest, p = 0.031; Cannibal, p = 0.017; Roach, p = 0.471).

Unsupported declarations
It was hypothesised that there would be significantly more unsupported declarations in response to the intuition stimuli than to the reasoning stimulus. Table 13 presents the mean and mean ranking of unsupported declarations across stimuli. A Friedman two-way ANOVA (α = 0.05) was used to compare unsupported declarations in response to Heinz to that of Incest, Cannibal, Roach, and Soul. Unsupported declarations present in the responses of participants varied significantly across the five presented stimuli, $X^2_{4, 44} = 67.58$ (corrected for ties), $p < 0.001$. 

52
Table 13
Mean and Mean Ranking of Unsupported Declarations in Response to the Dumbfounding Stimuli

<table>
<thead>
<tr>
<th>Stimuli</th>
<th>$M$</th>
<th>$SD$</th>
<th>Mean Rank</th>
</tr>
</thead>
<tbody>
<tr>
<td>Heinz</td>
<td>0.09</td>
<td>0.42</td>
<td>2.06</td>
</tr>
<tr>
<td>Incest</td>
<td>1.84</td>
<td>2.12</td>
<td>3.70*</td>
</tr>
<tr>
<td>Cannibal</td>
<td>2.09*</td>
<td>2.23</td>
<td>3.91*</td>
</tr>
<tr>
<td>Roach</td>
<td>0.32</td>
<td>0.96</td>
<td>2.28</td>
</tr>
<tr>
<td>Soul</td>
<td>0.91</td>
<td>1.22</td>
<td>3.05*</td>
</tr>
</tbody>
</table>

Note. $N = 44$. $M =$ mean; $SD =$ standard deviation.
*Significantly different to Heinz (i.e., the reasoning stimulus) at the at Bonferroni adjusted $\alpha$ of 0.0125.

The Wilcoxon Signed Rank Test (Bonferroni adjusted $\alpha = 0.0125$) was used for post-hoc comparisons. Participants offered unsupported declarations significantly more in response to three of the four intuition stimuli than to the reasoning stimulus. The only task to not differ significantly from Heinz was the Roach task ($p = .116$). Responses to the Incest vignette contained significantly more unsupported declarations than responses to the Heinz vignette, $z = -4.66$, $N - $Ties $= 28$, $p = <.001$, $r = .70$ (very large effect size). There were also significantly more unsupported declarations in response to the Cannibal vignette than Heinz, $z = -5.05$, $N - $Ties $= 33$, $p = <.001$, $r = .76$ (extremely large effect size). Finally, there were significantly more unsupported declarations in response to the Soul task than Heinz, $z = -3.34$, $N - $Ties $= 22$, $p = <.001$, $r = .50$ (large effect size).

**Dumbfounding statements**

It was hypothesised that there would be significantly more dumbfounding statements in response to the intuition stimuli than to the reasoning stimulus. Table 14 presents the mean and mean ranking of dumbfounding statements across stimuli. A Friedman two-way ANOVA ($\alpha = 0.05$) was used to compare dumbfounding statements in response to Heinz to that of Incest, Cannibal, Roach, and Soul. There was no difference in the frequency of dumbfounding statements across the stimuli, $\chi^2 (df = 4, N = 44) = 4.00$ (corrected for ties), $p = 0.556$.  

53
Table 14
Mean and Mean Ranking of Dumbfounding Statements in Response to the Dumbfounding Stimuli

<table>
<thead>
<tr>
<th>Stimuli</th>
<th>M</th>
<th>SD</th>
<th>Mean Rank</th>
</tr>
</thead>
<tbody>
<tr>
<td>Heinz</td>
<td>0.02</td>
<td>0.15</td>
<td>2.95</td>
</tr>
<tr>
<td>Incest</td>
<td>0.11</td>
<td>0.49</td>
<td>3.07</td>
</tr>
<tr>
<td>Cannibal</td>
<td>0.05</td>
<td>0.21</td>
<td>3.01</td>
</tr>
<tr>
<td>Roach</td>
<td>0.00</td>
<td>0.00</td>
<td>2.90</td>
</tr>
<tr>
<td>Soul</td>
<td>0.16</td>
<td>0.78</td>
<td>3.07</td>
</tr>
</tbody>
</table>

Note. N = 44. M = mean; SD = standard deviation. No stimulus was significantly different to Heinz (i.e., the reasoning stimulus) at the Bonferroni adjusted α of 0.0125.

Novel measures

Novel measures of linguistic responses (i.e., variables that were not measured by Haidt et al., 2000) include the composite variable of failure to provide a reason for judgment (comprising the novel measures of tautological reasons and illogical arguments, and the replicated measure of unsupported declarations), the composite variable of struggle to produce a reason for judgment (comprising the novel measure of unwillingness to support, and the replicated measure of dumbfounding statements), unique justifications, and puzzlement at an inability to produce a justification for judgment.

Failure to provide a reason

Table 15 presents the mean and mean ranking, across stimuli, of the composite variable of failure to provide a reason for judgment and two of the three components that comprise this variable, being tautological reasons and illogical arguments (with the other component, unsupported declarations, presented above). A series of Friedman two-way ANOVAs (α = 0.05) were used to compare three or more related samples (i.e., responses to Heinz, Incest, Cannibal, Roach, and Soul) for failure to provide a reason for judgment, tautological reasons, and illogical arguments, respectively. The Wilcoxon Signed Rank Test (Bonferroni adjusted α = 0.0125) was used for post-hoc comparisons.
Table 15
Mean and Mean Ranking of the Composite Variable of Failure to Provide a Reason for Judgment, and the Component Variables of Tautological Reasons and Illogical Arguments, in Response to the Dumbfounding Stimuli

<table>
<thead>
<tr>
<th>Stimuli</th>
<th>Failure to Provide a Reason</th>
<th>Tautological Reasons</th>
<th>Illogical Arguments</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( M )</td>
<td>( SD )</td>
<td>Mean Rank</td>
</tr>
<tr>
<td>Heinz</td>
<td>0.16</td>
<td>0.37</td>
<td>2.09</td>
</tr>
<tr>
<td>Incest</td>
<td>0.70</td>
<td>0.46</td>
<td>3.45*</td>
</tr>
<tr>
<td>Cannibal</td>
<td>0.84</td>
<td>0.37</td>
<td>3.80*</td>
</tr>
<tr>
<td>Roach</td>
<td>0.36</td>
<td>0.49</td>
<td>2.60</td>
</tr>
<tr>
<td>Soul</td>
<td>0.55</td>
<td>0.50</td>
<td>3.06*</td>
</tr>
</tbody>
</table>

Note. \( N = 44 \). \( M \) = mean; \( SD \) = standard deviation.
*Significantly different to Heinz (i.e., the reasoning stimulus) at the Bonferroni adjusted \( \alpha \) of 0.0125.

Failure to provide a reason for judgment varied significantly across the five presented stimuli, \( \chi^2 \) (\( df = 4 \), \( N = 44 \)) = 51.46 (corrected for ties), \( p < 0.001 \). Three of the four intuition stimuli differed significantly from the reasoning stimulus in a failure to provide a reason (Roach, \( p = 0.023 \), did not differ from Heinz). The Incest vignette was also significantly different from the Heinz vignette, \( z = -4.57 \), \( N - Ties = 28 \), \( p = <0.001 \), \( r = 0.68 \) (very large effect size). There was also significantly greater failure to provide a reason when responding to the Cannibal vignette than to the Heinz vignette; \( z = -5.48 \), \( N - Ties = 30 \), \( p = <0.001 \), \( r = 0.83 \) (extremely large effect size). Finally, the failure to provide a reason was a significantly greater when responding to the Soul task than to the Heinz vignette, \( z = -3.55 \), \( N - Ties = 23 \), \( p = <0.001 \), \( r = 0.53 \) (large effect size).

Tautological reasons
The offering of tautological reasons varied significantly across the five presented stimuli, \( \chi^2 \) (\( df = 4 \), \( N = 44 \)) = 31.70 (corrected for ties), \( p < 0.001 \). There were significantly more tautological reasons offered in response to the Incest vignette than to the Heinz vignette, \( z = -2.81 \), \( N - Ties = 20 \), \( p = .002 \), \( r = 0.42 \) (medium to large effect size). In addition, there were significantly more tautological reasons offered in response to the Cannibal vignette than during the Heinz vignette, \( z = -2.72 \), \( N - Ties = 19 \).
\[ p = .004, r = 0.41 \] (medium to large effect size). The difference between the behavioural tasks and the Heinz vignette in tautological reasons were non-significant (Roach, \( p = 0.210 \); Soul, \( p = 0.068 \)).

**Illogical arguments**

Illogical arguments, too, varied significantly across the five presented stimuli, \( \chi^2 (df = 4, N = 44) = 16.27 \) (corrected for ties), \( p = 0.003 \). Three of the intuition stimuli differed significantly to the reasoning measure in illogical arguments. The Incest vignette elicited significantly more illogical arguments than the Heinz vignette, \( z = -2.71, N - Ties = 8, p = .007, r = 0.41 \) (medium to large effect size). In addition, there were significantly more illogical arguments in response to the Soul task than to the Heinz vignette, \( z = -2.97, N - Ties = 10, p = .001, r = 0.45 \) (large effect size). The Roach task, too, elicited significantly more illogical arguments than the Heinz vignette, \( z = -2.64, N - Ties = 8, p = .003, r = 0.40 \) (medium to large effect size). Compared to the Heinz vignette, the difference in illogical arguments offered on the Cannibal vignette was non-significant (\( p = 0.50 \)).

**Struggle to produce a reason**

Table 16 presents the mean and mean ranking, across stimuli, of the composite variable of struggle to produce a reason and one of the two components that comprise this variable, being an unwillingness to support (with the other component, dumbfounding statements, presented above). A pair of Friedman two-way ANOVA (\( \alpha = 0.05 \)) were used to compare three or more related samples (i.e., responses to Heinz, Incest, Cannibal, Roach, and Soul) for struggle to produce a reason for judgment, and unwillingness to support, respectively. The Wilcoxon Signed Rank Test (Bonferroni adjusted \( \alpha = 0.0125 \)) was used for post-hoc comparisons.
Table 16
Mean and Mean Ranking of the Composite Variable of Struggle to Produce a Reason, and the Component Variable of Statements of Unwillingness to Support, in Response to the Dumbfounding Stimuli

<table>
<thead>
<tr>
<th>Stimuli</th>
<th>Struggle to Produce a Reason</th>
<th>Unwillingness to Support</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(M)</td>
<td>SD</td>
</tr>
<tr>
<td>Heinz</td>
<td>0.05</td>
<td>0.21</td>
</tr>
<tr>
<td>Incest</td>
<td>0.27</td>
<td>0.45</td>
</tr>
<tr>
<td>Cannibal</td>
<td>0.14</td>
<td>0.35</td>
</tr>
<tr>
<td>Roach</td>
<td>0.18</td>
<td>0.39</td>
</tr>
<tr>
<td>Soul</td>
<td>0.41</td>
<td>0.50</td>
</tr>
</tbody>
</table>

Note. \(N = 44\), \(M = \text{mean}\); SD = standard deviation.
*Significantly different to Heinz (i.e., the reasoning stimulus) at the Bonferroni adjusted \(\alpha = 0.0125\).

Struggle to produce a reason varied significantly across the five presented stimuli, \(X^2_F (df = 4, N = 44) = 20.96\) (corrected for ties), \(p = <0.001\). There was significantly greater struggle to produce a reason when responding to the Incest vignette than to the Heinz vignette, \(z = -2.89, N = \text{Ties} = 12, p = 0.003, r = 0.44\) (large effect size). There was also significantly greater struggle to produce a reason when responding to the Soul task than to the Heinz vignette, \(z = -3.77, N = \text{Ties} = 18, p = <0.001, r = 0.57\) (large effect size). The remaining two intuition stimuli did not differ from the reasoning stimulus in a struggle to produce a reason (Cannibal, \(p = 0.110\); Roach, \(p = 0.052\)).

Participant statements of unwillingness to support varied significantly across the five presented stimuli, \(X^2_F (df = 4, N = 44) = 25.03\) (corrected for ties), \(p = <0.001\). Three of the intuition stimuli differed significantly to the reasoning measure in an unwillingness to support. The Incest vignette elicited significantly more statements of unwillingness to support than the Heinz vignette, \(z = -2.65, N = \text{Ties} = 11, p = .004, r = 0.40\) (medium to large effect size). In addition, there were significantly more statements of unwillingness to support in response to the Roach task than to the Heinz vignette, \(z = -2.36, N = \text{Ties} = 9, p = .011, r = 0.36\) (medium effect size). Finally, the Soul task elicited significantly more statements of unwillingness to support than the Heinz vignette, \(z = -3.68, N = \text{Ties} = 19, p = <0.001, r = 0.56\) (large effect size). The Cannibal vignette did not differ significantly from the reasoning stimulus in statements of unwillingness to support (\(p = 0.158\)).
Unique justifications

Table 17 presents the mean and mean ranking of unique justifications offered for judgments across stimuli. A Friedman two-way ANOVA (α = 0.05) was used to compare unique justifications in response to Heinz to that of Incest, Cannibal, Roach, and Soul. Unique justifications offered by participants varied significantly across the five presented stimuli, $\chi^2_p (df = 4, N = 44) = 69.32$ (corrected for ties), $p < 0.001$.

Table 17
Mean and Mean Ranking of Unique Justifications for a Judgment of the Dumbfounding Stimuli

<table>
<thead>
<tr>
<th>Stimuli</th>
<th>$M$</th>
<th>$SD$</th>
<th>Mean Rank</th>
</tr>
</thead>
<tbody>
<tr>
<td>Heinz</td>
<td>3.57</td>
<td>1.78</td>
<td>2.81</td>
</tr>
<tr>
<td>Incest</td>
<td>4.80</td>
<td>1.73</td>
<td>3.68*</td>
</tr>
<tr>
<td>Cannibal</td>
<td>5.36</td>
<td>1.81</td>
<td>4.19*</td>
</tr>
<tr>
<td>Roach</td>
<td>2.05</td>
<td>1.51</td>
<td>1.76*</td>
</tr>
<tr>
<td>Soul</td>
<td>3.36</td>
<td>2.28</td>
<td>2.56</td>
</tr>
</tbody>
</table>

*Note. $N = 44$. $M =$ mean; $SD =$ standard deviation.

The Wilcoxon Signed Rank Test (Bonferroni adjusted $\alpha = 0.0125$) was used for post-hoc comparisons. Almost all intuition stimuli differed significantly from the reasoning stimulus. Participants made significantly more unique justifications for their judgments in response to both the intuition vignettes than to the Heinz vignette; Incest, $z = -2.99$, $N - \text{Ties} = 33, p = 0.002$, $r = 0.45$ (medium to large effect size); Cannibal, $z = -4.44$, $N - \text{Ties} = 40, p < 0.001$, $r = 0.67$ (very large effect size). Unique justifications for judgments also differed significantly between the Roach task and the Heinz vignette; however, the direction of the relationship was opposite to that of the other intuition stimuli, eliciting significantly less unique justifications than the reasoning stimulus, $z = -3.96$, $N - \text{Ties} = 38, p < 0.001$, $r = 0.60$, (very large effect size). There was no relationship between unique justifications offered for the Soul task and the Heinz vignette ($p = 0.628$).

The mean of the global unique justifications measure was $3.83$ ($SD = 1.11$). As unique justifications is a novel measure, it was necessary to test for the possibility of a significant relationship with moral dumbfounding in both directions; thus, a two-tailed $\alpha$ of 0.025 for each tail was set. Kendall’s tau-b was used to explore the relationship
between moral dumbfounding ($M = 2.09; SD = 1.14$) and unique justifications. No relationship was found; $\tau = 0.09, p = 0.468, N = 44$.

**Puzzlement**

Table 18 presents the mean and mean ranking of statements of puzzlement across stimuli. A Friedman two-way ANOVA ($\alpha = 0.05$) was used to compare puzzlement in response to *Heinz* to that of *Incest, Cannibal, Roach,* and *Soul*. Puzzlement at an inability to produce a justification for judgment did not vary significantly across the five presented stimuli, $\chi^2_F (df = 4, N = 44) = 8.25$ (corrected for ties), $p = 0.083$.

**Table 18**

*Mean and Mean Ranking of Statements of Puzzlement at an Inability to Justify a Judgment of the Dumbfounding Stimuli*

<table>
<thead>
<tr>
<th>Stimuli</th>
<th>$M$</th>
<th>$SD$</th>
<th>Mean Rank</th>
</tr>
</thead>
<tbody>
<tr>
<td>Heinz</td>
<td>0</td>
<td>0.0</td>
<td>2.91</td>
</tr>
<tr>
<td>Incest</td>
<td>4</td>
<td>9.1</td>
<td>3.14</td>
</tr>
<tr>
<td>Cannibal</td>
<td>1</td>
<td>2.3</td>
<td>2.97</td>
</tr>
<tr>
<td>Roach</td>
<td>0</td>
<td>0.0</td>
<td>2.91</td>
</tr>
<tr>
<td>Soul</td>
<td>3</td>
<td>6.8</td>
<td>3.08</td>
</tr>
</tbody>
</table>

*Note. N = 44. M = mean; SD = standard deviation. No stimulus was significantly different to *Heinz* (i.e., the reasoning stimulus) at the Bonferroni adjusted $\alpha$ of 0.00625 (two-tailed).*

Participants who explicitly stated puzzlement at an inability to provide justification typically offered one statement of puzzlement (the only exception was that one participant made two statements of puzzlement in response to the *Incest* vignette). As such, it made sense to transform this variable into a binary measure prior to analyses (i.e., ‘Puzzled During the Interview’ versus ‘Not Puzzled During the Interview’). There were eight participants (18.2%) who made a statement of puzzlement during their interview (*Incest = 4; Cannibal = 1; Soul = 3*), and 36 participants who did not explicitly state that they were puzzled. The McNemar test of change ($\alpha = 0.05$) was used to test the prevalence of moral dumbfounding in puzzled compared to non-puzzled participants. The tendency for those morally dumbfounded participants to be puzzled was significantly greater than for the non-dumbfounded; $\chi^2 (df = 1, N = 44) = 30.03, p < 0.001$. All eight participants who made a statement during the interview that they were
puzzled at their inability to produce a justification for their judgment were also
classified as dumbfounded. Figure 4 provides a graphical overview of the size and
direction of the significant effect.

![Clustered column chart](image)

**Figure 4.** Clustered column chart illustrating the significant effect between puzzlement and moral dumbfounding elicited using the dumbfounding stimuli.

**Discussion**

**Moral dumbfounding**

This study is the first full replication of the seminal moral dumbfounding study
conducted by Haidt et al. (2000) but it also involves a series of methodological
improvements to address limitations of the original research. As predicted, there was a
demonstration of moral dumbfounding (90.9%) in which actions were judged as wrong
without the provision of logical supporting reasons. The frequency of observed
dumbfounding was higher than that found by McHugh et al. (2017; i.e., 71%). Also as
predicted, moral dumbfounding was significantly more prevalent in response to the
majority of intuition stimuli than to the reasoning stimulus, mirroring the results of both
Haidt et al. and McHugh et al. The present study confirms that, at times, we cannot
articulate a reason for our moral judgment.

Haidt and colleagues (Haidt, 2001; Haidt et al., 2000; Haidt & Hersh, 2001)
argue that moral dumbfounding occurs because harmless yet offensive dilemmas (i.e.,
the intuition stimuli) provoke an automatic emotional reaction, and that dumbfounging occurs less with the Heinz dilemma (i.e., the reasoning measure) because participants are employing their impassive moral reasoning skills. When asked to defend their judgment on intuition stimuli, participants are unable to provide a reason as the source of moral judgment (i.e., affective intuition) is outside of conscious awareness. Subsequently, participants use their post-hoc reasoning process to search for a rationale for their decision and, in desperation, they may propose reasons that are easily discounted by the circumstantial details. Having had their justifications rejected, participants realise (with embarrassment) that they are without a reason (sometimes admitting as much) and simply reaffirm their initial, emotion-laden, intuitive judgment. Haidt claims moral dumbfounging as the only direct evidence for his SIM (2001, p. 817) in which moral judgment is an emotionally based intuitive process and reasoning plays no causal role. While it is not possible to find direct (even partial) support for SIM based on the current study (or Haidt et al., 2000), the findings here are discussed using the framework of Haidt’s claims.

**Claim 1: Differences between the reasoning stimulus and the intuition stimuli**

**Judgments of wrongness of an action**

While there was no difference between the behavioural tasks and the reasoning measure in judgments of wrongness of action, participants initially perceived the intuition vignettes as significantly more wrong than the reasoning vignette (as expected). Given that the differences in judgments of wrongness between each intuition vignette and the reasoning vignette were not trivial (large effect sizes were obtained for each comparison), and that confounding was minimised in the current study through holding the presentation of scenarios constant (i.e., video recorded as opposed to being read in each session by the experimenter or participant), it appears that there is a perceived distinction in permissibility of the content of the vignettes. This result supports Haidt et al. (2000) but is in contrast to McHugh et al. (2017) in which the frequency of judgments of wrongness was higher for Heinz than for the intuition vignettes (although no inferential analyses were conducted by either Haidt et al. nor McHugh et al. to quantify these results). Haidt et al. speculated that the difference in permissibility (and, indeed dumbfounging) between the vignettes was due to the nature of the stories (i.e., reasoning versus intuition). The reasoning dilemma is designed to elicit a considered response to prototypical concerns of harm versus rights, whereas the intuition dilemmas
are designed to evoke disgust. This makes drawing conclusions from the measures difficult as it is unclear if differences are due to the distinction between intuition and reasoning (as Haidt et al. claim) or if they are due to acceptability differences in moral content.

Change of judgment
Participants in the current study were no more likely to change their judgment in response to the reasoning measure than any of the intuition measures. Haidt et al. (2000) also reported that the rate of participants changing their decision did not vary significantly across stimuli, so it is curious that maintenance of judgment was included in their original definition of moral dumbfounding (p. 1). Nonetheless, some researchers (such as Haidt, 2001; Haidt & Björklund, 2008; Haidt & Hersh, 2001; Jacobson, 2012; Royzman et al., 2015) operationalise moral dumbfounding as stubbornly adhering to judgment despite the removal of reasons used to support condemnation. Collectively, the results from the current study, the original dumbfounding study (Haidt et al., 2000), and the partial replication of McHugh et al. (in which participants changed their minds more in response to Cannibal than Heinz - although not confirmed statistically; 2017, Study 1), suggest that definitions of the phenomenon excluding maintenance of a judgment (such as those of Cushman et al., 2010; Pizarro & Bloom, 2003; Sneddon, 2007) may more accurately represent the phenomenon than others. Researchers should, therefore, consider omitting this variable when operationalising moral dumbfounding in future investigations.

Confidence in judgments
Contrary to prediction, there was no difference in participants’ confidence in their judgments across tasks and there was no relationship between confidence in judgment and moral dumbfounding. In the original moral dumbfounding study (Haidt et al., 2000), participants demonstrated significantly more confidence in their judgment of the Heinz vignette (the only measure not to elicit moral dumbfounding in their study) than the intuition measures (except for the Roach task). McHugh and colleagues’ participants also reported more confidence in their judgment of Heinz than Cannibal (2017, Study 1). Together, these findings indicate that sureness in judgment may have little to do with moral dumbfounding. Nonetheless, as many researchers describe dumbfounding with reference to confidence in judgment (Cushman et al., 2010; Fedyk, 2017; Hauser, Cushman, Young, Kang-Xing Jin, & Mikhail, 2007; Hauser et al., 2008; and Pizarro &
Bloom, 2003; Russell & Giner-Sorolla, 2011; Sneddon, 2007), the prevalence of moral dumbfounding embodying confidence in a judgment should be further explored.

**Confusion and irritation**

It was hypothesised that participants would report significantly more confusion and irritation in response to the intuition stimuli than to the reasoning stimulus. Participants were significantly more confused during the *Soul* task than the *Heinz* vignette, with no other differences in confusion noted between intuition and reasoning tasks. In contrast, the *Soul* task was the only measure in Haidt et al. (2000) not to differ significantly from the *Heinz* vignette in confusion. Furthermore, while McHugh et al. (2017; Study 1) reported descriptive statistics suggesting greater confusion with the intuition vignettes (*Incest* and *Cannibal*) than with the *Heinz* dilemma, their second reasoning measure (*Trolley*) yielded the greatest confusion (but, again, this result was not quantified).

With respect to irritation, in contrast to the prediction, participants were significantly less irritated in the current study during the *Roach* task than the *Heinz* vignette, with no other differences in irritation noted between intuition and reasoning tasks. These results mirror those of Haidt et al. who also found the *Roach* task to have the lowest rating of irritation. Here, and as noted by Haidt et al., lack of irritation may reflect the perception of this task as amusing. For McHugh et al., the reasoning measure, *Trolley*, yielded the greatest irritation rating (not quantified). The conflicting results for confusion and irritation across the three studies, coupled with the lack of difference between the reasoning measure and the majority of intuition measures in the current study, suggest that conclusions about the importance of investigating confusion and irritation in relation to moral dumbfounding cannot be drawn at present. This is particularly so given probable validity problems with these items.

Specifically, the researcher in the current study noted that, during responding, some participants questioned whether confusion and irritation referred to the nature of the task, the content of the task, or their reaction to the ‘devil’s advocate’. As confusion and irritation could arise from many sources, the ambiguity of the items casts doubt on their validity as moral dumbfounding measures. The data for each empirical investigation in this dissertation were collected simultaneously, so improvement to these items for subsequent studies was not possible; consequently, confusion and irritation are not analysed past Study 1. It is recommended that these items are redesigned for future research.
Psychic discomfort

Participants were expected to exhibit significantly more psychic discomfort in response to the intuition stimuli than to the reasoning stimulus and, although an extremely large difference was found between the behavioural tasks and the reasoning measure in psychic discomfort, no significant difference emerged between intuition and reasoning vignettes. Haidt et al. (2000) report the same pattern (although they analysed their indicators of psychic discomfort separately), which reveals that, as anticipated, there is a distinction between the perception of third-person and first-person scenarios. As highlighted by Kennett (2012), being asked about another’s moral choices is different from being asked about one’s own. Moreover, verbal responses to hypothetical moral dilemmas have little predictive value for actual decisions made when responding with an action (Bostyn et al., 2018; FeldmanHall et al., 2012). Both McHugh et al. (2017) and Royzman et al. (2015) restricted their measures to third-person vignettes; thus, they neglected to investigate (and replicate) judgments relevant to personal decision-making (ordinarily considered to be primary instances of moral judgment; Kennett, 2012). The present result signals the importance for future studies of this moral dumbfounding to include behavioural measures, and also suggests that further investigation is needed into the relationship between task type and moral dumbfounding (see Study 3).

Unique justifications

The frequency of unique justifications was a novel measure developed for this study and was taken to indicate persistence in the line of defence of a judgment. Almost all intuition stimuli differed significantly from the reasoning stimulus in the provision of unique justifications for judgment (the Soul task was the only measure not to vary from Heinz). This result indicates that persistence in the line of defence may be better to include in the measurement of moral dumbfounding than the maintenance of a judgment (given that participants were no more likely to change their judgment in response to any of the measures). While there was no relationship between unique justifications and moral dumbfounding in the current study, this measure yielded interesting results regarding the direction of relationships between stimuli; the intuition vignettes elicited more unique justifications than the reasoning vignette, but the Roach task prompted less. Accordingly, the frequency of the unique justifications’ variable could benefit from further investigation (see Study 3).
Claim 2: Moral judgment is an intuitive response

It was expected that there would be a difference in the response timings of the presented stimuli such that participants would respond significantly faster when presented with the intuition stimuli than the reasoning stimulus. The Roach task was the only intuition stimulus to differ significantly from the reasoning measure in response time, with both faster judgments and justifications observed in response to this behavioural task than for Heinz. The rapid response time to the Roach task reflects the manner in which participants reacted to this task; specifically, many participants drank the juice immediately upon being asked without any discussion. In contrast to Haidt et al. (who report that participants responded faster when presented with the intuition measures than the reasoning measure; 2000), the lack of difference here in response time between the intuition stimuli and the reasoning measure (with the exception of the Roach task) suggests that the source of responses to all measures was equivalent; that is, the judgments and justifications came from an equally fast (or slow) process. Moreover, moral dumbfounding was not related to the speed with which participants offered their judgment; however, even if there was an association between moral dumbfounding and rate of response, this does not rule out that one arrived quickly at the decision due to repeated use of a similar judgment form in comparable scenarios (i.e., automatised prior judgment). If judgment represents a habitual unconscious response (Dwyer, 2009; Pizarro & Bloom, 2003), it is reasonable that participants would not have direct access to the specific reasons upon which their judgment is based. Accordingly, it is argued that the findings of the current study do not support the view of others (Haidt, 2001; Haidt et al., 2000; Haidt & Hersh, 2001; McHugh et al., 2017) that moral judgment is an intuitive automatic reaction devoid of reasoning.

Claim 3: Moral judgment is emotionally based

The basis underpinning participants’ initial judgment for all measures (including the supposed reasoning stimulus) was gut feeling (the mean of initial judgments was below the balance point of the gut feeling-reasoning continuum, indicating that more gut feeling and less reasoning was used as the basis of judgment). At first glance, these results appear consistent with Haidt’s (2001) theory that emotion-laden intuition (i.e., gut feeling) is the source of moral judgment; however, the basis underpinning judgment was self-reported so the limits of introspection undermine any conclusions. Moreover, moral dumbfounding was not related to the foundation upon which participants reportedly made their decision (i.e., reasoning or gut feeling). Consequently, rather than
offering support for the idea that moral dumbfounding is evidence that our moral judgments are grounded in how a moral dilemma makes us feel (rather than on a process of reason), it can only be concluded that participants report that the source of their initial response to a moral dilemma is gut feeling.

**Claim 4: Reasoning is post-hoc in the formation of moral judgment**

**Basis underpinning judgments**

While the (self-reported) basis underpinning participants’ initial judgment for all measures was gut feeling, by the end of the response time for each measure, participants had thought substantially about the presented case and reported basing their final judgment for each scenario significantly more on reasoning than they did their initial judgment (with the exception of the Soul task). While these results appear consistent with the claim that reasoning is post-hoc in the formation of moral judgment (Haidt, 2001), this would only be the case if participants did not change their minds, and reasons were generated subsequently to support their initial decision. That people do change their mind (in the current study, 52.3% of participants amended their initial judgment across tasks) suggests that moral judgment (even if not initially derived through conscious reasoning) is rationally evaluated and open to modification. As the final moral decision is established through rational means, reasoning is not merely post-hoc but plays a causal role in the formation of moral judgment.

**Response timing**

The sequence of responses to the presented measures was justification prior to judgment (except for the Incest vignette), which is in contrast to the results of Haidt et al. (2000) who observed their participant responses to the intuition cases (i.e., those that elicit moral dumbfounding) as being quick judgment followed by justification. Although the current results appear inconsistent with those of Haidt et al., they merely show that participants explicitly offered their rationalisation before their decision on almost all measures; they do not conclusively reveal an internal process of justification preceding judgment. This point is particularly relevant when considering the importance that Haidt places on the observations from the original moral dumbfounding study for his subsequent theory of the psychological processes underlying moral judgment (2001). Moreover, no relationship was found between moral dumbfounding and initial response time (both judgment and justification) in the current study (mirroring the results of McHugh et al., 2017). As such, the present research cannot be taken to support the position that moral dumbfounding demonstrates that reasoned justification is always
post-hoc and that moral judgment is intuitive (Haidt, 2001; Haidt et al., 2000; Haidt & Hersh, 2001).

**Claim 5: Moral dumbfounding is evidence of judgment without reason**

**Speech dysfluency**

The prediction of less speech dysfluency in responses to the intuition stimuli compared with the reasoning stimulus was partially supported, as responses to the behavioural tasks (i.e., Roach and Soul) contained significantly less speech dysfluency than the Heinz dilemma (mirroring the results of Haidt et al., 2000). It is probable that this result reflects the nature of the tasks; Roach and Soul are behavioural tasks and Heinz is a verbal task. Contrary to prediction, however, participants demonstrated greater speech dysfluency during the Cannibal vignette than during the Heinz vignette. According to McHugh et al. (2017), speech dysfluency indicates that the participant is working toward a reason for their judgment; less speech dysfluency was expected in responses to the intuition stimuli (compared to the reasoning measure) as judgments to these stimuli are (supposedly) without reason. The conflicting results of the current study help underscore the possible error arising from assuming conscious access to internal processes. Specifically, some moral dumbfounding researchers (Haidt et al., 2000; McHugh et al., 2017) argue that an observed behaviour (i.e., mere inarticulateness) represents a proposed internal state (i.e., being without reason for judgment); however, speech dysfluency may reflect an inability (for various reasons) to articulate our thoughts. Not every person can explain their moral reasoning clearly (and few can do so on all occasions); thus, difficulty in articulating a defence of judgment may have little to do with the internal process driving the decision. As the implications of speech dysfluency remains unclear, this variable is not analysed in this dissertation past the current study.

**Dead-end arguments**

Participants offered significantly more dead-end arguments in response to the intuition vignettes than to the reasoning stimulus (as expected), which makes sense when considering that the Heinz vignette represents a common concern (somewhat equivalent to stealing a loaf of bread to feed your family) whereas the intuition vignettes represent (disgust-inducing) scenarios that most likely would not have been considered before presentation. It is not unreasonable to verbalise while working towards a solution to a novel problem, stopping and adjusting lines of thought as required. Unlike the intuition
vignettes, the behavioural tasks did not differ significantly from the reasoning measure in dead-end arguments, which may be because a decision to engage in an imminent personal action is less likely to prompt external dialogue than careful consideration of a third-person action (once again highlighting the difference between first- and third-person measures). Nonetheless, while Haidt et al. (2000) argue that the rationale offered for a judgment is incomplete because there is no reasoning from which to extract a reason, it is more likely that dead-end arguments (like speech dysfluency) indicate that the participant is trying to articulate their reason for a judgment (rather than that they are without reason). As such, analysis of dead-end arguments is also omitted in the subsequent investigations reported in this dissertation.

**Statements of “I don’t know”**
The linguistic responses of participants to the intuition stimuli were hypothesised to contain significantly more statements of “I don’t know” than their responses to the reasoning stimulus. Participants did state “I don’t know” more in response to the *Soul* task than to the *Heinz* vignette. While some may argue that an explicit admission of not knowing why one has made a judgment is a definitive measure of moral dumbfounding (Haidt, 2001; Haidt et al., 2000; Haidt & Hersh, 2001; McHugh et al., 2017), this is not inevitably so given that such responses may include instances of low motivation, cognitive laziness, failure of introspection, and the like. Statements of “I don’t know”, therefore, reveal that the participant has failed to provide a reason rather than they are *without* reason. Even so, unlike the assumed implication of speech dysfluency and dead-end arguments, statements of “I don’t know” are an explicit acknowledgement signalling a *possible* lack of a reason behind a judgment; therefore, such statements are included in the analysis of subsequent studies.

**Failure to provide a reason**
The prediction that there would be a greater failure to provide a reason for judgments to the intuition stimuli than to the reasoning measure was mostly supported. A significantly higher failure of justification occurred when justifying judgments to *Incest, Cannibal, and Soul*, compared to *Heinz*; the magnitude of these differences ranged from large to extremely large (again, the non-verbal response to *Roach* explains the lack of distinction between this task and the *Heinz* vignette). That participants failed to provide a reason for their judgment for the majority of intuition stimuli may be relevant to understanding dumbfounding (given the phenomenon is described as a failure to give a reason for a judgment), but any conclusions in this regard should be moderated.
Taking the *Incest* vignette as an example, statements such as “That’s terrible” (unsupported declaration), “They’re related” (tautological reason), or “They might get pregnant” (illogical argument), did not satisfy the researcher’s criteria of having a logical basis for a judgment (sc., Mallon & Nichols, 2011; McHugh et al., 2017) and were subsequently taken as a failure to provide a reason. Nonetheless, participants most likely offered such statements believing them to be logical reasons for their judgment. While Haidt et al. (2000) maintain that each of the dumbfounding stimuli encompass the violation of a social taboo carefully designed to be perceived as harm-free, in real life, incest is a taboo *because* it carries with it many potentially harmful consequences (such as familial dysfunction and abuse, genetic abnormality of offspring, etc.). It seems unreasonable to expect that participants would discount their understanding of incest in everyday terms during the experiment (Kennett, 2011, 2012); therefore, statements such as, “That’s terrible”, “They’re related”, and “They might get pregnant” were most likely offered by participants as logical reasons for their judgment. Accordingly, a failure to provide a reason should not be taken as definitive evidence that participants are, in fact, without reason.

**Struggle to produce a reason**

The hypothesis that there would be significantly more dumbfounding statements contained in responses to the intuition stimuli than to the reasoning stimulus was not supported. This result was surprising given the importance placed on such statements as a measurable indicator of dumbfounding in existing empirical studies (Haidt et al., 2000; Haidt & Hersh, 2001; McHugh et al., 2017; Royzman et al., 2015). There was, however, a significant difference found in statements of *unwillingness to support* between the stimuli. Developed inductively during this study, the variable of *unwillingness to support* encompassed statements supporting an acceptability judgment for the scenario or task, but a reluctance to commit to such a view; the *Incest* vignette, and the behavioural tasks, elicited more unwillingness to support than *Heinz*. Given the differences found between the majority of intuition stimuli and the reasoning measure (but no such distinction in dumbfounding statements between any of the stimuli), this novel measure should be included in future research on moral dumbfounding.

Furthermore, integrating the two variables of dumbfounding statements and unwillingness to support (as a struggle to produce a reason for a judgment) resulted in *Roach* no longer being significantly different from the reasoning measure. This indicates that the variable of struggle to produce a reason is a better measure to
differentiate between moral dumbfounding stimuli than dumbfounding statements, and unwillingness to support, alone. While the dumbfounding statement hypothesis was not supported, the current findings are broadly compatible with Haidt et al. (2000) insofar as some participants explicitly struggled to provide a reason for judgment.

**Puzzlement**

In the current study, *all* participants who made a statement that they were puzzled by their inability to produce a justification for their judgment were independently classified as morally dumbfounded. That puzzlement did not differ significantly between the stimuli indicates that it is an essential element of moral dumbfounding; regardless of differences in presenting dilemma, if one is puzzled at an inability to produce a reason for one’s judgment they are, necessarily, dumbfounded. Despite omitting puzzlement in the measurement of dumbfounding for their partial replication study, McHugh et al. (2017) admit that this element may be “useful in differentiating between a failure to provide reasons and a refusal to provide reasons” (p. 2). Puzzlement is included in the original moral dumbfounding definition (Haidt et al., 2000, p. 1) and the current study supports this inclusion.

**Limitations**

A moral judgment can embody different forms; for instance, we can make a holistic judgment about an agent (e.g., “he is honest”, “he is deceitful”), or about a particular action of an agent (e.g., “she is being respectful”, “she deserved what is happening”). We can make a moral judgment about past actions (e.g., “Australia should not have participated in the invasion of Iraq”, “Australia gave generous relief to the bushfire victims”), or about future actions such as what ought to, or ought not to, happen (e.g., “the wallet he found should be handed in to the police”, “he should not steal the man’s wallet”). The measures in the original study of Haidt et al. (2000), upon which this research was based, elicit first- and third-person moral judgments; therefore, the current research is limited to moral judgments of present first-person actions (e.g., “I will/will not do that”) and past third-person actions (e.g., “What they did was/was not wrong”).

Like moral judgment, moral dumbfounding is not a unidimensional phenomenon (i.e., there are distinct but related phenomena that come under its rubric); for instance, one may find the arguments against eating meat to be compelling (e.g., that animals are sentient and will suffer in the production of meat) but may nevertheless judge eating meat to be permissible and not be able to provide justification. Nonetheless, moral dumbfounding is almost exclusively limited in extant research and literature to
situations in which one judges that an action is impermissible (e.g., mutually-consensual incest) but is unable to justify this view. Given that Haidt (Haidt, 2001; Haidt et al., 2000; Haidt & Hersh, 2001) restricts his discussion to the first type of moral dumbfounding, the current research is limited to situations in which participants judge an action to be wrong but are unable to provide a reason (or at least an acceptable reason) for this judgment. In future research, it would be of interest to explore other possible manifestations of moral dumbfounding.

Finally, coding of verbal and behavioural variables was conducted independently by the researcher and a research assistant. The two-person coding procedure offers increased confidence in the reliability of the results compared with the sole researcher coding of Haidt et al. (2000) and McHugh et al. (2017); however, both coders in this study were aware of the purpose of the study. While reliability techniques were applied (such as calculation of inter-rater reliability, consensus by discussion on all disagreements for the entire sample, and peer-review by a second research assistant for a random selection of interviews), it is possible that blind coding would have yielded different results and, therefore, should be considered for future research.

**Conclusion**

Moral dumbfounding was operationalised in this study according to the way it is most commonly described; that is, as the maintenance of a judgment without provision of supporting reasons (s.c., Astuti & Bloch, 2015; Dwyer, 2009; Gray et al., 2014; Kasachkoff & Saltzstein, 2008; Kennett, 2012; Kennett & Fine, 2009; McHugh et al., 2017; and Railton, 2014). The hypothesis that moral dumbfounding would be elicited in response to the stimuli presented was supported. The prevalence of the phenomenon using this definition (90.9%) seems to correspond with the description reported in Haidt et al.’s (2000) original dumbfounding study (although, as noted, the rate was not quantified by Haidt et al). Other hypotheses derived from the original dumbfounding study were also supported in the current research; specifically, the intuition stimuli elicited significantly more frequency of moral dumbfounding, and judgments of wrong, than the reasoning stimulus. Also as expected, judgments to the intuition stimuli were based significantly more on gut feeling compared to the reasoning stimulus; however, the basis underpinning judgment was self-reported, so the limits of introspection undermine any conclusions. Furthermore, moral dumbfounding was not related to the foundation upon which participants reportedly made their decision (i.e., reasoning or gut feeling). Consequently, rather than offering support for the idea that moral
dumbfounding is evidence that our moral judgments are grounded in how a moral dilemma makes us feel (rather than on a process of reason), it can only be concluded that participants report that the source of their initial response to a moral dilemma is gut feeling. The lack of support for other hypotheses also appear to undermine the broad claims of Haidt et al.

In contrast to expectations, responses to the intuition stimuli were not significantly faster than responses to the reasoning stimulus; thus, the findings of the current study do not support the view that moral judgment is an intuitive automatic reaction devoid of reasoning (even fast responding does not rule out that one arrived quickly at the decision due to repeated use of a similar judgment form in comparable scenarios, i.e., automatized prior judgment). There was also no difference between the intuition and reasoning stimuli in whether judgment or justification was offered first, failing to support the position that moral dumbfounding demonstrates that reasoned justification is always post-hoc and that moral judgment is intuitive (notably, offering an explicit judgment before justification, or vice versa, does not conclusively reveal an internal process of justification preceding judgment). Nonetheless, while the hypothesis that there would be significantly more dumbfounding statements in response to the intuition stimuli than to the reasoning stimulus was not supported, the current findings are broadly compatible with Haidt et al. (2000) insofar as some participants explicitly struggled to provide a reason for judgment. The variable of unwillingness to support (developed inductively during this study) encompassed statements supporting an acceptability judgment for the scenario or task, but a reluctance to commit to such a view. Given the differences found between the majority of intuition stimuli and the reasoning measure (but no such distinction in dumbfounding statements between any of the stimuli), it was noted that this novel measure should be included in future research on moral dumbfounding. Such further research will aid practical understanding of this clearly thought-provoking and important phenomenon.
Chapter 4: Study 2 – Defining Moral Dumbfounding

Study 1 provides evidence that moral dumbfounding is a genuine, replicable phenomenon. While the shortcomings of previous empirical studies (Haidt et al., 2000; Haidt & Hersh, 2001; McHugh et al., 2017; Royzman et al., 2015) were addressed, moral dumbfounding was not operationalised as it was originally defined (being “the stubborn and puzzled maintenance of a judgment without supporting reasons”; Haidt et al., 2000, p. 1). Instead, the phenomenon was operationalised as the maintenance of a judgment without provision of supporting reasons to correspond with the most common description of dumbfounding in the literature (i.e., without the requirement of stubbornness or puzzlement; s.c., Astuti & Bloch, 2015; Dwyer, 2009; Gray et al., 2014; Kasachkoff & Saltzstein, 2008; Kennett, 2012; Kennett & Fine, 2009; McHugh et al., 2017; and Railton, 2014). Still other definitions of moral dumbfounding do exist, and their specific details vary. Moreover, the different definitions have resulted in a discrepancy of the reported prevalence in extant research (i.e., 82% in Haidt & Hersh, 2001; 71% in McHugh et al., 2017; 1.88% in Royzman et al., 2015). In this chapter, an analysis of the overall prevalence of moral dumbfounding according to each definition is undertaken. Furthermore, a determination of the proportion of variance accounted for by proposed important indicators of the phenomenon is undertaken here. The aim of these analyses is to address the question ‘what is moral dumbfounding’ by providing a single empirically supported definition of the phenomenon that enables the identification, and prediction, of its occurrence.

Definitions of moral dumbfounding

Common definition

Gray et al. (2014) describe moral dumbfounding as occurring when “participants are rendered ‘dumb’ to explain their enduring moral judgments” (p. 1). This description, which can be broadly characterised as the maintenance of a judgment without supporting reasons, is widespread throughout the literature (s.c., Astuti & Bloch, 2015; Dwyer, 2009; Kasachkoff & Saltzstein, 2008; Kennett, 2012; Kennett & Fine, 2009; McHugh et al., 2017; and Railton, 2014). Although it is commonly accepted in the literature, this definition is not based on a thorough empirical analysis of the phenomenon of moral dumbfounding and, therefore, it is important to carefully analyse
all of the proposed features of moral dumbfounding in order to produce an empirically defensible definition.

**Original definition**

Some researchers (such as Björklund, 2000; Haidt, 2001; Haidt & Björklund, 2008; Jacobson, 2012; Royzman et al., 2015) operationalise moral dumbfounding according to the initial four components: (a) stubbornness, (b) puzzlement, (c) maintenance of a judgment, (d) without supporting reasons. While only the abstract of Haidt et al.’s (2000) seminal paper mentions stubbornness and puzzlement (the researchers do not specifically operationalise either of these components and merely report behaviour that they assume to be representative of these two states), their original definition of moral dumbfounding comprises all four elements (p. 1). So, too, does the description of the phenomenon given by Haidt for instances in which his participants were presented with the intuition measures:

People almost always start out by saying it’s wrong. Then they start to give reasons … When that reason is stripped from them, they give another reason. When the new reason is stripped from them, they reach for another reason. And it’s only when they reach deep into their pockets for another reason, and come up empty-handed, that they enter the state we call ‘moral dumbfounding’. Because they fully expect to find reasons. They’re surprised when they don’t find reasons. And so in some of the videotapes you can see, they start laughing. But it’s not an ‘it’s so funny’ laugh. It’s more of a nervous-embarrassment puzzled laugh. So it’s a cognitive state where you ‘know’ that something is morally wrong, but you can’t find reasons to justify your belief. Instead of changing your mind about what’s wrong, you just say: ‘I don’t know, I can’t explain it. I just know it’s wrong’. (Haidt, 2005, para 3)

Given the original definition of moral dumbfounding (Haidt et al., 2000), its endorsement by other researchers (such as Björklund, 2000; Haidt & Björklund, 2008; Jacobson, 2012; Royzman et al., 2015), and the way in which Haidt (2001, 2005) promotes the phenomenon, it is interesting that the prevalence according to the original definition is actually yet to be determined. As such, it is an aim of the current study to report the prevalence of moral dumbfounding when defined as the “stubborn and puzzled maintenance of a judgment without supporting reasons” (Haidt et al., 2000, p. 1).

**Judgment without justification definition**

McHugh et al. (2017) argue that stubbornness and puzzlement are not essential elements of moral dumbfounding but, instead, are consequences of upholding a judgment in the absence of reasons; consequently, they operationalise moral dumbfounding as the
maintenance of a judgment without supporting reasons. In computing the prevalence of the phenomenon using this definition, McHugh and colleagues applied criteria for distinguishing between being without a reason, and failing to provide a reason, for a judgment; specifically, they contend that a judgment can be said to be made without supporting reasons if a participant fails to offer supporting rationale other than an unsatisfactory statement (such as an ‘unsupported declaration’). McHugh et al. subsequently reported a dumbfounding estimate of 71%. As the study by McHugh and colleagues is one of the very few replications of moral dumbfounding, further evidence is required before accurate estimates of the rate of moral dumbfounding can be generated.

**Confidence definition**

Another definition of moral dumbfounding that omits any reference to stubbornness and puzzlement, but also maintenance of a judgment, is that of Sneddon (2007): “moral dumbfounding occurs when someone confidently pronounces a moral judgment, then finds that he has little or nothing to say in defense of it” (p. 731). Confidence in judgment was not an element included in the description of dumbfounding as described initially (Haidt et al., 2000) and there was no relationship between confidence in judgment and moral dumbfounding in Study 1 (there was also no difference in participants’ confidence in their judgment across tasks). Nonetheless, as many researchers describe dumbfounding in a manner similar to Sneddon (e.g., Cushman et al., 2010; Fedyk, 2017; Hauser et al., 2007; Hauser et al., 2008; and Pizarro & Bloom, 2003; Russell & Giner-Sorolla, 2011), the prevalence of moral dumbfounding embodying confidence in a judgment will be determined.

**Credulity definition**

In the original moral dumbfounding study (Haidt et al., 2000), the scenarios were designed to be perceived as harmless, and it was stipulated that no harm would result from the actions described. Some researchers (Gray et al., 2014; Huebner, 2011; Jacobson, 2012; Kennett, 2012; Royzman et al., 2015) assert that the apparent failure to provide a reason was an artefact of the measures used. In other words, participants did not believe the stipulation that the actions were harmless, but because the experimenter negated all harm-based reasons, they were left unable to justify their judgment based on the potential for harm. Royzman et al. (2015) conducted a partial replication in which moral dumbfounding was operationalised according to the original definition, but all participants who did not accept that the event was truly and credibly harm-free were
excluded. This procedure left one participant (resulting in a non-significant
dumbfounding rate of 1.88%). Accordingly, Royzman and colleagues contend that true
moral dumbfounding (in which a person resolutely condemns an act without supporting
reasons), is extraordinarily rare (if it exists at all). To test this conclusion, the prevalence
of moral dumbfounding according to the criterion of Royzman et al. will be ascertained.

**Puzzlement definition**

Moral dumbfounding, as described initially, includes the element of puzzlement at an
inability to justify a moral judgment (Haidt et al., 2000). Commonly, however, this
phenomenon is defined without this element and simply as judgment without
justification (e.g., Astuti & Bloch, 2015; Dwyer, 2009; Gray et al., 2014; Kasachkoff &
Saltzstein, 2008; Kennett, 2012; Kennett & Fine, 2009; McHugh et al., 2017; and
Railton, 2014). In Study 1 (reported in Chapter 3), all participants who made a
statement that they were puzzled by their inability to produce a justification for their
judgment (regardless of the presenting dilemma) were independently classified as
dumbfounded. In light of this, moral dumbfounding will be also measured using a
definition encompassing puzzlement at an inability to justify judgment.

**Evidence-based definition**

An evidence-based definition of moral dumbfounding can be derived by bringing
together the findings of Study 1 (reported in Chapter 3) and existing empirical research
(i.e., Haidt et al., 2000; Haidt & Hersh, 2001; McHugh et al., 2017; Royzman et al.,
2015). As stated above, some researchers (such as Björklund, 2000; Haidt, 2001; Haidt
& Björklund, 2008; Jacobson, 2012; Royzman et al., 2015) operationalise moral
dumbfounding according to the initial four components: (a) stubbornness, (b)
puzzlement, (c) maintenance of a judgment, (d) without supporting reasons. First,
stubbornness of judgment is essentially the same as maintenance of a judgment both
theoretically (McHugh et al., 2017) and empirically (i.e., precluding maintenance of
judgment from initial to final judgment leaves only an intermediate judgment; a single
measure is, for obvious reasons, not a valid measure of stubbornness) so this element
should be excluded from an evidence-based definition of the phenomenon. Furthermore,
participants of Study 1 were only more likely to maintain their decision in response to
the *Cannibal* vignette than to the *Heinz* vignette, and Haidt et al. (2000) state that the
rate of participants maintaining their judgment did not vary significantly across
measures. The lack of variation in maintaining judgment across the measures (in which
moral dumbfounding does vary) indicates that maintenance of a judgment should also
be an omitted element. Regarding puzzlement, while all eight puzzled participants in Study 1 were independently classified as dumbfounded, 32 of the 40 participants classified as morally dumbfounded were not puzzled. As puzzlement is a sufficient signifier, but not a necessary condition to moral dumbfounding, it will be omitted from this empirically defensible definition. Finally, the element of belief that the action is harm-free from Royzman et al. (2015) will be included, and the criterion for measuring judgment without justification from McHugh et al. (2017; i.e., failure to offer supporting rationale other than an unsatisfactory statement, such as an ‘unsupported declaration’) will be utilised. Altogether, this provides for a novel definition of moral dumbfounding as ‘an inability to justify disapproval for a seemingly harm-free action’. The prevalence of the phenomenon when employing this definition will be established.

The relative contribution of key indicators of moral dumbfounding

An analysis of the overall prevalence of moral dumbfounding according to various definitions is important to unify future moral dumbfounding literature under an ecologically valid definition. It is also important to analyse the relative contribution of key definitional elements of dumbfounding. This is because different elements of the definitions (formerly analysed as a whole) may explain a disproportionately large amount of variance in the phenomenon (i.e., are of value in predicting its occurrence), or may not contribute significantly to the variance at all (i.e., may be excluded from measurement). Two indicators theorised to be key to moral dumbfounding are a belief that an action is genuinely harm-free, and puzzlement at an inability to justify a judgment.

Belief that an action is harm-free

Royzman et al. (2015) contend that true moral dumbfounding can only occur when one is unable to justify condemnation of an action they believe to be genuinely harm-free. Only one participant was left after the researchers excluded those who did not accept that the event was truly and credibly harm-free (no explanation was offered for this participant’s dumbfounding). Nonetheless, Royzman et al.’s credulity check may have artificially reduced the prevalence of dumbfounding. Consider how we often suspend disbelief when watching movies to engage with implausible storylines. Suppose, for example, one was asked to judge the action of a villain who had detonated explosives in a crowded building. The judgment would almost certainly be that it was wrong; the
reason provided may be an ‘unsupported declaration’ (e.g., “because murder is wrong”). When reminded of the fictional nature of the scenario, and asked again why it is wrong, one might show the classic signs of dumbfounding (i.e., “it just is”). While we may not believe that, realistically, there will be harm, our judgment is based on our awareness that, given the perceived situation, harm is a likely result. Consequently, it may not be important whether or not participants accept the stipulation of lack of harm. Rather than speculating about the role of the credence of the harm-free stipulation, the proportion of variance in moral dumbfounding accounted for by belief that the moral situation upon which we are asked to judge is harm-free, will be explored.

**Puzzlement at an inability to justify**

The omission of puzzlement from the most common description of moral dumbfounding in the literature (s.c., Astuti & Bloch, 2015; Dwyer, 2009; Gray et al., 2014; Kasachkoff & Saltzstein, 2008; Kennett, 2012; Kennett & Fine, 2009; McHugh et al., 2017; and Railton, 2014) is problematic, as the inability to articulate a justification for one’s moral belief does not alone reveal moral dumbfounding. The failure to provide a reason does not, necessarily, equate to a judgment devoid of reason because cognitive laziness cannot be discounted as an explanation for the failure; there may be low motivation for effortful consideration of the rationale or to explain it. On the other hand, one would most likely only become puzzled if an effortful introspective search process has failed to produce a conscious reason for the judgment. Indeed, Haidt infers that true moral dumbfounding only occurs when one is surprised at this inability: “...it’s only when they reach deep into their pocket for another reason, and come up empty-handed, that they enter the state we call ‘moral dumbfounding’. Because they fully expect to find reasons. They’re surprised when they don’t find reasons” (2005, para 3). Given that such bewilderment may hold the key to the measurement of true moral dumbfounding, the amount of variance in moral dumbfounding that can be accounted for by puzzlement will be investigated.

With regard to the moral dumbfounding literature and the results of Study 1 (reported in Chapter 3), it is hypothesised that:

1. There will be significantly greater frequency of moral dumbfounding when the phenomenon is operationalised using more inclusive definitions (i.e., the Common definition, the Judgment Without Justification definition, and the Confidence definition) than more exclusive definitions (i.e., the Original definition, the Credulity definition, the Puzzlement definition, and the Evidence-based definition).
2. The definitional components of belief that an action is genuinely harm-free, and puzzlement at an inability to justify, will account for a significant proportion of unique variance in moral dumbfounding.

Methods

Participants
Studies 1 and 2 employ the same participants and methodology. Refer to the Methods section of Study 1 (p.18) for details.

Sample size and power
Sample size and power were calculated for all statistical tests used in the current study using G*power (version 3.1):

- Friedman two-way ANOVA: At a significance level of 0.05, small effect size (0.4), and 80% power, the required sample size was 22 (this was calculated by adding 15% to the resulting sample size as this is a non-parametric test; Lehmann, 2006).
- Wilcoxon signed ranks test: At a significance level of 0.05, medium effect size = 0.5, and 80% power, the required sample size was 28.
- Multiple regression: At a significance level of 0.05, and 80% power with two predictors, the required sample size was 29.

Thus, there was sufficient sample size and power for statistical analyses in the current study given \( n = 44 \).

Research design
This study is a frequency-based investigation into the various features of moral dumbfounding. The aim of the study was to (a) compare measures of moral dumbfounding to ascertain differences in prevalence, and (b) predict moral dumbfounding using elements theorised to be relevant to the phenomenon. The overarching purpose of the study was to develop an evidence-based definition of the phenomenon and provide measurable indicators. The qualitative and quantitative data collected during Study 1 (using a standardised open-ended interview to replicate the research design of the original moral dumbfounding study; Haidt et al., 2000) were utilised. Refer to the Method section of Study 1 (p. 20) for further information regarding the research design.
**Measures and procedure**

**Belief that the action is harm-free**

As stated in Study 1, participants were asked to complete a short questionnaire after each task comprising items pertaining to how they made their judgment and how they perceived the task (annexed as Appendix D.1). One of the items seeking a response from participants was: ‘I find it difficult to imagine that the nature of the story/task is harm-free’. Responses to this item were recorded on a 5-point rating scale ranging from 1 (strongly disagree) to 5 (strongly agree). This credulity check was asked to empirically test the criticism aimed at Haidt et al. (2000) that participants may not have perceived the scenarios to be harmless just because careful stipulations were included to create the perception of harmlessness (Gray et al., 2014; Jacobson, 2012; Kennett, 2012; Royzman et al., 2015).

The responses of participants to the item, ‘I find it difficult to imagine that the nature of the story/task is harm-free’ were imported into SPSS Statistics for analysis (refer to the Methods section of Study 1 on p. 33 for further information, as this item was downloaded from the GoogleForms collection format with the other End-of-Task Questionnaire responses). The data were inspected for completeness and consistency. There were no missing data. The ‘belief that the action is harm-free’ variable was transformed into a binary measure whereby if a participant’s response was recorded as less than or equal to the relevant mean for that item, they were considered to believe that the story/task was harm-free. A global measure was computed for each participant by averaging their frequency of belief that the depicted action is representative of a situation which may happen in the real world, across stimuli. Higher ratings were taken to indicate a greater belief that the actions undertaken in the scenarios were realistic.

**Definitions of moral dumbfounding**

**Common definition**

The global measure of moral dumbfounding utilised in Study 1 (i.e., the maintenance of a judgment without provision of supporting reasons) was used again in this study. To be classified as dumbfounded according to this definition (hereafter referred to as the Common definition), a participant needed to have a frequency greater than zero for the variables of ‘maintained judgment of wrong’ AND (‘failure to provide a reason’ OR ‘struggle to produce a reason’). A global measure was computed for each participant by aggregating the frequency of moral dumbfounding according to this criterion across stimuli.
Original definition

A variable was computed to measure moral dumbfounding for each stimulus according to the Original definition (i.e., “stubborn and puzzled maintenance of a judgment without supporting reasons”; Haidt et al., 2000, p. 1). To be classified as dumbfounded under this definition, a participant needed to have a frequency greater than zero for the variables of ‘stubborn judgment of wrong’ AND ‘puzzlement at inability to justify’ AND ‘maintained judgment of wrong’ (satisfied by the ‘stubborn judgment of wrong’ criterion) AND (‘failure to provide a reason’ OR ‘struggle to produce a reason’). A global measure was computed for each participant by aggregating the frequency of moral dumbfounding according to this criterion across stimuli.

Judgment without justification definition

The variable of ‘failure to provide a reason’ utilised in Study 1 was used as the basis of a new Judgment Without Justification variable created for the current study. To ascertain if participants who had been identified as failing to provide a reason for a judgment had, in fact, offered any other supporting reason for their judgment (other than the three measures that comprise ‘failure to provide a reason’; i.e., ‘unsupported declarations’, ‘tautological reasons’, or ‘illogical arguments’), a manual search process of each relevant scenario response was undertaken. The researcher and a research assistant independently conducted the coding. The two-person coding procedure resulted in a very high degree of inter-rater reliability between the five stimuli measurements. The average measure intraclass correlation coefficient was 0.997 with a 95% confidence interval from 0.967 to 1.000 ($F(4, 4) = 385.67, p < 0.001$). Further reliability was performed by the two coders collectively reviewing each response to the stimuli that had been independently coded as the participant not providing any other supporting reasons. Any differences in assignment of ‘no further reason’ were resolved through discussion. Finally, a subset of responses ($n = 11$) that had been identified as the participant not providing any other supporting reasons was peer-reviewed by a second research assistant. There was no discrepancy in the coding. A variable of ‘judgment without justification’ was subsequently computed for each participant (0 Provided Justification versus 1 Judgment Without Justification).

A variable was computed to measure moral dumbfounding for each stimulus according to the Judgment Without Justification definition (i.e., offering no other supporting reason for a judgment besides ‘unsupported declarations’ or ‘tautological reasons’; McHugh et al., 2017, or ‘illogical arguments’ inductively coded in Study 1).
To be classified as dumbfounded under this definition, a participant needed to have a frequency greater than zero for the variables of ‘maintained judgment of wrong’ AND (‘judgment without justification’ OR ‘struggle to produce a reason’). A global measure was computed for each participant by aggregating the frequency of moral dumbfounding according to this criterion across stimuli.

Confidence definition
A variable was computed to measure moral dumbfounding for each stimulus according to the Confidence definition (i.e., confidently pronouncing a moral judgment with little or nothing to say in defence of it; Sneddon, 2007, p. 731). To be classified as dumbfounded under this definition, a participant needed to have a frequency greater than zero for the variables of ‘confidence in judgment’ AND (‘failure to provide a reason’ OR ‘struggle to produce a reason’). A global measure was computed for each participant by aggregating the frequency of moral dumbfounding according to this criterion across stimuli.

Credulity definition
A variable was computed to measure moral dumbfounding for each stimulus according to the Credulity definition (i.e., stubborn and puzzled maintenance of a judgment of a harm-free dilemma without supporting reasons; e.g., Rozyman et al., 2015). To be classified as dumbfounded under this definition, a participant needed to have a frequency greater than zero for the variables of ‘stubborn judgment of wrong’ AND ‘puzzlement at inability to justify’ AND ‘maintained judgment of wrong’ (satisfied by the ‘stubborn judgment’ criterion) AND (‘failure to provide a reason’ OR ‘struggle to produce a reason’) AND ‘belief that the action is harm-free’. A global measure was computed for each participant by aggregating their frequency of moral dumbfounding according to this criterion across stimuli.

Puzzlement definition
The variable of ‘puzzlement at an inability to justify judgment’ utilised in Study 1 was used in the current study. A variable was computed to measure moral dumbfounding for each stimulus according to the Puzzlement definition (i.e., “confused inability to explain one’s position”; Haidt & Hersh, 2001, p. 209). To be classified as dumbfounded according to this definition a participant needed to have a frequency greater than zero for the variables of ‘puzzlement at an inability to justify’ AND (‘failure to provide a reason’ OR ‘struggle to produce a reason’). A global measure was computed for each
participant by aggregating the frequency of moral dumbfounding according to this criterion across stimuli.

*Evidence-based definition*

A variable was computed to measure moral dumbfounding for each stimulus according to an evidence-based definition (i.e., an inability to justify disapproval for a seemingly harm-free action; hereafter referred to as the *Evidence-based* definition). To be classified as dumbfounded under this definition, a participant needed to have a frequency greater than zero for the variables of 'belief that the action is harm-free' AND ('judgment without justification' OR 'struggle to produce a reason'). A global measure was computed for each participant by aggregating their frequency of moral dumbfounding according to this criterion across stimuli.

*Data analysis*

A Monte Carlo resampling method was applied to generate a large number of simulated samples (10,000 samples) to yield unbiased estimates about the characteristics of the population from which the sample at hand is drawn (sc., Field, 2009). All variable distributions were checked to assess the appropriateness of parametric tests for analyses.

*Comparison of moral dumbfounding definitions*

Normality (i.e., skewness and kurtosis less than ±1, and significance of the Shapiro-Wilk statistic) and sphericity (i.e., the significance of the Mauchly’s Test of Sphericity) were assessed using SPSS. As SPSS does not provide a test of homogeneity of variance for repeated measures variables, an $F_{\text{max}}$ test was computed by hand using the formula

$$F_{\text{max}} = \frac{\text{Largest Sample Variance}}{\text{Smallest Sample Variance}},$$

where the largest sample variance is simply the largest standard deviation squared, and the smallest sample variance is the smallest standard deviation squared (Allen & Bennett, 2008). Homogeneity of variance can be assumed when $F_{\text{max}}$ is less than 10 (Tabachnick & Fidell, 2001). The assumptions of normality, sphericity, and homogeneity of variance were violated for many of the variables and, therefore, equivalent non-parametric tests were used.

To maintain a family-wise alpha rate of 0.05, a one-tailed significance level was set at a Bonferroni adjusted $\alpha$ of 0.0024 (i.e., a standard level of .05 divided by the 21 planned comparisons between the 7 definitions of moral dumbfounding; sc., Allen & Bennett). Effect sizes are reported where possible, with the effect size formula being

$$r = \frac{\pi}{\sqrt{N}}.$$  

Descriptions of effect size were guided by J. D. Cohen (1988) in which $r = 0.1$
can be considered small, \( r = 0.3 \) can be considered medium, and \( r = 0.5 \) can be considered large.

**The relative contribution of key indicators of moral dumbfounding**

Multiple regression analysis (MRA) was chosen to test the hypothesis that the definitional components of belief that an action is genuinely harm-free, and puzzlement at an inability to justify, will account for a significant proportion of unique variance in moral dumbfounding. The assumptions underlying the multiple regression procedure were checked. Inspection of the normal probability plot of Regression Standardised Residuals was undertaken, which revealed that the assumption of normality was met (the points clustered reasonably tightly along the diagonal line). The intercorrelation between the predictor variables was evaluated and revealed relatively high tolerances for the predictors in the regression model (Tolerance was >0.2 at 0.986 and the variance inflation factor was 1.014) indicating that multicollinearity was not a concern. Multivariate outliers were not of concern as Mahalanobis distance did not exceed the critical \( \chi^2 \) for \( df = 2 \) (at \( \alpha = 0.01 \)) of 13.82 for any cases (the Maximum Mahalanobis Distance was 6.96). Homoscedasticity of residuals and linearity were examined via the Scatterplot of Standardised Residuals Against Standardised Predicted Values; an absence of any clear pattern in the spread of points indicated that the assumptions were met. Consequently, the data were suitable for multiple linear regression to be reliably interpreted.

Hierarchical MRA was chosen because a participant is presented with a moral dilemma (that they either will or will not perceive as harm-free) before possibly experiencing puzzlement at an inability to produce a justification for their judgment of the dilemma. Accordingly, the ‘belief that the action is harm-free’ predictor variable was added to the regression model in the first step, and the ‘puzzlement at an inability to justify’ predictor variable was added to the second step.

Although \( R^2 \) is an adequate index of effect size for multiple regression, to provide a description of effect size, Cohen’s \( f^2 \) was calculated using the formula

\[
f^2 = \frac{R^2}{1-R^2} \quad \text{(Allen & Bennett).}
\]

Descriptions of effect size are guided by J. D. Cohen (1988) in which \( f^2 = 0.02 \) can be considered small, \( f^2 = 0.15 \) can be considered medium, and \( f^2 = 0.35 \) can be considered large.
Results

Comparison of moral dumbfounding definitions

It was hypothesised that there would be significantly greater frequency of moral dumbfounding when the phenomenon is operationalised using more inclusive definitions (i.e., the Common definition, the Judgment Without Justification definition, and the Confidence definition) than more exclusive definitions (i.e., the Original definition, the Credulity definition, the Puzzlement definition, and the Evidence-based definition). Table 19 presents the frequency of participants classified as morally dumbfounded (at least once during the interview) according to each of the various definitions (the mean ranking, z scores, corrections for ties, and effect sizes for significant pairwise comparisons between the various other definitions of moral dumbfounding are presented in tables, whereas the non-significant results are discussed within the text). A Friedman two-way ANOVA ($\alpha = 0.05$) was used to compare the frequency of moral dumbfounding when operationalised according to the seven definitions. The frequency of the phenomenon differed significantly across the various definitions, $\chi^2_p (df = 6, N = 44) = 168.66$ (corrected for ties), $p < 0.001$. The Wilcoxon Signed Rank Test (Bonferroni adjusted $\alpha = 0.0024$) was used for post-hoc comparisons.
Table 19
Frequency of Participants Classified as Morally Dumbfounded (at Least Once) in Response to the Dumbfounding Stimuli, According to Each Dumbfounding Definition

<table>
<thead>
<tr>
<th>Definition</th>
<th>^Satisfy Criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Definition</strong></td>
<td><strong>n</strong></td>
</tr>
<tr>
<td>Common (i.e., the maintenance of a judgment without provision of supporting reasons)</td>
<td>40</td>
</tr>
<tr>
<td>Original (i.e., “stubborn and puzzled maintenance of a judgment without supporting reasons”; Haidt et al., 2000, p. 1)</td>
<td>5</td>
</tr>
<tr>
<td>Judgment Without Justification (i.e., offering no other supporting reason for a judgment besides ‘unsupported declarations’ or ‘tautological reasons’; McHugh et al., 2017, or ‘illogical arguments’)</td>
<td>34</td>
</tr>
<tr>
<td>Confidence (i.e., confidently pronouncing a moral judgment with little or nothing to say in defence of it; Sneddon, 2007, p. 731)</td>
<td>29</td>
</tr>
<tr>
<td>Credulity (i.e., stubborn and puzzled maintenance of a judgment of a harm-free dilemma without supporting reasons; Royzman et al., 2015)</td>
<td>2</td>
</tr>
<tr>
<td>Puzzlement (i.e., “confused inability to explain one’s position”; Haidt &amp; Hersh, 2001, p. 209)</td>
<td>8</td>
</tr>
<tr>
<td>Evidence-based (i.e., an inability to justify disapproval for a seemingly harm-free action)</td>
<td>10</td>
</tr>
</tbody>
</table>

Note. \( n = \) frequency of participants from \( N = 44 \).
\(^{a}\)Number of participants who were classified as morally dumbfounded at least once during the interview.

Inclusive definitions

**Common definition**

The *Common* definition variable ranged from 0 to 4 with a mean of 2.09 (SD = 1.14). A significantly greater number of participants were classified as morally dumbfounded (at some point during the interview) according to the *Common* definition than all other definitions (i.e., both inclusive and exclusive definitions). Table 20 presents the indices of pairwise comparisons. There were extremely large effect sizes for all comparisons except for between the *Common* definition and the *Confidence* definition (but this effect size can still be considered large).
Table 20
Mean Ranking, z Scores, Corrections for Ties, and Effect Sizes of Pairwise Comparisons Between the Common Definition and Other Definitions of Moral Dumbfounding

<table>
<thead>
<tr>
<th>Definition</th>
<th>Mean Rank</th>
<th>z Score</th>
<th>N - Ties</th>
<th>( r^2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Original</td>
<td>2.84</td>
<td>-5.57*</td>
<td>40</td>
<td>0.84</td>
</tr>
<tr>
<td>Judgment Without Justification</td>
<td>5.11</td>
<td>-4.63*</td>
<td>26</td>
<td>0.70</td>
</tr>
<tr>
<td>Confidence</td>
<td>4.92</td>
<td>-3.99*</td>
<td>24</td>
<td>0.60</td>
</tr>
<tr>
<td>Credulity</td>
<td>2.66</td>
<td>-5.56*</td>
<td>40</td>
<td>0.84</td>
</tr>
<tr>
<td>Puzzlement</td>
<td>3.05</td>
<td>-5.50*</td>
<td>39</td>
<td>0.83</td>
</tr>
<tr>
<td>Evidence-based</td>
<td>3.11</td>
<td>-5.47*</td>
<td>40</td>
<td>0.83</td>
</tr>
</tbody>
</table>

Note. \( N = 44 \).
\( *r = J. D. Cohen's (1988) effect size for pairwise comparison. \)
\( *\) All definitions differed significantly from the Common definition at the Bonferroni adjusted \( \alpha \) of 0.0024.

Judgment without justification definition

The Judgment Without Justification variable ranged from 0 to 3 with a mean of 1.25 (\( SD = 0.94 \)). The z scores, corrections for ties, and effect sizes of pairwise comparisons between the Judgment Without Justification definition and the other definitions of moral dumbfounding are presented in Table 20 (refer to Table 19 for the mean ranking of the various definitions). The Judgment Without Justification definition yielded a significantly greater number of morally dumbfounded participants than all except for the Confidence definition (\( p = 0.501 \)). Significant pairwise comparisons achieved extremely large effect sizes.
Table 21
z Scores, Corrections for Ties, and Effect Sizes of Pairwise Comparisons Between the Judgment Without Justification Definition and Other Definitions of Moral Dumbfounding

<table>
<thead>
<tr>
<th>Definition</th>
<th>z Score</th>
<th>N - Ties</th>
<th>r</th>
</tr>
</thead>
<tbody>
<tr>
<td>Original</td>
<td>-5.08*</td>
<td>32</td>
<td>0.77</td>
</tr>
<tr>
<td>Confidence</td>
<td>-0.023</td>
<td>29</td>
<td>na</td>
</tr>
<tr>
<td>Credulity</td>
<td>-5.13*</td>
<td>33</td>
<td>0.77</td>
</tr>
<tr>
<td>Puzzlement</td>
<td>-4.92*</td>
<td>30</td>
<td>0.74</td>
</tr>
<tr>
<td>Evidence-based</td>
<td>-4.93*</td>
<td>32</td>
<td>0.74</td>
</tr>
</tbody>
</table>

Note. N = 44.

*r = J. D. Cohen's (1988) effect size for pairwise comparison.
*Difference significantly from the Judgment Without Justification definition at the Bonferroni adjusted α of 0.0024.

Confidence definition

The Confidence variable ranged from 0 to 4 with a mean of 1.25 (SD = 1.18). The z scores, corrections for ties, and effect sizes of pairwise comparisons between the Confidence definition and the remaining four definitions are presented in Table 22 (again, refer to Table 19 for the mean ranking of the various definitions). There was a significantly greater number of participants classified as morally dumbfounded (at some point during the interview) according to the Confidence definition than each other definition (except for the Judgment Without Justification definition as stated above). All pairwise comparisons achieved large to extremely large effect sizes.
### Table 22

<table>
<thead>
<tr>
<th>Definition</th>
<th>Pairwise Comparison with the Confidence Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>z Score</td>
</tr>
<tr>
<td>Original</td>
<td>-4.45*</td>
</tr>
<tr>
<td>Credulity</td>
<td>-4.70*</td>
</tr>
<tr>
<td>Puzzlement</td>
<td>-4.21*</td>
</tr>
<tr>
<td>Evidence-based</td>
<td>-4.18*</td>
</tr>
</tbody>
</table>

*Note. N = 44.

\(^a_r\) = J. D. Cohen's (1988) effect size for pairwise comparison.

*Differed significantly from the Confidence definition at the Bonferroni adjusted \(\alpha\) of 0.0024.

### Exclusive definitions

**Original definition**

The *Original* definition ranged from 0 to 1 with a mean of 0.11 (SD = 0.32). The frequency of participants classified as morally dumbfounded according to this definition did not differ significantly to the frequency classified according to any of the other exclusive definitions (*Puzzlement, p = 0.126; Credulity, p = 0.131; Evidence-based, p = 0.047*).

**Credulity definition**

The *Credulity* definition ranged from 0 to 1 with a mean of 0.05 (SD = 0.21). The frequency of moral dumbfounding according to this definition did not differ significantly to that of the *Puzzlement* definition (\(p = 0.017\)), nor the *Evidence-based* definition (\(p = 0.004\)).

**Puzzlement definition**

The *Puzzlement* definition ranged from 0 to 1 with a mean of 0.18 (SD = 0.39). The frequency of moral dumbfounding according to this definition did not differ significantly that of the *Evidence-based* definition (\(p = 0.132\)).

**Evidence-based definition**

The *Evidence-based* definition ranged from 0 to 2 with a mean of 0.30 (SD = 0.59).
The relative contribution of key indicators of moral dumbfounding

A belief that an action is genuinely harm-free (which ranged from 0 to 1 with a mean of 0.58; \( SD = 0.32 \)) was hypothesised to account for a significant proportion of unique variance in moral dumbfounding (as commonly described in the literature). It was further hypothesised that puzzlement at an inability to produce justification for judgment would account for a significant proportion of unique variance in moral dumbfounding beyond that already accounted for by a belief that an action is truly harm-free. In combination, the two predictor variables explained a statistically significant 13.9\% of the variance in moral dumbfounding, \( F(1, 42) = 3.32, p = 0.046 \). By J. D. Cohen’s (1988) conventions, a combined effect of this magnitude can be considered medium (\( f^2 = 0.16 \)).

The step 1 variable of ‘belief that the action is harm-free’ emerged as the only significant unique predictor of moral dumbfounding, accounting for 12.7\% of the variability (\( R^2 = 0.127 \)). This proportion of variance was statistically significant, \( F(1, 42) = 6.13, p = 0.017 \). When ‘puzzlement at an inability to justify’ was entered into the regression model on step 2, \( R^2 \) increased to 0.139; thus, puzzlement at an inability to justify one’s judgment accounts for an additional 1.2\% of the variability in moral dumbfounding beyond the variance already accounted for by a belief that the action is harm-free. This incremental increase in \( R^2 \) on step 2 was, however, non-significant, \( \Delta F(1, 41) = 0.568, p = 0.455 \). Unstandardised beta (\( B \)) and standardised beta (\( \beta \)) regression coefficients, and squared semi-partial correlations (\( sr^2 \)) of each predictor, on each step of the hierarchical regression model, are reported in Table 23.

Table 23
Unstandardised and Standardised Beta Regression Coefficients, and Squared Semi-Partial Correlations, of Each Predictor Variable, on Each Step of a Hierarchical Multiple Regression Model Predicting Moral Dumbfounding

<table>
<thead>
<tr>
<th>Variable</th>
<th>( B )</th>
<th>( \beta )</th>
<th>( sr^2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Step 1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Belief that the Action is Harm-Free</td>
<td>1.290*</td>
<td>.357</td>
<td>0.127</td>
</tr>
<tr>
<td>Step 2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Belief that the Action is Harm-Free</td>
<td>1.336*</td>
<td>.370</td>
<td>0.135</td>
</tr>
<tr>
<td>Puzzlement at Inability to Justify</td>
<td>1.603</td>
<td>.110</td>
<td>0.012</td>
</tr>
</tbody>
</table>

Note. \( N = 44 \). \( B \) = unstandardised beta regression coefficient; \( \beta \) = standardised beta regression coefficient; \( sr^2 \) = squared semi-partial correlation.

*Significant at \( p < 0.05 \).
Discussion

Comparison of moral dumbfounding definitions

Although definitions of moral dumbfounding tend to include the notion of making a moral judgment without justification, the specific features included in each definition vary widely. Moreover, the reported rates of observed dumbfounding depend upon how it is operationalised, which leads to a discrepancy in empirical findings. This study is the first systematic quantification of the difference in prevalence of moral dumbfounding when applying the various definitions. It was expected that more inclusive definitions (i.e., Common, Judgment Without Justification, and Confidence definitions) would yield a significantly higher prevalence of moral dumbfounding than more exclusive definitions (i.e., Original, Credulity, Puzzlement, and Evidence-based definitions). This hypothesis was supported; the three inclusive definitions resulted in a significantly higher prevalence of moral dumbfounding than the four exclusive definitions, and all significant pairwise comparisons yielded large to very large effect sizes.

The Common definition is the most inclusive definition as there only needs to be perseverance in a moral judgment without justification to satisfy this criterion. This definition yielded a significantly higher prevalence of moral dumbfounding than all other definitions (90.9%). The ecological validity of measuring moral dumbfounding using this definition is questionable given the extremely high obtained frequency. Our lived experience is that we can provide reasons for our judgments most of the time, with very infrequent moral dumbfounding. Furthermore, merely offering a judgment without giving a supporting reason does not constitute true moral dumbfounding given there are many reasons why this may occur (i.e., cognitive laziness, being provided with insufficient time, etc.); thus, using the Common definition may not accurately reflect the true rate of moral dumbfounding.

A substantial number of participants were also classified as morally dumbfounded under the Judgment Without Justification definition. The Judgment Without Justification definition was modelled on the operationalisation of moral dumbfounding of McHugh et al. (i.e., offering no other supporting reason for a judgment besides ‘unsupported declarations’ or ‘tautological reasons’; 2017, or ‘illogical arguments’ inductively coded in Study 1). The prevalence obtained under this classification in the current study (77.3%) is akin to that obtained by McHugh et al., 2017 (71%) indicating accuracy of measurement.
The Confidence definition, too, yielded a substantial frequency of moral dumbfounding. The Confidence definition was derived from Sneddon (2007): “moral dumbfounding occurs when someone confidently pronounces a moral judgment, then finds that he has little or nothing to say in defense of it” (p. 731). Although many researchers describe dumbfounding in a manner similar to Sneddon (e.g., Cushman et al., 2010; Fedyk, 2017; Hauser et al., 2007; Hauser et al., 2008; Pizarro & Bloom, 2003; Russell & Giner-Sorolla, 2011), the current study is the first to determine the prevalence of moral dumbfounding embodying confidence in a judgment (65.9%).

The Judgment Without Justification and Confidence definitions did not differ from each other. Consequently, these inclusive definitions can be considered equivalent and characterised under a unified definition of ‘confidently pronouncing a moral judgment without supporting reasons’. Again, however, the relatively high prevalences obtained using this definition may be a misleading indication of the true rate of moral dumbfounding. On the other hand, each of the exclusive definitions indicated greater ecological validity.

The exclusive definitions (Original, Credulity, Puzzlement, and Evidence-based) produced far less prevalence of moral dumbfounding than the inclusive definitions (11.4%, 4.5%, 18.2%, and 22.7%, respectively). In addition, these four exclusive definitions did not differ from each other; as such, they can be considered equivalent measures and characterised under a unified definition as ‘puzzlement at an inability to provide supporting reasons for condemnation of an action which is perceived to be genuinely harm-free’. Two key definitional components of ‘belief in the nature of an action as genuinely harm-free’ and ‘puzzlement at an inability to provide supporting reasons for condemnation’ were then analysed to ascertain their relative contribution to moral dumbfounding.

**The Relative Contribution of Key Indicators of Moral Dumbfounding**

This study is the first to explore relationships between proposed elements of moral dumbfounding with the goal of developing a meaningful (ecologically valid) definition of this phenomenon and provide measurable indicators (optimally to predicting it in the general population). Drawing on the empirical findings of moral dumbfounding research (i.e., Haidt et al., 2000; Haidt & Hersh, 2001; McHugh et al., 2017; Royzman et al., 2015), it was hypothesised that belief in the nature of an action as genuinely harm-free, and puzzlement at an inability to produce a justification for one’s moral
judgment, would explain a significant proportion of the variance in moral
dumbfounding. This hypothesis was partially supported.

The belief that an action is harm-free accounted for a significant proportion of
unique variance in moral dumbfounding. This result supports the finding of Royzman et
al. (2015) that such a belief is important to moral dumbfounding; yet, it cannot be taken
to support their position that true moral dumbfounding can only occur when this
condition is satisfied. While 12.7% of the variability in moral dumbfounding was
accounted for by such a belief, 86.1% of the variance of the phenomenon was left
unexplained (aside from the 1.2% variability accounted for by puzzlement). It may be
that other aspects of an action (such as the degree of realism of the scenario, the
medium of presentation, or characteristics of the judge (such as age, gender, personality
traits) explain a proportion of the remaining variance. Further investigation is required
in this regard. While it cannot be said that belief in a lack of consequential harm wholly
explains the phenomenon, such a belief partially explains moral dumbfounding and
should be included in its operationalisation for future investigations.

Beyond that accounted for by the belief that an action is harm-free, puzzlement
did not explain variance in moral dumbfounding. While puzzlement has value as a clear
signal of dumbfounding (given all participants who made a puzzlement statement,
regardless of the presenting dilemma, were independently classified as dumbfounded),
the lack of unique variance confirms that puzzlement is not a necessary indicator of the
phenomenon (not all those classified as dumbfounded explicitly stated puzzlement at
their inability to justify). As such, puzzlement is not specifically included in the final
empirically derived definition of moral dumbfounding, being ‘an inability to justify
disapproval for a seemingly harm-free action’.

While the definition derived in the present study includes an inability to provide
supporting reasons for a judgment, this inability cannot be taken as supporting Haidt’s
(2001) inference that participants will never be able to offer a reason for their judgment
(because the judgment is not based on reason). It could be that if they were given ample
time to reflect, or were in different circumstances where reasons offered were not
challenged, they would be able to provide a justification for the judgement. Moreover,
even if participants are never able to offer a reason, this does not mean that the notion of
ourselves as rational moral agents is under threat (s.c., Kennett & Fine, 2009). If moral
judgment is informed by an unconscious habitual response (due to repeated exposure to
comparable scenarios in which prior reasoning has become automatised), it is
reasonable that participants would not have direct access to the specific rationale upon which their judgment is based (Dwyer, 2009; Pizarro & Bloom, 2003). Thus, even if we are consistently unable to provide reasons in these cases, this does not mean that, more generally, reasoning never plays a role in moral judgments. The results of this study merely show that an inability to justify disapproval for a seemingly harm-free action is currently a suitable description of empirically observed moral dumbfounding.

Limitations
Evidence that moral dumbfounding is a genuine, replicable phenomenon has been provided in this study using the original moral dumbfounding measures (Haidt et al., 2000). These measures have, however, been critiqued as unrealistic and extreme and, indeed, possibly the reason for the phenomenon (Jacobson, 2012; Kennett, 2012; Royzman et al., 2015). It may be the case that moral dumbfounding can only be elicited in response to the set of stimuli used here. Further research is essential to draw generalisations from these scenarios to more typical moral cases.

Conclusion
Moral dumbfounding, a thought-provoking and influential phenomenon, is described in various ways in the psychological and philosophical literature. In the present study, the various definitions, and the elements of dumbfounding they encompass, were analysed to make a case for a single, empirically-derived definition of the phenomenon that provides reliable signifiers of moral dumbfounding. While puzzlement can be taken as a sufficient (but not necessary) indicator of moral dumbfounding, an articulation of such bewilderment is not necessary to the prediction or measurement of its occurrence. At present, moral dumbfounding can be understood as an inability to justify disapproval for a seemingly harm-free action. Using this definition, a moral dumbfounding estimate of 22.7% was obtained. This prevalence has greater ecological validity than previous research (i.e., 90.9% obtained in Study 1; 71% in McHugh et al., 2017; and the observations reported in Haidt et al., 2000) as it is more representative of the phenomenon as it is typically experienced; that is, we can provide reasons for our judgments most of the time, with infrequent moral dumbfounding (rather than the other way around). With empirical evidence of moral dumbfounding and a greater understanding of the phenomenon at hand, the next obvious move is to determine whether this phenomenon can be elicited using alternative moral stimuli that vary from the original measures in features such as the degree of realism of the scenario, the medium of presentation, and the moral domain from which the scenario is drawn.
Chapter 5: Study 3 – Investigating the Moral Dumbfounding Measures

Moral dumbfounding has been confirmed as a genuine, replicable phenomenon (Study 1) and defined as an inability to justify disapproval for a seemingly harm-free action (Study 2). Some researchers (Jacobson, 2012; Kennett, 2012) may dismiss these findings as a reflection of the unrealistic stimuli employed to elicit the phenomenon. They argue that the participants of Haidt et al. (2000; and, by extension, participants of the Study 1 replication) were presented with situations that they are unlikely to ever encounter in real life and were most likely dumbfounded by what they were asked to consider rather than by their own judgments. Empirical research conducted after the original study (Haidt & Hersh, 2001; McHugh et al., 2017; Royzman et al., 2015) has employed only the original dumbfounding vignettes and, therefore, it may be the case that moral dumbfounding is only elicited in response to a peculiar set of stimuli. The aim of Study 3 is to see if moral dumbfounding only occurs under particular conditions. If moral dumbfounding can be elicited using alternative stimuli, it will provide a stronger case for the notion that dumbfounding is a genuine phenomenon that is likely to occur in situations one may encounter in everyday life.

Unrealistic scenarios

In Study 1, participants explicitly stated that they were struggling to provide a reason for their judgment significantly more in response to two of the four intuition stimuli (i.e., Incest and Soul) than to the reasoning measure. Haidt et al. (2000) would likely explain this difference in prevalence of moral dumbfounding as the intuition measures producing a quick emotional response and the reasoning measure (i.e., the Heinz dilemma) eliciting reasoning-based judgment. Some theorists (Jacobson, 2012; Kennett, 2012) argue, however, that moral dumbfounding in these cases is an artefact of the stimuli used. In particular, harm is not successfully removed from the scenarios, the measures elicit are third- versus actual transgressions, the intuition stimuli depict extreme and unfamiliar situations that would not have been considered prior to presentation in the experimental context whereas the Heinz dilemma reflects a common concern (somewhat equivalent to stealing a loaf of bread to feed your family), and. In the present study, depictions of more realistic scenarios are used to determine whether moral dumbfounding is merely a product of the ‘intuition’ stimuli.
Harm
The original intuition stimuli (Haidt et al., 2000) were intended to elicit quick emotional condemnation but be resistant to harm-based reasoning because no actual harm resulted from the depicted action. It is not, however, unreasonable for participants to impute harm in these cases. Taking the Incest vignette as an example, participants are asked to believe that Julie and Mark risk widespread contempt if their violation of the social taboo becomes known, just because they think having sex might be “interesting and fun” or at least “a new experience” (Haidt et al., 2000, p. 18; as cited in Jacobson, 2012). A similar criticism befalls Cannibal in that Jennifer risks her moral identity by abandoning her vegetarian principles and defiling a corpse simply because she thinks it is “irrational to waste perfectly edible meat” (Haidt et al., p. 18; as cited in Jacobson, p. 307). While harm had been artificially removed from these scenarios (by stipulation), the balance of probable costs versus benefits leads to the conclusion that the protagonists’ actions are “foolish and reckless” (Jacobson, p. 300); therefore, having no harm-based reason is not the same as having no reason at all. The alternative scenarios used here involve different levels of risk to the intuition stimuli and, hence, stipulations of harmlessness may be more believable.

Hypothetical versus actual transgressions
The Incest and Cannibal scenarios are typical of ‘dumbfounding’ stimuli with the type of measure employed in extant research almost wholly restricted to hypothetical vignettes (except for the Roach and Soul tasks used by Haidt et al., 2000). Hypothetical scenarios require participants to rely on mental simulation of the situation in the absence of contextual features (such as auditory-visual cues) that are typically accessible to people during real-life dilemmas (McCurrie, Crone, Bigelow, & Latham, 2018). For the present study, a series of tasks were incorporated to elicit moral judgments of actions that participants had performed (real, rather than hypothetical, actions) or witnessed (third-person judgments of actual behaviour).

First-person behavioural tasks
Moral action is profoundly influenced by tangible costs and benefits (FeldmanHall et al., 2012). In Soul, participants are asked to sign a non-legally binding ‘contract’ (which they could rip up immediately after signing) entitling the experimenter to own their soul in exchange for $2 (Haidt et al., 2000, and Study 1). Again, Jacobson suggests that the balance of probable costs and benefits of signing the Soul ‘contract’ leads to the conclusion that it would be wrong (“Speaking for myself, I would not sign some
manipulative experimenter’s bogus contract for my soul for $2, even though for $2000 I would laugh all the way to the bank” (2012, p. 313). It is difficult, however, to design an alternative dilemma to the Soul task because of its salient value.

It is a deep fact about human nature that we are symbol-making creatures, and that we imbue these symbols with meaning … Even pretending to sell your soul to someone is a symbolic act of subjection. The fact that the contract isn’t binding does not erase this symbolism. (Jacobson, 2012, pp. 313-314)

Although the effect of this symbolism is most likely enhanced for participants who hold religious beliefs (Weisbuch-Remington, Berry Mendes, Seery, & Blascovich, 2005), it also applies to those who do not subscribe to an idea of possessing a soul. Not all superstitions are based in religion, and non-religious people also subscribe to superstitious notions (Vyse, 2014). The results of Study 1 highlight the suitability of the Soul task as a stimulus to elicit dumbfounding. The Soul task was the only measure not to differ significantly from the reasoning stimulus in the foundation underpinning initial and final judgments; participants initially made both initial and final judgments based on gut feeling. Participants were confident in their decision, committed to their justifications (offering a smaller number of reasons than they did in defending their judgment to the intuition stories), and appeared to have less difficulty in articulating their defence (there was far less speech dysfluency during responses to Soul compared to all other measures). Notwithstanding that their justifications were delivered with conviction, the participants stated “I don’t know” more often, offered more unsupported declarations and illogical arguments, and made more statements admitting that they should find the task acceptable, but were unwilling to commit to such a judgment (compared to the reasoning measure). It appeared that this struggle to produce a reason for their decision made the participants comparatively more confused and psychically uncomfortable (than in response to the reasoning stimulus). Consequently, the tendency for participants to become morally dumbfounded was significantly higher on the Soul task than to the Heinz dilemma.

Given that the Soul task appears to be a particularly valuable measure to use when investigating moral dumbfounding, the original task was retained. To determine whether Jacobson’s (2012) assertion has broad application, participants who refuse to sign the contract will be asked: “Is there any amount of money that I could pay you to get you to sign the contract?” While it is not expected that the prevalence of moral dumbfounding in response to the Soul task will be influenced by this question (given
that the additional question will be asked after participants provide their final judgment), responses will provide useful information for future research.

In contrast to Soul, it is easier to design an alternative to the Roach task. In Roach, the participant is offered a sip of water into which they witness a sterilised cockroach being submerged and then removed (Haidt et al., 2000, and Study 1). If the participant refuses to drink the water on the grounds that it is disgusting, they are reminded that the sterilised cockroach does not carry germs and, thus, there is no harm (note, an assertion that something is disgusting is not necessarily the same as saying that it is morally wrong or harmful). While it is reasonable to assume that, at some stage, we may use a cup that a cockroach has run across without our knowledge, intentionally drinking ‘roach juice’ is unlikely to occur in the real world. On the other hand, most people have heard of the ‘five-second rule’ in which eating a piece of food that has been dropped on the ground is ‘permitted’ provided it is quickly retrieved. A task in which this ‘rule’ is violated (but the food remains edible) would retain the aversion toward eating undesirable food but present a more realistic behavioural dilemma than the Roach task. It is likely that the more realistic nature of this comparative behavioural task will reduce the prevalence of dumbfounding compared to the original Roach task.

It is reasonable to assume that behavioural tasks will be perceived as more realistic than hypothetical dilemmas and less moral dumbfounding will be elicited by the behavioural tasks (compared with the vignettes) given actual transgressions generate moral judgments that are more congruent with moral behaviour in real situations (Feldman-Hall et al., 2012). The behavioural tasks, however, elicit first-person judgments whereas the vignettes elicit third-person judgments. As previously stated, being asked about another’s moral choices is different from being asked about one’s own (Kennett, 2012). As such, in addition to first-person scenarios, visual depictions of harmless taboo violations are required to elicit third-person moral judgments in response to witnessed actions.

**Third-person witnessed transgressions**

For an action to be judged as immoral, there needs to be the perception of an agent causing harm to a victim (Gray, Waytz, & Young, 2012; Gray, Young, & Waytz, 2012). The victim does not need the capacity to experience suffering for the action to be perceived as wrong; sentence is not a requirement of the right to hold moral status (Torrance, 2013). Consider, for example, how forests, oceans, and other ecosystems cannot experience pain yet there is a sense of obligation for their treatment not only in
the interests of those that inhabit the areas, but because environments have moral interests in their own right (Curry, 2011). Moral engagement with ontological entities that were not formerly considered to hold intrinsic value demonstrates the expansion of the moral circle of concern over time (Crimston, Hornsey, Bain, & Brock, 2018; Laham, 2009). The ongoing expansion of moral concern relative to non-sentient aspects of the natural world is extended to designed features of our environment. Robots increasingly resemble living beings; indeed, humanoid robots have already been awarded citizenship status (Abbass, 2017). Importantly, mistreatment of a robot can elicit a negative emotional response in observers (Parke, 2015). That such a response is occasioned makes visual depictions of the mistreatment of a robot an ideal stimulus to elicit moral dumbfounding as such actions would be expected to evoke swift moral condemnation but be resistant to harm-based justifications. Two videos were selected for the present study that depict scenarios comparable to Haidt et al.’s (2000) ‘harmless’ yet emotionally arousing measures. The videos show the realistic movement abilities of Atlas, a humanoid robot, and Spot, a quadruped robot ‘dog’ (Boston Dynamics, February 2, 2016, February 9, 2015).

The Atlas and Spot videos address the criticism of Haidt et al.’s (2000) measures as unrealistic (Kennett, 2012; Royzman et al., 2015) as the videos are recordings of a real-life event. Furthermore, the depiction of the actions of robots (which is modern technology) means that, most likely, participants would be unfamiliar with the footage. The benefit of this is two-fold: the stimuli will elicit a fresh response, and participants cannot rely upon prior knowledge and established social opinions when making their judgments and justifications. It is expected that the videos will elicit significantly more moral dumbfounding than the vignettes given hypothetical transgressions are often judged to be more wrong, and more independent of rules, than matched witnessed violations (Smetana et al., 1999; Turmel, 2002). The more substantial negative value attributed to imagined immoral behaviour may be due to participants relying on their basic duty of being a moral agent rather than tangible aspects of the depicted action (Tversky & Kahneman, 1974, 1981). Consequently, responses to Atlas and Spot may inform whether moral judgment is an intuitive automatic reaction (i.e., devoid of reasoning; Haidt, 2001; Haidt et al., 2000; Haidt & Hersh, 2001; McHugh et al., 2017) as opposed to unconscious automated prior reasoning (Dwyer, 2009; Pizarro & Bloom, 2003).
With regard to the moral dumbfounding literature and the results of Study 1 (reported in Chapter 3), the following within-subject hypotheses are offered:

1. Moral dumbfounding (defined as ‘an inability to justify disapproval for a seemingly harm-free action’) will be elicited in response to the stimuli presented.
2. The intuition stimuli will elicit significantly more moral dumbfounding than the reasoning stimulus.
3. The vignettes will elicit moral dumbfounding with significantly more frequency than the behavioural tasks and videos.
4. The videos will be perceived as significantly more realistic than the vignettes and behavioural tasks.
5. There will be a negative correlation between belief that the scenario presented is realistic and moral dumbfounding.
6. The intuition stimuli will be judged as wrong with significantly more frequency than the reasoning stimulus.
7. Participants will respond significantly faster when presented with the intuition stimuli than the reasoning stimulus.
8. The intuition stimuli will elicit judgment prior to justification, whereas the reasoning stimulus will elicit justification prior to judgment.
9. Participants will report basing their judgments significantly more on gut feeling in response to the intuition stimuli than to the reasoning stimulus.
10. There will be significantly more psychic discomfort in response to the intuition stimuli than to the reasoning stimulus.
11. There will be significantly more statements of “I don’t know” in response to the intuition stimuli than to the reasoning stimulus.
12. There will be significantly more failure to provide a reason (i.e., unsupported declarations, tautological reasons, illogical arguments, and these variables combined) in response to the intuition stimuli than to the reasoning stimulus.
13. There will be significantly more struggle to produce a reason (i.e., dumbfounding statements, statements of unwillingness to support, and these variables combined) in response to the intuition stimuli than to the reasoning stimulus.
14. There will be significantly more unique justifications in response to the intuition stimuli than to the reasoning stimulus.

Furthermore, the following between-subjects hypotheses are offered:
15. The stimuli presented will be perceived as significantly more realistic than the original measures of Haidt et al. (2000) employed in Study 1.

16. The stimuli presented will elicit significantly less moral dumbfounding than the original measures of Haidt et al. (2000) employed in Study 1.

**Methods**

The data for all four studies undertaken and reported in this dissertation were collected simultaneously. Participants ($N = 88$) were randomly allocated to Study 1 (to receive the original measures of Haidt et al., 2000; $n = 44$) or the current study (to receive the alternative measures; $n = 44$). There were no age or gender differences detected between groups (age: $t (86) = 0.988, p = 0.326$; gender: $\chi^2 (df = 1, N = 88) = 0.051, p = 0.821$).

**Participants**

The sample for the current study consisted of Australian university students ranging in age between 18 and 57 years ($M = 25.75, SD = 9.53$). The male participants ($n = 14$) were aged between 19 and 44 years ($M = 23.79, SD = 6.66$). The female participants ($n = 30$) were aged between 18 and 57 years ($M = 26.67, SD = 10.59$). As an inferential analysis between demographic variables and moral dumbfounding was not conducted in the current study, full descriptive demographic statistics (for participants of Studies 1 and 3) and inferential analyses (for the full sample; $N = 88$) are presented in Chapter 6 (Study 4).

**Sampling procedure**

Participants in the current study were recruited at the same time and in the same manner as Study 1 participants. Participants were randomly allocated to Studies 1 or 3 using Random.org List Randomiser (https://www.random.org/lists/). Refer to the Methods section of Study 1 (reported in Chapter 3, see pp. 19-20) for information regarding sampling procedure and ethical considerations.

**Sample size and power**

Sample size and power were calculated for all statistical tests used in the current study using G*power (version 3.1):

- Cochran’s q - At a significance level of 0.05, odds ratio = 4, and 80% power, the required sample size was 29. As Cochran’s q is an extension of McNemar’s test, the sample size calculation for this test was used.
• Friedman two-way ANOVA: At a significance level of 0.05, small effect size (0.4), and 80% power, the required sample size was 22 (this was calculated by adding 15% to the resulting sample size as this is a non-parametric test; Lehmann, 2006).
• Wilcoxon signed ranks test: At a significance level of 0.05, medium effect size = 0.5, and 80% power, the required sample size was 28.
• Kendall’s tau-b: At a significance level of 0.05, correlation of .5, and 80% power, the required sample size was 35.
• Mann-Whitney U test: At a significance level of 0.05, medium effect size = .6, and 80% power, the required sample size was 37.
• Fisher’s exact test: At a significance level of 0.05, proportion 1 = .05, proportion 2 = .4, and 80% power, the required sample size was 40.
Thus, there was sufficient sample size and power for the statistical analyses in the current study given $n = 44$.

Research design

The current study replicated the research design of the original moral dumbfounding study (Haidt et al., 2000). Like Study 1, both qualitative and quantitative data were collected in a standardised open-ended interview. Refer to the Methods section of Study 1 (reported in Chapter 3, see p. 20) for details regarding the research design.

Apparatus and materials

Apparatus
The same apparatus was utilised in the current study that was used in Study 1. The equipment included hardware and software for vignette recording, stimuli presentation, and participant response recording. Refer to the Methods section of Study 1 (reported in Chapter 3, see p. 21) for details regarding the apparatus.

Materials
Tasks included in the current study were distinct but comparable to those employed by Haidt et al. (2000, pp. 18-19) and in Study 1. To ensure that the three alternative vignettes were presented identically to each participant, they were video recorded by the researcher who was seated in front of a black screen and read each of the stories verbatim from a teleprompter. Each recording began with the words, “Please let me know if the following story is familiar to you”. Refer to the Methods section of Study 1
(reported in Chapter 3, see p. 22) for further information regarding preparation of the materials.

Reasoning vignette

The researcher designed an alternative vignette to the Heinz dilemma employed in Study 1 (developed initially by Kohlberg, 1969; annexed as Appendix C1). Like the Heinz dilemma, the John story (annexed as Appendix F.1) is thought to elicit impassive consideration (i.e., reasoning) between prototypical concerns of harm and rights, and depicts a man stealing (for his terminally-ill wife) a life-saving but cost-prohibitive drug. To provide an updated measure with relevance for Australian participants, the name of the protagonist was modified, he stole the drug from the warehouse of a pharmaceutical company rather than from a sole pharmacist, and details of the subsidy application process for pharmaceuticals in Australia was provided for context.

Intuition vignettes

As in Study 1, all intuition stimuli (that represent taboo violations) were intended to elicit quick emotional condemnation but to be resistant to harm-based justifications. The researcher designed an alternative vignette to the Incest story which was employed in Study 1 (Haidt et al., 2000, p. 18; annexed as Appendix C.2) and is referred to as the Masturbation story. This scenario expands on the basis of stimulus used by Haidt and Hersh in which "A 25-year-old man likes to masturbate while his dog willingly licks his owner’s genitals and seems to enjoy it" (2001, p. 197). In the current study, the man allows his dog to aid his masturbation once but decides to never to let it happen again (annexed as Appendix F.2). Masturbation was designed to retain the abhorrence toward violating a sexual taboo but nullify criticisms of the original story; specifically, that in spite of stipulations, Incest is associated with a small risk of harm (i.e., extremely unlikely but possible pregnancy, and the capacity of both parties to tell the secret; Jacobson, 2012).

The researcher also designed an alternative to the Cannibal vignette employed in Study 1 (Haidt et al., 2000, p. 18; annexed as Appendix C.3). The alternative Liver story (annexed as Appendix F.3) depicts a non-vegetarian theatre nurse who can not afford the meat she needs to increase her dangerously low iron levels. The nurse takes home (and cooks) a piece of liver from the partial hepatectomy of a live donor, but which was unable to be used and is now considered waste. The Liver story was designed to retain the abhorrence toward eating human meat, but nullify criticisms of the original story;
specifically, the abandonment of vegetarian principles on a whim, and desecration of a corpse (Jacobson, 2012).

**Intuition behavioural tasks**

The researcher designed an alternative to the Roach task employed in Study 1 (developed initially by Rozin et al., 1986; annexed as Appendix C.4). In the alternative Lolly task, participants are asked to eat a lolly that has been dropped on a clean carpeted floor (annexed as Appendix F.4). The Lolly task was designed to retain the aversion toward eating undesirable food, but to provide a more realistic behavioural dilemma (eating food that has been dropped on the ground is arguably more common and, therefore, more realistic than knowingly drinking roach juice). A square of new carpet (1.5m x 1.5m) was purchased from a flooring shop, shampooed using a Vax Rapide Ultimate upright carpet cleaner, and treated with ultraviolet radiation by exposure to sunlight for six hours to eliminate any remaining pathogens. Once sterilised, the carpet was placed on the floor of the research laboratory room in a zero-traffic area. A sign was placed on the carpet prohibiting people from walking on the carpet (this sign was removed before participant arrival at each testing session so that attention was not drawn to the carpet). During the experiment, the researcher wore gloves to open and handle the individually wrapped food (to prevent bacterial transfer from the experimenter’s hands to the food).

The Soul task employed in Study 1 (Haidt et al., 2000, p. 19; annexed as Appendix C.5) was again employed in the current study due to the difficulty in designing a suitable substitute. Upon completion of this task, however, participants in the present study who did not sign the contract were asked: “Is there any amount of money that I could pay you to get you to sign the contract?” (open-ended response; annexed as Appendix F.5). This question was asked to determine whether Jacobson’s (2012, p. 313) assertion, that he would sign the contract for $2000 but not for $2, is more broadly applicable.

**Intuition videos**

Two novel intuition stimuli were included here as each involves a scenario which is comparable to Haidt et al.’s (2000) ‘harmless’ yet emotionally arousing measures (i.e., they show an unprovoked violent action toward a robot); consequently, they were expected to elicit quick moral condemnation but be resistant to harm-based justifications. The first video is referred to as Atlas (Boston Dynamics, February 2, 2016; still photographs of the video are annexed as Appendix F.6). The edited Atlas
video depicts movements of a life-size, bipedal, humanoid robot operating both inside buildings and outdoors. In the final scene, a man is seen pushing Atlas in the back with a pole and Atlas falls to the ground.

The second video is referred to as Spot (Boston Dynamics, February 9, 2015; still photographs of the video are annexed as Appendix F.7). The edited Spot video depicts the movements of a life-size, quadruped, animalian robot ‘dog’ operating both inside buildings and outdoors. In the final scene, a man is seen kicking Spot in the ‘stomach’ with his foot; Spot staggers to regain its footing after sustaining the blow but does not fall to the ground.

The Atlas and Spot videos were downloaded from the Boston Dynamics webpage using the open source software, Online Video Converter (version 2.0; https://www.onlinevideoconverter.com/video-converter) and then imported into Microsoft Movie Maker (version 2012). The videos were edited so that their duration was approximately the same (Atlas is 51 seconds in length; Spot is 58 seconds in length). Once the video editing was complete, a Windows Media Video (.wmv) file was created for each excerpt. Each video excerpt was embedded on an individual slide within each PowerPoint presentation (also embedded with the vignettes).

There were several reasons for including the video excerpts. First, they extend Haidt et al.’s (2000) study by providing additional third-person tasks that involve making moral judgments about witnessed actions, and also address the criticism that some of Haidt et al.’s measures are unrealistic (Kennett, 2012; Royzman et al., 2015) as the videos are recordings of a real-life event. Furthermore, they were expected to elicit an untrained response to a novel moral dilemma as the video excerpts were hopefully unfamiliar to the participants.

**Measures**

Several measures previously assessed in Studies 1 and 2 are not assessed in the current study (nor in remaining investigations of this dissertation). Confidence in judgment is not included as this measure did not play a role in differentiating between tasks or determining dumbfounding rates in Studies 1 and 2. Confused by task and irritated by task are also not included due to the ambiguity of the items casting doubt upon their validity as moral dumbfounding measures (see the discussion section of Study 1, p. 60). Finally, speech dysfluency and dead-end arguments most likely indicate that a participant is trying to articulate a reason for their judgment rather than they are without reason (again, see the discussion section of Study 1, p. 60), so these two measures are
also not included. The measures that are assessed in the current study are presented below (refer to the Methods section of Study 1, p. 23 for further information regarding each measure).

**Moral dumbfounding**

The *Evidence-based* definition of moral dumbfounding developed in Study 2 was utilised in the current study; that is, ‘an inability to justify disapproval for a seemingly harm-free action’. To satisfy this criterion, a participant needed to have a frequency greater than zero for the variables of ‘belief that the action is harm-free’ AND (‘judgment without justification’ OR ‘struggle to produce a reason’). A moral dumbfounding variable was also computed for each presentation medium by averaging the frequency of moral dumbfounding separately across vignettes (*John, Masturbation, and Liver*), behavioural tasks (*Lolly and Extended Soul*), and videos (*Atlas and Spot*). Finally, a global measure of moral dumbfounding was computed for each participant individually by aggregating the frequency of moral dumbfounding across stimuli. Higher numbers were taken to indicate greater frequency of moral dumbfounding during the interview. In addition to the features of the empirically supported definition, the following measures (the same as those tested in Study 1) are tested to allow for comparison of the performance of the alternative stimuli with the original moral dumbfounding stimuli.

**Judgments of wrongness of an action**

Judgments of action included initial judgment, intermediate judgment, and final judgment. All judgments were coded as either ‘wrong’, ‘undecided’, or ‘OK’. As participants provided both initial and final judgments, it was appropriate to analyse judgments of wrongness of action for initial and final judgments separately.

**Change of judgment**

Exploration of the relationship between the tendency of participants to change their judgment and moral dumbfounding was undertaken in the current study to verify prior argument (made in Study 2) that maintenance of judgment has little to do with the phenomenon. Such analysis deviated slightly from Study 1 and was able to be conducted in the present study as maintenance of judgment was no longer an element in the operationalisation of moral dumbfounding. In contrast to the ‘change of judgment’ variable used in Study 1 (wherein intermediate judgment was not included in the calculation), a ‘change of judgment’ variable was computed in which a participant was considered to have changed their judgment if they made at least one judgment that the
action was ‘wrong’, but at least one of the other judgments (i.e., initial, intermediate, or final) was ‘undecided’ or ‘OK’. A global variable of ‘change of judgment’ was computed for each participant by averaging the frequency across tasks, with higher numbers indicating greater change of judgment during the interview.

Response timing
The total response time for each task was measured from completion of the experimenter’s request for a first task initial judgment to the end of participant final justification for their last task judgment. As a response could take the form of a judgment or a justification, timings to the commencement of both responses were analysed. The order of responses (time difference between the start of participants’ initial judgment and that of their initial justification) was measured to test the assumption that participants typically provide a quick judgment followed by justification when presented with the intuition cases (i.e., the measures that elicit moral dumbfounding). Timings to the start of initial judgment and initial justification were used to test the assumption that moral dumbfounding occurs with quick judgment and slow justification.

The time-based variables reported in the original dumbfounding study (Haidt et al., 2000) were also measured here. The total response time for each task was measured from completion of the experimenter’s request for a first task initial judgment to the end of participant final justification for their last task judgment. According to Haidt et al. (2000), when presented with the intuition cases (i.e., the measures that often elicited moral dumbfounding), participants typically provide a quick judgment followed by justification. The order of responses for participants (judgment or justification given first) was measured using the time difference between the commencement of participants’ initial judgment and that of their initial justification. Timings to the start of initial judgment and initial justification were used to test the assumption that moral dumbfounding occurs with quick judgment and slow justification (Haidt et al.).

Basis underpinning judgments
The foundation underpinning judgments (specifically, a judgment made on gut feeling) is claimed to be central to moral dumbfounding (Haidt, 2001; Haidt et al., 2000; Haidt & Hersh, 2001). A foundation of judgment variable was computed using the difference in participant ratings of the degree to which they based their decision on reasoning versus gut feeling (with lower ratings indicating more gut feeling and less reasoning as the foundation underpinning judgment). The basis of initial judgment, and the basis of
final judgment, was computed for each stimulus. To explore a possible relationship using the current study measures, a global measure of the basis of final judgment was computed for each participant by averaging their basis of final judgment across stimuli (as outlined in Study 1, moral dumbfounding can only be demonstrated after a final judgment, so only the basis for final judgment measure was derived). Lower ratings were taken to indicate more gut feeling and less reasoning as the basis underpinning their decision.

**Psychic discomfort**
The total response time differed between tasks ranging from an average of approximately 33 seconds for the *Lolly* task (many participants ate the lolly immediately upon being asked without any discussion) to approximately three and a half minutes for the intuition vignettes. The variance in total task time meant that there was more or less time for participants to demonstrate behaviour (i.e., laughter, face touching, posture shifting, and fiddling). Accordingly, the behavioural variables were standardised by dividing the total number of each behavioural variable, per person per task, by the total task time (in seconds). As in Study 1, it was considered unnecessary to analyse individual behavioural variables, so the four behavioural variables were combined and collectively taken to indicate the physical manifestation of psychic discomfort.

**Linguistic responses**
Linguistic responses included statements of (a) “I don’t know”, (b) failure to provide a reason (i.e., unsupported declarations, tautological reasons, illogical arguments, and these variables combined), (c) struggle to produce a reason (i.e., dumbfounding statements, statements of unwillingness to support, and these variables combined), (d) unique justifications, and (e) puzzlement. To explore the relationship between moral dumbfounding and unique justifications, a global measure was computed for each participant by averaging the number of unique justifications across stimuli. The number of unique justifications offered during a task were taken as persistence in defence of a judgment; that is, the smaller the number of justifications offered, the greater the persistence in defence of that judgment. To explore the relationship between moral dumbfounding and puzzlement, a global measure was computed for each participant by averaging their number of puzzlement statements across stimuli. A greater number of statements were taken to indicate greater puzzlement at an inability to justify judgment.
Belief that an action is realistic

An additional measure was assessed in the current investigation that was measured, but not assessed, in Studies 1 and 2 (as it was not relevant to the aim of the respective studies). As stated in the previous studies, participants were asked to complete a short questionnaire after each task comprising items pertaining to how they made their judgment and how they perceived the task (annexed as Appendix D.1). One of the items seeking a response from participants was: ‘I believe that the story/task is representative of a situation which may happen in the real world’. Responses to this item were recorded on a five-point rating scale ranging from 1 (strongly disagree) to 5 (strongly agree). This measure was included to empirically test the criticism that moral dumbfounding results from the unrealistic nature of Haidt et al.’s (2000) intuition measures (Jacobson, 2012; Kennett, 2012).

The End-of-Task Questionnaire responses were downloaded from the GoogleForms collection format which resulted in an Excel spreadsheet format (refer to the Methods section of Study 1, p. 33 for details). The data were inspected for completeness and consistency. There were no missing data. The responses of participants to the believability item were imported into SPSS Statistics for analysis. A global measure was computed for each participant by aggregating their frequency of belief that the depicted action is representative of a situation which may happen in the real world, across stimuli. Higher numbers were taken to indicate a greater belief that the actions undertaken in the scenarios were realistic.

The video excerpts were argued to address the criticism that the measures of Haidt et al (2000) are unrealistic (Kennett, 2012; Royzman et al., 2015) given that they are recordings of a real-life event. It was, therefore, important to ascertain if these visual stimuli were perceived by the participants as more realistic than the vignettes and behavioural tasks. Global measures of belief that the scenarios were realistic, by presentation medium, were computed for each participant by averaging the frequency of realistic belief separately across intuition vignettes (Incest, Masturbation, Cannibal, and Liver), behavioural tasks (Roach, Lolly, Soul, and Extended Soul), and videos (Atlas and Spot).

Procedure

The additional video stimuli employed in the current study were always presented last so the order of task presentation reflected that of Study 1 and the original study of Haidt
et al. (2000). The addition of the video excerpt stimuli (Atlas followed by Spot; Spot
followed by Atlas) resulted in four task orders as follows:
1. Masturbation, Lolly, Liver, John, Soul, Atlas, Spot
2. Masturbation, Lolly, Liver, John, Soul, Spot, Atlas
3. John, Liver, Lolly, Masturbation, Soul, Atlas, Spot
4. John, Liver, Lolly, Masturbation, Soul, Spot, Atlas

The procedure for the current study was the same as Study 1 (refer to the
Methods section of Study 1 on p. 30 for further information).

Data analysis

As in Study 1, qualitative and quantitative analyses were applied in the current study
(refer to the Methods section of Study 1 on p. 32 for information regarding the
qualitative and quantitative analyses, as well as general data-analysis information).
Coding of verbal and behavioural variables was conducted independently by the
researcher and a research assistant. The two-person coding procedure resulted in a very
high degree of inter-rater reliability between 20 measurements. The average measure
intra-class correlation coefficient was 0.989 with a 95% confidence interval from 0.972
to 0.995 ($F(19,19) = 85.58, p < 0.001$). Further reliability was performed whereby the
two coders together reviewed each of the instances recorded under each variable node in
NVivo. The representativeness of the cases to each variable was discussed with
differences in node assignments resolved. Finally, a subset of interviews ($n = 22$),
together with the coding allocations, was peer-reviewed by a second research assistant.
There was no discrepancy in the coding.

Like Study 1, while the sample size of the current study ($N = 44$) was larger than
the original research ($N = 30$; Haidt et al., 2000), a Monte Carlo resampling method was
applied to generate a large number of simulated samples (10,000 samples) to yield
unbiased estimates about the characteristics of the population from which the sample at
hand is drawn (sc., Field, 2009). All variables were checked for normality and
sphericity using SPSS except for homogeneity of variance for repeated measures
variables which was computed by hand (using the formula

$$F_{\text{max}} = \frac{\text{Largest Sample Variance}}{\text{Smallest Sample Variance}},$$

where the largest sample variance is simply the largest
standard deviation squared, and the smallest sample variance is the smallest standard
deviation squared; homogeneity of variance can be assumed when $F_{\text{max}}$ is less than 10).
As in Study 1, the assumptions of normality, sphericity, and homogeneity of variance, were violated for many of the variables and, therefore, non-parametric tests were used.

A Bonferroni adjustment was used to maintain a family-wise alpha rate of 0.05 over multiple comparisons and, therefore, reduce the risk of false-positive results. Unless otherwise stated, a one-tailed significance level was set at $\alpha = 0.0083$ (i.e., a standard level of $0.05$ divided by the six planned comparisons between the reasoning stimulus and the six intuition stimuli; see, Allen & Bennett). Effect sizes are reported where possible, with the effect size formula for significant pairwise comparisons being:

$$ r = \frac{z}{\sqrt{N}} $$

Descriptions of effect size were guided by J. D. Cohen (1988) in which $r = 0.1$ can be considered small, $r = 0.3$ can be considered medium, and $r = 0.5$ can be considered large. Where the effect size is unable to be ascertained due to the limited availability of statistical techniques, a graphical overview is provided to give an adequate sense of the size and direction of the effect of significant comparisons.

The mean and standard deviation are presented in the current study as measures of central tendency and variation instead of the customary presentation of median and interquartile range for the non-parametric tests. Like Study 1, the nonconformity of presentation was required due to the narrow scale of self-response items (typically 0 to 5) and the small frequency of some variables (e.g., statements of puzzlement usually occurred once during an interview, if at all). Presenting the mean provided a nuanced gauge of difference between the measures (unable to be revealed in presenting the median); however, it should not be considered the most accurate representation of central tendency (given the assumption of normality was violated for many of the variables in the current study, which is why non-parametric tests were used). The median for each variable in the current study is available upon request, and the mean ranking is provided below for each variable (in the relevant sections), which is an accurate indicator of the performance of the presented stimuli. Differences between the mean ranks of pairwise comparisons that are equal to, or bigger than, the critical difference are interpreted as significant.

Results

Within-subject results of Study 3

Five participants indicated familiarity with the materials. Two participants indicated that the John vignette was similar to a story they had heard in a psychology class (i.e., the Heinz dilemma). One participant specified (in reference to the Masturbation vignette)
that they had heard a friend’s son had encouraged his dog to help him masturbate. One participant specified that they had previously viewed the Atlas video, but had not seen the end in which the man pushed Atlas in the back with a pole. Another participant had viewed the Spot video, but stated that it was a while ago. No familiarity with any of the other stimuli was reported. As the degree of familiarity with the stimuli for each of the five participants was either superficial or not recent, and due to the small frequency of familiarity in the total sample, it was not considered that exposure to, and previous contemplation of, the scenarios would influence the results.

**Moral dumbfounding**

It was hypothesised that moral dumbfounding would be elicited in response to the stimuli presented.

Table 24 presents the frequency of participants who were classified as morally dumbfounded (at least once during the interview) across the stimuli. The percentage of dumbfounded participants (across all tasks) was 27.3%. The global variable of moral dumbfounding ranged from 0 to 2 with a mean of 0.30 (SD = 0.51).

<table>
<thead>
<tr>
<th>Stimuli</th>
<th>Morally Dumbfounded</th>
<th>n</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>John</td>
<td></td>
<td>0</td>
<td>0.0</td>
</tr>
<tr>
<td>Masturbation</td>
<td></td>
<td>3</td>
<td>6.8</td>
</tr>
<tr>
<td>Liver</td>
<td></td>
<td>4</td>
<td>9.1</td>
</tr>
<tr>
<td>Lolly</td>
<td></td>
<td>1</td>
<td>2.3</td>
</tr>
<tr>
<td>Extended Soul</td>
<td></td>
<td>3</td>
<td>6.8</td>
</tr>
<tr>
<td>Atlas</td>
<td></td>
<td>2</td>
<td>4.5</td>
</tr>
<tr>
<td>Spot</td>
<td></td>
<td>0</td>
<td>0.0</td>
</tr>
</tbody>
</table>

*Note. n = frequency of participants from N = 44.*

**Reasoning versus intuition stimuli and moral dumbfounding**

It was hypothesised that the intuition stimuli would elicit significantly more moral dumbfounding than the reasoning stimulus. A Cochran’s $Q (\alpha = 0.05)$ revealed that the prevalence of dumbfounding did not vary significantly across stimuli, $Q (df = 6, N = 44) = 8.21, p = 0.276.$
**Presentation medium and moral dumbfounding**

It was further hypothesised that the vignettes would elicit moral dumbfounding with significantly more frequency than the behavioural tasks and videos. There were seven participants (15.91%) who were dumbfounded in response to at least one of the vignettes. There were four participants (9.09%) who were dumbfounded in response to at least one of the behavioural tasks. Finally, there were two participants (4.55%) who were dumbfounded in response to at least one of the videos. A Cochran’s $Q$ ($\alpha = 0.025$, two-tailed) was used to test the prevalence of moral dumbfounding across presentation medium. The prevalence of dumbfounding did not vary significantly across medium of presentation, $Q (df = 2, N = 44) = 3.17$, $p = 0.238$.

**Belief that an action is realistic**

**Presentation medium and belief that an action is realistic**

It was hypothesised that the videos would be perceived as significantly more realistic than the vignettes and behavioural tasks. Table 25 presents the mean and mean ranking of the ‘I believe that the story/task is representative of a situation which may happen in the real world’ item, across presentation medium, for both Study 1 and Study 2 participants. A Friedman two-way ANOVA ($\alpha = 0.05$) was used to compare belief that the scenario is realistic across presentation medium. Participant belief in the realistic nature of scenarios varied significantly across the three presentation media, $\chi^2_p (df = 2, N = 88) = 45.83$ (corrected for ties), $p = <0.001$. The Wilcoxon Signed Rank Test was used for post-hoc comparisons (Bonferroni adjusted $\alpha$ of 0.0083; i.e., an overall adjusted $\alpha$ of 0.017, divided by the three planned comparisons, divided by the two tails). Participants perceived the actions depicted in the video excerpts as significantly more realistic than the intuition vignettes, $z = -5.19$, $N - Ties = 74$, $p = <0.001$, $r = 0.55$ (large effect size). Participants also perceived the actions depicted in the video excerpts as significantly more realistic than the behavioural tasks, $z = -5.72$, $N - Ties = 70$, $p = <0.001$, $r = 0.61$ (large to very large effect size). There was no difference between intuition vignettes and behavioural tasks in the belief that the scenario is representative of a situation which may happen in the real world ($p = 0.058$).
Table 25

Mean and Mean Ranking of Belief that the Action Depicted in the Dumbfounding Stimuli, by Presentation Medium, is Realistic

<table>
<thead>
<tr>
<th>Stimuli</th>
<th>Mean (M)</th>
<th>Standard Deviation (SD)</th>
<th>Mean Rank</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intuition Vignettes</td>
<td>3.39</td>
<td>0.61</td>
<td>1.71*</td>
</tr>
<tr>
<td>Behavioural Tasks</td>
<td>3.16</td>
<td>0.99</td>
<td>1.74*</td>
</tr>
<tr>
<td>Video Excerpts</td>
<td>3.96</td>
<td>0.66</td>
<td>2.55</td>
</tr>
</tbody>
</table>

Note. N = 88. M = mean; SD = standard deviation. Responses recorded on a five-point rating scale ranging from 1 (strongly disagree) to 5 (strongly agree).
*Significantly different from the video excerpts at the Bonferroni adjusted α of 0.0083 (two-tailed).

Belief that an action is realistic and moral dumbfounding

It was hypothesised that there would be a negative correlation between belief that the scenario presented is realistic and moral dumbfounding. The global variable of belief that an action is realistic ranged from 0 to 1 with a mean of 0.42 (SD = 0.25). Kendall’s tau-b (α = 0.05) was used to explore the relationship between moral dumbfounding (M = 0.30; SD = 0.51) and belief that the action is realistic; there was no relationship; τ = -0.039, p = 0.381, N = 44.

Judgments of wrongness of an action

It was hypothesised that the intuition stimuli would be judged as wrong with significantly more frequency than the reasoning stimulus. Figure 5 presents a graphical overview depicting the frequency of participant initial and final judgments to the seven presented stimuli. As the relevant hypothesis focuses on wrongness of an action,
Table 26 presents only the frequency and mean ranking of initial and final judgments of ‘wrong’ to the presented stimuli. A Friedman two-way ANOVA ($\alpha = 0.05$) was used to compare judgments across stimuli (initial and final judgments were analysed separately). The Wilcoxon Signed Rank Test (Bonferroni adjusted $\alpha = 0.0083$) was used for post-hoc comparisons.

Figure 5. Clustered column chart illustrating the frequency of participant initial and final judgments to the alternative dumbfounding stimuli.
Table 26  
Frequency and Mean Ranking of Initial and Final Judgments of ‘Wrong’ in Response to the Alternative Dumbfounding Stimuli

<table>
<thead>
<tr>
<th>Judgment</th>
<th>Judgments of ‘Wrong’</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n</td>
<td>%</td>
<td>Mean Rank</td>
<td></td>
</tr>
<tr>
<td><strong>Initial</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>John</td>
<td>27</td>
<td>61.4</td>
<td>4.78</td>
<td></td>
</tr>
<tr>
<td>Masturbation</td>
<td>30</td>
<td>68.2</td>
<td>4.94</td>
<td></td>
</tr>
<tr>
<td>Liver</td>
<td>31</td>
<td>70.5</td>
<td>5.02</td>
<td></td>
</tr>
<tr>
<td>Lolly</td>
<td>6</td>
<td>13.6</td>
<td>2.57*</td>
<td></td>
</tr>
<tr>
<td>Extended Soul</td>
<td>27</td>
<td>61.4</td>
<td>4.43</td>
<td></td>
</tr>
<tr>
<td>Atlas</td>
<td>9</td>
<td>20.5</td>
<td>2.93*</td>
<td></td>
</tr>
<tr>
<td>Spot</td>
<td>10</td>
<td>22.7</td>
<td>3.32*</td>
<td></td>
</tr>
<tr>
<td><strong>Final</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>John</td>
<td>25</td>
<td>56.8</td>
<td>4.60</td>
<td></td>
</tr>
<tr>
<td>Masturbation</td>
<td>29</td>
<td>65.9</td>
<td>4.80</td>
<td></td>
</tr>
<tr>
<td>Liver</td>
<td>28</td>
<td>63.6</td>
<td>4.90</td>
<td></td>
</tr>
<tr>
<td>Lolly</td>
<td>6</td>
<td>13.6</td>
<td>2.81*</td>
<td></td>
</tr>
<tr>
<td>Extended Soul</td>
<td>23</td>
<td>52.3</td>
<td>4.30</td>
<td></td>
</tr>
<tr>
<td>Atlas</td>
<td>12</td>
<td>27.3</td>
<td>3.44</td>
<td></td>
</tr>
<tr>
<td>Spot</td>
<td>8</td>
<td>18.2</td>
<td>3.16</td>
<td></td>
</tr>
</tbody>
</table>

*Significantly different to John (i.e., the reasoning stimulus) at the Bonferroni adjusted α of 0.0083.

Note. n = frequency of participants from N = 44.

**Initial judgments of wrongness of an action**

Rankings of initial judgments varied significantly across the seven presented stimuli, $\chi^2 (df = 6, N = 44) = 83.96$ (corrected for ties), $p < 0.001$. In contrast to expectations, there were significantly fewer initial judgments of ‘wrong’ to the Lolly task than to the John vignette, $z = -4.62$, $N - Ties = 37$, $p = 0.001$, $r = 0.709$ (very large effect size). There were also significantly fewer initial judgments of ‘wrong’ to both intuition videos than to the John vignette (Atlas video: $z = -4.31$, $N - Ties = 32$, $p < 0.001$, $r = 0.65$, very large effect size; Spot video: $z = -3.32$, $N - Ties = 33$, $p < 0.001$, $r = 0.50$, large effect size). Neither of the intuition vignettes, nor the Soul task, differed significantly to John in initial judgments of wrongness (Masturbation, $p = 0.390$; Liver, $p = 0.266$; Soul, $p = 0.167$).
Final judgments of wrongness of an action

Rankings of final judgments also varied significantly across the seven presented stimuli, \( \chi^2_F (df = 6, N = 44) = 59.87 \) (corrected for ties), \( p < 0.001 \). Again, there were significantly fewer final judgments of ‘wrong’ to the Lolly task than to the John vignette, \( z = -4.06, N - Ties = 32, p < 0.001, r = 0.61 \) (very large effect size). In contrast to initial judgments, final judgments of wrongness for the Atlas and Spot videos approached, but did not reach, a significant difference from the John vignette (Atlas, \( p = 0.001 \); Spot, \( p = 0.001 \)). None of the other intuition tasks differed significantly to John in final judgments of wrongness (Masturbation, \( p = 0.266 \); Liver, \( p = 0.236 \); Soul, \( p = 0.178 \)).

Change of judgment

No prediction was made regarding differences between the measures in a change of judgment given no such differences were found in either Haidt et al. (2000) nor Study 1. Table 27 presents the frequency, across stimuli, of participants who changed their judgment (initial to final judgment). A Cochran’s \( Q (\alpha = 0.05) \) was used to test the prevalence of changing judgment (from initial to final) across tasks. The tendency for participants to change their judgment did not vary significantly across the seven presented stimuli, \( Q (df = 6, N = 44) = 11.39, p = 0.076 \).

Table 27
Frequency of Participants Who Changed Their Judgment of ‘Wrong’ to the Alternative Dumbfounding Stimuli

<table>
<thead>
<tr>
<th>Stimuli</th>
<th>Change of Judgment</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( n )</td>
</tr>
<tr>
<td>John</td>
<td>2</td>
</tr>
<tr>
<td>Masturbation</td>
<td>7</td>
</tr>
<tr>
<td>Liver</td>
<td>7</td>
</tr>
<tr>
<td>Lolly</td>
<td>0</td>
</tr>
<tr>
<td>Extended Soul</td>
<td>4</td>
</tr>
<tr>
<td>Atlas</td>
<td>3</td>
</tr>
<tr>
<td>Spot</td>
<td>6</td>
</tr>
</tbody>
</table>

Note. \( n = \) frequency of participants from \( N = 44 \). No stimulus was significantly different to John (i.e., the reasoning stimulus) the Bonferroni adjusted \( \alpha \) of 0.00415 (two-tailed).
No prediction was made regarding the relationship between the tendency to change judgment and moral dumbfounding given that participants were no more likely to change their judgment in response to any of the measures in Study 1. The global variable of ‘change judgment of wrong’ ranged from 0 to 1 with a mean of 0.37 ($SD = 0.21$). Kendall’s tau-b ($\alpha = 0.025$; two-tailed) was used to explore the relationship between a change of judgment and moral dumbfounding ($M = 0.30; SD = 0.51$); there was no relationship; $\tau = -0.119, p = 0.376, N = 44$.

**Response timing**

The total response time for interviews ranged from 5 minutes and 56 seconds through to 37 minutes and 46 seconds. The average response time for interviews was 17 minutes and 23 seconds ($SD = 6$ minutes and 37 seconds). Table 28 displays the range and mean response time (in seconds) to the presented stimuli.

**Table 28**

*Range and Mean of Total Response Time to the Alternative Dumbfounding Stimuli*

<table>
<thead>
<tr>
<th>Stimuli</th>
<th>Minimum</th>
<th>Maximum</th>
<th>$M$</th>
<th>$SD$</th>
</tr>
</thead>
<tbody>
<tr>
<td>John</td>
<td>36</td>
<td>399</td>
<td>146.30</td>
<td>79.81</td>
</tr>
<tr>
<td>Masturbation</td>
<td>86</td>
<td>488</td>
<td>205.18</td>
<td>98.93</td>
</tr>
<tr>
<td>Liver</td>
<td>70</td>
<td>345</td>
<td>206.43</td>
<td>95.65</td>
</tr>
<tr>
<td>Lolly</td>
<td>1</td>
<td>256</td>
<td>33.59</td>
<td>57.00</td>
</tr>
<tr>
<td>Extended Soul</td>
<td>4</td>
<td>413</td>
<td>104.14</td>
<td>97.93</td>
</tr>
<tr>
<td>Atlas</td>
<td>42</td>
<td>424</td>
<td>166.41</td>
<td>78.66</td>
</tr>
<tr>
<td>Spot</td>
<td>48</td>
<td>420</td>
<td>180.80</td>
<td>70.42</td>
</tr>
</tbody>
</table>

*Note. $N = 44$, $M =$ mean; $SD =$ standard deviation.*

It was hypothesised that participants would respond significantly faster when presented with the intuition stimuli than the reasoning stimulus. Table 29 presents the mean (in seconds) and mean ranking of time to initial judgment, time to initial justification, and the difference in time between judgment and justification, to the presented stimuli. A series of Friedman two-way ANOVA ($\alpha = 0.05$) were used to compare three or more related samples (i.e., response timings to each of the seven stimuli) for time to initial judgment, time to initial justification, and order of responses, respectively. The Wilcoxon Signed Rank Test (Bonferroni adjusted $\alpha = 0.0083$) was used for post-hoc comparisons.
Table 29  
Mean Response Times and Mean Ranking of Time to Initial Judgment, Time to Initial Justification, and Time Difference Between Judgment and Justification, Across the Alternative Dumbfounding Stimuli  

<table>
<thead>
<tr>
<th>Stimuli</th>
<th>Initial Judgment</th>
<th>Initial Justification</th>
<th>Judgment Precedes Justification</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M</td>
<td>SD</td>
<td>Mean Rank</td>
</tr>
<tr>
<td>John</td>
<td>14.27</td>
<td>17.60</td>
<td>4.61</td>
</tr>
<tr>
<td>Masturbation</td>
<td>14.09</td>
<td>16.01</td>
<td>4.78</td>
</tr>
<tr>
<td>Liver</td>
<td>11.34</td>
<td>12.49</td>
<td>4.51</td>
</tr>
<tr>
<td>Lolly</td>
<td>4.64</td>
<td>12.39</td>
<td>1.82*</td>
</tr>
<tr>
<td>Extended Soul</td>
<td>10.25</td>
<td>9.61</td>
<td>4.68</td>
</tr>
<tr>
<td>Atlas</td>
<td>11.11</td>
<td>17.57</td>
<td>3.92</td>
</tr>
<tr>
<td>Spot</td>
<td>8.80</td>
<td>17.21</td>
<td>3.67</td>
</tr>
<tr>
<td>Total</td>
<td>10.64</td>
<td>5.60</td>
<td>-</td>
</tr>
</tbody>
</table>

Note. N = 44. M = mean; SD = standard deviation.  

*Time is in seconds.  

*Negative sign indicates that justification precedes judgment (of 'wrong').  

*Significantly different to John (i.e., the reasoning stimulus) at the Bonferroni adjusted α of 0.0083.
Time to initial judgment

The time to initial judgment varied significantly across the seven presented stimuli, \( \chi^2 \) \(_F\) \((df = 6, N = 44) = 64.11 \) (corrected for ties), \( p < 0.001 \). Time to initial judgment for the Lolly task (mean rank = 1.82) was significantly faster than for the John vignette (mean rank = 4.61), \( z = -4.15, N - Ties = 42, p = <0.001, r = 0.63 \) (very large effect size). The difference between the Spot video (mean rank = 3.67) and the reasoning stimulus approached, but did not reach, significance (\( p = 0.009 \)). None of the other intuition stimuli differed significantly to the reasoning stimulus in time to initial judgment (Masturbation, \( p = 0.324 \); Liver, \( p = 0.370 \); Soul, \( p = 0.271 \); Atlas, \( p = 0.174 \)).

Time to initial justification

The time to initial justification also varied significantly across the seven presented stimuli, \( \chi^2 \) \(_F\) \((df = 6, N = 44) = 85.14 \) (corrected for ties), \( p < 0.001 \). Time to initial justification for the Liver vignette (mean rank = 3.89) was significantly faster than for the John vignette (mean rank = 4.84), \( z = -2.44, N - Ties = 40, p = 0.007, r = 0.37 \) (medium effect size). There was also significantly faster justification to the Lolly task (mean rank = 1.91) than for the John vignette, \( z = -5.18, N - Ties = 43, p = <0.001, r = 0.78 \) (extremely large effect size). In addition, initial justification to the Atlas video (mean rank = 3.07) was significantly faster than for the John vignette, \( z = -4.40, N - Ties = 42, p = <0.001, r = 0.66 \) (very large effect size). The remaining three intuition stimuli were non-significant (Masturbation, \( p = 0.204 \); Soul, \( p = 0.296 \); Spot, \( p = 0.024 \)).

Order of responses

It was hypothesised that the intuition stimuli would elicit judgment prior to justification, whereas the reasoning stimulus would elicit justification prior to judgment. While the order of judgment and justifications varied significantly across the seven presented stimuli, \( \chi^2 \) \(_F\) \((df = 6, N = 44) = 18.96 \) (corrected for ties), \( p = 0.003 \), no comparison was significant (Masturbation, \( p = 0.388 \); Liver, \( p = 0.292 \); Lolly, \( p = 0.409 \); Soul, \( p = 0.218 \); Atlas, \( p = 0.176 \); and Spot, \( p = 0.171 \)).

Response timing and moral dumbfounding

Separate inferential analyses between moral dumbfounding and time to initial judgment, and time to initial justification, were undertaken. No prediction was made regarding response timing and moral dumbfounding given that no relationship was found in Study 1. Kendall’s tau-b (\( \alpha = 0.05 \)) was used to explore possible relationships between these
variables and moral dumbfounding ($M = 0.30; SD = 0.51$). There was no correlation between moral dumbfounding and ranked time to initial judgment; $\tau = 0.034, p = 0.391, N = 44$. There was, however, a negative, weak correlation between moral dumbfounding and ranked time to initial justification; $\tau = -0.206, p = 0.050, N = 44$.

**Basis underpinning judgments**

It was hypothesised that participants would report basing their judgments significantly more on gut feeling in response to the intuition stimuli than to the reasoning stimulus. Given the assumed differences in the use of reason and gut feeling in making initial and final judgments, within-subject analyses were conducted using a series of Wilcoxon Signed Ranks Tests (a two-tailed $\alpha$ of 0.025 was set). Table 30 presents the mean and mean ranking of the basis underpinning initial and final judgments, across stimuli. Lower ratings indicate more gut feeling and less reasoning as the foundation underpinning judgment.

**Table 30**

*Man and Mean Ranking, of the Basis (i.e., Reasoning or Gut Feeling) Underpinning Initial and Final Judgments of the Alternative Dumbfounding Stimuli*

<table>
<thead>
<tr>
<th>Stimuli</th>
<th>Basis of Initial Judgment</th>
<th>Basis of Final Judgment</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$M$</td>
<td>$SD$</td>
</tr>
<tr>
<td>John</td>
<td>$-0.75$</td>
<td>1.43</td>
</tr>
<tr>
<td>Masturbation</td>
<td>$-0.41$</td>
<td>1.32</td>
</tr>
<tr>
<td>Liver</td>
<td>$-0.07$</td>
<td>1.77</td>
</tr>
<tr>
<td>Lolly</td>
<td>0.16</td>
<td>1.63</td>
</tr>
<tr>
<td>Extended Soul</td>
<td>$-0.45$</td>
<td>1.61</td>
</tr>
<tr>
<td>Atlas</td>
<td>0.32</td>
<td>1.67</td>
</tr>
<tr>
<td>Spot</td>
<td>0.27</td>
<td>1.91</td>
</tr>
</tbody>
</table>

*Note. $N = 44$. $M =$ mean; $SD =$ standard deviation. Lower ratings indicate more gut feeling and less reasoning as the basis of judgment.

*Significantly different to John (i.e., the reasoning stimulus) at the Bonferroni adjusted $\alpha$ of 0.0083.*

**Comparison of original and alternative dumbfounding stimuli**

The only stimulus to not differ significantly between the self-reported basis of initial and final judgments was the *Lolly* task (Mean Rank$_{initial} = 8.50$, Mean Rank$_{final} = 8.50$; $p = 0.027$). For the remaining stimuli, participants were significantly more likely to base their final judgment on reasoning than they were their initial judgment. Reasoning was used significantly more for the final judgment of the *John* vignette (mean rank = 19.06)
than for the initial judgment (mean rank = 12.44), z = -3.40, N - Ties = 34, p = <0.001, r = 0.51 (large effect size). For the Masturbation vignette, the final judgment involved significantly more reasoning (mean rank = 9.67) than the initial judgment (mean rank = 8.67), z = -2.66, N - Ties = 18, p = 0.008, r = 0.40 (medium to large effect size).

Regarding the Liver vignette, the final judgment (mean rank = 11.92) was based significantly more on reasoning than the initial judgment (mean rank = 5.50), z = -3.50, N - Ties = 21, p = <0.001, r = 0.53 (large effect size). For the Soul task, the final judgment (mean rank = 10.75) and the initial judgment (mean rank = 6.00) differed significantly in the basis underpinning judgment, z = -3.17, N - Ties = 19, p = 0.001, r = 0.48 (large effect size). Reasoning was also used significantly more for the final judgment of the Atlas video (mean rank = 13.59) than for the initial judgment (mean rank = 7.50), z = -2.91, N - Ties = 21, p = 0.002, r = 0.44 (medium to large effect size).

Finally, for the Spot video, there was a significant difference between final judgment (mean rank = 7.77) and initial judgment (mean rank = 4.00), z = -3.09, N - Ties = 14, p = 0.001, r = 0.47 (large effect size). Given that almost all of the seven within-subject analyses confirm that the foundation of the moral judgment varies across time, it was necessary to keep initial and final judgments separate when comparing the basis underpinning judgment between measures.

Comparison between alternative dumbfounding stimuli
A pair of Friedman two-way ANOVA (α = 0.05) was used to compare the stimuli in regard to the self-reported basis of initial judgment, and final judgment. The Wilcoxon Signed Rank Test (Bonferroni adjusted α = 0.0083) was used for post-hoc comparisons.

The self-reported basis of initial judgment varied across the seven presented stimuli, $\chi^2_F$ (df = 6, N = 44) = 13.73, p = 0.033. Against prediction, participants were significantly more likely to base their initial judgment on gut feeling when responding to the John vignette (i.e., the reasoning stimulus) than to the Atlas video, z = -2.54, N - Ties = 36, p = 0.005, r = 0.38 (medium to large effect size). None of the other stimuli differed significantly from the reasoning stimulus in the basis underpinning initial judgments (Masturbation, p = 0.176; Liver, p = 0.061; Lolly, p = 0.014; Soul, p = 0.356; Spot, p = 0.015). A significant difference was also found between the stimuli in the self-reported basis underpinning final judgments, $\chi^2_F$ (df = 6, N = 44) = 15.78 (corrected for ties), p = 0.015. Participants were significantly more likely to base their final judgment on gut feeling when responding to the Masturbation vignette than to the John vignette, z = -2.60, N - Ties = 32, p = 0.003, r = 0.39 (medium to large effect size). None of the
other stimuli differed significantly from the reasoning stimulus in the basis underpinning final judgments (Liver, \( p = 0.254 \); Lolly, \( p = 0.065 \); Soul, \( p = 0.024 \); Atlas, \( p = 0.499 \); Spot, \( p = 0.469 \)).

No prediction was made regarding the foundation underpinning judgments and moral dumbfounding given that no relationship was found between these variables in Study 1. The mean of basis for final judgment (\( M = 0.65 \), \( SD = 1.59 \)) was a positive integer indicating that, on average, this basis was more reasoning than gut feeling. Kendall’s \( \tau-b \) (\( \alpha = 0.05 \)) was used to explore the relationship between moral dumbfounding (\( M = 0.30 \); \( SD = 0.51 \)) and the basis underpinning judgments. No relationship was found; \( \tau = -0.096, p = 0.227, N = 44 \).

**Psychic discomfort**

It was hypothesised that there would be significantly more psychic discomfort in response to the intuition stimuli than to the reasoning stimulus. Table 31 presents the mean and mean ranking of observed manifestation of psychic discomfort across stimuli. A Friedman two-way ANOVA (\( \alpha = 0.05 \)) was used to compare psychic discomfort in response to John to that of Masturbation, Liver, Lolly, Soul, Atlas, and Spot. Psychic discomfort varied significantly across the seven presented stimuli, \( \chi^2_F (df = 6, N = 44) = 113.36 \) (corrected for ties), \( p < 0.001 \).

**Table 31**

*Mean and Mean Ranking of Psychic Discomfort in Response to the Alternative Dumbfounding Stimuli*

<table>
<thead>
<tr>
<th>Stimuli</th>
<th>( M )</th>
<th>( SD )</th>
<th>Mean Rank</th>
</tr>
</thead>
<tbody>
<tr>
<td>John</td>
<td>3.33</td>
<td>2.58</td>
<td>2.75</td>
</tr>
<tr>
<td>Masturbation</td>
<td>4.02</td>
<td>2.95</td>
<td>3.20</td>
</tr>
<tr>
<td>Liver</td>
<td>4.27</td>
<td>3.83</td>
<td>3.69*</td>
</tr>
<tr>
<td>Lolly</td>
<td>55.71</td>
<td>57.97</td>
<td>6.73*</td>
</tr>
<tr>
<td>Extended Soul</td>
<td>9.32</td>
<td>10.69</td>
<td>5.08*</td>
</tr>
<tr>
<td>Atlas</td>
<td>2.35</td>
<td>0.60</td>
<td>3.10</td>
</tr>
<tr>
<td>Spot</td>
<td>2.26</td>
<td>0.75</td>
<td>3.44</td>
</tr>
</tbody>
</table>

*Note. N = 44. M = mean; SD = standard deviation. *Significantly different to John (i.e., the reasoning stimulus) at the Bonferroni adjusted \( \alpha \) of 0.0083.*
The Wilcoxon Signed Rank Test (Bonferroni adjusted $\alpha = 0.0083$) was used for post-hoc comparisons. Psychic discomfort during the Liver vignette was significantly higher than during the John vignette, $z = -2.66$, $N - \text{Ties} = 43$, $p = 0.004$, $r = 0.40$ (medium to large effect size). In addition, participants exhibited significantly more psychic discomfort during the Lolly task than during the John vignette, $z = -5.68$, $N - \text{Ties} = 44$, $p < 0.001$, $r = 0.86$ (extremely large effect size). There was also significantly more psychic discomfort displayed during the Soul task than during the John vignette, $z = -4.62$, $N - \text{Ties} = 44$, $p < 0.001$, $r = 0.70$ (very large effect size). Neither the intuition video, nor the Masturbation vignette, differed significantly to the reasoning stimulus in the observed manifestation of psychic discomfort (Masturbation, $p = 0.020$; Atlas, $p = 0.088$; Spot, $p = 0.035$).

Linguistic responses

Statements of "I don't know"

It was hypothesised that there would be significantly more statements of "I don’t know" in response to the intuition stimuli than to the reasoning stimulus. Table 32 presents the mean and mean ranking of "I don’t know" statements across stimuli. A Friedman two-way ANOVA ($\alpha = 0.05$) was used to analyse statements of "I don’t know" in response to John to that of Masturbation, Liver, Lolly, Soul, Atlas, and Spot. Statements of "I don’t know" present in the responses of participants varied significantly across the seven presented stimuli, $\chi^2_F (df = 6, N = 44) = 31.77$ (corrected for ties), $p < 0.001$.

<table>
<thead>
<tr>
<th>Stimuli</th>
<th>$M$</th>
<th>$SD$</th>
<th>Mean Rank</th>
</tr>
</thead>
<tbody>
<tr>
<td>John</td>
<td>0.23</td>
<td>0.61</td>
<td>3.59</td>
</tr>
<tr>
<td>Masturbation</td>
<td>1.18</td>
<td>1.92</td>
<td>4.70*</td>
</tr>
<tr>
<td>Liver</td>
<td>0.66</td>
<td>1.33</td>
<td>4.19</td>
</tr>
<tr>
<td>Lolly</td>
<td>0.02</td>
<td>0.15</td>
<td>3.23</td>
</tr>
<tr>
<td>Extended Soul</td>
<td>2.17</td>
<td>2.23</td>
<td>4.36*</td>
</tr>
<tr>
<td>Atlas</td>
<td>0.30</td>
<td>0.80</td>
<td>3.65</td>
</tr>
<tr>
<td>Spot</td>
<td>0.75</td>
<td>1.38</td>
<td>4.27*</td>
</tr>
</tbody>
</table>

Note. $N = 44$. $M = \text{mean}; SD = \text{standard deviation}$.

*Significantly different to John (i.e., the reasoning stimulus) at the Bonferroni adjusted $\alpha$ of 0.0083.
The Wilcoxon Signed Rank Test (Bonferroni adjusted $\alpha = 0.0083$) was used for post-hoc comparisons. Participants stated “I don’t know” significantly more in response to the Masturbation vignette than to the John vignette, $z = -3.16$, $N - Ties = 21$, $p = .001$, $r = 0.48$ (large effect size). There were also significantly more statements of “I don’t know” in response to the Soul task than to the reasoning stimulus, $z = -2.74$, $N - Ties = 17$, $p = 0.002$, $r = 0.41$ (medium to large effect size). Finally, participants stated “I don’t know” significantly more in response to the Spot video than to the John vignette, $z = -2.46$, $N - Ties = 16$, $p = 0.005$, $r = 0.37$ (medium effect size). None of the responses to the remaining intuition stimuli differed significantly from the reasoning stimulus in statements of “I don’t know” ($Liver, p = 0.014$; $Lolly, p = 0.026$; $Atlas, p = 0.284$).

**Failure to provide a reason**

It was hypothesised that there would be significantly more failure to provide a reason in response to the intuition stimuli than to the reasoning stimulus. Table 33 presents the mean and mean ranking, across stimuli, of the composite variable of failure to provide a reason for judgment and the three components that comprise this variable (being unsupported declarations, tautological reasons, and illogical arguments). A series of Friedman two-way ANOVAs ($\alpha = 0.05$) were used to compare three or more related samples (i.e., responses to John, Masturbation, Liver, Lolly, Soul, Atlas, and Spot) for failure to provide a reason for judgment, unsupported declarations, tautological reasons, and illogical arguments, respectively. The Wilcoxon Signed Rank Test (Bonferroni adjusted $\alpha = 0.0083$) was used for post-hoc comparisons.
<table>
<thead>
<tr>
<th>Stimuli</th>
<th>Failure to Provide a Reason</th>
<th>Unsupported Declarations</th>
<th>Tautological Reasons</th>
<th>Illogical Arguments</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$M$</td>
<td>$SD$</td>
<td>Mean Rank</td>
<td>$M$</td>
</tr>
<tr>
<td>John</td>
<td>0.09</td>
<td>0.29</td>
<td>2.99</td>
<td>0.11</td>
</tr>
<tr>
<td>Masturbation</td>
<td>0.70</td>
<td>0.46</td>
<td>5.14*</td>
<td>1.61</td>
</tr>
<tr>
<td>Liver</td>
<td>0.70</td>
<td>0.46</td>
<td>5.14*</td>
<td>1.73</td>
</tr>
<tr>
<td>Lolly</td>
<td>0.14</td>
<td>0.35</td>
<td>3.15</td>
<td>0.16</td>
</tr>
<tr>
<td>Extended Soul</td>
<td>0.45</td>
<td>0.50</td>
<td>4.26*</td>
<td>0.91</td>
</tr>
<tr>
<td>Atlas</td>
<td>0.18</td>
<td>0.39</td>
<td>3.31</td>
<td>0.07</td>
</tr>
<tr>
<td>Spot</td>
<td>0.39</td>
<td>0.49</td>
<td>4.02*</td>
<td>0.14</td>
</tr>
</tbody>
</table>

Note. $N = 44$. $M =$ mean; $SD =$ standard deviation.

*Significantly different to John (i.e., the reasoning stimulus) at the Bonferroni adjusted $\alpha$ of 0.0083.
Failure to provide a reason for judgment varied significantly across the seven presented stimuli, $\chi^2_p (df = 6, N = 44) = 76.06$ (corrected for ties), $p = <0.001$. Four of the six intuition stimuli differed significantly from the reasoning stimulus in a failure to provide a reason. The only two intuition stimuli to not differ from John were Lolly, $p = 0.379$, and Atlas, $p = 0.171$. There was significantly greater failure to provide a reason when responding to the Masturbation vignette than to the John vignette, $z = -5.01$, $N - Ties = 29$, $p = <0.001$, $r = 0.76$ (extremely large effect size). The Liver vignette was also significantly different from the John vignette in failure to provide a reason, $z = -5.20$, $N - Ties = 27$, $p = <0.001$, $r = 0.78$ (extremely large effect size). In addition, there was a significantly greater failure to provide a reason when responding to the Soul task than to the John vignette, $z = -3.77$, $N - Ties = 18$, $p = <0.001$, $r = 0.57$ (large effect size). Finally, the failure to provide a reason was significantly greater when responding to the Spot video than to the John vignette, $z = -3.15$, $N - Ties = 17$, $p = 0.001$, $r = 0.48$ (large effect size).

It was also hypothesised that there would be significantly more unsupported declarations in response to the intuition stimuli than to the reasoning stimulus. Unsupported declarations present in the responses of participants varied significantly across the seven presented stimuli, $\chi^2_p (df = 6, N = 44) = 101.20$ (corrected for ties), $p = <0.001$. Participants offered unsupported declarations significantly more in response to three of the six intuition stimuli than in response to the reasoning stimulus. Responses to the Masturbation vignette contained significantly more unsupported declarations than to the John vignette, $z = -4.75$, $N - Ties = 29$, $p = <0.001$, $r = 0.72$ (very large effect size). There were also significantly more unsupported declarations in response to the Liver vignette than John, $z = -4.49$, $N - Ties = 26$, $p = <0.001$, $r = 0.68$ (very large effect size). Finally, there were significantly more unsupported declarations in response to the Soul task than Heinz, $z = -3.67$, $N - Ties = 17$, $p = <0.001$, $r = 0.55$ (large effect size). The unsupported declarations in response to the Lolly task ($p = 0.416$), the Atlas video ($p = 0.499$), and the Spot video ($p = 0.492$) did not differ significantly from John.

It was further hypothesised that there would be significantly more tautological reasons in response to the intuition stimuli than to the reasoning stimulus. The offering of tautological reasons varied significantly across the seven presented stimuli, $\chi^2_p (df = 6, N = 44) = 39.19$ (corrected for ties), $p < 0.001$. There were significantly more tautological reasons offered during the Liver vignette than to the John vignette, $z = -3.21$, $N - Ties = 13$, $p = <0.001$, $r = 0.48$ (large effect size). While the difference in
tautological reasons offered for judgments to the Spot video compared to the John vignette approached, but did not reach, significance (p = 0.010), the differences in tautological reasons between the remaining tasks and reasoning measure were clearly non-significant (Masturbation, p = 0.076; Lolly, p = 0.245; Soul, p = 0.245; Atlas, p = 0.228).

In addition, it was hypothesised that there would be significantly more illogical arguments in response to the intuition stimuli than to the reasoning stimulus. Illogical arguments varied significantly across the seven presented stimuli, $\chi^2_F (df = 6, N = 44) = 36.56$ (corrected for ties), $p < 0.001$. Only the Soul task differed significantly to the reasoning measure; there were significantly more illogical arguments offered in response to the Soul task than to the John vignette, $z = -2.71$, $N - Ties = 8$, $p = 0.008$, $r = 0.41$ (medium to large effect size). The remaining measures yielded the same means and mean rankings as the reasoning measure except for the Lolly task ($p = 0.244$).

**Struggle to produce a reason**

It was hypothesised that there would be significantly more struggle to produce a reason in response to the intuition stimuli than to the reasoning stimulus. Table 34 presents the mean and mean ranking, across stimuli, of the composite variable of struggle to produce a reason and the two components that comprise this variable (being dumbfounding statements and unwillingness to support). A series of Friedman two-way ANOVAs ($\alpha = 0.05$) were used to compare three or more related samples (i.e., responses to John, Masturbation, Liver, Lolly, Soul, Atlas, and Spot) for struggle to produce a reason for judgment, dumbfounding statements, and unwillingness to support, respectively. The Wilcoxon Signed Rank Test (Bonferroni adjusted $\alpha = 0.0083$) was used for post-hoc comparisons.
Table 34
Mean and Mean Ranking of the Composite Variable of Struggle to Produce a Reason, and the Component Variables of Dumbfounding Statements and Statements of Unwillingness to Support, in Response to the Alternative Dumbfounding Stimuli

<table>
<thead>
<tr>
<th>Stimuli</th>
<th>Struggle to Produce a Reason</th>
<th>Dumbfounding Statements</th>
<th>Unwillingness to Support</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$M$</td>
<td>$SD$</td>
<td>Mean Rank</td>
</tr>
<tr>
<td>John</td>
<td>0.00</td>
<td>0.00</td>
<td>3.51</td>
</tr>
<tr>
<td>Masturbation</td>
<td>0.18</td>
<td>0.39</td>
<td>4.15*</td>
</tr>
<tr>
<td>Liver</td>
<td>0.14</td>
<td>0.35</td>
<td>3.99</td>
</tr>
<tr>
<td>Lolly</td>
<td>0.07</td>
<td>0.26</td>
<td>3.75</td>
</tr>
<tr>
<td>Extended Soul</td>
<td>0.43</td>
<td>0.50</td>
<td>5.02*</td>
</tr>
<tr>
<td>Atlas</td>
<td>0.02</td>
<td>0.15</td>
<td>3.59</td>
</tr>
<tr>
<td>Spot</td>
<td>0.14</td>
<td>0.35</td>
<td>3.99</td>
</tr>
</tbody>
</table>

Note. $N = 44$, $M =$ mean; $SD =$ standard deviation.

*Significantly different to John (i.e., the reasoning stimulus) at the Bonferroni adjusted $\alpha$ of 0.0083.
Struggle to produce a reason varied significantly across the seven presented stimuli, $\chi^2_F (df = 6, N = 44) = 43.59$ (corrected for ties), $p = <0.001$. There was significantly greater struggle to produce a reason when responding to the Masturbation vignette than to the John vignette, $z = -2.83, N - Ties = 8, p = 0.004, r = 0.43$ (large effect size). There was also significantly greater struggle to produce a reason when responding to the Soul task than to the John vignette, $z = -4.36, N - Ties = 19, p = < 0.001, r = 0.66$ (extremely large effect size). The remaining four intuition stimuli did not differ from the reasoning stimulus in struggle to produce a reason ($Liver, p = 0.013; Lolly, p = 0.129; Atlas, p = 0.509; and Spot, p = 0.013$).

It was also hypothesised that there would be significantly more dumbfounding statements in response to the intuition stimuli than to the reasoning stimulus. While dumbfounding statements present in the responses of participants varied significantly across the seven presented stimuli, $\chi^2_F (df = 6, N = 44) = 12.48$ (corrected for ties), $p = 0.044$, post-hoc comparisons revealed that no comparison achieved statistical significance at the Bonferroni adjusted $\alpha$ of 0.0083 (one-tailed). The comparison closest to reaching significance was the Masturbation vignette and the John vignette ($p = 0.015$). The observed tendency for participants to make dumbfounding statements in response to the remaining intuition stimuli compared with the reasoning measure were non-significant ($Liver, p = 0.498; Lolly, p = 0.498; Soul, p = 0.063; Atlas, p = 0.498; and Spot, p = 0.130$).

It was further hypothesised that there would be significantly more statements of unwillingness to support in response to the intuition stimuli than to the reasoning stimulus. Participant statements of unwillingness to support varied significantly across the seven presented stimuli, $\chi^2_F (df = 6, N = 44) = 48.39$ (corrected for ties), $p = < 0.001$. The Soul task elicited significantly more statements of unwillingness to support than the John vignette, $z = -3.66, N - Ties = 17, p = < 0.001, r = 0.55$ (large effect size). The remaining stimuli did not differ significantly from the reasoning stimulus in statements of unwillingness to support ($Masturbation, p = 0.034; Liver, p = 0.040; Lolly, p = 0.131; Atlas, p = 1.0; and Spot, p = 0.131$).

*Unique justifications*

It was hypothesised that there would be significantly more unique justifications in response to the intuition stimuli than to the reasoning stimulus. Table 35 presents the mean and mean ranking of unique justifications offered for judgments across stimuli. A
Friedman two-way ANOVA ($\alpha = 0.05$) was used to compare unique justifications in response to *John* to that of *Masturbation, Liver, Lolly, Soul, Atlas, and Spot*. Unique justifications offered by participants varied significantly across the seven presented stimuli, $\chi^2_F(6, N = 44) = 129.67$ (corrected for ties), $p < 0.001$.

Table 35

<table>
<thead>
<tr>
<th>Stimuli</th>
<th>$M$</th>
<th>$SD$</th>
<th>Mean Rank</th>
</tr>
</thead>
<tbody>
<tr>
<td>John</td>
<td>4.37</td>
<td>1.94</td>
<td>4.66</td>
</tr>
<tr>
<td>Masturbation</td>
<td>5.30</td>
<td>2.32</td>
<td>5.59*</td>
</tr>
<tr>
<td>Liver</td>
<td>6.09</td>
<td>2.78</td>
<td>5.91*</td>
</tr>
<tr>
<td>Lolly</td>
<td>1.42</td>
<td>1.05</td>
<td>1.70*</td>
</tr>
<tr>
<td>Soul</td>
<td>3.09</td>
<td>2.07</td>
<td>3.60*</td>
</tr>
<tr>
<td>Atlas</td>
<td>2.70</td>
<td>1.52</td>
<td>3.06*</td>
</tr>
<tr>
<td>Spot</td>
<td>3.02</td>
<td>1.55</td>
<td>3.48*</td>
</tr>
</tbody>
</table>

*Note. $N = 44$. $M =$ mean; $SD =$ standard deviation.

*Significantly different to *John* (i.e., the reasoning stimulus) at the Bonferroni adjusted $\alpha$ of 0.0083.

The Wilcoxon Signed Rank Test (Bonferroni adjusted $\alpha = 0.0083$) was used for post-hoc comparisons. All intuition stimuli differed significantly from the reasoning stimulus. Participants made significantly more unique justifications for their judgments in response to both the intuition vignettes than to the *John* vignette; *Masturbation*, $z = -2.48$, $N - \text{Ties} = 38$, $p = 0.007$, $r = 0.37$ (medium effect size); *Liver*, $z = -3.75$, $N - \text{Ties} = 40$, $p < 0.001$, $r = 0.56$ (large effect size). Unique justifications for judgments also differed significantly between the remaining stimuli and the *John* vignette; however, the direction of the relationships were opposite to that predicted. The behavioural tasks elicited significantly fewer unique justifications than the reasoning stimulus; *Lolly*, $z = -5.52$, $N - \text{Ties} = 42$, $p < 0.001$, $r = 0.83$ (extremely large effect size); *Soul*, $z = -2.85$, $N - \text{Ties} = 39$, $p < 0.001$, $r = 0.43$ (medium to large effect size). There were also significantly fewer unique justifications in response to both video excerpts compared with *John; Atlas*, $z = -3.98$, $N - \text{Ties} = 39$, $p < 0.001$, $r = 0.60$; large effect size); *Spot* ($z = -3.92$, $N - \text{Ties} = 41$, $p < 0.001$, $r = 0.59$; large effect size).

No prediction was made regarding the number of unique justifications offered for a judgment and moral dumbfounding given that no relationship was found between
the variables in Study 1. The mean of the unique justifications measure was 3.73 (SD = 1.10). Kendall’s tau-b (α = 0.05) was used to explore the relationship between moral dumbfounding (M = 0.30; SD = 0.51) and unique justifications. No relationship was found; τ = 0.076, p = 0.545, N = 44.

Puzzlement

No prediction was made regarding differences in puzzlement at an inability to produce a justification for a judgment between the measures given that no differences were found in Study 1. Table 36 presents the mean and mean ranking of statements of puzzlement across stimuli. A Friedman two-way ANOVA (α = 0.05) was used to compare puzzlement in response to John to that of Masturbation, Liver, Lolly, Soul, Atlas, and Spot. Puzzlement at an inability to produce a justification for judgment did not vary significantly across the seven presented stimuli, \( \chi^2_p (df = 6, N = 44) = 8.00 \) (corrected for ties), p = 0.332.

Table 36

<table>
<thead>
<tr>
<th>Stimuli</th>
<th>M</th>
<th>SD</th>
<th>Mean Rank</th>
</tr>
</thead>
<tbody>
<tr>
<td>John</td>
<td>0.00</td>
<td>0.00</td>
<td>3.92</td>
</tr>
<tr>
<td>Masturbation</td>
<td>0.05</td>
<td>0.21</td>
<td>4.08</td>
</tr>
<tr>
<td>Liver</td>
<td>0.00</td>
<td>0.00</td>
<td>3.92</td>
</tr>
<tr>
<td>Lolly</td>
<td>0.02</td>
<td>0.15</td>
<td>4.00</td>
</tr>
<tr>
<td>Extended Soul</td>
<td>0.09</td>
<td>0.36</td>
<td>4.16</td>
</tr>
<tr>
<td>Atlas</td>
<td>0.00</td>
<td>0.00</td>
<td>3.92</td>
</tr>
<tr>
<td>Spot</td>
<td>0.02</td>
<td>0.15</td>
<td>4.00</td>
</tr>
</tbody>
</table>

Note. N = 44. M = mean; SD = standard deviation. No stimulus was significantly different to John (i.e., the reasoning stimulus) at the Bonferroni adjusted α of 0.00415 (two-tailed).

Participants who explicitly stated puzzlement at an inability to provide justification typically offered one statement of puzzlement (the only exception was that one participant made two statements of puzzlement during the Extended Soul task). As such, it made sense to transform this variable into a binary measure (i.e., ‘Puzzled During the Interview’ versus ‘Not Puzzled During the Interview’) prior to analyses. There were seven participants who made an explicit statement of puzzlement during their interview (Masturbation = 2; Lolly = 1; Extended Soul = 3; Spot = 1). The
McNemar test of change ($\alpha = 0.05$) was used to test the prevalence of moral dumbfounding in puzzled compared to non-puzzled participants. As predicted, the tendency for morally dumbfounded participants to be puzzled was significantly greater than for the non-dumbfounded; $\chi^2 (df = 1, N = 44) = 20.45, p = <0.001$. Of the seven participants (15.9%) who made a statement during the interview that they were puzzled at their inability to produce a justification for their judgment, six were independently classified as dumbfounded. Figure 6 provides a graphical overview of the size and direction of the significant effect.

![Figure 6. Clustered column chart illustrating the significant effect between puzzlement and moral dumbfounding elicited using the alternative dumbfounding stimuli.](image)

The responses of the seven participants who made a puzzlement statement were further analysed to ascertain how one could be puzzled by an inability to justify a judgment, yet considered able to provide justification (i.e., not morally dumbfounded). The participant in question was puzzled but not dumbfounded in response to the Spot video. Despite making statements satisfying most of the criteria of moral dumbfounding (i.e., they had a frequency greater than zero for the variables of ‘belief that the action is harm-free’, ‘failure to provide a reason’, and ‘struggle to produce a reason’), they did not judge the action to be wrong. Their puzzlement statement (“It doesn’t make any sense to me”; Line 343) was in reference to the different reaction they experienced to the Spot video compared with the Atlas video:
“I mean, there’s nothing wrong with it. I’ve just got slightly different feelings on how I perceived it. It was funny when it was a robot human type thing, but it wasn’t as funny with a robot reindeer type thing. I think that’s just, it emulates an animal rather than a person, and so it’s slightly more wrong, but not wrong if that makes sense?” (Participant #160616105240, Lines 344-348)

Given the restriction of moral dumbfounding here to situations in which participants judge an action to be wrong but are unable to provide a reason (or at least an acceptable reason) for this judgment, Participant #160616105240 was not eligible for classification as morally dumbfounded (this participant is presented in the last column of Figure 6. Consequently, all participants who made a statement during the interview that they were puzzled at their inability to produce a justification for their judgment of wrong were independently classified as morally dumbfounded.

**Between-subject results of Studies 1 and 3**

**Belief that an action is realistic**

It was hypothesised that the stimuli presented would be perceived as significantly more realistic than the original measures of Haidt et al. (2000) employed in Study 1. Table 37 presents the mean and mean ranking of the ‘I believe that the story/task is representative of a situation which may happen in the real world’ item, across stimuli, for both Study 1 and Study 3 participants (Study 1 participants were also presented with Atlas and Spot; results were not reported in Chapter 3 for exposition consistency with the original study of Haidt et al.). A series of Mann-Whitney U Tests ($\alpha = 0.025$, two-tailed) were used to compare two independent samples (Study 1, $n = 44$; and Study 3, $n = 44$) of ordinal data (0 to 5 ratings of belief that the scenario is realistic). Participants perceived the Lolly task as significantly more realistic than the Roach task, $U (N = 88) = 388.50, z = -5.04, p = <0.001, r = 0.54$ (large effect size). Participants also perceived the alternative Extended Soul task as significantly more realistic than the original Soul task, $U (N = 88) = 686.00, z = -2.43, p = 0.015, r = 0.26$ (medium effect size). Participants did not perceive any differences in realism between Heinz and John ($p = 0.99$), Incest and Masturbation ($p = 0.86$), Cannibal and Liver ($p = 0.41$), Atlas (Study 1) and Atlas (Study 3; $p = 0.79$), nor Spot (Study 1) and Spot (Study 3; $p = 0.98$).
Table 37
Mean and Mean Ranking of Belief that the Action Depicted in the Dumbfounding
Stimulus is Realistic

<table>
<thead>
<tr>
<th>Stimuli</th>
<th>Believes that the Action is Realistic</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$M$</td>
</tr>
<tr>
<td>Study 1</td>
<td></td>
</tr>
<tr>
<td>Heinz</td>
<td>4.32</td>
</tr>
<tr>
<td>Incest</td>
<td>3.93</td>
</tr>
<tr>
<td>Cannibal</td>
<td>2.70</td>
</tr>
<tr>
<td>Roach</td>
<td>3.05</td>
</tr>
<tr>
<td>Soul</td>
<td>2.34</td>
</tr>
<tr>
<td>Atlas</td>
<td>4.00</td>
</tr>
<tr>
<td>Spot</td>
<td>3.93</td>
</tr>
<tr>
<td>Study 3</td>
<td></td>
</tr>
<tr>
<td>John</td>
<td>4.30</td>
</tr>
<tr>
<td>Masturbation</td>
<td>4.00</td>
</tr>
<tr>
<td>Liver</td>
<td>2.91</td>
</tr>
<tr>
<td>Lolly</td>
<td>4.34</td>
</tr>
<tr>
<td>Extended Soul</td>
<td>2.93</td>
</tr>
<tr>
<td>Atlas</td>
<td>3.93</td>
</tr>
<tr>
<td>Spot</td>
<td>3.98</td>
</tr>
</tbody>
</table>

*Note. Each stimulus was presented to $N = 44$. $M$ = mean; $SD$ = standard deviation. Responses recorded on a five-point rating scale ranging from 1 (strongly disagree) to 5 (strongly agree).
**Significantly different from the Roach task (i.e., the comparative task to Lolly) at $p < 0.025$ (two-tailed). **Significantly different from the Soul task (i.e., the comparative task to Extended Soul) at $p < 0.025$ (two-tailed).

Moral dumbfounding

It was hypothesised that the stimuli presented would elicit significantly less moral dumbfounding than the original measures of Haidt et al. (2000) employed in Study 1. The frequency of participants classified as morally dumbfounded in Study 1 ($n = 44$) and Study 3 ($n = 44$) according to both definitions (i.e., the Study 1 operationalisation of ‘the maintenance of a judgment without provision of supporting reasons’, and the current study operationalisation of ‘an inability to justify disapproval for a seemingly harm-free action’), across tasks, was determined. Table 38 presents the frequency of participants who were classified as morally dumbfounded (at least once during the

135
interview) across stimuli according to the *Common* definition (used in Study 1) and the *Evidence-based* definition (developed in Study 2 and used in the current study).

Table 38

<table>
<thead>
<tr>
<th>Stimuli</th>
<th>Morally Dumbfounded (Common Definition)</th>
<th>Morally Dumbfounded (Evidence-based Definition)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td><em>n</em></td>
<td>%</td>
</tr>
<tr>
<td>Study 1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Heinz</td>
<td>4</td>
<td>9.1</td>
</tr>
<tr>
<td>Incest</td>
<td>26</td>
<td>59.1</td>
</tr>
<tr>
<td>Cannibal</td>
<td>28</td>
<td>63.6</td>
</tr>
<tr>
<td>Roach</td>
<td>13</td>
<td>29.5</td>
</tr>
<tr>
<td>Soul</td>
<td>21</td>
<td>47.7</td>
</tr>
<tr>
<td>Atlas</td>
<td>3</td>
<td>6.8</td>
</tr>
<tr>
<td>Spot</td>
<td>5</td>
<td>11.4</td>
</tr>
<tr>
<td>Study 3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>John</td>
<td>4</td>
<td>9.1</td>
</tr>
<tr>
<td>Masturbation</td>
<td>22</td>
<td>50.0</td>
</tr>
<tr>
<td>Liver</td>
<td>18</td>
<td>40.9</td>
</tr>
<tr>
<td>Lolly</td>
<td>6</td>
<td>13.6</td>
</tr>
<tr>
<td>Extended Soul</td>
<td>19</td>
<td>43.2</td>
</tr>
<tr>
<td>Atlas</td>
<td>2</td>
<td>4.5</td>
</tr>
<tr>
<td>Spot</td>
<td>3</td>
<td>6.8</td>
</tr>
</tbody>
</table>

*Note. n = frequency of participants from N = 44.*

*Significantly different from the Roach task (i.e., the comparative task to Lolly) at p < 0.025 (two-tailed). **Significantly different from the Soul task (i.e., the comparative task to Extended Soul) at p < 0.025 (two-tailed).

The McNemar test of change (α = 0.05) was used to see if the lower rate of moral dumbfounding observed in the current study compared to Study 1 is due to the difference in operationalisation. Participants (N = 88) were classified as morally dumbfounded with significantly greater frequency under the *Common* definition than the *Evidence-based* definition; \( \chi^2 (df = 1, N = 88) = 34.03, p = <0.001 \). Of the 37 participants who were classified as morally dumbfounded according to either definition
(42%), 36 were classified as dumbfounded under the Common definition but not the Evidence-based definition. In contrast, there were no participants classified as dumbfounded under the Evidence-based definition but not the Common definition. Figure 7 provides a graphical overview of the size and direction of the significant effect.

![Bar chart](image)

*Figure 7. Clustered column chart illustrating the significant effect between moral dumbfounding when defined according to the Common definition compared to the Evidence-based definition.*

A series of Fisher’s exact tests \( (\alpha = 0.05) \) were used to see if the lower rate of moral dumbfounding observed in the current study \( (N = 44) \) compared to Study 1 \( (N = 44) \) is due to the difference in presented stimuli. Given concern regarding unnecessary inflation of error rates, and the significant difference found between the prevalence of moral dumbfounding obtained under the two definitions, only the Evidence-based definition was utilised to compare the matched tasks. The Evidence-based definition was chosen as a case has been made for greater ecological validity of this definition compared to the Common definition (see Study 2). Fisher’s exact test is used for computing the exact probability of the chi-square statistic (as opposed to the approximate distribution of the chi-square) when there are expected cell frequencies less than 5 (making significance tests of the approximate chi-square distribution inaccurate; Field, 2009, p. 690).

There was significantly more moral dumbfounding elicited in response to the Roach task than to the Lolly task; \( \chi^2 (1) = 12.23, p = 0.001 \). Overall, 14 participants (15.9% of the total sample) were morally dumbfounded; of these, 13 were dumbfounded in response to the Roach task (92.9% of the total dumbfounded) and one was dumbfounded in response to the Lolly task (7.1% of the total dumbfounded). The
odds ratio was calculated as a measure of effect size. The odds of a participant being classified as morally dumbfounded when presented with the Roach task are 18.03 times higher than if they were presented with the Lolly task. There was also significantly more moral dumbfounding elicited in response to the original Soul task than to the Extended Soul task; $\chi^2 (1) = 7.64, p = 0.011$. In total, 16 participants (18.2% of the total sample) were morally dumbfounded; of these, 13 were dumbfounded in response to the original Soul task (81.3% of the total dumbfounded) and three were dumbfounded in response to the Extended Soul task (18.8% of the total dumbfounded). Based on the odds ratio, a participant presented with the original Soul task is 5.73 times more likely to be classified as morally dumbfounded than if they are presented with the Extended Soul task. There was no significant difference in frequency of moral dumbfounding between Heinz and John ($p = 0.494$), Incest and Masturbation ($p = 0.068$), Cannibal and Liver ($p = 1.000$). Further, Atlas did not elicit a significantly different frequency of moral dumbfounding in participants from Studies 1 and 3 ($p = 0.494$) and neither did Spot ($p = 0.494$).

**Discussion**

This study is the first to investigate whether moral dumbfounding can be elicited using different stimuli to those employed in extant research (Haidt et al., 2000; Haidt & Hersh, 2001; McHugh et al., 2017; Royzman et al., 2015). As predicted, some participants demonstrated moral dumbfounding such that they offered judgments of ‘wrong’ without providing logical supporting reasons. The prevalence of dumbfounding (which did not vary significantly across stimuli) was 27.3%; this frequency is smaller than that found in Study 1 (90.9%), by Haidt and Hersh (82%; 2001), and by McHugh et al. (71%; 2017; Study 1). The smaller prevalence observed in the current study may be due to differences between stimuli presented in Studies 1 and 3 or the operationalisation of moral dumbfounding.

The alternative stimuli of the current study were designed to be comparable to those employed by Haidt et al. (2000) as both stimulus sets invoke a judgment that the action depicted is wrong that is not amenable to harm-based justifications. Unlike those adopted by Haidt et al., however, it was speculated that these stimuli would be perceived as more representative of situations that may happen in the real world. Against prediction, participants did not perceive any differences in realism between the majority of original and alternative stimuli; the behavioural tasks deviated from this trend. The Lolly task was perceived as significantly more realistic than the Roach task.
This result is unsurprising given that eating food that has been dropped onto clean carpet is arguably more common and, therefore, more realistic, than knowingly drinking roach juice. The difference between the ethereal tasks (whereby the Extended Soul task was perceived as significantly more realistic than the original Soul task) is somewhat harder to explain.

Due to the difficulty in designing a suitable substitute, the original Soul task (from Haidt et al., 2000, and Study 1) was employed again here but, upon completion of this task, participants in the current study who did not sign the contract were asked: “Is there any amount of money that I could pay you to get you to sign the contract?” Introducing a price negotiation may prompt more effortful consideration of the Extended Soul task than the original Soul task (participants in the current study were significantly more likely to base their final judgment for the Extended Soul task on reasoning than they were their initial judgment whereas, in Study 1, there was no such difference for the Soul task in self-reported basis of initial and final judgments). It could be that Jacobson (2012, p. 313) is correct — that most people would consider $2 insufficient in exchange for one’s soul. Prompting refusing participants to consider what they would need to be offered, in real life, for them to sell their soul may have made the real world salient and, therefore, influenced responses. This influence of realism on moral dumbfounding was further considered given that one is 5.73 times less likely to be classified as morally dumbfounded when presented with the more realistic Extended Soul task than the original Soul task. Furthermore, judgments of task realism mirrored the observed prevalence of moral dumbfounding for each task.

As stated above, the Lolly task was perceived as significantly more realistic than the Roach task and there was significantly less moral dumbfounding elicited in response to the Lolly task than the Roach task. There were also no differences in realism reported between the reasoning vignettes (Heinz and John), sexual taboo vignettes (Incest and Masturbation), or the cannibalism vignettes (Cannibal and Liver), and there was no difference in moral dumbfounding between these matched stimuli. Finally, participants in Studies 1 and 3 did not report any difference in realism for the Atlas video nor the Spot video and, again, there was no difference in dumbfounding in response to either video between Studies 1 and 3. Despite the trend, overall, there was no relationship between belief that the action is realistic and moral dumbfounding. Nonetheless, the perception of presented stimuli as representative of a situation that may happen in the real world is an element that should be further explored empirically. This is especially
so given that the actions depicted in the video excerpts were perceived as significantly more realistic than both the intuition vignettes and the behavioural tasks.

The rationale for having participants witness actual harmless yet offensive moral transgressions (i.e., Atlas and Spot; Boston Dynamics, February 2, 2016; Boston Dynamics, February 9, 2015) was to circumvent criticism of the stimuli typically employed to elicit moral dumbfounding being unrealistic (s.e., Kennett, 2012; Royzman et al., 2015). Not only was the footage perceived as significantly more realistic than the other stimuli, but that moral dumbfounding was elicited by Atlas and Spot (4.55%) substantiates these videos as suitable stimuli in this context. In addition, 20.45% of participants initially judged the action depicted in Atlas as ‘wrong’ (final judgments of ‘wrong’ comprised 27.27%) and 22.73% of participants initially judged the action depicted in Spot as ‘wrong’ (final judgments of ‘wrong’ comprised 18.18%). These results confirm that presentation medium affects both moral judgment (i.e., if an action is perceived as wrong) and justification (i.e., why a transgression is wrong; McCurrie et al., 2018). In addition, they validate that there is an ongoing expansion of moral concern (Crimston et al., 2018; Laham, 2009), the capacity to experience suffering has been decoupled from the right to hold moral status (Torrance, 2013), and some people feel a morally obligated to treat robots humanely.

It was also claimed that the prevalence of dumbfounding to such cases (i.e., witnessed third-party behaviour) would provide a more accurate reflection of the rate of moral dumbfounding in the real world (compared with those elicited by hypothetical vignettes). This was suggested because enhancing available contextual information reduces mental simulation and brings moral judgments more in line with real-life moral behaviour (FeldmanHall et al., 2012). Contrary to prediction, the prevalence of dumbfounding did not vary significantly between witnessed (4.55%) and hypothetical third-party actions (15.91%). Nonetheless, that moral dumbfounding was elicited using a different presentation medium suggests that this phenomenon can be as reliably elicited by hypothetical and actual behaviour.

At any rate, differences in perception that the stimuli are realistic does not seem to account for the different rates of dumbfounding seen in Studies 1 and 3. Instead, the drop in the dumbfounding rate from Study 1 to Study 3 appears due to differences in operationalisation of the phenomenon. Keeping the stimuli constant, there was a difference in the rate of moral dumbfounding as a function of definition. Participants were classified as morally dumbfounded with significantly greater frequency under the
Common definition (i.e., maintained judgment without provision of supporting reasons) than under the Evidence-based definition (an inability to justify disapproval for a seemingly harm-free action). That moral dumbfounding was found (albeit at a lower rate) with a more precise operationalisation to that typically utilised, and using alternative measures to those employed in extant research, shows that dumbfounding is a genuine, replicable empirical phenomenon.

**Replication analyses of Haidt et al. (2000) and Study 1**

For ease of comparison between the original and alternative measures, the remaining discussion preserves the presentation format of Study 1.

**Claim 1: Differences between the reasoning stimulus and the intuition stimuli**

**Judgments of wrongness of an action**
Contrary to prediction, there were fewer judgments that the action depicted was wrong to the food consumption task, *Lolly*, than to the reasoning measure, *John*. The rarity of refusal to *Lolly* (only six participants would not eat the lolly) explains this result. Also against expectation, the *Soul* task and the intuition vignettes did not differ significantly from the reasoning vignette in judgments of wrongness (both initial and final judgments). Given this, there does not appear to be a perceptual permissibility distinction in the nature of tasks (i.e., intuition and reasoning) as Haidt et al. (2001) claim.

**Change of judgment**
In the current study (as in Haidt et al., 2000, and Study 1) there was no difference in the tendency for participants to change their mind from initial to final judgment. The relationship between maintenance of judgment and moral dumbfounding was further explored here to verify the argument (made in Study 2) that the definition of moral dumbfounding as simply the maintenance of a judgment without provision of supporting reasons (i.e., the *Common* definition) is an unacceptable characterisation of the phenomenon. Analysis showed that there is no relationship between the maintenance of a judgment of ‘wrong’ and moral dumbfounding. This suggests that the definition of moral dumbfounding derived in Study 2 (i.e., an inability to justify disapproval for a seemingly harm-free action) more accurately represents the phenomenon than when it is defined as stubbornly adhering to a judgment despite the
removal of reasons used to support condemnation (sc., Haidt, 2001; Haidt & Björklund, 2008; Haidt & Hersh, 2001; Jacobson, 2012; and Royzman et al., 2015).

**Psychic discomfort**

In the current study, there was an extremely large differential relationship between the reasoning measure and the behavioural tasks (*Lolly* and *Soul*) in manifestations of psychic discomfort. Neither intuition video, nor the *Masturbation* vignette, differed significantly to the reasoning stimulus in psychic discomfort. These results support conclusions drawn from Study 1 that being asked to respond with an imminent personal action (i.e., presented with a behavioural task) is vastly different from simply providing a third-person moral judgment (i.e., the *John* and *Masturbation* vignettes, and both video excerpts). In contrast to Study 1 (in which there was no difference in psychic discomfort between the reasoning measure and the *Cannibal* vignette), the *Liver* vignette elicited significantly greater psychic discomfort than the current study reasoning measure, *John*. The reason for the discrepancy is unclear. It may be that participants are more uncomfortable with a theatre nurse who violates her workplace protocol (*Liver*) than a scientist (*Cannibal*), or perhaps there is more discomfort from the thought of eating human meat from a live donor (*Liver*) as opposed to that from a corpse (*Cannibal*). Redesigning the *Liver* vignette to bring it more in line with the original *Cannibal* story may reduce the discrepancy.

**Unique justifications**

Unique justifications were taken to indicate persistence in defence of a judgment. All intuition stimuli differed significantly from the reasoning stimulus in the number of unique justifications offered to support a judgment of ‘wrong’. While participants made significantly more unique justifications for their judgments in response to the intuition vignettes than to the reasoning vignette (like Study 1), the remaining intuition stimuli elicited significantly fewer unique justifications than *John*. There was, however, no correlation between unique justifications and moral dumbfounding, suggesting that persistence in defence of a judgment is not, as tentatively proposed in Study 1, a suitable alternative indicator of moral dumbfounding than is the maintenance of a judgment. Further exploration of the association between moral dumbfounding and unique justifications is warranted.
Claim 2: Moral judgment is an intuitive response

Response timing
Time to initial judgment, and time to initial justification, for the food consumption task was significantly faster than for the reasoning stimulus (like Study 1). Corresponding to the Roach task, the fast response time to the Lolly task reflects the manner in which the vast majority of participants immediately ate the lolly without any discussion. No other intuition stimuli differed significantly from the reasoning measure in time to initial judgment. The lack of difference in response time between the intuition stimuli (temporarily discounting the Lolly task) and the reasoning measure suggests (once again) that the source of responses to all measures was equivalent; that is, the judgments and justifications came from an equally fast (or slow) process. Moreover, moral dumbfounding was (for a second time) not related to the speed with which participants offered their judgment. Accordingly, neither the current study nor Study 1 can be taken as straightforward confirmation of the view that moral judgment is a quick, intuitive reaction.

Basis underpinning judgments
Participants were significantly more likely to base their initial judgment on reasoning when responding to the Atlas video than to the John vignette (although non-significant, the Spot video demonstrated the same pattern). Participants (except for two) were unfamiliar with the video footage and the depicted technology meant that they could not rely upon prior knowledge and established social opinions (i.e., unconscious automatised prior reasoning) when making their judgments and justifications. Haidt (2001) asserts that reasoning is most likely to be used when we are exposed to novel moral situations, and the results regarding Atlas and Spot appear support this notion. Nonetheless, reiterating from Study 1, the basis underpinning judgment was self-reported and, therefore, any conclusions are undermined by the limits of introspection.

Claim 3: Moral judgment is emotionally based
Like Study 1, the basis underpinning participants’ initial judgment for all measures was gut feeling (the mean of initial judgments was below the balance point of the gut feeling and reasoning continuum indicating that more gut feeling and less reasoning was used as the basis of judgment). Once again, however, moral dumbfounding was not related to the basis upon which participants reportedly made their judgment (i.e., reasoning or gut feeling). Consequently, rather than confirming that moral dumbfounding is evidence of the process underlying our moral judgments (sc., Haidt, 2001), it can only be concluded
that participants report (or at best, believe) that the source of their initial response to the majority of presented moral dilemma is feeling.

**Claim 4: Reasoning is post-hoc in the formation of moral judgment**

**Basis underpinning judgments**

Corresponding to Study 1, participants reported basing their final judgment for each stimulus significantly more on reasoning than they did their initial judgment (with the exception of the Lolly task). The efficacy of reasoning in moral judgment cannot be excluded given that 34.1% of participants changed their judgment from initial to final judgment (across tasks). That people do change their mind suggests that moral judgment (even if not initially derived through conscious reasoning) is rationally evaluated and open to modification; thus, reasoning plays a causal role in the formation of moral judgment and is not merely post-hoc (in opposition to the claim of Haidt, 2001).

**Response timing**

The sequence of responses to all presented measures was justification prior to judgment (except for the Soul task). In addition, the time to initial justification for the Liver vignette, the Lolly task, and the Atlas video was significantly faster than John. These results support the findings of Study 1 but are in contrast to those of Haidt et al. (2000), who observed that participant responses to the intuition cases (i.e., those that elicit moral dumbfounding) typically followed a pattern of quick judgment followed by justification. Furthermore, in the current study, swift justification was associated with less likelihood of being dumbfounded by one’s own judgment (indeed there was no relationship between time to initial judgment and dumbfounding). These results challenge the claim that moral dumbfounding reveals that reasons given to support a judgment are rationalisation of an already made intuitive judgement. (Haidt, 2001; Haidt et al., 2000; Haidt & Hersh, 2001; McHugh et al., 2017). It is important to remember, however, that observations of response timing do not conclusively reveal an internal process (as Haidt proposes); thus, despite these results, the current study cannot be taken as evidence against Haidt’s position that intuitive judgment precedes reasoned justification (Haidt, 2001; Haidt et al., 2000; Haidt & Hersh, 2001).
Claim 5: Moral dumbfounding is evidence of judgment without reason

Statements of "I don’t know"
The linguistic responses of participants to the intuition stimuli were hypothesised to contain significantly more statements of "I don’t know" than their responses to the reasoning stimulus. Participants did state "I don’t know" more in response to the Extended Soul task than to the reasoning stimulus (like Study 1). There were also more statements of "I don’t know" in response to Masturbation and Spot compared with John. Again, while some may argue that an explicit admission of not knowing why one has made a judgment is a definitive measure of moral dumbfounding (Haidt, 2001; Haidt et al., 2000; Haidt & Hersh, 2001; McHugh et al., 2017), this is not inevitably so given that such responses may include instances of low motivation, failure of introspection, and the like. Statements of "I don’t know", therefore, reveal the participant has failed to provide a reason rather than they are necessarily without reason for their judgment.

Failure to provide a reason
Participant responses were found to have a significantly greater failure to provide a reason for a judgment for the same three abovementioned intuition stimuli (i.e., Masturbation, Soul, and Spot), and the Liver vignette, compared with John. These results mirror Study 1 in which there was more failure to provide a reason for a judgment of ‘wrong’ to the majority of intuition stimuli compared with the reasoning stimulus (with the exception of the food consumption task; the Atlas video was not analysed in Study 1 and did not differ significantly from John in the current study). The results of the individual elements that comprise the composite ‘failure to provide a reason’ variable (i.e., unsupported declarations, tautological reasons, and illogical arguments) were also comparable to their Study 1 counterparts. The only differences between the studies were that, in Study 1, there were significantly more tautological reasons offered in response to the sexual taboo vignette than to the reasoning vignette, and there were significantly more illogical arguments offered in response to both the sexual taboo vignette and the food consumption task than to the reasoning vignette. These differences did not occur in Study 3. Notwithstanding these differences, it appears that the reasoning and intuition measures from Studies 1 and 3 performed comparably in occasioning a failure to provide a reason for judgment.
Concerning such a failure, it appears that the Spot video performs similarly to the other intuition stimuli whereas the Atlas video does not. The Spot and Atlas videos are analogous except that Spot is an animalian robot ‘dog’ whereas Atlas is a humanoid robot, and Spot does not fall to the ground whereas Atlas does. Whether the performance difference of the two videos is due to a perceptual difference between the form of the robot or the result of the action (or, indeed, some other factor) are questions for future research. Such research is becoming increasingly important given the rapid advancement of technology; robots are beginning to act and look progressively more like living beings, so it is increasingly hard not to perceive them in this way. That some participants exhibited moral dumbfounding in response to depictions of the mistreatment of a robot further confirms that these videos are ideal additions to the set of stimuli employed to elicit this phenomenon.

Struggle to produce a reason
There was a significantly greater struggle to produce a reason when responding to the sexual taboo vignette and the Soul task, than to the reasoning vignette (mirroring Study 1). The replication of results from Study 1 continued as the remaining intuition stimuli did not differ from the reasoning stimulus with respect to the variable ‘struggle to produce a reason’ and there was no difference in dumbfounding statements between the stimuli. The variable ‘unwillingness to support’ (encompassing statements supporting an acceptability judgment for the scenario or task, but reluctance to commit to such a judgment) was again shown to be a better measurable indicator than dumbfounding statements, as there were significantly more statements of unwillingness to support in response to the Soul task compared with the reasoning vignette in both the current study and Study 1. Unlike Study 1 (in which the sexual taboo vignette and the food consumption task elicited significantly more statements of unwillingness to support than the reasoning vignette), there were no further differences between any of the other measures and the reasoning measure in the current study. Nonetheless, the current findings are compatible with Haidt et al. (2000) and Study 1 insofar as some participants explicitly struggled to provide a reason for judgment.

Puzzlement
Like Study 1, puzzlement at an inability to produce justification for judgment in the current study did not differ significantly between the stimuli, and all participants who made a statement that they were puzzled by their inability to produce a justification for their judgment of ‘wrong’ were independently classified as morally dumbfounded.
Results from both studies indicate that puzzlement is a sufficient (but not necessary) condition for moral dumbfounding; regardless of differences in the presenting stimulus; if one is puzzled at an inability to produce a reason for one’s judgment they are, necessarily, dumbfounded. Indeed, Haidt suggests that true moral dumbfounding only occurs when one is bewildered: “...it’s only when they reach deep into their pocket for another reason, and come up empty-handed, that they enter the state we call ‘moral dumbfounding’. Because they fully expect to find reasons. They’re surprised when they don’t find reasons” (2005, para 3). Nonetheless, in Studies 1, 2, and 3, such bewilderment was not correlated with moral dumbfounding (although this result may be due to the small number of participants who articulated puzzlement). The majority of morally dumbfounded participants did not report being puzzled by their inability to produce a justification for their judgment. It makes sense that puzzlement will signify dumbfounding for some people whereas others will not articulate or explicitly demonstrate this internal state. As such, while puzzlement has value as a clear signal of dumbfounding, the presence of such bewilderment is not a necessary defining feature of the phenomenon.

**Limitations**

Haidt (Haidt, 2001; Haidt et al., 2000; Haidt & Hersh, 2001) confines his discussion of moral dumbfounding to situations in which participants judge an action to be wrong but are unable to provide a reason (or at least an accepted reason) for this judgment. To confine discussion within the scope set by Haidt, this dissertation also limits moral dumbfounding to judgments of wrongness of an action. That a participant in the current study demonstrated all the hallmarks of moral dumbfounding, but was not classified as such because they did not judge the situation to be wrong, emphasises the need for future research to extend the scope of moral dumbfounding to illuminate situations pertaining to judgments of permissibility made without justification.

**Conclusion**

Moral dumbfounding, defined as an inability to justify disapproval for a seemingly harm-free action, was elicited in the present study using a range of novel scenarios demonstrating that the phenomenon does not only occur under particular conditions (i.e., in response to the limited set of stimuli employed in extant research; Haidt et al., 2000; McHugh et al., 2017; Royzman et al., 2015). The prevalence of moral dumbfounding (27.3%) corresponds to that obtained in Study 2 (22.7% elicited using the original stimuli), which provides further empirical evidence that we are unable to
justify our own moral judgments (of 'wrong') approximately one-quarter of the time. While this study provides a stronger case than existing research for the notion that moral dumbfounding is likely to occur in situations one may encounter in everyday life, it is noted that participants did not perceive the majority of novel stimuli as any more realistic than the original stimuli that have been criticised for being extreme and unusual (s.e., Jacobson, 2012; Kennett, 2012). It may be that explicit judgments of realism do not reflect implicit judgments of realism. That perceived realism does play a key role in moral dumbfounding is reflected in the fact that prevalence of moral dumbfounding did not vary significantly between the reasoning and intuition stimuli, and two videos depicting real-life scenarios (in which harm was not artificially removed) were not only perceived as significantly more realistic than the other stimuli, but they elicited moral dumbfounding.

The inability of some participants to justify disapproval to the novel cases presented here substantiates that our moral judgments to everyday dilemmas may reflect responses that are difficult to defend. Prior to defending this conclusion, an investigation of individual proneness toward moral dumbfounding is warranted, as it may be that participants with certain demographic or personality characteristics are responsible for a disproportionate amount of the observed prevalence. Such an investigation will further aid understanding of this important phenomenon.
Chapter 6: Study 4 – Exploring Individual Differences in Moral Dumbfounding

Haidt and colleagues (Haidt, 2001; Haidt et al., 2000) assume that moral dumbfounding is typical of moral judgment, but little attention is paid to individual who were not morally dumbfounded. It may be the case that individual differences predict moral dumbfounding. Given that moral judgment is affected by both internal and external factors (Karamavrou, Mouratidou, Evaggelinos, Koidou, & Parisi, 2016), it is plausible that demographic and personality characteristics predispose individuals to dumbfounding. Participants in moral dumbfounding studies are typically pressured to respond quickly (perhaps precluding deliberation) such that their moral judgment is elicited under conditions unrepresentative of the usual scope of moral reasoning (Jacobson, 2012; Kasachkoff & Saltzstein, 2008; Kennett, 2012; Kennett & Fine, 2009; Mackenzie, 2012; Royzman et al., 2015; Saltzstein & Kasachkoff, 2004). Those who are more susceptible to such pressure may, therefore, be more likely to manifest moral dumbfounding (as an artefact of the experimental paradigm). The results of Study 3 confirm that the phenomenon does not only occur in response to a small subset of stimuli, but do particular conditions make moral dumbfounding more likely to occur? By examining the effect of both static and dynamic factors (i.e., demographic and personality characteristics, and susceptibility to situational influences, respectively) on the phenomenon, we get closer to establishing whether this phenomenon is typical of everyday moral judgments.

**Individual characteristics and moral dumbfounding**

Some researchers who were capable of addressing the possible influence of personological variables on moral dumbfounding neglected to do so. McHugh et al. (2017) report descriptive statistics for gender, age, and level of education, but they did not investigate the extent to which these variables were related to the phenomenon. Haidt et al. (2000) asked participants to indicate their religious ideology and political preference, but did not explore the relationship between these variables and dumbfounding any further. In the very scant research to explicitly investigate moral dumbfounding and address demographic factors, no relationship was found between moral dumbfounding and gender (Haidt et al., 2000; Haidt & Hersh, 2001; Royzman et al., 2015), religious ideology (McHugh et al., 2017), or political preference (Haidt & Hersh, 2001).
Nonetheless, past research has found that gender, age, ethnicity, socioeconomic status, religiosity, and political preference are all associated with moral concerns (Koleva, Graham, Iyer, Ditto, & Haidt, 2012). Reactions to moral transgressions are influenced by both cultural and individual factors (Laham, Chopra, Lalljee, & Parkinson, 2010) and research using measures of ‘harmless’ taboo violation stories (but not specifically investigating moral dumbfounding) has found some such differences in moral judgment. Relevant research is discussed below; included is research which addresses the impact of individual variables on what researchers claim to be moral dumbfounding. These studies were previously excluded from discussion because they fail to provide empirical evidence of the existence of this phenomenon; that is moral dumbfounding is inadequately defined (as, for example, simply the provision of ‘unsupported declarations’ which should not be used as the sole measure of an inability to justify judgment; s.e., McHugh et al., 2017; Royzman et al., 2015). Such studies are only presented here as peripheral research upon which to derive possible hypotheses.

**Gender**

Gilligan (1982) famously suggested that gender differences exist in moral judgment; however, neither Hauser et al. (2007) nor Haidt, Koller, and Dias (1993) found gender differences in the moral judgments of participants presented with harmless taboo violations. In contrast, Tepe, Piyale, Sirin, and Sirin (2016) found female participants to be less permissive than males in moral judgments made in response to the same five stories employed by Haidt et al. (1993). Royzman et al. (2015), too, found females significantly more likely to indicate that the depicted action in Incest was wrong compared to males. Nonetheless, as stated above, variance in moral dumbfounding due to gender has not yet been found (Haidt et al., 2000; Haidt & Hersh, 2001; Royzman et al., 2015); thus, it is important to further investigate whether males or females are prone to moral dumbfounding.

**Age**

Just as Gilligan (1982) suggested gender differences, Kohlberg (1969) proposed lifespan disparities in moral judgment. Nonetheless, it appears that age has no effect on moral judgments elicited using harmless taboo violation stories (Haidt et al., 1993; Hauser et al., 2007; Uhlmann & Zhu, 2013). Whether age has any influence on moral dumbfounding is yet to be addressed.
Nationality and ethnicity
Haidt et al. (1993) conclude that morality varies across cultures. American participants were more permissive of harmless taboo violations than their Brazilian counterparts, who found it harder than the Americans to justify their moral judgment. Given concerns about relying solely on unsupported declarations as a measure of moral dumbfounding (McHugh et al., 2017; Royzman et al., 2015), evidence for their claim is weak. Furthermore, nationality had no effect on judgments of harmless taboo violation stories for Piazza and Sousa’s participants (from America and India; 2014), nor those of Hauser et al. (from Australia, Brazil, Canada, India, United States, and United Kingdom; 2007). There was also no relationship between ethnicity and moral judgments for the participants of Hauser and colleagues (i.e., American Indian, Alaskan Native, Asian, Pacific Islander, Black Non-Hispanic, Hispanic, and White Non-Hispanic). Based on these studies, it appears unlikely that nationality and ethnicity would have an influence on moral dumbfounding; however, further research is required.

Education
While educational level is not associated with moral dumbfounding (elicited using web-based technology; Hauser et al., 2007; Uhlmann & Zhu, 2013), type of education may have an influence on the phenomenon. Hauser et al. (2007) found that, from over 5,000 participants, only those who had philosophical exposure were able to provide a sufficient justification for their moral judgments. The participants in the original dumbfounding study (Haidt et al., 2000) were first-year psychology students and may have been ill-equipped to argue the nuances of their justifications as they were without formal training in philosophy. Consequently, those studying philosophy or a related course with explicit logical reasoning exposure may be less likely to become morally dumbfounded. Hauser and colleagues collected data online in which the accuracy of provided demographic information cannot be verified (i.e., some participants may have falsely indicated they were PhD or philosophy students for social desirability). Whether education type has any influence on moral dumbfounding during face-to-face interviews (or, indeed, at all) remains undetermined.

Socioeconomic status
Haidt et al. (1993) report a large and consistent effect of social class (SES) in which participants with low SES were less permissive, and responded with significantly more ‘unsupported declarations’, than those with high SES. Nonetheless, all of the low-SES participants were black, and all of the high-SES participants were white (regardless of
their nationality). While Haidt et al. argue that the characteristics of participants such as SES, race, and type of education are not necessarily confounds ("It is impossible to have two groups of people who vary only by SES, or only by westernization, without also varying on dozens of other variables", p. 617), the association should not be so easily dismissed as it influences how research results are interpreted. With only one study addressing moral dumbfounding and SES, further empirical research is warranted.

Religious ideology
While the type of religion practiced by participants was found to have no effect on moral judgments of harmless taboo violations (Hauser et al., 2007), those with a stronger belief in religion (regardless of denomination) are more likely to evaluate these scenarios as wrong (Gladden, Welch, Figueredo, & Jacobs, 2009; Haidt & Hersh, 2001; Tepe et al., 2016). One possible explanation for this finding is that those with stronger religious beliefs may adhere to the moral rules espoused by their religion and be less likely to engage in a careful, cost-benefit analysis of moral situations (s.e., Pennycook, Ross, Koehler, & Fugelsang, 2016). Indeed, the religious participants of Piazza and Sousa (2014) were more likely to persist in their condemnation of the Incest and Cannibalism scenarios than less religious participants, but this relationship was mediated by a philosophical reasoning style. Specifically, the Incest and Cannibalism scenarios are portrayed as having no negative consequences, but religious participants disregarded this outcome-based consideration and judged the actions as wrong by appealing to rules dictating that such actions are never morally permissible. In moral dumbfounding studies, statements such as, “it’s just wrong” are coded as an unsupported declaration yet such statements may reflect an appreciation of a religious rule. It is worthwhile here to remember that moral dumbfounding is moral judgment without justification, not moral judgment with poorly articulated justification (s.e., Jacobson, 2012). Accordingly, religious strength may be associated with moral dumbfounding; however, with only one study having explicitly investigated the relationship between the phenomenon and religiosity (and finding no relationship; McHugh et al., 2017), a definitive conclusion cannot yet be drawn.

Political preference
Piazza and Sousa (2014) again found philosophical reasoning style (i.e., consequentialist versus deontological) to have a mediating role, this time between impermissibility judgments (to the Incest and Cannibalism scenarios) and political preference; compared to liberal participants, political conservatives were more likely to
condemn the depicted actions because they were resistant to consequentialist reasoning. Royzman et al. (2015) also found that politically conservative participants were more likely to indicate that the *incest* scenario is wrong than were liberal participants. Setting aside judgments, no relationship has yet been found between moral dumbfounding and political preference (Haidt & Hersh, 2001; Uhlmann & Zhu, 2013). With only two studies from which to draw conclusions, more research is needed to ascertain if political preference has an influence on moral dumbfounding.

**Personality**

To date, no research has explored whether personality influences moral dumbfounding. Personality traits (such as openness to experience and extroversion) have been associated with beliefs about the nature of morality (i.e., if morality is objective or relative; Knobe, 2011) as well as various moral topics (such as free will and responsibility; Feltz & Cokely, 2009). A relationship between personality and broad patterns of moral responding has also been found (Graham et al., 2013) including moral judgment (Bartels & Pizarro, 2011). Consequently, moral dumbfounding may depend upon established patterns of responding by the moral judge such that individuals with certain personality traits are more or less likely to become morally dumbfounded. It is through the lens of the five broad domains of openness to experience, conscientiousness, agreeableness, extraversion, and neuroticism (i.e., the Five Factor Model of Personality proposed by Costa & McCrae, 1992), that the possible influence of personality traits on moral dumbfounding is investigated.

**Demand characteristics and moral dumbfounding**

Psychological experiments are a social situation, as there is an interaction between participant and researcher (Rosenthal, 1969/2009). In social situations, norms guide behaviour and are enforced by social sanctions (Goffman, 1963, 1971). The strong natural desire for social acceptance has been demonstrated in many social psychological studies (for a popular example see Solomon Asch’s line judgment experiment; 1951). The social norm in psychological studies relates to the experimenter as an authority figure and the participant as a compliant subject (Milgram, 1974). It is reasonable to assume that the original dumbfounding study by Haidt et al. (2000) would have been no exception; to comply with expectation, participants would have looked for social cues.

Participants use cues from the experimenter as guidance on how to fulfil their role as a good participant (sc., Ome, 1962). The cues from Haidt et al.’s (2000) experimenter would have included the insistence that the infractions embodied in
scenarios are harmless and, therefore, the depicted actions should be permissible. Not only does compliance (and noncompliance) with social norms reveal the moral character of individuals (Goffman, 1971), but making a moral judgment (especially about disgusting scenarios upon which Haidt et al. mostly relied) is an expression to others of who you are as a moral person (Giner-Sorolla & Chapman, 2016). Consequently, some participants may have been unwilling to provide harm-based justifications (despite their judgment being based on considerations of harm) because they yielded to pressure from the experimenter to fulfil the role of ‘good participant’.

The most famous demonstration of pressure from the experimenter is Milgram’s obedience study (1974). While Milgram’s study is typically understood to demonstrate a natural tendency for people to obey the instructions of those in authority, Haslam and colleagues (Haslam, Reicher, & Birney, 2014; Haslam, Reicher, & Millard, 2015; Reicher, Haslam, & Smith, 2012) have found that it is not a natural tendency, nor the instruction as such, that predicts willingness to obey, but the person’s identification with the researcher and the research endeavour. The psychology undergraduates in the study of Haidt et al. (2000) were rewarded for participating (i.e., received course credit). Given this, and the topic of the study, identification with the research was likely amplified and, hence, so was willingness to comply with experimenter expectation. Coupled with strong natural desires to be socially accepted and provide a favourable moral impression, the participants may have been highly susceptible to cues provided by the experimenter. The moral dumbfounding observed by Haidt et al. may (at least partly) reflect an inability to articulate justification in light of experimenter pressure. This is evidenced by the research of Royzman et al. (2015).

Despite retracting their verbal justifications for judgments to the Incest story, some of Royzman and colleagues’ (2015) participants reported that they did not believe that there would be no harmful consequences deriving from the situation (in spite of the experimenter’s insistence). Upon being asked to account for their response inconsistency, all relevant participants acknowledged that they should not have recanted their justification with the majority stating that they felt pressured to “respond ‘under the assumption’ that no harm has occurred or will occur” (Royzman et al., p. 304). The researchers concluded that it was the influence of the social dynamics of the interview process (i.e., experimenter pressure) that led to the observed moral dumbfounding for all except one participant (no explanation was offered for this participant’s dumbfounding).
It is, therefore, hypothesised that participants susceptible to experimenter pressure (i.e., high in identification with the experiment) who judge the depicted action as wrong will be more prone to moral dumbfounding than others because cues from the experimenter (i.e., stipulation that there is no harm) will render them unwilling to provide harm-based justifications. Given limited extant research, no a priori hypotheses are made regarding other individual differences. Relationships between moral dumbfounding and each of the independent variables of gender, age, nationality and ethnicity, education, SES, religious ideology, political preference, and personality traits will be explored.

Method

Participants

The sample consisted of 88 Australian university students ranging in age between 18 and 63 years ($M = 26.97$, $SD = 11.54$). Male participants were aged between 18 and 63 years ($M = 24.52$, $SD = 10.31$). Female participants were aged between 18 and 57 years ($M = 28.17$, $SD = 12.01$). Table 39 presents the demographic characteristics for the sample (refer to the Measures and Procedure section below for information regarding the items utilised to elicit the data, as well as information on political preference and personality traits that do not appear in the following table as they are reported using means rather than frequencies).

Table 39
Descriptive Demographic Statistics of the Sample for Study 4 (and Studies 1 and 3 that Comprise this Sample)

<table>
<thead>
<tr>
<th>Demographic Variable</th>
<th>Study 1</th>
<th></th>
<th>Study 3</th>
<th></th>
<th>Study 4</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$n$</td>
<td>%</td>
<td>$n$</td>
<td>%</td>
<td>$n$</td>
<td>%</td>
</tr>
<tr>
<td>Gender</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>15</td>
<td>34.1</td>
<td>14</td>
<td>31.8</td>
<td>29</td>
<td>33.0</td>
</tr>
<tr>
<td>Female</td>
<td>29</td>
<td>65.9</td>
<td>30</td>
<td>68.2</td>
<td>59</td>
<td>67.0</td>
</tr>
<tr>
<td>Nationality</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Australian</td>
<td>42</td>
<td>95.5</td>
<td>41</td>
<td>93.2</td>
<td>83</td>
<td>94.3</td>
</tr>
<tr>
<td>Mauritian</td>
<td>1</td>
<td>2.3</td>
<td>0</td>
<td>0.0</td>
<td>1</td>
<td>1.1</td>
</tr>
<tr>
<td>New Zealander</td>
<td>0</td>
<td>0.0</td>
<td>1</td>
<td>2.3</td>
<td>1</td>
<td>1.1</td>
</tr>
<tr>
<td>Sri Lankan</td>
<td>0</td>
<td>0.0</td>
<td>1</td>
<td>2.3</td>
<td>1</td>
<td>1.1</td>
</tr>
<tr>
<td>British, Welsh, Scottish (i.e., from United Kingdom)</td>
<td>1</td>
<td>2.3</td>
<td>1</td>
<td>2.3</td>
<td>2</td>
<td>2.3</td>
</tr>
<tr>
<td>Ethnicity</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Aboriginal or Torres Strait Islander</td>
<td>1</td>
<td>2.3</td>
<td>1</td>
<td>2.3</td>
<td>2</td>
<td>2.3</td>
</tr>
<tr>
<td>Australian</td>
<td>40</td>
<td>90.8</td>
<td>39</td>
<td>88.6</td>
<td>78</td>
<td>88.6</td>
</tr>
<tr>
<td>Demographic Variable</td>
<td>Study 1</td>
<td>Study 3</td>
<td>Study 4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>--------------------------------</td>
<td>---------</td>
<td>---------</td>
<td>---------</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>$n$</td>
<td>$%$</td>
<td>$n$</td>
<td>$%$</td>
<td>$n$</td>
<td>$%$</td>
</tr>
<tr>
<td>Chinese</td>
<td>1</td>
<td>2.3</td>
<td>0</td>
<td>0.0</td>
<td>1</td>
<td>1.1</td>
</tr>
<tr>
<td>Dutch-Peruvian</td>
<td>0</td>
<td>0.0</td>
<td>1</td>
<td>2.3</td>
<td>1</td>
<td>1.1</td>
</tr>
<tr>
<td>Maltese</td>
<td>0</td>
<td>0.0</td>
<td>1</td>
<td>2.3</td>
<td>1</td>
<td>1.1</td>
</tr>
<tr>
<td>New Zealander</td>
<td>0</td>
<td>0.0</td>
<td>1</td>
<td>2.3</td>
<td>1</td>
<td>1.1</td>
</tr>
<tr>
<td>Multietnic</td>
<td>1</td>
<td>2.3</td>
<td>0</td>
<td>0.0</td>
<td>1</td>
<td>1.1</td>
</tr>
<tr>
<td>Scottish</td>
<td>1</td>
<td>2.3</td>
<td>0</td>
<td>0.0</td>
<td>1</td>
<td>1.1</td>
</tr>
<tr>
<td>Sinhalese</td>
<td>0</td>
<td>0.0</td>
<td>1</td>
<td>2.3</td>
<td>1</td>
<td>1.1</td>
</tr>
<tr>
<td><strong>University Enrolment</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Australian National University</td>
<td>1</td>
<td>2.3</td>
<td>0</td>
<td>0.0</td>
<td>1</td>
<td>1.1</td>
</tr>
<tr>
<td>C3 College</td>
<td>1</td>
<td>2.3</td>
<td>0</td>
<td>0.0</td>
<td>1</td>
<td>1.1</td>
</tr>
<tr>
<td>Charles Sturt University</td>
<td>41</td>
<td>93.2</td>
<td>36</td>
<td>81.8</td>
<td>77</td>
<td>87.5</td>
</tr>
<tr>
<td>Edith Cowan University</td>
<td>0</td>
<td>0.0</td>
<td>1</td>
<td>2.3</td>
<td>1</td>
<td>1.1</td>
</tr>
<tr>
<td>University of New England</td>
<td>0</td>
<td>0.0</td>
<td>1</td>
<td>2.3</td>
<td>1</td>
<td>1.1</td>
</tr>
<tr>
<td>University of New South Wales</td>
<td>0</td>
<td>0.0</td>
<td>2</td>
<td>4.5</td>
<td>2</td>
<td>2.3</td>
</tr>
<tr>
<td>University of Newcastle</td>
<td>1</td>
<td>2.3</td>
<td>2</td>
<td>4.5</td>
<td>3</td>
<td>3.4</td>
</tr>
<tr>
<td>University of Technology Sydney</td>
<td>0</td>
<td>0.0</td>
<td>1</td>
<td>2.3</td>
<td>1</td>
<td>1.1</td>
</tr>
<tr>
<td>University of Wollongong</td>
<td>0</td>
<td>0.0</td>
<td>1</td>
<td>2.3</td>
<td>1</td>
<td>1.1</td>
</tr>
<tr>
<td><strong>Level</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Undergraduate</td>
<td>35</td>
<td>79.5</td>
<td>36</td>
<td>81.8</td>
<td>71</td>
<td>80.7</td>
</tr>
<tr>
<td>Postgraduate</td>
<td>9</td>
<td>20.5</td>
<td>8</td>
<td>18.2</td>
<td>17</td>
<td>19.3</td>
</tr>
<tr>
<td><strong>Course</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bachelor of Advertising</td>
<td>1</td>
<td>2.3</td>
<td>1</td>
<td>2.3</td>
<td>2</td>
<td>2.3</td>
</tr>
<tr>
<td>Bachelor of Arts</td>
<td>2</td>
<td>4.5</td>
<td>4</td>
<td>9.1</td>
<td>6</td>
<td>6.8</td>
</tr>
<tr>
<td>Bachelor of Business</td>
<td>1</td>
<td>2.3</td>
<td>1</td>
<td>2.3</td>
<td>2</td>
<td>2.3</td>
</tr>
<tr>
<td>Bachelor of Clinical Practice (Paramedic)</td>
<td>1</td>
<td>2.3</td>
<td>2</td>
<td>4.5</td>
<td>3</td>
<td>3.4</td>
</tr>
<tr>
<td>Bachelor of Criminal Justice</td>
<td>3</td>
<td>6.8</td>
<td>0</td>
<td>0.0</td>
<td>3</td>
<td>3.4</td>
</tr>
<tr>
<td>Bachelor of Diagnostic Radiography</td>
<td>0</td>
<td>0.0</td>
<td>1</td>
<td>2.3</td>
<td>1</td>
<td>1.1</td>
</tr>
<tr>
<td>Bachelor of Education</td>
<td>4</td>
<td>9.1</td>
<td>3</td>
<td>6.8</td>
<td>7</td>
<td>8.0</td>
</tr>
<tr>
<td>Bachelor of Engineering / Bachelor of Business</td>
<td>0</td>
<td>0.0</td>
<td>1</td>
<td>2.3</td>
<td>1</td>
<td>1.1</td>
</tr>
<tr>
<td>Bachelor of Health and Rehabilitation Science</td>
<td>0</td>
<td>0.0</td>
<td>1</td>
<td>2.3</td>
<td>1</td>
<td>1.1</td>
</tr>
<tr>
<td>Bachelor of Mechatronic Engineering / Master of</td>
<td>0</td>
<td>0.0</td>
<td>1</td>
<td>2.3</td>
<td>1</td>
<td>1.1</td>
</tr>
<tr>
<td>Demographic Variable</td>
<td>Study 1</td>
<td>Study 2</td>
<td>Study 3</td>
<td>Study 4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-----------------------------------------------------------</td>
<td>---------</td>
<td>---------</td>
<td>---------</td>
<td>---------</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>n</td>
<td>%</td>
<td>n</td>
<td>%</td>
<td>n</td>
<td>%</td>
</tr>
<tr>
<td>Biomedical Engineering</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bachelor of Nursing</td>
<td>2</td>
<td>4.5</td>
<td>3</td>
<td>6.8</td>
<td>5</td>
<td>5.7</td>
</tr>
<tr>
<td>Bachelor of Physical Education</td>
<td>1</td>
<td>2.3</td>
<td>0</td>
<td>0.0</td>
<td>1</td>
<td>1.1</td>
</tr>
<tr>
<td>Bachelor of Psychology</td>
<td>8</td>
<td>18.2</td>
<td>8</td>
<td>18.2</td>
<td>16</td>
<td>18.2</td>
</tr>
<tr>
<td>Bachelor of Rural Science</td>
<td>0</td>
<td>0.0</td>
<td>1</td>
<td>2.3</td>
<td>1</td>
<td>1.1</td>
</tr>
<tr>
<td>Bachelor of Social Science (Psychology)</td>
<td>6</td>
<td>13.6</td>
<td>9</td>
<td>20.5</td>
<td>15</td>
<td>17.0</td>
</tr>
<tr>
<td>Bachelor of Social Science / Bachelor of Business</td>
<td>3</td>
<td>6.8</td>
<td>0</td>
<td>0.0</td>
<td>3</td>
<td>3.4</td>
</tr>
<tr>
<td>Bachelor of Speech Pathology</td>
<td>1</td>
<td>2.3</td>
<td>0</td>
<td>0.0</td>
<td>1</td>
<td>1.1</td>
</tr>
<tr>
<td>Bachelor of Stage and Screen (Television Production)</td>
<td>1</td>
<td>2.3</td>
<td>0</td>
<td>0.0</td>
<td>1</td>
<td>1.1</td>
</tr>
<tr>
<td>Bachelor of Theology</td>
<td>1</td>
<td>2.3</td>
<td>0</td>
<td>0.0</td>
<td>1</td>
<td>1.1</td>
</tr>
<tr>
<td>Diploma of Leadership and Ministry</td>
<td>1</td>
<td>2.3</td>
<td>0</td>
<td>0.0</td>
<td>1</td>
<td>1.1</td>
</tr>
<tr>
<td>Doctor of Philosophy (Arts)</td>
<td>0</td>
<td>0.0</td>
<td>1</td>
<td>2.3</td>
<td>1</td>
<td>1.1</td>
</tr>
<tr>
<td>Doctor of Philosophy (Business Management)</td>
<td>0</td>
<td>0.0</td>
<td>1</td>
<td>2.3</td>
<td>1</td>
<td>1.1</td>
</tr>
<tr>
<td>Doctor of Philosophy (Law)</td>
<td>1</td>
<td>2.3</td>
<td>1</td>
<td>2.3</td>
<td>2</td>
<td>2.3</td>
</tr>
<tr>
<td>Doctor of Philosophy (Psychology)</td>
<td>1</td>
<td>2.3</td>
<td>1</td>
<td>2.3</td>
<td>2</td>
<td>2.3</td>
</tr>
<tr>
<td>Doctor of Philosophy (Public Health)</td>
<td>0</td>
<td>0.0</td>
<td>1</td>
<td>2.3</td>
<td>1</td>
<td>1.1</td>
</tr>
<tr>
<td>Master of Nursing</td>
<td>1</td>
<td>2.3</td>
<td>0</td>
<td>0.0</td>
<td>1</td>
<td>1.1</td>
</tr>
<tr>
<td>Master of Social Work</td>
<td>0</td>
<td>0.0</td>
<td>1</td>
<td>2.3</td>
<td>1</td>
<td>1.1</td>
</tr>
<tr>
<td>Master of Paramedical Science (Clinical Care)</td>
<td>0</td>
<td>0.0</td>
<td>1</td>
<td>2.3</td>
<td>1</td>
<td>1.1</td>
</tr>
<tr>
<td>Post-Graduate Diploma of Advanced Educational Leadership</td>
<td>1</td>
<td>2.3</td>
<td>0</td>
<td>0.0</td>
<td>1</td>
<td>1.1</td>
</tr>
<tr>
<td>Post-Graduate Diploma of Psychology</td>
<td>3</td>
<td>6.8</td>
<td>1</td>
<td>2.3</td>
<td>4</td>
<td>4.5</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Socioeconomic Status</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Parental Class</td>
<td>Working class</td>
<td>3</td>
<td>6.8</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>Lower-middle class</td>
<td>2</td>
<td>4.5</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>Middle class</td>
<td>23</td>
<td>52.3</td>
<td>25</td>
</tr>
</tbody>
</table>

157
<table>
<thead>
<tr>
<th>Demographic Variable</th>
<th>Study 1</th>
<th>Study 3</th>
<th>Study 4</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(n)</td>
<td>(%)</td>
<td>(n)</td>
</tr>
<tr>
<td>Upper-middle class</td>
<td>16</td>
<td>36.4</td>
<td>11</td>
</tr>
<tr>
<td>Financially supported</td>
<td>7</td>
<td>15.9</td>
<td>4</td>
</tr>
<tr>
<td>Full-time employment</td>
<td>9</td>
<td>20.5</td>
<td>5</td>
</tr>
<tr>
<td>Part-time employment</td>
<td>18</td>
<td>40.9</td>
<td>20</td>
</tr>
<tr>
<td>Government benefits</td>
<td>8</td>
<td>18.2</td>
<td>11</td>
</tr>
<tr>
<td>Scholarship</td>
<td>0</td>
<td>0.0</td>
<td>4</td>
</tr>
<tr>
<td>Other</td>
<td>2</td>
<td>4.6</td>
<td>0</td>
</tr>
<tr>
<td>Yes Religious / Spiritual</td>
<td>24</td>
<td>54.5</td>
<td>19</td>
</tr>
<tr>
<td>Anglican</td>
<td>1</td>
<td>2.3</td>
<td>4</td>
</tr>
<tr>
<td>Baptist</td>
<td>1</td>
<td>2.3</td>
<td>0</td>
</tr>
<tr>
<td>Buddhism</td>
<td>2</td>
<td>4.5</td>
<td>3</td>
</tr>
<tr>
<td>Catholicism</td>
<td>8</td>
<td>18.2</td>
<td>3</td>
</tr>
<tr>
<td>Islam</td>
<td>1</td>
<td>2.3</td>
<td>0</td>
</tr>
<tr>
<td>Pentecostal</td>
<td>4</td>
<td>9.1</td>
<td>1</td>
</tr>
<tr>
<td>Presbyterian</td>
<td>1</td>
<td>2.3</td>
<td>0</td>
</tr>
<tr>
<td>Uniting Church</td>
<td>1</td>
<td>2.3</td>
<td>1</td>
</tr>
<tr>
<td>Other Christian, Not Further Defined</td>
<td>4</td>
<td>9.1</td>
<td>2</td>
</tr>
<tr>
<td>Other Non-Christian, Not Further Defined</td>
<td>1</td>
<td>2.3</td>
<td>5</td>
</tr>
<tr>
<td>No Religious / Spiritual</td>
<td>20</td>
<td>45.5</td>
<td>25</td>
</tr>
<tr>
<td>Agnostic</td>
<td>1</td>
<td>2.3</td>
<td>6</td>
</tr>
<tr>
<td>Atheist</td>
<td>6</td>
<td>13.6</td>
<td>5</td>
</tr>
<tr>
<td>Humanist</td>
<td>3</td>
<td>6.8</td>
<td>3</td>
</tr>
<tr>
<td>Rationalist</td>
<td>1</td>
<td>2.3</td>
<td>4</td>
</tr>
<tr>
<td>No Religion, Not Further Defined</td>
<td>9</td>
<td>20.5</td>
<td>7</td>
</tr>
</tbody>
</table>

Note. \(n\) = frequency of participants from \(N = 44\) participants each for Studies 1 and 3, and \(N = 88\) participants for Study 4.
Sample size and power

Sample size and power were calculated for all statistical tests used in the current study using G*power (version 3.1):

- Kendall’s tau-b: At a significance level of 0.05, correlation of .4, and 80% power, the required sample size was 55.
- Chi square test of contingencies: At a significance level of 0.05, medium effect size (0.5), and 80% power, the required sample size was 52.
- Binomial logistic regression: At a significance level of 0.05, and 80% power with 5 predictor variables, the required sample size was 58 (this was calculated by adding 15% to the resulting sample size for multiple regression as this is a non-parametric equivalent.)

Thus, there was sufficient sample size and power for the statistical analyses in the current study given \( N = 88 \).

Research design

This an investigation into the various individual factors that may influence moral dumbfounding; data collected during Studies 1 and 3 were utilised. Both qualitative and quantitative data were collected in a standardised open-ended interview (replicating the research design of the original moral dumbfounding study; Haidt et al., 2000). Refer to the Methods section of Study 1 (p. 20) for further information regarding the research design.

Measures and procedure

The procedure for the current study is the same as Studies 1 and 3 (refer to the Methods section of the studies for further information – pp. 30 and 105, respectively).

Moral dumbfounding

The Evidence-based definition of moral dumbfounding developed in Study 2 was utilised in the current study (i.e., an inability to justify disapproval for a seemingly harm-free action). A moral dumbfounding variable was computed for each stimulus according to the following criteria: a participant needed to have a frequency greater than zero for the variables of ‘belief that the action is harm-free’ AND (‘judgment without justification’ OR ‘struggle to produce a reason’). A global measure of moral dumbfounding was computed for each participant by aggregating the frequency of moral dumbfounding across stimuli. Higher numbers were taken to indicate greater frequency of moral dumbfounding during the interview.
Nationality and ethnicity

Participants were asked their nationality (open-ended response). Participants were also asked two items to ascertain their ethnicity: (a) ‘Are you of Aboriginal or Torres Strait Islander origin?’ (yes, no), and (b) ‘With which racial or ethnic group(s) do you most identify?’ (open-ended response). Finally, with concerns regarding possible differences in the ability to express justification for moral judgment (due to language barriers), participants were asked: ‘Is English your first language?’ (yes, no). Those who responded ‘no’ were also asked: ‘How long have you been speaking English?’ (open-ended response) because extensive English experience would suggest sufficient proficiency to express justification for moral judgments.

Education

To allow for variation in the sample, (rather than restricting it to psychology students from a single university like Haidt et al., 2000), individuals eligible to participate in the present study were enrolled in any tertiary education course at any Australian tertiary institution. Participants were asked, ‘At which university are you enrolled as a student?’ (open-ended response). Participants were also asked, ‘What is the name of the course in which you are currently enrolled?’ (open-ended response).

Socioeconomic status

It is difficult to measure the SES of university students because common measures of SES (such as employment status) may not be as relevant to higher education students compared with a working population (Australian Bureau of Statistics, 2011). A proxy measure of occupation and education level of the students’ parents has been suggested as an indication of household wealth given the greater economic capacity of some parents to support their children through study (Department of Education Employment and Workplace Relations, 2009). As such, participants were asked, ‘In terms of education and income, would you say your parents are...’ (categorical selection, e.g., ‘Working Class’, ‘Lower-middle Class’...). To ensure the relevance of a common indicator to the two broad categories of students (i.e., school-leavers and mature-aged students), participants were also asked, ‘What is your major source of income?’ (categorical selection, e.g., ‘Government or university funded scholarship’, ‘Wages: full-time employment’... including open-ended response if category absent).

Religious ideology

The confound between religious and moral views has been previously recognised by moral dumbfounding researchers (Haidt, 2007; Haidt & Hersh, 2001; McHugh et al.,
2017) and so religious ideology and strength of commitment was measured. The first item was ‘Do you consider yourself to be a religious person?’ (yes, no). Those participants who identified as religious or spiritual were asked their religious affiliation (categorical selection, e.g., ‘Buddhism’, ‘Catholicism’, ‘Islam’... including open-ended response if category absent). They were then asked, ‘How important an influence is your religious or spiritual faith in your life?’ Responses to this item were recorded on a five-point rating scale ranging from 1 (most important) to 5 (not at all important). Participants who did not identify as religious or spiritual were asked if they identify as ‘Agnostic’, ‘Atheist’, ‘Humanist’, ‘Rationalist’, or ‘No Religion – Not Further Defined’.

Political preference
While some moral dumbfounding studies use the labels ‘liberal’ (‘left’) and ‘conservative’ (‘right’) to capture political preference (sc., Haidt & Hersh, 2001; Royzman et al., 2015), Dawson and Tyson (2012) point out that one of the conservative coalition parties in Australia is called the Liberal Party, which may cause confusion for Australian participants. Furthermore, due to the young age of the cohort for this study, the labels ‘left’ and ‘right’ may be unfamiliar. Political preference was, therefore, measured using three proxy items. Responses to each of the items were recorded on a five-point rating scale ranging from 1 (strongly disagree) to 5 (strongly agree). These were: (a) ‘An individual’s rights should be sacrificed for the well-being of society in certain situations’, (b) ‘The role of the criminal justice system should be to punish versus rehabilitate offenders’, and (c) ‘I support the legalisation of same-sex marriage in Australia’. After reversal of item (b), higher scores on each of the three items were taken to indicate more liberal, left-wing political preference.

Personality
The Five Factor Model of Personality (Costa & McCrae, 1992) is the most utilised personality model when exploring correlations with political ideology and religiosity. The Big Five Aspect Scales (BFAS; DeYoung, Quilty, & Peterson, 2007, annexed as Appendix G) determines the five broad dimensions of the Five Factor Model of Personality, but it provides for a more fine-grained analysis of the relationship between personality and variables of interest, as each dimension can be separated into two sub-factors (called ‘aspects’). The 100 items of the BFAS comprise 20-item scales for the five dimensions and 10-item scales for the 10 aspects. Responses to each of the 100 items are recorded on a five-point rating scale ranging from 1 (strongly disagree) to 5
(strongly agree). Scores for each aspect are computed by taking the mean of the corresponding items (with appropriate reversals). Scores for each dimension are computed by taking the mean of the two aspect scores. Higher scores in each scale indicate a greater propensity toward behavior consistent with that personality trait across time and situations. The BFAS scales are reliable (all $a > 0.73$) and have good test-retest reliability (all $r > 0.72$); convergent and discriminant validity has also been demonstrated (DeYoung et al., 2007).

**Susceptibility to experimenter pressure**

To ascertain the effect of the experimental situation on individuals (and, therefore, possibly moral dumbfounding), a measure of susceptibility to experimenter pressure was included. In the absence of an existing scale, this was measured using items drawn from Haslam et al. (2014), as well as items designed by the researcher. Given the potential for influence of the experimenter-participant relationship on participant response, participants responded to this measure using the online demographic questionnaire.

Questions drawn from Haslam et al. (2014) were altered to reflect the nature of the current study. Specifically, ‘neural networking’ was changed to ‘moral judgment’, and the rating scale was changed from a seven-point scale to a five-point scale ($1 = \text{strongly disagree}; 5 = \text{strongly agree}$) for consistency with other rating scales in the current study. Two items measured ‘Identification with Experimental Goals’: (a) ‘I think research on moral judgment is important’ and (b) ‘I want to help scientists understand the nature of moral judgment’. The remaining two items measured ‘Identification with Psychological Science’: (a) ‘It is important to me that I contribute to psychological knowledge through research participation’ and (b) ‘I feel good about furthering psychological science’. Higher scores were taken to indicate greater identification with experimental goals, and psychological science, respectively. Haslam et al. report that the two items assessing each construct form reliable scales (‘Identification with Experimental Goals’: $M = 4.96, r = 0.53, p < 0.001$; ‘Identification with Psychological Science’: $M = 5.41, r = 0.78, p < 0.001$). These statistics are based on a subsample of 68 participants drawn from a sample of 100 University of Exeter students (79 females and 19 males, two gender unknown) with a mean age of 29.7 years ($SD = 10.1$) all from a white British background. Together, the two subscales were combined in the current study to form a variable of ‘Identification with the Experiment’.  

162
In addition to the 'Identification with the Experiment' variable, Bieri and Lobeck's (1959) construct of 'Degree of Involvement' was measured using three items designed by the researcher. The first item was: 'How much did you enjoy taking part in this study?' Responses to this item were recorded on a five-point numerical rating scale ranging from 1 (I did not enjoy it at all) to 5 (I enjoyed it very much). The second item was: 'Would you recommend participating in this study to your friends?' Responses to this item were recorded on a five-point numerical rating scale ranging from 1 (I definitely would not recommend this study to my friends) to 5 (I would definitely recommend this study to my friends). The final scale item was: 'How important is it to you that you answered the experimenter’s questions to the best of your ability?' Responses to this item were recorded on a five-point numerical rating scale ranging from 1 (not at all important) to 5 (very important). Higher scores on each of the three items were taken to indicate a greater degree of involvement in the experiment. As this is a novel scale, no validity or reliability statistics are available. The 'Identification with the Experiment' and 'Degree of Involvement' variables were combined (after being assessed for internal reliability – see demand characteristics and moral dumbfounding in the data-analysis section below). The seven items were taken as a collective measure of 'Susceptibility to Experimenter Pressure'. The constructed scale is annexed as Appendix H.

Data analysis

The same methods of qualitative and quantitative analyses from Studies 1 and 3 were utilised in the current study (refer to the Methods section of Study 1 on p. 32 for information regarding the qualitative and quantitative analyses, as well as general data-analysis information). The two-person coding procedure resulted in a very high degree of inter-rater reliability between 20 measurements. The average measure intra-class correlation coefficient was 0.988 with a 95% confidence interval from 0.971 to 0.995 \((F(19,19) = 84.79, p < 0.001)\).

While the sample size of the current study \((N = 88)\) was substantially larger than that of the original study \((N = 30;\) Haidt et al., 2000), a Monte Carlo resampling method was applied to generate a large number of simulated samples \((10,000\) samples) to yield unbiased estimates about the characteristics of the population from which the sample at hand is drawn (see, Field, 2009). All variable distributions were checked to assess the appropriateness of parametric tests for analyses.
Individual characteristics and moral dumbfounding

Normality (i.e., skewness and kurtosis less than ±1, and significance of the Shapiro-Wilk statistic) and sphericity (i.e., the significance of the Mauchly’s Test of Sphericity) were assessed using SPSS. As SPSS does not provide a test of homogeneity of variance for repeated measures variables, an $F_{\text{max}}$ test was computed by hand using the formula

$$F_{\text{max}} = \frac{\text{Largest Sample Variance}}{\text{Smallest Sample Variance}},$$

where the largest sample variance is simply the largest standard deviation squared, and the smallest sample variance is the smallest standard deviation squared (Allen & Bennett, 2008). Homogeneity of variance can be assumed when $F_{\text{max}}$ is less than 10 (Tabachnick & Fidell, 2001). As in Studies 1, 2, and 3, the assumptions of normality, sphericity, and homogeneity of variance, were violated for many of the variables and, therefore, equivalent non-parametric tests were used.

Unless otherwise stated, the significance level was set at $\alpha = 0.05$. Effect sizes are reported where possible, with the effect size formula being $r = \frac{z}{\sqrt{N}}$. Descriptions of effect size were guided by J. D. Cohen (1988) in which $r = 0.1$ can be considered small, $r = 0.3$ can be considered medium, and $r = 0.5$ can be considered large. Where the effect size is unable to be ascertained due to the limited availability of statistical techniques, a graphical overview is provided to give an adequate sense of the size and direction of the effect of significant comparisons.

In relation to personality and moral dumbfounding, analysis for the (BFAS; DeYoung et al., 2007) began with recoding of the atypically worded items. Table 40 presents the five broad dimensions, the two aspects of each dimension, sample items for each scale, and internal consistency reliability information for the current study. All BFAS scales reached an acceptable level (all $\alpha \geq 0.73$); subsequently, a total score was calculated by taking the mean of the corresponding items. Higher scores were taken to indicate a greater propensity toward behaviour consistent with that personality trait across time and situations.
<table>
<thead>
<tr>
<th>Personality Scale</th>
<th>Personality Subscale</th>
<th>Example Item</th>
<th>$\alpha$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Neuroticism</td>
<td>Withdrawal</td>
<td>Get upset easily</td>
<td>0.90</td>
</tr>
<tr>
<td></td>
<td>Volatility</td>
<td>Feel threatened easily</td>
<td>0.80</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Get angry easily</td>
<td>0.87</td>
</tr>
<tr>
<td>Agreeableness</td>
<td>Compassion</td>
<td>Like to do things for others</td>
<td>0.83</td>
</tr>
<tr>
<td></td>
<td>Politeness</td>
<td>Feel others’ emotions</td>
<td>0.77</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Hate to seem pushy</td>
<td>0.80</td>
</tr>
<tr>
<td>Conscientiousness</td>
<td>Industriousness</td>
<td>Follow a schedule</td>
<td>0.82</td>
</tr>
<tr>
<td></td>
<td>Orderliness</td>
<td>Carry out my plans</td>
<td>0.81</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Keep things tidy</td>
<td>0.73</td>
</tr>
<tr>
<td>Extraversion</td>
<td>Enthusiasm</td>
<td>See myself as a good leader</td>
<td>0.82</td>
</tr>
<tr>
<td></td>
<td>Assertiveness</td>
<td>Make friends easily</td>
<td>0.82</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Take charge</td>
<td>0.85</td>
</tr>
<tr>
<td>Openness</td>
<td>Intellect</td>
<td>Love to reflect on things</td>
<td>0.83</td>
</tr>
<tr>
<td></td>
<td>Openness</td>
<td>Have a rich vocabulary</td>
<td>0.82</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Enjoy the beauty of nature</td>
<td>0.78</td>
</tr>
</tbody>
</table>

*Note. $N = 88$.*  
*α* Estimate of internal consistency was Cronbach’s alpha.

Given the large number of subscales comprising the BFAS (DeYoung et al., 2007), the medium sample size of the current study ($N = 88$), and the concern regarding unnecessary inflation of error rates, it was deemed beneficial to use only the scales (rather than the subscales) in inferential analyses. Prior to regression analysis, an inspection of the normal probability plot of Regression Standardised Residuals was undertaken which revealed that the assumption of normality was not met (the points were very distant from the diagonal line indicating a large deviation from normality). Binomial logistic regression was, therefore, used to overcome many of the restrictive assumptions of multiple linear regression (Field, 2009). A binary measure of moral dumbfounding was computed (i.e., ‘Dumbfounded During the Interview’ versus ‘Not Dumbfounded During the Interview’) and, prior to inferential analyses, the remaining assumptions of binomial logistic regression were assessed.
Multivariate outliers were a concern as Mahalanobis distance exceeded the critical $\chi^2$ for $df = 5$ (at $\alpha = 0.01$) of 15.09 (the Maximum Mahalanobis Distance was 17.03). Two participants had Mahalanobis distances exceeding the critical value; the first was presented with the original stimuli and was classified as dumbfounded, and the second was presented with the alternative stimuli and was not classified as dumbfounded. As removing these two participants from regression analyses would not alter the ratio of participants according to stimuli condition, nor the proportion of dumbfounded participants to non-dumbfounded, the model was reanalysed without these two outliers (rather than transforming all participant data). Without the data of the two participants, multivariate outliers were no longer a concern as the Maximum Mahalanobis Distance was 14.99.

The intercorrelation between the predictor variables was evaluated and revealed relatively high tolerances for the predictors in the regression model (Tolerance was $\geq 0.65$ for all predictors and the variance inflation factor was $\geq 1.02$ for all predictors) indicating that multicollinearity was not an issue. Linearity with the continuous predictor variables and the odds ratio was checked by assessing the interaction between each predictor and the log of itself. All five interactions had significance values greater than 0.05 indicating that the assumption of linearity of the logit had been met (Neuroticism $p = 0.63$; Agreeableness $p = 0.44$; Conscientiousness $p = 0.95$; Extraversion $p = 0.46$; Openness $p = 0.62$).

**Demand characteristics and moral dumbfounding**

Table 41 presents the items for the ‘Susceptibility to Experimenter Pressure’ scale and internal consistency reliability information for the current study. Each item reached an acceptable level with no item falling below $\alpha \leq 0.35$, indicating that the items share common variance and may be represented as a single measure. Further, the combined Cronbach’s alpha was 0.79 indicating high internal reliability across the seven items. Subsequently, a total score for each participant on the scale was calculated by taking the mean of the seven items, with higher scores indicating greater susceptibility to experimenter pressure.
Table 41  
*Items and Internal Consistency Reliabilities for the Susceptibility to Experimenter Pressure Scale*  

<table>
<thead>
<tr>
<th>Susceptibility to Experimenter Pressure Scale Item</th>
<th>α</th>
</tr>
</thead>
<tbody>
<tr>
<td>I think research on moral judgment is important</td>
<td>0.55</td>
</tr>
<tr>
<td>I want to help scientists understand the nature of moral judgment</td>
<td>0.61</td>
</tr>
<tr>
<td>It is important to me that I contribute to psychological knowledge through research participation</td>
<td>0.52</td>
</tr>
<tr>
<td>I feel good about furthering psychological science</td>
<td>0.54</td>
</tr>
<tr>
<td>How much did you enjoy taking part in this study?</td>
<td>0.53</td>
</tr>
<tr>
<td>Would you recommend participating in this study to your friends?</td>
<td>0.59</td>
</tr>
<tr>
<td>How important is it to you that you answered the experimenter’s questions to the best of your ability?</td>
<td>0.35</td>
</tr>
</tbody>
</table>

*Note. N = 88.  
*α* Estimate of internal consistency was Cronbach’s alpha.*

As the relevant hypothesis focuses on participants who were more susceptible to experimenter pressure than others (i.e., higher in identification with the experiment), this variable was transformed into a binary measure prior to analysis. Participants were classified as ‘High in Identification with the Experiment’ if their ‘Susceptibility to Experimenter Pressure’ score was at least one standard deviation higher than the mean. All other participants were classified as having an average or less susceptibility to experimenter pressure.

**Results**

**Moral dumbfounding**

Overall, 18 participants (20.5%) were classified as morally dumbfounded (according to the *Evidence-based* definition) in response to one of their presented stimuli (i.e., original or alternative). There were three participants (3.4%) who were dumbfounded in response to two of the stimuli, and one participant who was dumbfounded in response to three tasks. In total, 22 participants (25%) were classified as dumbfounded at least once during the interview. The global measure of moral dumbfounding ranged between 0 and 2 with a mean of 0.31 (*SD = 0.59*).
Individual characteristics and moral dumbfounding

Gender
There were 59 female participants (67%) of which 15 (25.4%) were classified as morally dumbfounded at least once during the interview. There were 29 male participants (33%) of which seven (24.1%) were classified as dumbfounded. Chi-square test of contingency (α = 0.05) was used to assess whether dumbfounding is related to gender. Female participants were no more likely to be morally dumbfounded than male participants; $\chi^2 (df = 1, N = 88) = 0.017, p = 0.896$.

Age
The sample ranged in age between 18 and 63 years ($M = 26.97, SD = 11.54$). Kendall’s tau-b ($\alpha = 0.05$) was used to explore the relationship between moral dumbfounding ($M = 31; SD = 0.59$) and age. No relationship was found; $\tau = 0.016, p = 0.854, N = 88$.

Nationality and ethnicity
Five participants (5.7%) nominated a nationality other than Australian. Given the small frequency in each cell other than Australian, the participants who nominated a different nationality were combined. Of the 83 participants (94.3%) who indicated they were Australian, 22 (26.5%) were classified as morally dumbfounded at least once during the interview. None of the five participants of a nationality other than Australian were classified as dumbfounded. Chi-square test of contingency (α = 0.05) was used to assess whether dumbfounding is related to nationality. Participants born in Australia were no more likely to be morally dumbfounded than those born in another country; $\chi^2 (df = 1, N = 88) = 1.77, p = 0.184$.

Ten participants nominated an ethnicity other than Anglo-Australian (11.4%). Given the small frequency in each cell for ethnicity other than Anglo-Australian, the participants who nominated a different ethnicity were combined. Of the 78 Anglo-Australian participants (88.6%), 20 (25.6%) were classified as morally dumbfounded at least once during the interview. Of the 10 participants of another ethnicity, two (20%) were classified as dumbfounded. Chi-square test of contingency (α = 0.05) was used to assess whether dumbfounding is related to ethnicity. The participants of Anglo-Australian ethnicity were no more likely to be morally dumbfounded than participants from another racial or ethnic group; $\chi^2 (df = 1, N = 88) = 0.15, p = 0.698$.

Two participants (2.3%) indicated that English was not their first language; however, both had been speaking English for a minimum of 20 years (range = 20–27
years). The extensive English experience suggests sufficient proficiency to express justification for moral judgments.

**Education**

Most of the participants were enrolled at CSU ($n = 77; 87.5\%$). Given small frequencies in the remaining cells, participants who nominated a university other than CSU were combined. Of the participants enrolled at CSU, 21 (27.3\%) were classified as morally dumbfounded at least once during the interview. Of the 11 participants (12.5\%) enrolled at another Australian university, one (9.1\%) was classified as dumbfounded. Analysis was conducted using a Chi-square test of contingency ($\alpha = 0.05$); moral dumbfounding was not contingent on the university at which a participant is enrolled; $\chi^2 (df = 1, N = 88) = 1.70, p = 0.193$.

Thirty-one different courses of enrolment were reported by participants; given small cell frequencies, participants who nominated a course other than psychology were combined. Of the 42 participants (47.7\%) enrolled in a psychology course, 11 (26.2\%) were classified as morally dumbfounded at least once during the interview. Of the 46 participants (52.3\%) enrolled in a course other than psychology, 11 (23.9\%) were classified as dumbfounded. Analysis was conducted using a Chi-square test of contingency; participants studying psychology were no more likely to be morally dumbfounded than those studying another discipline; $\chi^2 (df = 1, N = 88) = 0.06, p = 0.805$.

Finally, the postgraduate education level variable also showed variation with small numbers in each cell; thus, participants who indicated that they were postgraduate-level students were combined. Of the 71 undergraduate students (80.7\%), 17 (23.9\%) were classified as morally dumbfounded at least once during the interview. Of the 17 postgraduate-level students (19.3\%), five (29.4\%) were classified as dumbfounded. Analysis was conducted using a Chi-square test of contingency; undergraduate participants were no more likely to be morally dumbfounded than postgraduate participants; $\chi^2 (df = 1, N = 88) = 0.219, p = 0.640$.

**Socioeconomic status**

Internal consistency was calculated for the two SES items. Cronbach’s alpha for the total scale did not reach an acceptable level of 0.7 ($\alpha = 0.04$). Accordingly, the two items of ‘Parental Class’ and ‘Income’ were utilised as separate variables for subsequent analyses. The frequency of participants who were classified as morally dumbfounded during the interview (at least once) across the two proxy SES measures is presented in
Table 42 (descriptive statistics of the SES items are duplicated from Table 39 for ease of reference). A chi-square test of contingency was used to assess whether dumbfounding is related to parental class; no relationship was found; $\chi^2 (df = 3, N = 88) = 6.27, p = 0.097$. A chi-square test of contingency was also used to assess whether dumbfounding is related to income. There was no association between dumbfounding and major source of income; $\chi^2 (df = 5, N = 88) = 8.74, p = 0.118$.

Table 42  
Frequency of Participants who were Classified as Morally Dumbfounded at Least Once in Response to the Stimuli Across Socioeconomic Proxy Measures of Parental Class and Income

<table>
<thead>
<tr>
<th>Parental Class</th>
<th>Sample</th>
<th>Classified as Dumbfounded</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$n$</td>
<td>%</td>
</tr>
<tr>
<td>Working class</td>
<td>5</td>
<td>5.7</td>
</tr>
<tr>
<td>Lower-middle class</td>
<td>8</td>
<td>9.1</td>
</tr>
<tr>
<td>Middle class</td>
<td>48</td>
<td>54.5</td>
</tr>
<tr>
<td>Upper-middle class</td>
<td>27</td>
<td>30.7</td>
</tr>
<tr>
<td>Major Source of Income</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Financially supported</td>
<td>11</td>
<td>12.5</td>
</tr>
<tr>
<td>Full-time employment</td>
<td>14</td>
<td>15.9</td>
</tr>
<tr>
<td>Part-time employment</td>
<td>38</td>
<td>43.2</td>
</tr>
<tr>
<td>Government benefits</td>
<td>19</td>
<td>21.6</td>
</tr>
<tr>
<td>Scholarship</td>
<td>4</td>
<td>4.5</td>
</tr>
<tr>
<td>Other</td>
<td>2</td>
<td>2.3</td>
</tr>
</tbody>
</table>

Note. $n$ = frequency of participants from $N = 88$. There was no association between Moral Dumbfounding and Parental Class, nor Major Source of Income, at $\alpha = 0.05$.

Religious ideology

The vast majority of religious or spiritual participants identified as Christian ($n = 31; 72.1\%$). Given the relatively small frequency in each cell other than Christian, these cells were combined. Of those who identified as Christian, eight (25.8%) were classified as morally dumbfounded at least once during the interview. Of the 12 non-Christian participants (27.9%), two (16.7%) were classified as dumbfounded. A chi-square test of contingency was used to assess whether dumbfounding is related to religion. Christians were no more likely to be morally dumbfounded than non-Christian religious participants; $\chi^2 (df = 1, N = 43) = 0.41, p = 0.525$.  

170
The measure of the importance of faith to the religious participants ranged from 0 to 5 with a mean of 2.95 (SD = 1.36). Kendall’s tau-b (α = 0.05) was used to explore the relationship between moral dumbfounding (M = 31; SD = 0.59) and the importance of faith to religious participants. No relationship was found; \( \tau = -0.13, p = 0.333, N = 43 \).

Descriptive statistics of the 45 participants (51.1%) who indicated that they did not hold religious belief are presented in Table 43 (duplicated from Table 39 for ease of reference) along with the frequency of participants who were classified as morally dumbfounded during the interview (at least once) across non-religious belief. A chi-square test of contingency was used to assess whether dumbfounding is related to type of non-religious identification. Moral dumbfounding was not contingent on type of non-religious belief; \( \chi^2 (df = 4, N = 45) = 1.72, p = 0.808 \).

Table 43
Frequency of Participants who were Classified as Morally Dumbfounded at Least Once in Response to the Stimuli Across Non-Religious Belief

<table>
<thead>
<tr>
<th>Sample</th>
<th>n</th>
<th>%</th>
<th>Classified as Dumbfounded</th>
<th>% within Variable</th>
<th>% Total Sample</th>
</tr>
</thead>
<tbody>
<tr>
<td>Agnostic</td>
<td>7</td>
<td>8.0</td>
<td>1</td>
<td>14.3</td>
<td>2.2</td>
</tr>
<tr>
<td>Atheist</td>
<td>11</td>
<td>12.5</td>
<td>2</td>
<td>18.2</td>
<td>4.4</td>
</tr>
<tr>
<td>Humanist</td>
<td>6</td>
<td>6.8</td>
<td>2</td>
<td>33.3</td>
<td>4.4</td>
</tr>
<tr>
<td>Rationalist</td>
<td>5</td>
<td>5.7</td>
<td>2</td>
<td>40.0</td>
<td>4.4</td>
</tr>
<tr>
<td>No Religion, Not Further Defined</td>
<td>16</td>
<td>18.2</td>
<td>5</td>
<td>31.3</td>
<td>11.1</td>
</tr>
</tbody>
</table>

*Note. n = frequency of participants from N = 45. There was no association between Moral Dumbfounding and Type of Non-Religious Belief at α = 0.05.*

A chi-square test of contingency was used to assess whether dumbfounding is contingent upon holding religious belief. Religious participants (regardless of type of belief) were no more likely to be morally dumbfounded than non-religious participants (regardless of type of identification); \( \chi^2 (df = 1, N = 88) = 0.136, p = 0.712 \).

**Political preference**
Reliability analysis of the three measures of political preference was conducted. Cronbach’s alpha for the items did not reach an acceptable level of 0.7 (α = 0.028). Furthermore, an examination of the questionnaire item-total statistics indicated that the
alpha would remain unacceptable if any of the items were removed. Accordingly, the three items of ‘Individual Rights vs Society’, ‘Punish vs Rehabilitate Offenders’, and ‘Support for Same-sex Marriage’ were utilised as separate variables for subsequent analyses. Kendall’s tau-b ($\alpha = 0.05$) was used to explore the relationship between the global dumbfounding variable ($M = 31; SD = 0.59$) and each of the three interval political items. The first political preference item, ‘An individual’s rights should be sacrificed for the well-being of society in certain situations’, yielded a mean of 2.93 ($SD = 1.16$). There was no relationship with moral dumbfounding; $\tau = -0.07, p = 0.442, N = 88$. The second item, ‘The role of the criminal justice system should be to punish versus rehabilitate offenders’, yielded a mean of 3.99 ($SD = 1.01$). Again, there was no relationship with moral dumbfounding; $\tau = -0.09, p = 0.356, N = 88$. The final political preference item, ‘I support the legalisation of same-sex marriage in Australia’, yielded a mean of 4.32 ($SD = 1.21$). Once more, there was no relationship with moral dumbfounding; $\tau = -0.11, p = 0.275, N = 88$.

**Personality**

Binomial logistic regression was performed to ascertain the predictive power of personality on moral dumbfounding. All variables were entered at Step 1 as appropriate when dealing with a small set of predictors and without a theory to guide the best prediction equation (Tabachnick & Fidell, 2001). The value of the constant was equal to $-1.13$ and the residual chi-square statistic was 4.81 which was not significant at $p < 0.05$ ($p = 0.440$); thus, entering all variables at Step 1 did not make a significant contribution to its predictive power.

The model explained between 56% and 83% of the variance in moral dumbfounding (Cox & Snell $R^2$, and Nagelkerke $R^2$ methods, respectively) and correctly classified 75.6% of cases. The model correctly predicted all of the 65 participants who were not dumbfounded. The model correctly predicted 20 cases of dumbfounding; however, one participant was misclassified as not dumbfounded when they were dumbfounded at least once during the interview. The predictive power of personality on moral dumbfounding was not significant; $\chi^2(5) = 4.92, p = 0.425$. Unstandardised beta regression coefficients ($B$), Wald statistics, and odds ratio ($\text{Exp}[\beta]$) of each predictor are reported in Table 44.
Table 44
Unstandardised Beta Regression Coefficients, Wald Statistics, and Odds Ratios of Each Predictor Variable in a Model Predicting Moral Dumbfounding

<table>
<thead>
<tr>
<th>Variables in the Equation</th>
<th>B</th>
<th>S.E.</th>
<th>Wald</th>
<th>Sig.</th>
<th>Exp(β)</th>
<th>95% C.I. for Exp(β)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Openness to Experience</td>
<td>-0.005</td>
<td>0.594</td>
<td>0.000</td>
<td>0.993</td>
<td>0.995</td>
<td>0.310 – 3.188</td>
</tr>
<tr>
<td>Conscientiousness</td>
<td>0.566</td>
<td>0.624</td>
<td>0.822</td>
<td>0.364</td>
<td>1.761</td>
<td>0.518 – 5.985</td>
</tr>
<tr>
<td>Extraversion</td>
<td>-0.372</td>
<td>0.731</td>
<td>0.259</td>
<td>0.611</td>
<td>.689</td>
<td>0.164 – 2.891</td>
</tr>
<tr>
<td>Agreeableness</td>
<td>1.275</td>
<td>0.758</td>
<td>2.831</td>
<td>0.092</td>
<td>3.580</td>
<td>0.810 – 15.815</td>
</tr>
<tr>
<td>Neuroticism</td>
<td>0.415</td>
<td>0.553</td>
<td>0.564</td>
<td>0.453</td>
<td>1.515</td>
<td>0.512 – 4.483</td>
</tr>
</tbody>
</table>

Note. N = 86. B = unstandardised beta regression coefficient; S.E. = standard error; Exp(β) = odds ratio; C.I. = confidence interval. No variable in the model was a significant predictor of moral dumbfounding at p < 0.05.

**Demand characteristics and moral dumbfounding**

It was hypothesised that participants susceptible to experimenter pressure (i.e., high in identification with the experiment) who judge the depicted action as wrong would be more prone to moral dumbfounding than others. There were 18 participants (20.5%) who were classified as highly susceptible to experimenter pressure, of which seven (38.9%) were classified as morally dumbfounded at least once during the interview. There were 70 participants (79.5%) who were classified as having an average or less susceptibility to experimenter pressure, of which 15 (21.4%) were classified as dumbfounded. Chi-square test of contingency (α = 0.05) was used to assess whether dumbfounding is related to demand characteristics. Participants susceptible to experimenter pressure (i.e., high in identification with the experiment) were no more likely to be morally dumbfounded than others; $\chi^2 (df = 1, N = 88) = 2.33, p = 0.127$.

Given that dichotomising the initially continuous ‘Susceptibility to Experimenter Pressure’ variable to a binary outcome resulted in a loss of power (see, J. Cohen, 1983), it was prudent to assess the relationship strength between susceptibility to experimenter pressure (non-transformed) and moral dumbfounding ($M = 31; SD = 0.59$). The global ‘Susceptibility to Experimenter Pressure’ variable had a mean of 4.35 (SD = 0.46). Analysis was conducted using Kendall’s tau-b (α = 0.05); there was no relationship; $\tau = 0.14, p = 0.132, N = 88$.  

173
Discussion

This study is the first to explicitly explore the association between individual differences and moral dumbfounding. Overall, 22 participants (25%) were classified as dumbfounded (i.e., unable to justify disapproval for a seemingly harm-free action) in response to at least one of their presented stimuli (i.e., original or alternative). No individual differences were found in the prevalence of moral dumbfounding across a wide range of variables including age, gender, nationality, ethnicity, religiosity, political preference, education, and SES. Moral dumbfounding could not be predicted using the personality traits of openness to experience, conscientiousness, agreeableness, extraversion, and neuroticism. Although not significant, the pattern of results for personality traits and moral dumbfounding indicate a fruitful direction for future research.

Participants high in agreeableness, conscientiousness, or neuroticism were more likely to be classified as morally dumbfounded compared with participants low in these personality traits. Specifically, the odds of being classified as morally dumbfounded were 3.58 times higher for agreeable participants than for less agreeable participants, 1.76 times higher for conscientious participants than for less conscientious participants, and 1.52 times higher for neurotic participants than for less neurotic participants. On the other hand, participants high in openness to experience and extraversion were less likely to be classified as morally dumbfounded compared with participants low in these traits (0.995 and 0.689, respectively). Nonetheless, again it must be stressed that these trends were not significant; further research is needed to ascertain if personality traits increase or decrease the odds of being dumbfounded.

As it currently stands, it does not appear that some individuals are more prone to moral dumbfounding than others. These results support studies in which no relationship has been found between moral dumbfounding and gender (Haidt et al., 2000; Haidt & Hersh, 2001; Royzman et al., 2015), religious ideology (McHugh et al., 2017), or political preference (Haidt & Hersh, 2001). It appears that situational forces are likely the primary drivers of the phenomenon. Such situational forces may include the unique conditions under which empirically observed moral dumbfounding is elicited.

This study is the first to explicitly explore the association between social dynamics of the interview context and moral dumbfounding. Contrary to prediction, participants high in identification with the experiment who judged the depicted action as wrong were no more prone to moral dumbfounding than others. Participants indicated
fairly high susceptibility to the pressure applied by the experimenter (measured on a response scale of 1 to 5, the ‘Susceptibility to Experimenter Pressure’ variable had a mean of 4.35; $SD = 0.45$). That experimenter pressure was present and perceived by the participants is illustrated by most (77.3%) complying with instruction on the food consumption task (and either drinking the roach juice or eating the lolly) despite the majority (88.7%) indicating that they believed the task presented to them involved harm.

Experimenter pressure has been proposed as an explanation for moral dumbfounding (Jacobson, 2012; Royzman et al., 2015). Royzman et al. (2015) found that some participants who retracted their verbal justifications for judgments to the Incest story (and were subsequently classified as dumbfounded) later provided written justification. When asked to account for response inconsistency, all relevant participants acknowledged that they should not have recanted their justification with some explaining that they felt pressured by the experimenter that rendered them unwilling to defend a justification for their judgment. The result of the current study does not necessarily contradict the findings of Royzman and colleagues. Despite improvements to methodology in this research to minimise such pressure (i.e., giving participants the opportunity to respond in their own words using open-ended survey items), moral dumbfounding was still elicited under experimental conditions. In addition, susceptibility to experimenter pressure was measured in degree (relative presence) and no participant endorsed a total absence. Consequently, while there was no relationship between susceptibility to experimenter pressure and moral dumbfounding, the ceiling effect could be obscuring the relationship (Type II error); the phenomenon observed here (and by other moral dumbfounding researchers; Haidt et al., 2000; Haidt & Hersh, 2001; McHugh et al., 2017) may reflect (at least partly) an inability to articulate justification in light of experimenter pressure.

It may not be possible to ascertain the effect of demand characteristics on moral dumbfounding using a face-to-face experimental design given the influence of research participation on behaviour (sc., Asch, 1951; Haslam et al., 2014; Haslam et al., 2015; Milgram, 1974; Orne, 1962; Reicher et al., 2012; Rosenthal, 1969/2009). The only researchers to date to employ an alternative methodology are McHugh et al. (2017; Studies 2, 3a, & 3b) who found that moral dumbfounding can be reliably elicited via a computer task; yet, their design also does not reflect everyday moral reasoning situations. An observational study would be ideal; however, it would be very difficult to
predict naturally occurring moral judgment. Furthermore, the lack of individual differences in dumbfounding means that a target sample is unlikely to increase the probability of observing the phenomenon above chance. Creating an environment in which the phenomenon is likely to be elicited, but without participant knowledge of intervention, would also be difficult but would yield greater ecologically valid results than extant research.

**Limitations**

While a Monte Carlo resampling method was applied to generate a large number of simulated samples (10,000 samples), the small-to-medium sample size of the current study \(N = 88\) does not provide a stringent test of potential individual differences in moral dumbfounding. For instance, the five scales (rather than the ten subscales) of The Big Five Aspect Scales (DeYoung, Quilty, & Peterson, 2007) were utilised in inferential analyses of personality due to concern regarding inflation of error rates. A more fine-grained analysis of the relationship between personality and variables of interest could have been conducted with a larger sample size. There was also a lack of variance in the current sample for responses to some items on the demographic questionnaire. Responses to the *Parental Class* item, for example, were indicated on a five-point categorical scale; however, 75 participants (82.5%) indicated they were from either middle class or upper-middle class. It is possible that the restricted range of responses on some independent variables in the present study resulted in Type II errors. Future research should encompass a larger, more variable sample size.

In addition, at the time the study was being designed, there was some concern about requiring participants to complete an unnecessarily long questionnaire, so proxy measures of SES and political preference were chosen. Due to a lack of internal reliability for items comprising the SES and political preference measures, the individual proxy measures were utilised as separate single item variables in analyses. Further research with more reliable and valid measures of SES and political preference is required. In particular, given the problems encountered with the choice of political ideology measure, and in light of the probable relevance of conservatism to moral dumbfounding (Haidt & Hersh, 2001; Piazza & Sousa, 2014; Royzman et al., 2015), the potential relationship between the tendency to hold attitudes of maintaining status-quo and tradition (characteristic of politically conservative individuals) and moral dumbfounding is highlighted as an important avenue for future research.
Conclusion

The present study supports and extends limited research (Haidt et al., 2000; Haidt & Hersh, 2001; McHugh et al., 2017; Royzman et al., 2015) by demonstrating that people are equally susceptible to experiencing moral dumbfounding despite a range of individual differences. This suggests that situational factors are likely key in explaining this phenomenon. Such factors include the nature of the scenario presented, and the context in which it is presented. An explanation for moral dumbfounding is unlikely to lie exclusively with the nature of the dilemma as, in Study 3, the phenomenon was found to occur in response to a range of stimuli (stories, visual scenes, and tasks) and the prevalence did not vary significantly between the reasoning and intuition stimuli. On the other hand, context in which the dilemma is presented may play a large role (Jacobson, 2012; Royzman et al., 2015). Despite improvements to methodology in this research to minimise pressure (i.e., giving participants the opportunity to respond in their own words using open-ended survey items), moral dumbfounding was still elicited under experimental conditions. As such, research with greater ecological validity is warranted (i.e., within the context of our ordinary moral reasoning situations) to establish whether this phenomenon provides important insights into the nature of everyday moral judgment.
Chapter 7: Does Moral Dumbfounding Reveal the Nature of Moral Judgment?

The findings presented in the current dissertation support the notion that moral dumbfounding is a genuine phenomenon. Simply, moral dumbfounding occurs when people make moral judgments that they find themselves unable to justify. While the frequency of dumbfounding was found not to be as common as reported in extant research (Haidt et al., 2000; Haidt & Hersh, 2001; McHugh et al., 2017), the phenomenon can be elicited by realistic scenarios (including actual witnessed moral transgressions) and, therefore, is not (as some have argued; Jacobson, 2012; Kennett, 2012) merely an artefact of the particular ‘moral intuition’ cases used by Haidt et al. (2000). Indeed Monin, Pizzaro, and Beer (2007) argue that different scenarios require different moral mechanisms (i.e., reasoning or intuition); thus, while cases that are framed in terms of a tension between conflicting moral principles invoke reasoning, witnessing other’s transgressions invokes intuitive processes. Moral dumbfounding was also not found to vary as a function of individual differences (although, given noted the limitations of sample size and variability, only limited conclusions can be drawn with regard to such differences).

Haidt (2001) takes moral dumbfounding to be evidence for his Social Intuitionist Model (SIM). To recap, according to SIM, moral judgments are quick and intuitive while reasoning (if it occurs at all) is typically post-hoc and is generated either to rationalise a judgment or to convince others to change their judgment. A key implication of SIM is that reasoning does not play a causal role in the production of moral judgments. But is moral dumbfounding actually evidence for SIM? In particular, does moral dumbfounding really show that reasoning plays no causal role in the formation of moral judgments?

In this chapter, it is argued that moral dumbfounding does not support SIM; specifically, it does not prove that reasoning plays no causal role in moral judgment. Even though the dumbfounding phenomenon is consistent with SIM, it is also consistent with various alternative theories about the causal production of moral judgments, as will be shown. A number of hypotheses will be outlined and each propose that reasoning does play a causal role in the production of moral judgments; thus, the dumbfounding phenomenon, as such, does not give us any better reason to accept SIM rather than one of these alternative hypotheses.

179
Ultimately, whether we have reason to think that reason plays a causal role in the production of moral judgments depends on which theory of moral judgment is independently best supported. There will not be an attempt, however, to fully evaluate the evidence for and against each hypothesis discussed here, beyond a few observations, nor will there be a comprehensive survey of theories of moral judgment. The aim is more limited: it will simply be shown that the reality of the dumbfounding phenomenon, as such, gives us no more reason to accept SIM rather than any of a number of alternative accounts that each give moral reasoning a genuine role to play in the production of moral judgments.

In each of the following sections, a hypothesis is described about the causal process involved in the production of moral judgments. None of these hypotheses constitutes a full theory of moral judgment. The hypotheses described are also (largely) mutually consistent, so a full theory could possibly incorporate them all. While a full defence of any hypothesis will not be given, there will be an attempt to show that (a) each is a viable hypotheses, (b) each hypothesis deserves careful scrutiny, (c) each hypothesis is consistent with the dumbfounding phenomenon, and (d) according to each hypothesis, moral reasoning does play a causal role in the production of moral judgments. The intention, in outlining these hypotheses, is simply to show that the dumbfounding phenomenon, as such, does not give us any reason to accept SIM and, thereby, to doubt that reasoning plays a role in the production of moral judgments.

**Hypothesis 1: Unconscious Reasoning**

It is widely accepted that we do not have conscious access to the higher order cognitive processes which cause complex behaviours and, when we when are asked to explain our evaluations, judgments, or choices, we report a causal process that could plausibly cause our behaviour, but not the process that actually caused the behaviour. Nisbett and Wilson (1977) argue that when we are asked to explain our behaviour, we reconstruct our cognitive processes, rather than using introspection to directly access them. In other words, we can report the product, but not the process, accurately; therefore, the mere fact that one cannot articulate reasons underlying any judgment (including a moral judgment) does not mean that a rational process did not occur in the formation of that judgment. Thus, the following hypothetical account of the moral judgment process is a genuine possibility that needs to be taken seriously:

We generate and evaluate reasons whenever we make moral judgments. Even though reasons inform our moral judgments, we lack conscious access to those
reasons and to the reasoning process and are, therefore, unable to articulate these reasons.

**Hypothesis 2: Automaticity**

A second viable hypothesis that can explain dumbfounding is the automatisation of behaviour. Performance of unpractised tasks is typically slow and effortful, while skilled behaviour is fast and makes small demands on our limited information processing capacity (Mowbray & Rhoades, 1959). Compare a novice and an expert chess player. The novice considers each play carefully before moving, is able to focus only on individual chess pieces, is able to remember a limited number of past moves, and can plan ahead only by a few moves; the expert, on the other hand, will typically make her moves quickly and can keep in mind all past moves and predict a vast number of possible moves that may occur in the present game (Saling & Phillips, 2007). The expert has the capacity to recognise patterns in the game and, therefore, does not need to think carefully about each move. With sufficient practice, the novice may become an expert and this shift reflects our general capacity to automatise behaviour. All of our behaviours, even very complex social behaviours (Bargh & Williams, 2006), can become automatic after sufficient repetition or practice. Given this, the following description of moral judgment is highly plausible:

Before we make a moral judgment for the first time, we engage in a process of reasoning: we evaluate reasons for and against the judgment in order to determine whether it is defensible. Over time, we make similar moral judgments, or at least judgments that follow a similar pattern. So, for instance, when I first consider the matter, I may judge that abortion is morally permissible because the mother’s right to autonomy trumps the foetus’s right to life. Later, when I first consider whether it is morally permissible to force a rational person to have medical treatment, I judge that it is not permissible because one’s right to autonomy trumps any (paternalistic) considerations. At this stage, I may self-consciously see these judgments as being similarly grounded by similar moral reasons. As time goes on, however, I may develop the implicit capacity to recognise a pattern of similar situations, without consciously reflecting on them. This may go hand-in-hand with an instinctive disposition to make similar judgments about these cases, whereby I take the right to autonomy as trumping other values. Crucially, however, I may not see the pattern of my instinctive judgments so, at this stage, I make judgments of a particular form _because_ of a
process of reasoning I previously engaged in. But I may not, at this stage, be able to trace the connection. The reason for this is that I may not be consciously aware of the way in which I am instinctively patterning cases, so I may not understand how I am disposed to judge cases which conform to the pattern, nor to be able to identify the reasoning process that led me to form this disposition in the first place.

When one is able to rely on an automatic moral judgment, one does not need to engage in reasoning. It does not mean though that the moral judgments were initially formed in the absence of a reasoning process but, rather, that now that the judgments have become automated, there is no need to engage in further reasoning. This manifests as an inability to provide justifications for our judgments, not because we never engaged in a process of reasoning but because we did so at a much earlier point in time. When the automatic moral judgment is activated, the reasons that caused that moral judgment are not activated and, therefore, we are unable to articulate our reasons.

The automaticity of moral judgment can explain the dumbfounding phenomenon in the following way. As we have seen, when people are dumbfounded, they may maintain a moral judgment even when presented with reasons that undercut that judgment. The puzzle is to explain why people don’t readily revise their judgments in cases like these. The hypothesis that moral judgments are the product of automatic processes can explain this. Automatic processes have a number of distinctive features, including a kind of ‘ballisticity’ (Bargh, 1999). This refers to the way in which an automatic process, once triggered, has a ‘momentum’ of its own that is not always directly subject to the will; for instance, an expert chess player may automatically ‘see’ that a particular position is strong even when given evidence that the position is bad. Now, he may or may not trust the automatic process, and it is often quite reasonable not to trust, and even to override, automatic processes in some circumstances. So, let’s imagine that the chess player does override the process, and chooses to avoid the seemingly strong position. The ballisticity of his ability to see the strength of a position is nevertheless manifested in the fact that the chess player continues to ‘see’ the position as strong, even though he fails to believe that it is.

The case of automatic judgments (as per our hypothesis) is a little different because one cannot easily come to believe that the process is unreliable on an occasion; thus, automatic judgments are harder to manage than automatic intuitions or automatic
perceptions. The ballisticity of an automatic judgment will also manifest slightly differently. Unlike the case of an automatic intuition, which persists despite a judgment that it is unreliable, an automatic judgment persists despite one coming to know of evidence that the judgment is false or unreliable. This kind of ballisticity nicely matches the way in which dumbfounded individuals discount or ignore evidence to the contrary of their moral judgments. Note though that even in the case of moral judgment, ballisticity is not inevitable, so just because the automatic process typically occurs unmodified, does not imply that it is not possible to modify through intention (Saling & Phillips, 2007).

**Hypothesis 3: People Aren’t Actually Dumbfounded**

Another plausible hypothesis is that the dumbfounding behaviour exhibited in experimental settings is not, in fact, evidence of genuine dumbfounding at all. This hypothesis has been foreshadowed in this dissertation when talking about pressure from the experimenter to ‘perform’ in a particular way, and the believability of stipulations of harmlessness. On this view, while people appear to maintain their moral judgments even when they remain incapable of justifying them, these appearances are misleading. Rather, the experimental participants in fact hold views that, by their own perspective, do justify their judgments.

Consider the following example: I am asked whether it is wrong for siblings to have sex and I judge that it is wrong. In generating this judgment, I engage in a reasoning process that, in this case, involves appealing to familiar harm-based reasons. I articulate my judgment and the associated justification but am told that, in this particular instance, there is no harm associated with the siblings’ action. I do not, however, actually believe that there is no harm in this case because I strongly believe that incest is always associated with potential harm. This case is too close to a real-world scenario for me to suspend disbelief and to accept arbitrary stipulations. For this reason, I fail to internalise the stipulation that there is no harm in this case (indeed, I might ‘refuse’ to accept the stipulation) but I may not feel that I can argue because my interlocutor is better at arguing than me (this would easily have been the case in the original study of Haidt et al., 2000, where a well-prepared researcher played ‘devil’s advocate’). So, when I say, “I don’t know why it is wrong, I just know that it is wrong”, what I’m actually thinking is, “I don’t know how to argue this, but I simply can’t accept your stipulation that there’s no harm in the scenario under discussion. So, I’m confident that the behaviour is wrong, and I know why it’s wrong: because it causes harm;
however, you’ve set things up so that I’m not entitled to say this, so there’s nothing more for me to say”. Therefore, while I appear, on the face of it, to be morally dumbfounded (as my overt statement reflects that I am left without reasons to support my judgment), I am, in fact, just opting out of an argument.

Of course, I may think that the risk of harm is very small in the Incest scenario (despite the experimenter’s insistence that there is no risk of harm at all). Nevertheless, for both prudential and moral reasons, I should still judge that the siblings’ actions are wrong. This is because the siblings would still be foolish to take what (given their knowledge of the balance of probable costs versus benefits) is still a small risk of a significant harm (Jacobson, pp. 299-301). It is, therefore, perfectly rational to judge the siblings’ actions to be impermissible as long as I think that there is even a small risk of harm.

To illuminate this argument, think of two cases of drunk driving. Harry has a blood alcohol level of 0.09 and drives home on a deserted country road late at night where the chance of encountering another driver or pedestrian is incredibly small. Contrast this with Larry with the same blood alcohol level who drives home late at night in a busy city where the chance of encountering other drivers and pedestrians is extremely likely. In the first case, the risk to the driver and others is very small but in the second case the risk of significant harm to the driver and to other road users is very high. Despite the stipulations in Harry’s case, one should still judge that his actions are impermissible as it is still wrong to act in a way that has a small chance of causing a significant harm.

**Hypothesis 4: Resoluteness in Moral Judgment**

Remaining resolute in one’s moral judgment is not necessarily irrational. Rather, there are good reasons for maintaining a judgment even when the justification for the judgment is undercut. These are as follows:

- I resist intellectual bullying. In Haidt et al. (2000), the experimenter was the ‘expert’ and the participants were ‘novices’. As such, the participants were susceptible to changing their judgment out of deference to the expertise of the experimenter, rather than because they were convinced by the experimenter to support an alternative judgment; but it is not obvious that we should assume that other people’s moral judgments are more credible than our own. In other words, it is unclear that it makes sense to talk about expertise in moral judgment; therefore, it is rational to resist other people’s attempts to manipulate our judgments. This is akin to perceptual
judgements. I judge that a car is red, my friend (looking at the same car) judges that the car is brown. I have no reason to assume that my friend’s perceptual judgement is more credible than mine (as presumably we are both equally susceptible to perceptual illusions) and so I will be unlikely to change my judgement of the car’s colour in light of my friend’s judgment. In other words, it seems non-sensical to think that perception relies on particular expertise. Of course, the degree of trust that I place in the two judgments would change if I (or my friend) were colour blind. Contrast this with another case, that of mathematical expertise. If there is a conflict between my solution to a complex maths problem and the solution generated by my mother, a mathematician, I should defer to her because she clearly has expertise that I lack.

- I maintain my authenticity. These are my moral judgments and I cannot be pressured into changing them. Consider this analogous case: I decide to go on a diet because my doctor has told me that I need to lose weight to lower my blood pressure. I don’t actually follow through, however, because I don’t feel like ‘me’ when I am on a diet. I maintain my authenticity by refusing to go on a diet.

Given these observations, we might hypothesise that dumbfounded individuals are, in a similar fashion, simply being resolute in their moral judgments. On this hypothesis, there is no reason to doubt that dumbfounded individuals made their original judgments as a result of a process of reasoning. But once they ‘dig their heels in’, they maintain these judgments despite being offered reasons to revise their judgments. Resoluteness might consist in ignoring such contrary reasons. When I make a resolute judgment, I am unwilling to continue deliberating about its validity because, if I do, I run the risk of entering into an argument that I may not win and that may ultimately result in me having to relinquish my judgment. On this hypothesis, when a dumbfounded individual says that he “doesn’t know” what justifies his judgment, what he means is that he no longer prepared to question its justification. This means that he will refuse to countenance arguments against his judgment, or even to reconstruct the reasons that led him to make it in the first place.

Why is Moral Dumbfounding Important?
The moral dumbfounding phenomenon, first identified by Haidt et al. (2000) is clearly important even if it does not support the SIM (Haidt, 2001). Moral dumbfounding is broadly consistent with a number of theories of moral judgment, including rationalist accounts and, therefore, moral dumbfounding does not provide evidence for (or against)
any particular theory of moral judgment. Nevertheless, it is a barrier to moral discourse, or at least poses a challenge to genuine moral discourse. This is because, at the point where we are unable or unwilling to articulate reasons for a judgment, we are no longer able to defend or re-evaluate our judgments (for reasons articulated in this chapter). Therefore, an impasse is reached, and no further discussion can ensue. This has particularly negative ramifications in political discourse as individuals at different ends of the political spectrum consistently make and commit to different moral judgments, often making conversation untenable (Frimer, Skitka, & Motyl, 2017; Haidt & Graham, 2007; Skitka & Morgan, 2014). This makes understanding the causes of moral dumbfounding important. Specifically, if we know what causes dumbfounding (which is ultimately an empirical question), we can know how to modify institutions and practices to moderate its negative effects.
Chapter 8: Conclusion

Instances in which we firmly judge something to be wrong, coupled with an inability to articulate plausible (or indeed, any) reasons to justify the judgment is a phenomenon that has been termed moral dumbfounding (Haidt et al., 2000). Moral dumbfounding is certainly familiar but a number of outstanding questions needed to be addressed such as (a) is moral dumbfounding a genuine phenomenon; (b) what is moral dumbfounding (i.e., how can we accurately measure it); (c) even if moral dumbfounding is a genuine phenomenon that can be accurately measured, does it only occur under particular conditions, or do particular conditions make it more likely to occur; and (d) do individual differences predict the rate of moral dumbfounding? This dissertation comprises the first full replication of the seminal moral dumbfounding study conducted by Haidt et al. (2000), but it also involves a series of methodological improvements to address limitations of the original research.

New evidence for moral dumbfounding as a genuine and replicable phenomenon is provided in Studies 1 and 3. In Study 2, the following empirically supported definition of the phenomenon was derived: moral dumbfounding is an inability to justify disapproval for a seemingly harm-free action. Study 3 confirms that moral dumbfounding can be elicited (albeit at a much lower frequency than originally proposed by Haidt) in response to a range of scenarios (including third-person stories and visual scenes) and first-person behavioural tasks. Given the limitations of the current sample size and variability, only limited conclusions can be drawn with regard to individual differences at this stage; however, the phenomenon was found in Study 4 not to be a product of age, gender, nationality, ethnicity, religiosity, political preference, education, SES, nor personality traits. Given the current lack of evidence for influence of static factors (i.e., individual characteristics) on the phenomenon, it appears that dynamic factors (i.e., situational forces) may be the primary drivers of moral dumbfounding. Such situational forces may include the unique conditions under which empirically observed moral dumbfounding is elicited. While there was no relationship between susceptibility to experimenter pressure and moral dumbfounding (Study 4), the phenomenon was elicited under experimental conditions in this dissertation, and all participants reported feeling some degree of pressure. Consequently, the social dynamics of the interview context is still a possible mechanism underpinning moral dumbfounding.
Although the manifestation of moral dumbfounding is a failure to articulate reasons for a moral judgment, this is not evidence for the absence of reasons in the formation of a moral judgement or for the impotence of a reasoning process in moral judgement. Indeed as shown here, reasoning can modify a moral judgement as demonstrated by a change from initial to final judgement. Rather, as outlined in Chapter 7, there are a number of other plausible processes that can explain moral dumbfounding that are consistent with the causal role of reason in moral judgment. Moral dumbfounding is, therefore, compatible with a number of accounts of moral judgment, including rationalist accounts, and does not provide specific evidence for or against any theory of moral judgment.
References


193


Appendices

Appendix A  Ethics Documents

Appendix A.1  CSU Human Research Ethics Committee Approval

14 April 2016

Ms Tammy Orreal-Scarborough
8 McCarthy Close
EGLINTON NSW 2795

Dear Ms Orreal-Scarborough,

Charles Sturt University’s (CSU) Human Research Ethics Committee (HREC) reviewed your research proposal entitled “Does intuition find no reason? A replication and modification of Haidt, Björklund, and Murphy’s (2000) moral dumbfounding study” on 4 April 2016.

The CSU HREC operates in accordance with the National Health and Medical Research Council’s National Statement on Ethical Conduct in Research Involving Humans.

I am pleased to advise that the project meets the requirements of the National Statement and ethical approval for this research is granted for a twelve-month period from 14/4/2016.

The protocol number issued with respect to this project is 2016/045. Please be sure to quote this number when responding to any request made by the Committee.

Please note the following conditions of approval:

- all Consent Forms and Information Sheets are to be printed on Charles Sturt University letterhead. Students should liaise with their Supervisor to arrange to have these documents printed;
- you must notify the Committee immediately in writing should your research differ in any way from that proposed. Forms are available at [http://www.csu.edu.au/research/ethics_safety/human/hrec_managing](http://www.csu.edu.au/research/ethics_safety/human/hrec_managing);
- you must notify the Committee immediately if any serious and or unexpected adverse events or outcomes occur associated with your research, that might affect the participants and therefore ethical acceptability of the project. An Adverse Incident Form is available from the website as above;
- amendments to the research design must be reviewed and approved by the Human Research Ethics Committee before commencement. Forms are available at the website above;
- if an extension of the approval period is required, a request must be submitted to the Human Research Ethics Committee. Forms are available at the website above;


197
• you are required to complete a Report On Research Project, which can be downloaded as above, by 12/4/2017 if your research has not been completed by that date;
• you are required to submit a final report, the form is available from the website above.

YOU ARE REMINDED THAT AN APPROVAL LETTER FROM THE CSU HREC CONSTITUTES ETHICAL APPROVAL ONLY.

If your research involves the use of radiation, biological materials, chemicals or animals a separate approval is required from the appropriate University Committee.

The Committee wishes you well in your research and please do not hesitate to contact the Executive Officer on telephone (02) 6338 4628 or email ethics@csu.edu.au if you have any enquiries.

Yours sincerely

Regan McIntosh
Executive Officer
Human Research Ethics Committee
Direct Telephone: (02) 6338 4628
Email: ethics@csu.edu.au

This HREC is constituted and operates in accordance with the National Health and Medical Research Council’s (NHMRC) National Statement on Ethical Conduct in Human Research (2007)
Appendix A.2  Information Statement

4. Are there risks and benefits to me in taking part in this study?

While there are no known risks for you in participating in this study, due to the possibly confronting nature of the situations, you may experience some discomfort. You do, however, have the right to refuse to respond to any question or request that is asked of you.

While you will not receive any direct benefits as a result of participating in this study, you will gain the indirect benefit of experiencing how behavioural studies are conducted through a University.

5. What if I don’t want to take part in this study?

Participation in this study is entirely your choice. Only those people who give their informed consent will be included in the project. There will be no consequences for you should you choose not to participate.

6. What if I participate and want to withdraw later?

Participation in this study is completely voluntary and you may decide to stop being a part of the study at any time without explanation. There will be no consequences for withdrawing from the study at any stage. Should you wish to withdraw, you may request that any data collected to that point be destroyed or not included in the results. Once you have completed participation, however, your data will be combined with other participant data. At that point it will not be possible to identify individual contributions and, therefore, destruction or withdrawal of your data will not be possible.

7. How will my confidentiality be protected?

Your personal information will remain confidential to the researchers. Your consent form will be stored in a sealed envelope, in a locked filing cabinet, in the Principal Researcher’s office on Charles Sturt University Bathurst campus. The video recording of your interview and the answers to the questionnaire will be stored on the Principal Researcher’s computers, which are password protected. All information you provide, including the video recording of your interview, will be destroyed five years after completion of the study.

8. What will happen to the information that I give you?

Only the researchers will have access to any information you provide, including the video recording of your interview. You may review the recording upon request.

You will not be identified in the results, which will only be presented as group data in the Principal Researcher’s doctoral thesis and relevant scientific journals and conferences.

9. What should I do if I want to discuss this study further before I decide?

If you would like further information, please contact the Principal Researcher, Ms Tammy Oreal-Scarborough, or either of her supervisors, Dr Andrew McGrath or Dr Lauren Salting. Should you choose to participate, please ensure that any questions you may have are answered to your satisfaction before the interview begins.
10. Who should I contact if I have concerns about the conduct of this study?

Charles Sturt University’s Human Research Ethics Committee has approved this project (#2018/045). If you have any complaints or reservations about the ethical conduct of this project, you may contact the Committee through the Executive Officer.

The Executive Officer
Human Research Ethics Committee
Office of Governance and Corporate Affairs
Charles Sturt University
Private Bag 29
Panorama Avenue
Bathurst NSW 2795
Tel. (02) 6338 4828
Email: ethics@csu.edu.au

Any issues you raise will be treated in confidence and investigated fully and you will be informed of the outcome.

Thank you for considering this invitation.
This information sheet is for you to keep.

Tammy Oreal-Scarborough
CONSENT FORM

Principal Researcher
Tammy Oreal-Scarborough, BPsych (Hons)
Charles Sturt University
School of Psychology, Bathurst
Doctor of Philosophy (Psychology) Candidate
Phone: 0431 988 459
Email: terrisa@scarborough@gmail.com

Supervisors
Dr. Andrew McGrath, BA (Hons), PhD
Charles Sturt University
School of Psychology, Bathurst
Phone: (02) 6338 4591
Email: amcgrath@csu.edu.au

Dr. Lauren Ralston, BA, BSc (Hons), PhD
Charles Sturt University
School of Psychology, Wagga Wagga
Phone: (02) 6933 2670
Email: lralston@csu.edu.au

Moral Judgments of Unfamiliar Situations

☐ I understand that the above research will be conducted as described in the Information Statement, a copy of which I have retained.

☐ I agree to participate in the above research and give my consent freely.

☐ I consent to participating in an interview and having it video recorded.

☐ I understand the video recording of my interview will be stored on the hard-drive of the researcher’s computers, which are password protected.

☐ I understand the forms I complete during the interview will be stored in a sealed envelope, in a locked filing cabinet, in the Principal Researcher’s office on Charles Sturt University Bathurst campus.

☐ I understand that my personal information will remain confidential to the researchers and I will not be identified in the results, which will only be presented as group data.

☐ I understand that only the researchers will have access to the data I have supplied and my data will be destroyed, as recommended by CSU, after five years.

☐ I understand I can withdraw from this research at any time and do not have to give any reason for withdrawing.

☐ I understand that there will be no consequences for withdrawing from this research at any stage.
I understand that I have the right to ask that any data I have supplied to the point of withdrawal be withdrawn/destroyed.

I understand that once I have completed participation, my data will be combined with other participant data, at which point it will not be possible to identify individual contributions and, therefore, destruction or withdrawal of my data will not be possible.

I understand that there are no known benefits or risks for me in participating in this research.

I have had the opportunity to have questions answered to my satisfaction.

Participant's Name (Printed)  
Participant's Signature  
Date

Charles Sturt University's Human Research Ethics Committee has approved this research (#2016/045). If you have any complaints or reservations about the ethical conduct of this research, you may contact the Committee through the Executive Officer:

The Executive Officer  
Human Research Ethics Committee  
Office of Governance and Corporate Affairs  
Charles Sturt University  
Private Bag 29  
Panorama Avenue  
Bathurst NSW 2795  
Tel: (02) 6338 4628  
Email: ethics@csu.edu.au

Any issues you raise will be treated in confidence and investigated fully and you will be informed of the outcome.
Appendix B  Photographs of the Data Collection Laboratory

Appendix B.1  Equipment View from Participant Seat

Appendix B.2  Equipment View from Experimenter Seat
Appendix B.3  Data Collection Laboratory Set-Up

Appendix B.4  Participant and Experimenter Positioning

Note. The people in this photograph are the author (right) and her research assistant and not participants of the studies undertaken and reported in this dissertation.
Appendix C  Stimuli Employed in Study 1

Appendix C.1  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 10 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. Was there anything wrong with what he did?
Appendix C.2  The Incest Story

Julie and Mark, who are brother and sister, are travelling together in France. They are both 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 it, but they decide not to do it again. They keep that night as a special secret between them, which makes them feel even closer to each other. So, what do you think about this? Was it wrong for them to have sex?
Appendix C.3  The Cannibal Story

Jennifer works in a medical school pathology lab as a research assistant. The lab prepares human cadavers that are used to teach medical students about anatomy. The cadavers come from people who had donated their body to science for research. One night, Jennifer is leaving the lab when she sees a body that is going to be discarded the next day. Jennifer was a vegetarian, for moral reasons. She thought it was wrong to kill animals for food. But then, when she saw a body about to be cremated, she thought it was irrational to waste perfectly edible meat. So, she cut off a piece of flesh, and took it home and cooked it. The person had died recently of a heart attack, and she cooked the meat thoroughly, so there was no risk of disease. Is there anything wrong with what she did?
Appendix C.4  The Roach Task

Experimenter: Do you like apple juice?
Participant: Yes.
Participant: No.
Experimenter: OK, then, I have some water.
Experimenter takes the appropriate beverage, a napkin, a cup, the roach container, and the tweezers to table.
Experimenter: OK, I have here a bottle of apple juice (water), which I’m going to pour into this glass.
Experimenter pours beverage into glass.
Experimenter: Would you be willing to take a sip of the juice (water)?
Experimenter waits for the participant to take a sip.
Experimenter: OK, now I have here in this container some sterilised cockroaches. I bought these cockroaches from a cockroach supply company.
Experimenter shows the box and receipt to the participant.
Experimenter: The roaches were raised in a clean environment, but just to be certain, we sterilised the roach again in an autoclave, which heats everything so hot that no germs can survive. I’m going to dip this cockroach into the juice (water), like this.
Experimenter picks up the cockroach with the tweezers and immerses it in the juice (water).
Experimenter: Now, would you take a sip of the juice (water)?
Experimenter waits for the participant to respond.
Appendix C.5  The *Soul* Task

Experimenter: I have a piece of paper here. If you agree to sign it, I'll give you two dollars. If you sign it, you can then rip up the paper immediately, and keep the pieces yourself.

Experimenter places the ‘contract’ on the table in front of the participant along with a pen.

Experimenter: Would you be willing to sign that piece of paper?

Experimenter waits for the participant to respond.

```
I, ________________

hereby sell my soul, after my death,

to Tammy Orreal Scarborough,

for the sum of $2.

____________________
(signed)

Note: This form is part of a psychology experiment.
It is NOT a legal or binding contract, in any way.
```
Appendix D  Questionnaires Employed in the Current Study

Appendix D.1  After Each Task Questionnaire

Here are some questions about the story or task that you just completed. Please fill in the number that best indicates the extent to which you agree or disagree with each statement listed below. Be as honest as possible but rely on your initial feeling and do not think too much about each item. Please use the following scale:

<p>| | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>Strongly Disagree</td>
<td>Disagree</td>
<td>Neither Agree nor Disagree</td>
<td>Agree</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Strongly Agree</td>
</tr>
</tbody>
</table>

1. I felt confident in my initial judgment about the story/task.

2. My initial judgment was based on careful reasoning about the facts involved in the story/task.

3. My initial judgment was based on a ‘gut feeling’ about the story/task.

4. I felt confident in my final judgment about the story/task.

5. My final judgment was based on careful reasoning about the facts involved in the story/task.

6. My final judgment was based on a ‘gut feeling’ about the story/task.

7. I was confused by this story/task.

8. I was irritated by this story/task.

9. I believe that the story/task is representative of a situation which may happen in the real world.
10. I find it difficult to imagine that the nature of the story/task is harm-free.

If the response option for item 10 was a 4 (Agree) or 5 (Strongly Agree), the participant was asked:

I believe the nature of the story/task is not harm-free because:
Appendix D.2  Demographic Questionnaire

1. What is your age? __________ years

2. Please place a tick in the box next to your gender: male □ female □

3. In which country were you born? _____________________________

4. Are you of Aboriginal or Torres Strait Islander origin? yes □ no □

5. With which racial or ethnic group(s) do you most identify? _________

6. Is English your first language? yes □ no □

   If NO, how long have you been speaking English? ____________ years

7. What is your major source of income?

   Financial support from family or others □

   Government benefits (e.g. ABSTUDY, Austudy, Youth Allowance, etc.) □

   Government or university funded scholarship □

   Other government income support □

   Wages: full-time employment □

   Wages: part-time employment □

   Salary □

   Other □

   If OTHER, please specify: ________________________________
8. In terms of education and income, would you say your parents are:

Upper class
Upper-middle class
Middle class
Lower-middle class
Working class

9. Do you consider yourself to be a religious person? yes ☐ no ☐

If YES:
a) What religion are you affiliated with?

Anglican
Baptist
Buddhism
Catholicism
Eastern Orthodox
Hinduism
Islam
Judaism
Lutheran
Pentecostal
Presbyterian
Uniting Church
Other Christian, not further defined
Other non-Christian, not further defined
b) How important an influence is your religious faith in your life?

- Most Important
- Very Important
- Somewhat Important
- Not too Important
- Not at All Important

If NO, do you consider yourself to be:

- Agnostic
- Atheist
- Humanist
- Rationalist
- No religion, not further defined

10. Please indicate to what extent you agree or disagree with the following statement:
   “An individual’s rights should be sacrificed for the well-being of society in certain situations”.
   - Strongly Disagree
   - Disagree
   - Unsure / No Opinion
   - Agree
   - Strongly Agree

11. Please indicate to what extent you agree or disagree with the following statement:
   “The role of the criminal justice system should be to punish versus rehabilitate offenders”.
   - Strongly Disagree
   - Disagree
   - Unsure / No Opinion
   - Agree
   - Strongly Agree

12. Please indicate to what extent you agree or disagree with the following statement:
   “I support the legalisation of same-sex marriage in Australia”.
   - Strongly Disagree
   - Disagree
   - Unsure / No Opinion
   - Agree
   - Strongly Agree
13. What is the name of the course in which you are currently enrolled?


14. At which university are you enrolled as a student?
### Appendix E  Hierarchical Coding Scheme for Variables Derived from Data

<table>
<thead>
<tr>
<th>Variable</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Judgment</strong></td>
<td>This category contains the judgments of participants regarding the presented stimuli.</td>
</tr>
<tr>
<td>Initial Judgment</td>
<td>Initial judgment as to the acceptability of the target’s behaviour.</td>
</tr>
<tr>
<td>Intermediate Judgment</td>
<td>Intermediate judgment as to the acceptability of the target’s behaviour.</td>
</tr>
<tr>
<td>Final Judgment</td>
<td>Final judgment as to the acceptability of the target’s behaviour.</td>
</tr>
<tr>
<td><strong>Justification</strong></td>
<td>This category contains the justifications offered by the participants in support of their judgment to the presented stimuli.</td>
</tr>
<tr>
<td>Initial Judgment Justification</td>
<td>Offered justification for initial judgment as to the acceptability of the target’s behaviour.</td>
</tr>
<tr>
<td>Intermediate Judgment Justification</td>
<td>Offered justification for intermediate judgment as to the acceptability of the target’s behaviour.</td>
</tr>
<tr>
<td>Final Judgment Justification</td>
<td>Offered justification for final judgment as to the acceptability of the target’s behaviour.</td>
</tr>
<tr>
<td><strong>Verbal</strong></td>
<td>This category contains verbal markers of dumbfounding according to Haidt et al. (2000); that is, dead-end arguments, dumbfounding statements, statements of “I don’t know”, and unsupported declarations. It also contains novel verbal markers found in Study 1 and utilised in the subsequent studies; that is, illogical arguments, puzzlement, tautological reasons, and unwillingness to support. This verbal category also contains speech dysfluencies (false starts, fillers, long pauses, repaired utterances).</td>
</tr>
<tr>
<td>Replication Verbal Markers</td>
<td>This category contains verbal markers of dumbfounding according to Haidt et al. (2000); that is, dead-end arguments, dumbfounding statements, statements of “I don’t know”, and unsupported declarations.</td>
</tr>
<tr>
<td>Dead-End Argument</td>
<td>An incomplete offered justification for an unacceptability judgment of the target's behaviour; for example, after talking the participant appears to realise that the argument is not going to work and stops in the middle of giving their justification, without prompting from the interviewer.</td>
</tr>
<tr>
<td>Dumbfounding Statement</td>
<td>Direct statements of dumbfounding offered as justification for an unacceptability judgment of the target's behaviour; for example: &quot;I believe X is wrong, but I can't find a reason&quot;; &quot;I can't find the words to explain it&quot;; &quot;I don't know&quot;.</td>
</tr>
<tr>
<td>I Don't Know</td>
<td>&quot;I don't know&quot; used as the offered justification for an unacceptability judgment of the target's behaviour.</td>
</tr>
<tr>
<td>Unsupported Declaration</td>
<td>Unsupported declarations used as the offered justification for an unacceptability judgment of the target’s behaviour; for example: “It’s just wrong to do that”; “That’s terrible”.</td>
</tr>
<tr>
<td>Novel Verbal Markers</td>
<td>This category contains novel verbal markers found in Study 1 and utilised in the subsequent studies; that is, illogical arguments,</td>
</tr>
<tr>
<td>Variable</td>
<td>Description</td>
</tr>
<tr>
<td>------------------------------</td>
<td>-----------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Illogical Argument</td>
<td>An illogical argument used as the offered justification for an unacceptability judgment of the target’s behaviour; for example: “They might get pregnant” when responding to the Incest story in which two forms of birth control were used.</td>
</tr>
<tr>
<td>Puzzlement</td>
<td>A statement of puzzlement at an inability to provide a reason for their own judgment as to the acceptability of the target’s behaviour; for example: “I don’t know. That’s really weird isn’t it?”</td>
</tr>
<tr>
<td>Tautological Reason</td>
<td>Restatement of facts from the task used as the offered justification for an unacceptability judgment of the target’s behaviour; for example: “It’s cannibalism”; “It’s incest”.</td>
</tr>
<tr>
<td>Unwillingness to Support</td>
<td>Statements in support of acceptability of the target’s behaviour but an unwillingness to make such a judgment; for example: “I understand that there is no harm and it should be OK, but I just still think it is wrong”.</td>
</tr>
<tr>
<td>Speech Dysfluency</td>
<td>This category contains breaks, irregularity, or non-lexical vocable that occurs within the flow of otherwise fluent speech; for example, false starts, fillers, long pauses, repaired utterances.</td>
</tr>
<tr>
<td>False Start</td>
<td>Words and sentences that are cut off mid-utterance; phrases that are restarted or repeated and repeated syllables.</td>
</tr>
<tr>
<td>Filler</td>
<td>Grunts or non-lexical utterances such as “uh”, “erm”, “um”, “well”, “you know”.</td>
</tr>
<tr>
<td>Long Pause</td>
<td>A long pause before answering or a long pause between words.</td>
</tr>
<tr>
<td>Repaired Utterance</td>
<td>Instances of participants correcting their own slips of the tongue or mispronunciations (before the interviewer gets a chance to).</td>
</tr>
<tr>
<td>Behavioural</td>
<td>This category contains the observed physical behaviour of the participants thought to indicate psychic discomfort.</td>
</tr>
<tr>
<td>Face Touching</td>
<td>Touching of the face or head during a conversational turn.</td>
</tr>
<tr>
<td>Fiddling</td>
<td>Manipulating hands or some object with hand or hands.</td>
</tr>
<tr>
<td>Laughter</td>
<td>Laughter during a conversational turn.</td>
</tr>
<tr>
<td>Posture Shifting</td>
<td>Moving around in the seat, leaning forward, leaning backwards.</td>
</tr>
<tr>
<td>Temporal</td>
<td>This category contains the records for the time-based variables of the participants’ judgments and justifications.</td>
</tr>
<tr>
<td>Total Task Time</td>
<td>The difference (in seconds) between the completion of the experimenter’s request for a judgment and completion of all responding.</td>
</tr>
<tr>
<td>Seconds to Initial Judgment</td>
<td>The difference (in seconds) between the completion of the experimenter’s request for a judgment and the start of the participant’s initial judgment as to the acceptability of the target’s behaviour.</td>
</tr>
<tr>
<td>Seconds to Initial Justification</td>
<td>The difference (in seconds) between the completion of the experimenter’s request for a judgment and the start of the participant’s initial justification for their judgment of the</td>
</tr>
<tr>
<td>Variable</td>
<td>Description</td>
</tr>
<tr>
<td>----------------------------</td>
<td>-----------------------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>acceptability of the target's behaviour.</td>
<td></td>
</tr>
<tr>
<td>Seconds Judgment Precedes Justification</td>
<td>The difference (in seconds) between the start of the participant's initial judgment as to the acceptability of the target's behaviour and the start of their initial justification for their judgment.</td>
</tr>
</tbody>
</table>
Appendix F  Stimuli Employed in Study 3

Appendix F.1  The John Dilemma

In Australia, new medicines need to be approved and listed by the Therapeutic Goods Administration (TGA) before they can be lawfully supplied. Some medicines, especially cancer medications, have TGA approval but have very high costs. The Pharmaceutical Benefits Scheme (PBS) subsidises the cost of TGA approved medications to make them affordable. For a medicine to be subsidised by the PBS, it must be approved by the Pharmaceutical Benefits Advisory Committee (PBAC). It takes an average of 1½ years from when the application is made to the PBAC to when a PBS subsidy is granted, if it is approved at all.

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 chemical compound that had TGA approval but had not yet been granted a PBS subsidy (although an application had recently been made). The cost of the drug was 100 times what the drug would cost if subsidised.

The sick woman’s husband, John, went to everyone he knew to borrow the money, but he could only get together about half of what it cost. He contacted the head of the pharmaceutical company that manufactured the drug and told him that his wife was dying and asked him to sell it cheaper or let him pay later. But the head of the company said, “No, my company manufactures the drug, and we want to profit from it”. So, John got desperate and broke into the warehouse of the pharmaceutical company to steal the drug for his wife. Was there anything wrong with what he did?
Appendix F.2  The Masturbation Story

Paul was alone in his house and decided to masturbate. He was laying on his bed with his pants down when suddenly, his dog Benny, walked into his bedroom. Benny jumped on the bed and started sniffing Paul’s private area. Paul tried several times to push his dog away but, each time, Benny went to bite his hand. Benny then started licking Paul’s private area. Paul found himself aroused despite his misgivings. Even though they both enjoyed it, Paul decides to never let it happen again. From then on, whenever Paul wanted to masturbate, he shut Benny out of his room. Paul never tells anyone about what happened, and the special secret between him and his dog makes him feel even closer to Benny. So, what do you think about this? Was it wrong for Paul to have allowed Benny to help him masturbate?
Appendix F.3 The *Liver* Story

Jennifer works as a theatre nurse in a rural hospital and was always feeling very tired despite getting enough sleep. After getting some blood tests, Jennifer was diagnosed with very low iron levels. The doctor advised Jennifer to increase her intake of iron-laden foods and suggested lambs fry as a delicious option. Lambs fry consists of tender pieces of lamb’s liver and bacon in a red wine gravy. It is a highly nutritious meal that is rich in iron, vitamin A, zinc, and folate. Unfortunately, Jennifer did not have very much money after paying her living expenses and so, for months, she had been having soup and toast for every evening meal.

One day at work, Jennifer was assisting in an operation to replace a patient’s diseased liver with a partial liver from a healthy live donor. The surgeon was attempting to remove the patient’s diseased liver when a large amount of unexpected bleeding occurred. The surgeon was unsuccessful in controlling the bleeding, and the patient died. The surgeon instructed Jennifer to dispose of the partial donor liver because by the time another recipient was found, and the organ was transported to their location, it would no longer be a viable organ for transplantation.

When Jennifer realised that the organ was just going to be disposed of, she thought it was irrational to waste perfectly edible meat. So, she took it home and cooked it. The liver donor had undergone the required blood tests to ensure the health of the organ, and Jennifer cooked the meat thoroughly, so there was no risk of disease. Is there anything wrong with what she did?
Appendix F.4  The Lolly Task

Experimenter: Do you like lollies?
Participant: Yes.  Participant: No.

Experimenter: OK, then, I have some biscuits.
Experimenter takes the appropriate food and a napkin to table.
Experimenter: OK, I have here a lolly (biscuit), which I'm going to unwrap.
Experimenter takes the individual wrapping off the food and places it on the napkin.
Experimenter: Would you be willing to eat the lolly (biscuit)?
Experimenter waits for the participant to eat the food.
Experimenter: OK, now I have here a square of new carpet. I bought this carpet from a
flooring supply company.
Experimenter shows the receipt to the participant.
Experimenter: The carpet is brand new but, just to be certain, I shampooed the carpet
and then treated it with ultraviolet radiation for six hours to eliminate any germs. I'm
going to unwrap this second lolly (biscuit), like this.
Experimenter takes the individual wrapping off the food.
Experimenter: I'm now going to drop the lolly (biscuit) and drop it onto the carpet and
leave it for 10 seconds.
Experimenter drops the food onto the carpet and waits 10 seconds.
Experimenter: Now, would you eat the lolly (biscuit)?
Experimenter waits for the participant to respond.
Appendix F.5 The *Extended Soul* Task

Experimenter: I have a piece of paper here. If you agree to sign it, I'll give you two dollars. If you sign it, you can then rip up the paper immediately, and keep the pieces yourself.

Experimenter places the 'contract' on the table in front of the participant along with a pen.

Experimenter: Would you be willing to sign that piece of paper?
Experimenter waits for the participant to respond.

If the participant's final judgment is a refusal to sign:

Experimenter: Is there any amount of money that I could pay you to get you to sign the contract?

Experimenter waits for the participant to respond.

---

I, ____________________________

hereby sell my soul, after my death,

       to Tammy Omeat Scarborough,  

       for the sum of $2.

                      (signed)

---

Note: This form is part of a psychology experiment. It is NOT a legal or binding contract, in any way.
Appendix F.6  The Atlas Video

This video depicts a life-size, bipedal, humanoid robot operating both inside buildings and outdoors. Atlas is seen opening doors, walking through the snow, walking next to a human, and working in a warehouse shifting boxes. In the final scene, a man is seen pushing Atlas in the back with a pole. Atlas falls to the ground.
Appendix F.7  The *Spot* Video

This video depicts a life-size, quadruped, animalian robot ‘dog’ operating both inside buildings and outdoors. *Spot* is seen navigating through an open-plan office space, climbing outside stairs and hills, walking next to another robot ‘dog’, and running next to a human. In the final scene, a man is seen kicking *Spot* in the ‘stomach’ with his foot. *Spot* staggers to regain its footing after sustaining the blow but does not fall to the ground.
Appendix G  The Big Five Aspect Scales Personality Questionnaire

Here are a number of characteristics that may or may not describe you. For example, do you agree that you seldom feel blue, compared to most other people? Please fill in the number that best indicates the extent to which you agree or disagree with each statement listed below. Be as honest as possible but rely on your initial feeling and do not think too much about each item. Please use the following scale:

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Strongly Disagree</td>
<td>Disagree</td>
<td>Neither Agree nor Disagree</td>
<td>Agree</td>
<td>Strongly Agree</td>
</tr>
</tbody>
</table>

1. ___Seldom feel blue.
2. ___Am not interested in other people’s problems.
3. ___Carry out my plans.
4. ___Make friends easily.
5. ___Am quick to understand things.
6. ___Get angry easily.
7. ___Respect authority.
8. ___Leave my belongings around.
9. ___Take charge.
10. ___Enjoy the beauty of nature.
11. ___Am filled with doubts about things.
12. ___Feel others’ emotions.
13. ___Waste my time.
14. ___Am hard to get to know.
15. ___Have difficulty understanding abstract ideas.
16. ___Rarely get irritated.
17. ___Believe that I am better than others.
18. ___Like order.
19. ___Have a strong personality.
20. ___Believe in the importance of art.
21. ___Feel comfortable with myself.
22. ___Inquire about others’ well-being.
23. ___Find it difficult to get down to work.
24. ___Keep others at a distance.
25. ___Can handle a lot of information.
26. ___Get upset easily.
27. ___Hate to seem pushy.
28. ___Keep things tidy.
29. ___Lack the talent for influencing people.
30. ___Love to reflect on things.
31. ___Feel threatened easily.
32. ___Can’t be bothered with other’s needs.
33. ___Mess things up.
34. ___Reveal little about myself.
35. ___Like to solve complex problems.
36. ___Keep my emotions under control.
37. ___Take advantage of others.
38. ___Follow a schedule.
39. ___Know how to captivate people.
40. ___Get deeply immersed in music.
41. ___Rarely feel depressed.
42. ___Sympathize with others’ feelings.
43. Finish what I start. 71. Am not embarrassed easily.
44. Warm up quickly to others. 72. Take an interest in other people’s
45. Avoid philosophical discussions. lives.
46. Change my mood a lot. 73. Always know what I am doing.
47. Avoid imposing my will on others. 74. Show my feelings when I’m happy.
48. Am not bothered by messy people. 75. Think quickly.
49. Wait for others to lead the way. 76. Am not easily annoyed.
50. Do not like poetry. 77. Seek conflict.
51. Worry about things. 78. Dislike routine.
52. Am indifferent to the feelings of 79. Hold back my opinions.
others. 80. Seldom get lost in thought.
53. Don’t put my mind on the task at 81. Become overwhelmed by events.
hand. 82. Don’t have a soft side.
54. Rarely get caught up in the 83. Postpone decisions.
excitement. 84. Have a lot of fun.
55. Avoid difficult reading material. 85. Learn things slowly.
56. Rarely lose my composure. 86. Get easily agitated.
57. Rarely put people under pressure. 87. Love a good fight.
58. Want everything to be ‘just right’. 88. See that rules are observed.
59. See myself as a good leader. 89. Am the first to act.
60. Seldom notice the emotional 90. Seldom daydream.
aspects of paintings and pictures. 91. Am afraid of many things.
61. Am easily discouraged. 92. Like to do things for others.
62. Take no time for others. 93. Am easily distracted.
63. Get things done quickly. 94. Laugh a lot.
64. Am not a very enthusiastic person. 95. Formulate ideas clearly.
65. Have a rich vocabulary. 96. Can be stirred up easily.
66. Am a person whose moods go up 97. Am out for my own personal gain.
and down easily. 98. Want every detail taken care of.
67. Insult people. 99. Do not have an assertive
68. Am not bothered by disorder. personality.
69. Can talk others into doing things. 100. See beauty in things that others
70. Need a creative outlet. might not notice.
Reverse response scores for items followed by “R” (i.e., 1 = 5, 2 = 4, 4 = 2, 5 = 1).

To compute scale scores, average completed items within each scale.

To compute Big Five scores, average scores for the two aspects within each domain.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Agreeableness:</td>
<td>Compassion:</td>
<td>2R, 12, 22, 32R, 42, 52R, 62R, 72, 82R, 92</td>
</tr>
<tr>
<td></td>
<td>Openness:</td>
<td>10, 20, 30, 40, 50R, 60R, 70, 80R, 90R, 100</td>
</tr>
</tbody>
</table>
Appendix H  Susceptibility to Experimenter Pressure

1. Please indicate to what extent you agree or disagree with the following statement: “I think research on moral judgment is important”.
   
   [ ] Strongly Disagree  [ ] Disagree  [ ] Unsure/ No Opinion  [ ] Agree  [ ] Strongly Agree

2. Please indicate to what extent you agree or disagree with the following statement: “I want to help scientists understand the nature of moral judgment”.
   
   [ ] Strongly Disagree  [ ] Disagree  [ ] Unsure/ No Opinion  [ ] Agree  [ ] Strongly Agree

3. Please indicate to what extent you agree or disagree with the following statement: “It is important to me that I contribute to psychological knowledge through research participation”.
   
   [ ] Strongly Disagree  [ ] Disagree  [ ] Unsure/ No Opinion  [ ] Agree  [ ] Strongly Agree

4. Please indicate to what extent you agree or disagree with the following statement: “I feel good about furthering psychological science”.
   
   [ ] Strongly Disagree  [ ] Disagree  [ ] Unsure/ No Opinion  [ ] Agree  [ ] Strongly Agree

5. How much did you enjoy taking part in this study?
   
   [ ] I Did Not Enjoy it at All  [ ] It was Somewhat Enjoyable  [ ] It was Fairly Enjoyable  [ ] It was Enjoyable  [ ] I Enjoyed it Very Much

229
6. Would you recommend participating in this study to your friends?

[ ] I Definitely Would Not Recommend this Study to my Friends
[ ] I Might Recommend this Study to my Friends
[ ] I Probably Would Recommend this Study to my Friends
[ ] I Would Recommend this Study to my Friends
[ ] I Definitely Would Recommend this Study to my Friends

7. How important is it to you that you answered the experimenter’s questions to the best of your ability?

[ ] Not at All Important
[ ] Somewhat Important
[ ] Fairly Important
[ ] Important
[ ] Very Important