

1 Searching for Moral Dumbfounding: Identifying Measurable Indicators of Moral
2 Dumbfounding

3 Cillian McHugh¹, Marek McGann¹, Eric R. Igou², & Elaine L. Kinsella²

4 ¹ Mary Immaculate College ~ University of Limerick

5 ² University of Limerick

6 Author Note

7 All procedures performed in studies involving human participants were approved by
8 institutional research ethics committee and conducted in accordance with the Code of
9 Professional Ethics of the Psychological Society of Ireland, and with the 1964 Helsinki
10 declaration and its later amendments or comparable ethical standards. Informed consent was
11 obtained from all individual participants included in the study. The authors declare that
12 there are no potential conflicts of interest with respect to the research, authorship, and/or
13 publication of this article. All authors consented to the submission of this manuscript.

14 Correspondence concerning this article should be addressed to Cillian McHugh, Mary
15 Immaculate College, South Circular road, Limerick, Ireland. E-mail:

16 cillian.mchugh@mic.ul.ie

Abstract

17

18 Moral dumbfounding is defined as maintaining a moral judgement, without supporting
19 reasons. The most cited demonstration of dumbfounding, does not identify a specific
20 measure of dumbfounding, and has not been published in peer-review form, or directly
21 replicated. Despite limited empirical examination, dumbfounding has been widely discussed
22 in moral psychology. The present research examines the reliability with which dumbfounding
23 can be elicited, and aims to identify measureable indicators of dumbfounding. Study 1 aimed
24 at establishing the effect that is reported in the literature. Participants read four scenarios
25 and judged the actions described. An Interviewer challenged participants' stated reasons for
26 judgements. Dumbfounding was evoked, as measured by two indicators admissions of not
27 having reasons (17%); unsupported declarations (9%), with differences between scenarios.
28 Study 2 measured dumbfounding as the selecting of an unsupported declaration as part of a
29 computerised task. We observed high rates of dumbfounding across all scenarios. Studies 3a
30 (college sample) and 3b (MTurk sample) addressing limitations in Study 2, replaced the
31 unsupported declaration with an admission of having no reason, and included open-ended
32 responses which were coded for unsupported declarations. As predicted, lower rates of
33 dumbfounding were observed (3a 20%; 3b 16%; or 3a 32%; 3b 24% including unsupported
34 declarations in open-ended responses). Two measures provided evidence for dumbfounding
35 across three studies; rates varied with task type (interview/computer task), and with
36 measure being employed (admissions of not having reasons/unsupported declarations).
37 Possible cognitive processes underlying dumbfounding, and limitations of methodologies
38 used, are discussed as a means to account for this variability.

39

Keywords: Morality, Dumbfounding, Judgement, Intuitions, Reasoning

40

Word count: 15,766

41 Searching for Moral Dumbfounding: Identifying Measurable Indicators of Moral 42 Dumbfounding

43 Moral dumbfounding occurs when people stubbornly maintain a moral judgement, even
44 though they can provide no reason to support their judgements (Haidt, 2001; Haidt,
45 Björklund, & Murphy, 2000; Prinz, 2005). It typically manifests as a state of confusion or
46 puzzlement coupled with (a) an admission of not having reasons or (b) the use of
47 unsupported declarations (“It’s just wrong!”) as justification for a judgement (Haidt &
48 Hersh, 2001; Haidt et al., 2000), particularly, when people encounter taboo behaviours that
49 do not result in any harm. The classic and most commonly cited example involves an act of
50 consensual incest between a brother and sister with the use of contraceptive (*Incest*).
51 Another example (*Cannibal*) involves an act of cannibalism with a body that is already dead
52 and is due to be incinerated the next day (Haidt et al., 2000).¹

53 Defining and Measuring Moral Dumbfounding

54 Definitions of moral dumbfounding vary within the moral psychology literature. It was
55 originally defined as “the stubborn and puzzled maintenance of a judgment without
56 supporting reasons” (Haidt & Björklund, 2008, p. 197; see also, Haidt & Hersh, 2001, p. 194;
57 Haidt et al., 2000, p. 2). Some authors cite the original definition verbatim (e.g., Jacobson,
58 2012; Royzman, Kim, & Leeman, 2015); others include the maintenance of a moral
59 judgement despite the absence of supporting reason, but omit any reference to stubbornness
60 or puzzlement (e.g., Cushman, Young, & Hauser, 2006; Dwyer, 2009; Gray, Schein, & Ward,
61 2014; Haidt, 2007; Wielenberg, 2014); and some refer to confidence in the judgement, but
62 again, omit any reference to stubbornness or puzzlement (e.g., Cushman, Young, & Greene,
63 2010; Hauser, Cushman, Young, Kang-Xing Jin, & Mikhail, 2007; Hauser, Young, &
64 Cushman, 2008; Pizarro & Bloom, 2003; Sneddon, 2007).

65 It is apparent from the literature that there is no single, agreed definition of moral
66 dumbfounding. That said, an absence of reasons for, or an inability to justify or defend, a

67 moral judgement, is consistently identified across definitions. However, even despite this
68 apparent consistency, there remains considerable variation in the language used to describe
69 this “failure to provide reasons for a moral judgement”. Indeed, the lack of definitional
70 specificity has led to differing interpretations of moral dumbfounding. It also allows for the
71 possibility of disagreement relating to the implications, both theoretical and practical, of
72 moral dumbfounding.

73 According to the original definition, moral dumbfounding is “the stubborn and puzzled
74 maintenance of a judgment without supporting reasons” (Haidt et al., 2000, p. 2). This
75 definition contains four separate elements: (i) stubbornness; (ii) puzzlement; (iii)
76 maintaining of the judgement; and (iv) the absence of supporting reasons. Of these
77 individual elements, stubbornness and puzzlement, arguably, emerge as consequences of the
78 combination of the maintenance of the judgement in the absence of supporting reasons. If a
79 person maintains a judgement in the absence of reasons (and this absence of reasons has
80 been pointed out to them) they will be perceived as stubborn; and, if a person becomes
81 aware that they do not have reasons for their judgement, they may become puzzled.

82 Following this, and in line with the wider literature, the combination of elements (iii)
83 and (iv), the maintenance of the judgement in the absence of supporting reasons are
84 identified as essential elements of dumbfounding. This does not mean that stubbornness and
85 puzzlement should be ignored entirely; accounting for them may be useful in differentiating
86 between a failure to provide reasons and a refusal to provide reasons. However, viewing
87 stubbornness and puzzlement as consequences of the maintenance of a judgement in the
88 absence of supporting reasons, indicates that they are subsequent to, and not a necessary
89 part of, moral dumbfounding.

90 This view of dumbfounding includes the elements of the phenomenon that are
91 mentioned the most frequently within the wider literature. It is also consistent with the way
92 dumbfounding is described in the original study by Haidt et al. (2000). They report
93 interesting variation in a number of non-verbal behaviours that may be linked with

94 stubbornness or puzzlement, but beyond these, they do not offer a specific indication of how
95 stubbornness and puzzlement are operationalised. Furthermore, other than appearing in the
96 introductory definition for dumbfounding, in the abstract, (Haidt et al., 2000, p. 2), the
97 terms “stubborn” and “puzzled” do not appear again for the remainder of the paper,
98 suggesting that they are not core elements of the phenomenon.

99 Haidt et al. (2000) report a range of responses that may illustrate a state of
100 dumbfoundedness (admissions of not having reasons and unsupported declarations), however,
101 they do not provide details of the numbers of participants they classified as dumbfounded, or
102 specific response that may be used to make such a classification. The numbers of
103 participants who provided admissions of not having reasons are reported, however it is
104 unclear whether or not this may be taken as a specific measure of dumbfounding or even if
105 such a measure exists. This vagueness in the initial operationalisation of dumbfounding is
106 reflected in the wider literature, whereby evidence of, or, illustrations of, dumbfounding
107 include unsupported declarations (Haidt, 2001, p. 817; Prinz, 2005, p. 101), and tautological
108 reasons (“because it’s incest”; Mallon & Nichols, 2011, p. 285). The current research aims
109 identify specific measurable responses that may be used as indicators of dumbfounding.

110 Drawing on the work of Haidt et al. (2000), and the wider literature, the absence of
111 supporting reasons appears to present in two distinct ways. Firstly, and non-controversially,
112 participants may become aware that they do not have reasons and acknowledge this
113 (admissions of not having reasons). Secondly, participants may fail to provide reasons.
114 Measuring this failure to provide reasons is more problematic; if a participant does not admit
115 to not having reasons, they attempt to disguise their failure to identify reasons. The use of
116 unsupported declarations or tautological reasons as justifications for a judgement may be
117 identified as a failure to provide reasons. Stating “it’s just wrong” or “because it’s wrong”
118 does not answer the question “do you have a reason for your judgement?” (Mallon & Nichols,
119 2011, p. 285).

(The Short) History of Moral Dumbfounding

120
121 The earliest evidence for moral dumbfounding emerged indirectly as a result of a study
122 by Haidt, Koller, and Dias (1993). This was a cross-cultural study examining the variability
123 of the moral judgements of participants depending on age, socio-economic status, and
124 nationality (USA or Brazil). Participants were presented with a range of moral scenarios,
125 some of which were offensive, but harmless; for example, cutting up a national flag (Brazil or
126 USA, matched to sample) and using it to clean the bathroom; a family eating their dog after
127 it was killed by a car; and, a brother and sister kissing each other on the mouth. When asked
128 to justify their condemnation of certain actions, some participants (from both countries)
129 used unsupported declarations as a reason; for example, “Because it’s wrong to eat your dog”
130 or “Because you’re not supposed to cut up the flag” (Haidt et al., 1993, p. 632). This study
131 was not a direct study of moral dumbfounding, rather it was investigating differences in the
132 way people reason about moral scenarios. The use of unsupported declarations in response
133 to some moral scenarios was noted among a range of responses (Haidt et al., 1993).

134 A later study, by Haidt et al. (2000), directly investigated the phenomenon of moral
135 dumbfounding. In their study two moral scenarios (*Incest* and *Cannibal*: see Appendix A)
136 designed to elicit strong emotional reactions, but with no identifiable harmful consequences
137 (emotional intuition scenarios), were contrasted against a traditional moral judgement
138 scenario (*Heinz*) that involved balancing the interests of two people (reasoning scenario).
139 They observed differences in responses between the two types of scenarios, participants were
140 better at defending their judgement for the reasoning scenario than for the emotional
141 intuition scenarios. It appeared that these emotional intuition scenarios could elicit
142 dumbfounding as evidenced by significant increases in (a) admissions of having no reasons
143 for a judgement, or (b) the use of unsupported declarations (“it’s just wrong”) as a
144 justification for a judgement (Haidt et al., 2000, p. 12). Although interesting, that study
145 (consisting of a final sample of thirty participants) has not been published in peer reviewed
146 form and has not been replicated.²

147 The following year, Haidt and Hersh (2001) investigated differences between
148 conservatives and liberals, across a range of responses to moral issues, and found that
149 conservatives produced more dumbfounded type responses (e.g., stuttering, stating “I don’t
150 know”, admitting they could not explain their answers (Haidt & Hersh, 2001, p. 200), than
151 liberals when discussing particular issues. Although this study did not investigate
152 dumbfounding directly, the findings indicate that there may be individual differences that
153 drive moral judgements which have not yet been fully investigated.

154 The phenomenon of moral dumbfounding has been widely discussed in the moral
155 psychology literature (e.g., Cushman, 2013; Cushman et al., 2010, 2006; Hauser et al., 2007;
156 Prinz, 2005; Royzman et al., 2015), but there is limited available empirical information
157 about the nature of moral dumbfounding and the reliability with which it can be elicited in
158 everyday human behaviour. Some authors have argued that moral dumbfounding does not
159 really exist (Gray et al., 2014; Jacoby, 1983; see also Royzman et al., 2015; Sneddon, 2007;
160 Wielenberg, 2014).³ The studies described in the present paper aim to replicate the initial
161 interview study of Haidt et al. (2000), and to explore practicable methods for testing the
162 phenomenon, and its variability, in larger sample sizes. This will allow for more detailed
163 study of the phenomenon. A deeper understanding of dumbfounding will inform the
164 continuing development of theories of moral judgement, furthering our understanding of the
165 interactions between intuitions and reasoned judgements in the way in which people make
166 moral evaluations.

167 **Moral Dumbfounding and Moral Intuitions**

168 Moral dumbfounding is used as supporting evidence for a range of “intuitionist”
169 theories of moral judgement (e.g. Cushman et al., 2010; Haidt, 2001; Prinz, 2005). According
170 to these intuitionist theories, our moral judgements are grounded in an emotional or intuitive
171 automatic response rather than slow deliberate reasoning (Cameron, Payne, & Doris, 2013;
172 Crockett, 2013; Cushman, 2013; Cushman et al., 2010; Greene, 2008; Haidt, 2001; Prinz,

173 2005). Two of the most influential such theories of moral judgement have been Haidt's social
174 intuitionist model (Haidt, 2001; Haidt & Björklund, 2008) and Greene's dual processes
175 model (Greene, 2008, 2013; Greene, Sommerville, Nystrom, Darley, & Cohen, 2001). Haidt
176 (2001) in his social intuitionist model likens the distinction between fast moral intuitions and
177 slow moral reasoning to the distinction between fast and slow thinking that appears in dual
178 systems theories of cognition (Chaiken, 1980; see also Chaiken & Trope, 1999; Epstein, 1994;
179 Haidt, 2001; Kahneman, 2011; Zajonc, 1980). In introducing and defending this model,
180 Haidt makes specific reference to one of the dumbfounding scenarios, and the findings from
181 the unpublished manuscript relating to this dilemma (Haidt, 2001; see also Haidt &
182 Björklund, 2008; Haidt & Hersh, 2001). Greene draws heavily on Haidt's work in defending
183 his dual-process model of moral judgement (Greene, 2008). In more recent years, Cushman
184 (2013; Cushman et al., 2010) and Crockett (2013), building on the work of Haidt and Greene
185 have continued the development intuitionist/dual-process theories of moral judgement
186 (Crockett, 2013; Cushman, 2013; Greene, 2008; Haidt, 2001).

187 The current research, following from Cushman (2013) and Crockett (2013), takes moral
188 intuitions as "model-free" (Crockett, 2013, p. 364; Cushman, 2013, p. 284) or habitual
189 responses, emerging through a long history of reinforcement learning. According to this
190 approach, consistent with other research on implicit learning (Barsalou, 2003, see also 2008,
191 2009; Berry & Dienes, 1993; Evans, 2003; Reber, 1989; Sun, Slusarz, & Terry, 2005), the
192 learning of a moral norm, leading to the emergence an associated moral intuition, can occur
193 independently of the learning of the reasons for, or explicit rules surrounding the norm.
194 Attributing moral judgements to intuitions in this way also means that moral reasoning does
195 not necessarily cause moral judgements, rather, at least in some circumstances, reasoning is
196 likely to occur post-hoc.

197 However, the claim that reasons for intuitions are learned independently of the
198 intuition does not necessarily imply that there are no reasons for a given intuition. This
199 leads to two difficulties in demonstrating this separation between intuitions and reasons for

200 the intuition have been identified. Firstly, in many circumstances, it is possible to trace the
201 emergence of a given social or moral norm to particular reasons. Pizarro and Bloom (2003)
202 defend the claim that moral intuitions may be rational, and informed by prior reasoning or
203 deliberation. A related, more general claim is that deliberative (model-based) responses can,
204 over time, become automatic or habitual (e.g., Barsalou, 2003; Cushman, 2013; H. L.
205 Dreyfus & Dreyfus, 1990). Secondly, in many cases, after an intuitive judgement is made,
206 reasons that are consistent with the judgement may be through post-hoc rationalisation (e.g.,
207 Cushman et al., 2006). This means that, although there is a clear theoretical case for a
208 separation between intuitions and reasons for these intuitions, demonstrating this separation
209 is problematic.

210 Moral dumbfounding, however, is a phenomenon that may demonstrate this separation
211 between an intuition and reasons for the intuition. In certain cases, people maintain an
212 intuition even though they cannot provide reasons for the intuitions. It is this standing, as a
213 rare demonstration of a crucial theoretical point, that makes moral dumbfounding so
214 interesting. Moral dumbfounding therefore, provides evidence in support of the claim that
215 moral intuitions are habitual and “model-free” (Crockett, 2013, p. 364; Cushman, 2013, p.
216 284). Demonstrating this separation between intuitions and reasons for the intuitions also
217 demonstrates a separation between intuitions and the reasoning process, providing evidence
218 for the suggestion that moral judgements are not necessarily dependent upon moral reasoning
219 and by extension, providing implicit evidence that moral reasoning occurs post-hoc.

220 The existence of moral dumbfounding, therefore, is compelling evidence for intuitionist
221 theories of moral judgement. These theories are supported by a large body of other empirical
222 evidence, however, they are also either directly (e.g., Cushman et al., 2010; Haidt, 2001;
223 Hauser et al., 2008; Prinz, 2005) or indirectly (e.g., Crockett, 2013; Cushman, 2013; Greene,
224 2008, 2013) grounded in the assumption that moral dumbfounding is a real phenomenon.
225 The present research aims, to test the validity of the claim that moral dumbfounding is a
226 real phenomenon through an attempted replication of the widely-cited unpublished study by

227 Haidt et al. (2000). This will also test the strength of existing moral theories grounded in its
228 existence. In addition to this, we aim to identify specific, measurable indicators of
229 dumbfounding and develop practicable methods for eliciting and measuring dumbfounding in
230 larger samples. These may be used to explore the phenomenon in greater depth, informing
231 the further development of moral theory.

232 **Challenges to Moral Dumbfounding**

233 In recent years moral dumbfounding has been challenged by a number of authors (e.g.,
234 Gray et al., 2014; Jacobson, 2012; Sneddon, 2007; Wielenberg, 2014), arguing, in line with
235 rationalist theories of moral judgement (Kohlberg, 1971; Narvaez, 2005; Topolski, Weaver,
236 Martin, & McCoy, 2013), that moral judgements are grounded in reasons. Recent work by
237 Royzman, Kim, and Leeman (2015), involving a series of studies focusing on the Incest
238 dilemma, identified two reasons that may be guiding participants' judgements. The reasons
239 identified were: (a) potential harm – where participants believed that harm could arise as a
240 result of the actions of the characters in the scenario despite the vignette stating that no
241 harm arose; and (b) normativity – where citing a moral norm is seen as sufficient
242 justification for making a judgement consistent with that norm. They found, that, when
243 participants who endorsed either of these reasons were excluded from analysis, there were
244 only four participants (from a sample of fifty-three) who rated the behaviour as wrong
245 without offering a reason. Following a subsequent interview, two of these participants
246 changed their judgement, and one changed her response to the question relating to
247 normative reasons. This left just one participant who maintained that the behaviour was
248 wrong without valid reason and, in their view, could be truly identified as dumbfounded.
249 Consequently, they argue that dumbfounding is not as prevalent a phenomenon as portrayed
250 by Haidt et al. (2000; Royzman et al., 2015, p. 310). In identifying reasons that appear to
251 be guiding people's judgements, they claim to have found evidence for rationalist theories of
252 moral judgement (Royzman et al., 2015, p. 311) over intuitionist theories. They argue that

253 the dumbfounded behaviours observed by Haidt et al. (2000) can be attributed to social
254 pressure that exists in an interview setting, whereby participants accept the
255 counter-arguments offered by the interviewer, even if they disagree, in order to appear
256 cooperative (Royzman et al., 2015, p. 299).

257 Royzman et al. (2015) successfully identified reasons (harm-based reasons; normative
258 reasons) that may underlie moral judgements in the case of the Incest dilemma, showing
259 that, in the vast majority of cases, participants who rate the behaviour as wrong also
260 endorse these reasons if given the opportunity. It is not surprising that instances of moral
261 dumbfounding – defined as the maintaining a moral judgement without providing supporting
262 reasons – can be dramatically reduced by providing participants with reasons for them to
263 endorse (particularly in view of the extensive literature on confabulation, e.g., Evans &
264 Wason, 1976; Gazzaniga & LeDoux, 2013; Johansson, Hall, Sikström, & Olsson, 2005;
265 Nisbett & Wilson, 1977; Wilson & Bar-Anan, 2008). If a participant endorses a reason that
266 is consistent with their judgement this does necessarily not mean that this reason
267 contributed to the making of the judgement. Whether or not participants are able to
268 articulate or volunteer these reasons, without external prompts, has not been the subject of
269 careful empirical investigation. The degree to which people falsely attribute every-day
270 judgements to reasons, that are more accurately described as post-hoc rationalisations, is
271 well documented (Greene, 2008; Johansson et al., 2005; Nisbett & Wilson, 1977).

272 The inability of people to articulate principles that are consistent with, and therefore
273 may arguably be guiding moral judgements has been documented in a study by Cushman et
274 al. (2006). They identified three distinct principles that appear to guide moral judgements;
275 these are: (a) harm caused by action is worse than harm caused by omission; (b) harm
276 intended is worse than harm foreseen; (c) harm involving physical contact is worse than
277 harm without physical contact. They conducted a series of studies in which participants'
278 judgements were largely consistent with these principles. Interestingly, however, when
279 questioned afterwards, participants were only reliably able to articulate two of these

280 principles (a) and (c). Principle (b), while consistent with the judgements made, was not
281 well articulated by participants. It appears that, making judgements consistent with a
282 principle does not imply that participants can articulate this principle. It is this inability to
283 articulate principles or reasons for a moral judgement that is the hallmark of moral
284 dumbfounding and is of key interest in the current research.

285

The Current Research

286 In response to the limited number of demonstrations of, and related uncertainty
287 surrounding moral dumbfounding, the primary aims of the current research are to (a) to
288 identify specific measurable indicators of moral dumbfounding; and (b) use these measures to
289 examine the reliability with which dumbfounded responding can be evoked. We conducted
290 four studies, each of which is a modified replication attempt of the original moral
291 dumbfounding study (Haidt et al., 2000). In these studies, dumbfounding is measured
292 according to two sets of responses: (a) an admission of having no reasons for a judgement (a
293 measure of self-reported dumbfounding) and, (b) use of unsupported declarations ("it's just
294 wrong") or tautological reasons ("because it's incest") as a justification for a judgement
295 (measures of a failure to provide reasons). Study 1 was designed to replicate Haidt et al.'s
296 (2000) initial study using the original methods (face to face interview). In Study 2 we piloted
297 alternative methods (a computer-based task) in an attempt to evoke moral dumbfounding in
298 a systematic way with a larger sample. In Study 3a and 3b the materials that were piloted in
299 Study 2 were refined and administered to a larger sample in an attempt to systematically
300 evoke dumbfounded responding.

301

Study 1: Interview

302 The primary aim of Study 1 was to replicate the original dumbfounding study (Haidt
303 et al., 2000). Four moral judgement vignettes were used (Appendix A). Three of these
304 vignettes (*Heinz*, *Incest*, and *Cannibal*) were taken from Haidt et al. (2000). A fourth
305 vignette (*Trolley*) was adapted Greene et al. (2001). Haidt et al. (2000) contrasted *Heinz*, a

306 so-called reasoning scenario, against *Cannibal* and *Incest*, so-called intuition scenarios. Their
307 study also included two tasks that did not have any moral content. For the purposes of
308 consistency and balance, the non-moral tasks were omitted from the present study, and a
309 second moral reasoning vignette was included in their stead, such that two reasoning
310 vignettes (*Heinz* and *Trolley*) were contrasted against two intuition vignettes (*Incest* and
311 *Cannibal*). We hypothesised that dumbfounding would be elicited and that rates of
312 dumbfounded responding would vary depending on the content of the dilemma, with the
313 intuition scenarios eliciting more dumbfounded responses than the reasoning scenarios. Two
314 measures of dumbfounding were taken reflecting the two distinct ways in which absence of
315 reasons may present: admissions of not having reasons (self-reported dumbfounding), and the
316 use of an unsupported declaration (it's just wrong) as a justification for a judgement, with a
317 failure to provide any alternative reason when the unsupported declaration was questioned (a
318 failure to provide reasons). As in the original study (Haidt et al., 2000), various non-verbal
319 measures were also recorded in an attempt to account for stubbornness and puzzlement.

320 Method

321 **Participants and design.** Study 1 was a frequency based attempted replication.
322 The aim was to identify if dumbfounded responding could be evoked. All participants were
323 presented with the same four moral vignettes. Results are primarily descriptive. Any further
324 analysis tested for differences in responding depending on the vignette, or type of vignette,
325 presented.

326 A sample of 31 participants (15 female, 16 male) with a mean age of $M_{\text{age}} = 28.83$
327 (min = 19, max = 64, $SD = 10.99$) took part in this study. Participants were undergraduate
328 students, postgraduate students, and alumni from Mary Immaculate College (MIC), and
329 University of Limerick (UL). Participation was voluntary and participants were not
330 reimbursed for their participation.

331 **Procedure and materials.** Four moral judgement vignettes were used (Appendix
332 A). Three of the vignettes (*Heinz*, *Incest*, and *Cannibal*) were taken from Haidt et al. (2000).
333 *Incest* was taken directly from the original study however *Cannibal* and *Heinz* were modified
334 slightly, following piloting.

335 The original version of *Cannibal* stated that people had “donated their body to science
336 for research”; participants during piloting were able to argue that eating does not constitute
337 “research”. In order to remove this as a possible argument, the modified version stated that
338 bodies had been donated for “the general use of the researchers in the lab” and that the
339 “bodies are normally cremated, however, severed cuts may be disposed of at the discretion of
340 lab researchers”.

341 Similarly, piloting suggested that participants agreed with the actions of Heinz and
342 condemned the actions of the druggist. The original wording of *Heinz* suggested that any
343 discussion related to Heinz as opposed to the druggist meaning that, for *Heinz*, participants
344 would typically be defending an approval of the character’s actions. However, for *Incest* and
345 *Cannibal* participants generally condemn the actions of the character and as such are
346 defending a judgement of “morally wrong”. In order to ensure that participants were
347 consistently defending a judgement of “morally wrong” across all scenarios, *Heinz* was
348 modified to include “The druggist had Heinz arrested and charged”. Any discussion on *Heinz*
349 then related to the character whose behaviour participants thought was wrong.

350 In the original study by Haidt et al. (2000), *Incest* and *Cannibal* are presented as
351 “intuition” stories, and contrasted against a single “reasoning” dilemma: *Heinz*. In order for
352 a more balanced comparison, a bridge variant of the classic trolley dilemma (*Trolley*) was
353 included as a second “reasoning” dilemma. In this vignette, participants judge the actions of
354 Paul, who pushes a large man off a bridge to stop a trolley and save five lives. The inclusion
355 of *Trolley* meant that there were two “reasoning” dilemmas to be contrasted with the two
356 “intuition” stories.

357 Sample counter arguments were prepared for each scenario. To ensure that participants

358 were only pushed to defend a judgement of “morally wrong” these counter arguments
359 exclusively defended the potentially questionable behaviour of the characters. A list of
360 prepared counter arguments can be seen in Appendix B. A post-discussion questionnaire,
361 taken from Haidt et al. (2000) was administered after discussion of each scenario (Appendix
362 C).

363 Two other measures were also taken for exploratory purposes.: Firstly, in response to a
364 possible link between meaning and morality (e.g., Bellin, 2012; Schnell, 2011), the Meaning
365 in Life questionnaire (MLQ; Steger, Kashdan, Sullivan, & Lorentz, 2008) was included. This
366 ten item scale, is made up of two five item sub scales: presence (e.g., “I understand my life’s
367 meaning”) and search (e.g., “I am looking for something that makes my life feel
368 meaningful”). Responses were recorded using a seven point Likert scale ranging from 1
369 (*strongly disagree*) to 7 (*strongly agree*). Secondly, in line with Haidt’s (2007; see also, Haidt
370 & Hersh, 2001) work, describing a link between religious conservatism and moral views, it
371 was hypothesised that incidences of dumbfounding may be moderated by individual
372 differences in religiosity . As such, the seven item CRSi7 scale, taken from The Centrality of
373 Religiosity Scale (S. Huber & Huber, 2012) was also included. Participants responded to
374 questions relating to the frequency with which they engage in religious or spiritual activity
375 (e.g., “How often do you think about religious issues?”). Responses were recorded using a
376 five point Likert scale ranging from 1 (*never*) to 5 (*very often*).

377 The interviews took place in a designated psychology lab in MIC and were recorded on
378 a digital video recording device. Participants were presented with an information sheet and a
379 consent form. The consent form required two signatures: firstly, participants consented to
380 take part in the study (including consent to be video recorded); the second signature related
381 to use of the video for any presentation of the research (with voice distorted and face
382 pixelated). Only two participants opted not to sign the second part.

383 Participants read brief vignettes describing each scenario, and were subsequently
384 interviewed regarding the protagonists. All four scenarios were discussed in a single interview

385 session, with a brief pause between each discussion for the participant to complete a
386 questionnaire about their judgements, and to read the next scenario. The conversation
387 continued when they were happy to do so. Each of the four moral dilemmas *Heinz*, *Trolley*,
388 *Cannibal* and *Incest* (Appendix A) were presented in this way and participants asked to
389 judge the behaviour of the characters in the dilemmas. The order of presenting the scenarios
390 was randomised. Judgements made by participants were challenged by the experimenter
391 (“Nobody was harmed, how can there be anything wrong?”; “Do you still think it was
392 wrong? Why?”; “Why do you think it is wrong?”; “Have you got a reason for your
393 judgement?”). The resulting discussion continued until participants could not articulate any
394 further arguments. Participants filled in a brief questionnaire after discussing each dilemma.
395 In this they were asked to rate, on a seven point Likert scale, how right/wrong they thought
396 the behaviour was; how confident they were in their judgement, how confused they were; how
397 irritated they were; how much their judgement had changed; how much their judgement was
398 based on reason; and how much their judgement was based on “gut” feeling. Participants
399 completed a longer questionnaire at the end of the interview. This contained the MLQ
400 (Steger et al., 2008), the Centrality of Religiosity Scale (S. Huber & Huber, 2012), and some
401 questions relating to demographics. The entire study lasted approximately 20 to 25 minutes.
402 The videos were analysed using BORIS – Behavioural Observation Research Interactive
403 Software (Friard & Gamba, 2015). All statistical analysis was conducted using R (3.4.0, R
404 Core Team, 2017b)⁴; SPSS (IBM Corp, 2015) was also used.

405 **Results and Discussion**

406 The videos of the interviews were analysed and participants were identified as
407 dumbfounded if they (a) admitted to not having reasons for their judgements; or (b) resorted
408 to using unsupported declarations (“It’s just wrong!”) as justification for their judgements,
409 and subsequently failed to provide reasons when questioned further. Table 1 shows the initial
410 and revised ratings of the behaviours for each scenario.

Table 1

Ratings of each scenario for each study

Study	Judgement	Heinz		Cannibal		Incest		Trolley	
		N	percent	N	percent	N	percent	N	percent
Study 1	Initial: Wrong	27	12.9%	25	80.65%	5	83.87%	23	74.19%
	Initial: Neutral	0	0%	0	0%	0	0%	0	0%
	Initial: OK	4	87.1%	6	19.35%	26	16.13%	8	25.81%
	Revised: Wrong	26	83.87%	23	74.19%	20	64.52%	22	70.97%
	Revised: Neutral	0	0%	0	0%	0	0%	1	3.23%
	Revised: OK	5	16.13%	8	25.81%	11	35.48%	8	25.81%
Study 2	Initial: Wrong	53	13.89%	68	94.44%	6	87.5%	50	69.44%
	Initial: Neutral	9	12.5%	3	4.17%	3	4.17%	6	8.33%
	Initial: OK	10	73.61%	1	1.39%	63	8.33%	16	22.22%
	Revised: Wrong	51	70.83%	67	93.06%	66	91.67%	48	66.67%
	Revised: Neutral	7	9.72%	3	4.17%	3	4.17%	9	12.5%
	Revised: OK	14	19.44%	2	2.78%	3	4.17%	15	20.83%
Study 3a	Initial: Wrong	54	16.67%	67	93.06%	4	84.72%	48	66.67%
	Initial: Neutral	6	8.33%	3	4.17%	7	9.72%	10	13.89%
	Initial: OK	12	75%	2	2.78%	61	5.56%	14	19.44%
	Revised: Wrong	53	73.61%	67	93.06%	57	79.17%	43	59.72%
	Revised: Neutral	11	15.28%	4	5.56%	12	16.67%	15	20.83%
	Revised: OK	8	11.11%	1	1.39%	3	4.17%	14	19.44%
Study 3b	Initial: Wrong	81	10.89%	85	84.16%	10	70.3%	66	65.35%
	Initial: Neutral	9	8.91%	13	12.87%	20	19.8%	14	13.86%
	Initial: OK	11	80.2%	3	2.97%	71	9.9%	21	20.79%
	Revised: Wrong	87	86.14%	82	81.19%	73	72.28%	59	58.42%
	Revised: Neutral	10	9.9%	15	14.85%	19	18.81%	17	16.83%
	Revised: OK	4	3.96%	4	3.96%	9	8.91%	25	24.75%

411 Twenty two of the 31 participants (70.97%) produced a dumbfounded response
 412 (admission of having no reasons; or the use of an unsupported declaration as a justification
 413 for a judgement, with a failure to provide any alternative reason when the unsupported
 414 declaration was questioned) at least once. Examples of such responses included “It just
 415 seems wrong and I cannot explain why, I don’t know”, “because I just think it’s wrong, oh
 416 God, I don’t know why, it’s just [pause] wrong”. Table 2 shows the number, and percentage,
 417 of participants who displayed dumbfounded responses and non-dumbfounded responses for
 418 each dilemma. The rates of each type of dumbfounded response are also displayed. Figure 1
 419 shows the percentage of participants displaying dumbfounded responses for each dilemma.
 420 Table 3 shows the responses to the questionnaires presented between dilemmas.

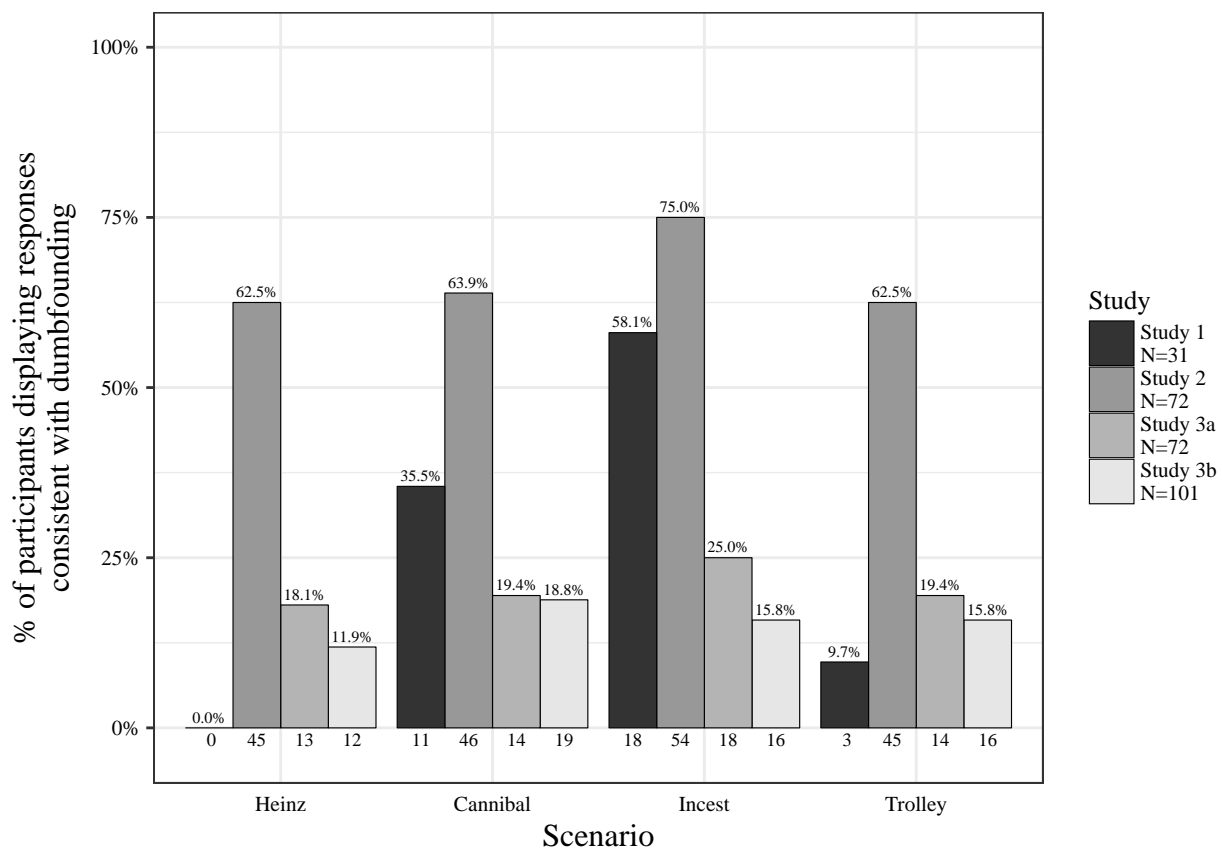


Figure 1. Rates of observed dumbfounding for each scenario across each study.

421 In line with the original study (Haidt et al., 2000), the videos were also coded, by the
 422 primary researcher, across a range of measures. Haidt et al. (2000) report differences,

Table 2

Observed frequency and percentage of each of the responses: dumbfounded, nothing wrong, and reasons provided

		Heinz		Cannibal		Incest		Trolley	
		N	percent	N	percent	N	percent	N	percent
Study 1	Nothing wrong	6	19.35%	8	25.81%	11	35.48%	8	25.81%
	Dumbfounded	0	0%	11	35.48%	18	58.06%	3	9.68%
	(admissions)	0	0%	8	25.81%	10	32.26%	3	9.68%
	(declarations)	0	0%	3	9.68%	8	25.81%	0	0%
	Reasons	25	80.65%	12	38.71%	2	6.45%	20	64.52%
Study 2	Nothing wrong	8	11.11%	4	5.56%	2	2.78%	10	13.89%
	Dumbfounded	45	62.5%	46	63.89%	54	75%	45	62.5%
	Reasons	19	26.39%	22	30.56%	16	22.22%	17	23.61%
Study 3a (critical slide)	Nothing wrong	14	19.44%	4	5.56%	12	16.67%	15	20.83%
	Dumbfounded	13	18.06%	14	19.44%	18	25%	14	19.44%
	Reasons	45	62.5%	54	75%	42	58.33%	43	59.72%
Study 3a (coded)	Nothing wrong	14	19.44%	4	5.56%	12	16.67%	15	20.83%
	Dumbfounded	19	26.39%	21	29.17%	31	43.06%	22	30.56%
	Reasons	39	54.17%	47	65.28%	29	40.28%	35	48.61%
Study 3b (critical slide)	Nothing wrong	21	20.79%	10	9.9%	31	30.69%	24	23.76%
	Dumbfounded	12	11.88%	19	18.81%	16	15.84%	16	15.84%
	Reasons	68	67.33%	72	71.29%	54	53.47%	61	60.4%
Study 3b (coded)	Nothing wrong	21	20.79%	10	9.9%	31	30.69%	24	23.76%
	Dumbfounded	16	15.84%	30	29.7%	28	27.72%	22	21.78%
	Reasons	64	63.37%	61	60.4%	42	41.58%	55	54.46%

Table 3

Responses to post-discussion questionnaire questions

Study	Question	Heinz	Cannibal	Incest	Trolley
Study 1	Changed mind	2.87	3.40	2.63	2.60
	Confidence	5.30	4.77	5.40	5.07
	Confused	3.00	3.67	3.33	3.70
	Irritated	3.00	3.33	3.13	3.37
	'Gut'	5.23	5.20	4.97	5.07
	'Reason'	4.83	4.40	4.43	4.77
	Gut minus Reason	0.40	0.80	0.53	0.30
Study 2	Confidence	6.10	5.86	5.62	5.26
	Confused	2.40	3.08	4.14	3.17
	Irritated	4.58	4.68	4.32	4.28
	'Gut'	5.29	5.54	5.82	4.96
	'Reason'	4.89	5.19	4.89	4.93
	Gut minus Reason	0.40	0.35	0.93	0.03
Study 3a	Changed mind	2.38	1.67	2.00	2.00
	Confidence	5.22	5.50	5.38	4.81
	Confused	2.75	2.96	3.25	2.89
	Irritated	3.94	4.64	4.07	3.60
	'Gut'	4.78	5.44	5.44	4.92
	'Reason'	5.07	5.26	5.11	5.06
	Gut minus Reason	-0.29	0.18	0.33	-0.14
Study 3b	Changed mind	1.74	1.60	1.57	1.83
	Confidence	5.78	6.16	5.81	5.36
	Confused	2.06	2.07	2.12	2.22
	Irritated	4.42	4.01	3.56	3.39
	'Gut'	4.42	4.43	4.47	4.01
	'Reason'	5.46	5.69	5.26	5.58
	Gut minus Reason	-1.04	-1.27	-0.79	-1.57

423 between intuition and reasoning scenarios. They do not, however, report comparisons
424 between participants identified as dumbfounded and participants not identified as
425 dumbfounded. The current research, aiming to identify measurable indicators of
426 dumbfounding, categorised participants as dumbfounded according to the two types of verbal
427 responses (admissions and unsupported declaration) and compared these groups with
428 participants who were not identified as dumbfounded, across a range of measures. There
429 were two stages in this analysis. Firstly, all participants identified as dumbfounded were
430 compared against participants who provided reasons only. Secondly, participants identified
431 as dumbfounded were grouped according to type of dumbfounded response, and participants
432 who did not rate the behaviour as wrong were also included in the analysis.

433 Judgement variables reported by Haidt et al. (2000) included the length of time until
434 the first argument, the length of time until the first evaluation, the length of time between
435 the first evaluation and the first argument. The current research reports the same judgement
436 variables.

437 A range of “argument variables” were also reported. Identifying specific objectively
438 verifiable measurable indicators for some of the “argument variables” reported by Haidt et al.
439 (2000) was problematic (e.g., “dead-ends”, “argument kept”, “argument dropped”). The
440 current research coded each verbal utterance according to a relevance for forming an
441 argument. As such some of the argument variables reported by Haidt et al. (2000) are not
442 reported here in the same way, however, related measures are reported.

443 Paralinguistic variables reported by Haidt et al. (2000) include frequency (per minute)
444 of: “ums, uhs, hmms”, “turns with laughter”, “turns with face touch”, “doubt faces”, and
445 “turns with pen fiddle”. As with the argument variables, the coding of the
446 non-verbal/paralinguistic responses also varies slightly from what was reported by Haidt et
447 al. (2000). We coded for both verbal hesitations (“um/em/uh”) and non-verbal
448 hesitations/stuttering. “Turns” was coded independently of other behaviours as changing
449 position. Laughter was coded for independently of changing position. The coding of hands

450 touching the self was not limited to the face. Participants did not have pens to fiddle with,
451 however we coded for generic fidgeting. The term “doubt faces” presented as problematic to
452 code for rigorously across different individuals. As such, two distinctive and opposing facial
453 expressions were coded for: smiling and frowning.

454 **Dumbfounded versus reasons.** Fifty nine cases of participants providing reasons,
455 were compared with 32 cases of dumbfounded responding. There was no difference in time
456 until first judgement between the dumbfounded group, ($M = 14.89$, $SD = 20.41$) and the
457 group who provided reasons ($M = 15.19$, $SD = 40.54$), $p = .969$. Similarly, there was no
458 difference in time until first argument between the dumbfounded group, ($M = 39.20$, $SD =$
459 28.90) and the group who provided reasons ($M = 30.49$, $SD = 32.30$), $F(1, 81) = 1.42$, $p =$
460 $.237$, partial $\eta^2 = .017$. There was no difference in time from first judgement to time of first
461 argument between the dumbfounded group, ($M = 20.60$, $SD = 36.76$) and the group who
462 provided reasons ($M = 15.65$, $SD = 46.42$), $p = .634$.

463 There was a significant difference in frequency (per minute) of utterances whereby
464 participants were working towards a reason between the dumbfounded group, ($M = 1.47$, SD
465 $= 1.45$) and the group who provided reasons ($M = 2.70$, $SD = 1.53$), $F(1, 89) = 13.82$, $p <$
466 $.001$, partial $\eta^2 = .134$. There was no difference in frequency (per minute) of irrelevant
467 arguments between the dumbfounded group, ($M = 1.03$, $SD = .74$) and the group who
468 provided reasons ($M = .86$, $SD = .77$), $F(1, 89) = 1.05$, $p = .308$, partial $\eta^2 = .012$. There
469 was a significant difference in frequency (per minute) of expressions of doubt between the
470 dumbfounded group, ($M = .63$, $SD = .65$) and the group who provided reasons ($M = .31$,
471 $SD = .58$), $F(1, 89) = 5.87$, $p = .017$, partial $\eta^2 = .062$.

472 A one-way ANOVA revealed a significant difference in number of times per minute
473 participants laughed between the dumbfounded group, ($M = 2.81$, $SD = 2.84$) and the group
474 who provided reasons ($M = 1.18$, $SD = 1.25$), $F(1, 89) = 14.35$, $p < .001$, partial $\eta^2 = .139$.
475 Similarly, a one-way ANOVA revealed a significant difference relative amount of time spent
476 smiling (as a proportion of the total time spent on the given scenario) between the

477 dumbfounded group, ($M = .32$, $SD = .15$) and the group who provided reasons ($M = .16$,
478 $SD = .14$), $F(1, 89) = 25.24$, $p < .001$, partial $\eta^2 = .221$. Consistent with the results
479 reported by Haidt et al. (2000), a series of one-way ANOVAs revealed no differences in verbal
480 hesitations, $F(1, 89) = 2.35$, $p = .129$, partial $\eta^2 = .026$, non-verbal hesitations, $p = .074$,
481 changing posture, $p = .485$, hands on the self, $p = .864$, frowning, $p = .958$, and fidgeting,
482 $F(1, 89) = 1.66$, $p = .201$, partial $\eta^2 = .018$. A one-way ANOVA revealed a significant
483 difference relative amount of time spent in silence (as a proportion of the total time spent on
484 the given scenario) between the dumbfounded group, ($M = .14$, $SD = .08$) and the group
485 who provided reasons ($M = .09$, $SD = .06$), $F(1, 89) = 9.72$, $p = .002$, partial $\eta^2 = .098$.

486 From the above analysis, it appears that, working towards reasons, expressions of
487 doubt, laughter, smiling, and silence were the only measures that varied significantly
488 depending on whether a person was identified as dumbfounded or provided reasons. Having
489 identified differences between dumbfounded participants and participants providing reasons,
490 the following analysis investigates if there are differences depending the type of dumbfounded
491 response provided. participants who did not rate the behaviour as wrong are also included in
492 the following analysis.

493 **Variation between different types of dumbfounded responses.** Four groups,
494 based on overall reaction to scenarios, were identified: participants who did not rate the
495 behaviour as wrong, participants who provided reasons, participants who provided
496 unsupported declarations, and participants who admitted to not having reasons.

497 A one-way ANOVA revealed a significant difference in relative frequency of utterances
498 whereby participants were working towards a reason depending on overall reaction to
499 scenarios, $F(3, 120) = 7.54$, $p < .001$, partial $\eta^2 = .159$. Tukey's post-hoc pairwise
500 comparison revealed that participants who provided reasons were identified as working
501 towards a reason significantly more frequently ($M = 2.70$, $SD = 1.53$) than participants who
502 did not rate the behaviour as wrong ($M = 1.76$, $SD = 1.48$), $p = .021$, and more frequently
503 than participants who provided unsupported declarations as justifications ($M = .64$, $SD =$

504 .72), $p < .001$. There was no difference between participants who admitted to not having
505 reasons ($M = 1.90$, $SD = 1.56$) and any of the other groups. A one-way ANOVA revealed no
506 significant difference in relative frequency of expressions of doubt depending on overall
507 reaction to scenarios, $F(3, 120) = 2.17$, $p = .096$, partial $\eta^2 = .051$.

508 A one-way ANOVA revealed a significant difference in relative frequency laughter
509 depending on overall reaction to scenarios, $F(3, 120) = 8.27$, $p < .001$, partial $\eta^2 = .171$.
510 Tukey's post-hoc pairwise comparison revealed that participants who admitted to not having
511 reasons laughed significantly more frequently ($M = 2.41$, $SD = 2.00$), than participants who
512 provided reasons ($M = 1.18$, $SD = 1.25$), $p = .039$, and more frequently than participants
513 who provided did not rate the behaviour as wrong ($M = .97$, $SD = 1.29$), $p = .025$.
514 Similarly, participants who provided unsupported declarations laughed significantly more
515 frequently ($M = 3.57$, $SD = 4.00$), than participants who provided reasons, $p < .001$, and
516 more frequently than participants who did not rate the behaviour as wrong, $p < .001$. There
517 was no difference between participants who provided reasons, and participants who did not
518 rate the behaviour as wrong $p = .951$. Interestingly, there was no difference between
519 participants who admitted to not having reasons and participants who provided unsupported
520 declarations, $p = .305$.

521 A similar pattern of results was found for time spent smiling. A one-way ANOVA
522 revealed a significant difference in relative time spent smiling depending on overall reaction
523 to scenarios, $F(3, 120) = 9.97$, $p < .001$, partial $\eta^2 = .200$. Tukey's post-hoc pairwise
524 comparison revealed that participants who admitted to not having reasons spent significantly
525 more time smiling ($M = .33$, $SD = .14$), than participants who provided reasons ($M = .16$,
526 $SD = .14$), $p < .001$, and more time smiling than participants who provided did not rate the
527 behaviour as wrong ($M = .16$, $SD = .13$), $p < .001$. Participants who provided unsupported
528 declarations spent significantly more time smiling ($M = .31$, $SD = .17$), than participants
529 who provided reasons, $p = .008$, and participants who did not rate the behaviour as wrong, p
530 $= .014$. There was no difference between participants who provided reasons, and participants

531 who did not rate the behaviour as wrong, $p = 1.000$. Again, there was no difference between
532 participants who admitted to not having reasons and participants who provided unsupported
533 declarations, $p = .996$.

534 A one-way ANOVA revealed a significant difference in relative amount of time spent in
535 silence depending on overall reaction to scenarios, $F(3, 120) = 3.31$, $p = .023$, partial $\eta^2 =$
536 $.076$. Mean proportion of interview time spent in silence are as follows: participants
537 providing reasons, $M = .09$, $SD = .06$; participants not rating the behaviour as wrong, $M =$
538 $.12$, $SD = .07$; participants admitting to not having reasons, $M = .14$, $SD = .09$; and
539 participants providing unsupported declarations, $M = .14$, $SD = .05$. Tukey's post-hoc
540 pairwise comparison did not reveal any significant differences between specific groups.

541 **Further analyses.** An exploratory analysis revealed no association between number
542 of times dumbfounded and score on either measures from the MLQ: Presence, $r(31) = 0.74$,
543 $p = .466$, or Search, $r(31) = 1.38$, $p = .179$, or the Centrality of Religiosity Scale $r(31) =$
544 0.35 , $p = .726$. There was no difference in observed rates of dumbfounded responses
545 depending on the order of scenario presentation, $\chi^2(6, N = 124) = 4.01$, $p = .676$. Rates of
546 dumbfounded responses varied depending on which moral dilemma was being discussed,
547 $\chi^2(6, N = 124) = 46.82$, $p < .001$. The highest rate of dumbfounding was recorded for
548 *Incest*, with 18 of the 31 (58.06%) participants displaying dumbfounded responses. Eleven
549 participants (35.48%) displayed dumbfounded responses for *Cannibal* and three participants
550 (9.68%) displayed dumbfounded responses for *Trolley*. The lowest recorded rate of
551 dumbfounded response was for the Heinz dilemma, with no participants resorting to
552 unsupported declarations as justification or admitting to not having reasons for their
553 judgement. This trend is generally consistent with that which emerged in the original study
554 (with the exception of *Trolley*, which was not used in the original study). Furthermore, rates
555 of dumbfounded responding varied depending on which type of moral scenario was being
556 discussed. *Heinz* and *Trolley*, identified as reasoning scenarios, were contrasted against the
557 intuition scenarios *Incest* and *Cannibal*. There was significantly more dumbfounded

558 responding for the intuition scenarios (29 instances) than for the reasoning scenarios (3
559 instances), $\chi^2(2, N = 124) = 38.17, p < .001$.

560 The aim of Study 1 was to examine the replicability of moral dumbfounding as
561 identified by Haidt et al. (2000), and identify specific measurable responses that may be
562 indicative of dumbfounding. The overall pattern of responses, and pattern of inter-scenario
563 variability in responding resembled that observed in the original study. As such, Study 1
564 successfully replicated the findings of the original moral dumbfounding study (Haidt et al.,
565 2000). Participants were identified as dumbfounded according to two specific measures,
566 admissions of having no reasons, and unsupported declarations followed by a failure to
567 provide reasons when questioned further. Both of these responses were accompanied by
568 similar increases incidences of laughter, and time spent smiling, when compared to
569 participants providing reasons, and participants not rating the behaviour as wrong. When
570 taken together, these responses were also accompanied by more silence during the interview,
571 when compared with participants who provided reasons. As such, it appears that identifying
572 incidences of dumbfounding according to unsupported declarations or admissions of not
573 having reasons largely capture dumbfounding as described by Haidt et al. (2000).

574 Study 1 provides evidence supporting the view that moral dumbfounding is a genuine
575 phenomenon and can be elicited in an interview setting when participants are pressed to
576 justify their judgements of particular moral scenarios. Two key limitations have been
577 identified as a result of conducting studies in an interview setting. Firstly, conducting
578 video-recorded interviews, and the accompanying analyses, is particularly labour intensive,
579 which leads to a smaller sample size. The aims of the present research were to examine the
580 replicability of dumbfounding, and to identify specific measurable indicators of
581 dumbfounding. A sample size of thirty-one is not sufficient in fulfilling the first aim.
582 Secondly, an interview setting introduces a social context that may influence the responses of
583 participants, in that, participants may feel a social pressure to behave in a particular way
584 (e.g., Royzman et al., 2015). Alternative methods are required to examine dumbfounding

585 with a larger sample, and whether it still occurs in the absence of the social pressure that is
586 present in an interview setting. Two responses have been identified as indicators of
587 dumbfounding. The degree to which each of these responses can be elicited in a setting other
588 than an interview is investigated in Studies 2 and 3.

589 **Study 2: Initial Computerised Task**

590 Having successfully elicited dumbfounded responses in a video recorded interview with
591 a small sample, the aim of Study 2 was to devise methods that might elicit dumbfounding in
592 a systematic way, using standardised materials and procedure that can be administered
593 without the need for an interviewer. This will eliminate participant-interviewer interaction as
594 a source of possible variability, remove the social pressure associated with an interview
595 setting, and enable the study to be conducted with a larger sample. It was hypothesised that
596 presenting participants with the same dilemmas and counter-arguments as in Study 1 as part
597 of a computer task, as opposed to in an interview, would lead to a similar state of
598 dumbfoundedness as found in Study 1. However, a major challenge to this alternative
599 medium of conducting the study is identifying specific behavioural responses that are
600 indicative of a state of dumbfoundedness that can be elicited and recorded. Without the
601 benefit of an experimenter to guide the discussion, and a video recording that can be
602 analysed, this challenge was addressed by developing a *critical slide* (described below).
603 Scenarios and counter-arguments to commonly made judgements were presented on a
604 sequence of slides before participants were asked to describe their judgement on a forced
605 choice critical slide. Participants were identified as dumbfounded if they selected an
606 unsupported declaration from a selection of three possible responses present on the critical
607 slide, or if they provided an unsupported declaration as a reason.

608 **Method**

609 **Participants and design.** Study 2 was a frequency-based, conceptual replication of
610 Study 1. The aim was to identify if dumbfounded responding could be evoked via a

611 computer-based task. All participants were presented with the same four moral vignettes.
612 Results are primarily descriptive. Further analysis tested for differences in responding
613 depending on the vignette, or type of vignette, presented.

614 A sample of of 72 participants (52 female, 20 male; $M_{\text{age}} = 21.18$, min = 18, max = 50,
615 $SD = 5.18$) took part in this study. Participants were undergraduate students and
616 postgraduate students from MIC. Participation was voluntary and participants were not
617 reimbursed for their participation.

618 **Procedure and materials.** This study used largely the same materials as in Study
619 1. The four vignettes from Study 1 *Heinz*, *Incest*, *Cannibal*, and *Trolley* (Appendix A) along
620 with the same prepared counter arguments (Appendix B) were used. Dumbfounding was
621 measured using the critical slide. The critical slide contained a statement defending the
622 behaviour and a question as to how the behaviour could be wrong (e.g., “Julie and Mark’s
623 behaviour did not harm anyone, how can there be anything wrong with what they did?”).
624 There were three possible answer options: (a) “There is nothing wrong”; (b) an unsupported
625 declaration, naming the specific behaviour described in the scenario (e.g., “Incest is just
626 wrong”); and finally a judgement with accompanying justification (c) “It’s wrong and I can
627 provide a valid reason”. The order of these response options was randomised. Participants
628 who selected (c) were then prompted on a following slide to type a reason. The selecting of
629 option (b), the unsupported declaration, was taken to be a dumbfounded response, as was
630 the use of an unsupported declaration as a justification for option (c).

631 This study made use of the same post-discussion questionnaire as in Study 1
632 (Appendix C). This was administered after the critical slide for each scenario. There was a
633 change to one of the questions on this post-discussion questionnaire: the question asking if
634 participants had changed their judgements was changed from “how much did your judgement
635 change?” with a seven point Likert scale response to “did your judgement change?” with a
636 binary “yes/no” response option. Both MLQ (Steger et al., 2008) and CRSi7 taken from The
637 Centrality of Religiosity Scale (S. Huber & Huber, 2012) were also used.

638 OpenSesame was used to present the vignettes and collect responses (Mathôt, Schreij,
639 & Theeuwes, 2012). The same four moral dilemmas (Appendix A) as in Study 1 were
640 presented to participants (in randomised order). Following the presentation of each dilemma,
641 participants were asked to judge, on a seven point Likert scale how right or wrong they
642 would rate the behaviour of the characters in the given scenario. After making a judgement
643 participants were then presented with a series of counter-arguments. Following these
644 counter-arguments, participants were presented with the critical slide. Following the critical
645 slide participants completed the same brief questionnaire as in Study 1 (between scenarios)
646 in which they were asked to rate, on a seven point Likert scale, how right/wrong they
647 thought the behaviour was; how confused they were; how irritated they were; how much
648 their judgement had changed; how much their judgement was based on reason; and how
649 much their judgement was based on “gut” feeling. When participants had completed all
650 questions relating to all four dilemmas they completed the same longer questionnaire as in
651 Study 1 containing the MLQ (Steger et al., 2008), the Centrality of Religiosity Scale (S.
652 Huber & Huber, 2012), and some questions relating to demographics. The entire study
653 lasted approximately fifteen to twenty minutes.

654 **Results and Discussion**

655 Participants who selected the unsupported declaration on the critical slide were
656 identified as dumbfounded. Table 1 shows the ratings of the behaviours across each scenario.
657 Table 2 shows the number, and percentage, of participants who displayed “dumbfounded”
658 responses (identified as the selecting of an unsupported declaration) and non-dumbfounded
659 responses for each dilemma. Figure 1 shows the percentage of participants displaying
660 dumbfounded responses for each dilemma. Table 3 shows the responses to the questionnaires
661 presented between dilemmas. The open-ended responses provided by participants who
662 selected option (c) “It’s wrong and I can provide a valid reason” were analysed and coded, by
663 the primary researcher, and unsupported declarations provided here were also identified as

664 dumbfounded responses. Following this coding, one additional participant was identified as
665 dumbfounded for *Trolley*. Sixty eight of the 72 participants (94%) selected the unsupported
666 declaration at least once. There was no statistically significant difference in responses to the
667 critical slide depending on the order of scenario presentation, $\chi^2(6, N = 288) = 4.13, p =$
668 $.659$. There was no statistically significant difference in responses to the critical slide
669 depending on scenario presented, $\chi^2(6, N = 288) = 9.00, p = .173$. Rates of dumbfounded
670 responding did not vary with type of moral scenario (100 instances for intuition scenarios, 90
671 instances for reasoning scenarios) being discussed, $\chi^2(2, N = 288) = 6.58, p = .037$. Forty
672 five participants (62.50%) selected the unsupported for *Heinz*. Forty six participants
673 (63.89%) selected (or provided) the unsupported declaration for *Cannibal* and *Trolley*. Fifty
674 four participants (75%) selected the unsupported declaration for *Incest*. There was no
675 association between number of times dumbfounded and score on either measure on the
676 Meaning and Life questionnaire; Presence $r(72) = -0.44, p = .662$, or Search, $r(72) = 1.12, p$
677 $= .268$, or the Centrality of Religiosity Scale $r(72) = 1.24, p = .220$.

678 The most striking result from this study was the willingness of participants to select
679 the unsupported declaration in response to a challenge to their judgement. This is
680 inconsistent with what was found in in both Study 1 and in the original study by Haidt et al.
681 (2000). In these studies, participants did not readily offer an unsupported declaration as
682 justification for their judgement, rather it was a last resort following extensive
683 cross-examining. The exceptionally high rates of dumbfounding observed in Study 2 do not
684 appear to be representative of the phenomenon more generally. There is, therefore, clearly a
685 difference between offering an unsupported declaration as a justification for a judgement
686 during an interview and selecting an unsupported declaration from a list of possible response
687 options during a computerised task. It is possible that, during the interview, participants
688 experienced a social pressure to successfully justify their judgement. This social pressure
689 may also have made participants were more aware of the illegitimacy of using an
690 unsupported declaration as a justification for their judgement. It is also possible that, seeing

691 it written down as a possible answer legitimises selecting it as a justification for the
692 judgement. The unsupported declaration does not provide an acceptable answer to the
693 question on the critical slide, however, its presence in the list of possible response options
694 may imply to participants that it is an acceptable answer, particularly if they do not put too
695 much thought into it. By selecting the unsupported declaration participants can move
696 quickly along to the next stage in the study without necessarily acknowledging any
697 inconsistency in their reasoning, avoiding potentially dissonant cognitions (e.g., Case,
698 Andrews, Johnson, & Allard, 2005; E. Harmon-Jones & Harmon-Jones, 2007; see also Heine,
699 Proulx, & Vohs, 2006). Selecting the unsupported declaration may also allow the participant
700 to proceed without expending effort trying to think of reasons for their judgement beyond
701 the intuitive justifications that had already been de-bunked.

702 Rates of dumbfounded responding in Study 2 were higher than expected. Possible
703 reasons for this could be (a) reduced social pressure to appear to have reasons for
704 judgements; (b) a failure of participants to comprehend that the unsupported declaration
705 does not provide a logically justifiable response to the question asked in the critical slide; (c)
706 the apparent legitimising of the unsupported declaration by its inclusion in the list of
707 possible response options; or (d) the selecting by participants of an “easy way out” option
708 without thinking about it fully (through carelessness/laziness/eagerness to move on to a less
709 taxing task). It appears that the selecting of unsupported declarations is not an accurate
710 measure of dumbfounding. In Study 1, participants were only identified as dumbfounded
711 based on the providing of an unsupported declaration if they subsequently failed to provide
712 further reasons when the unsupported declaration was questioned. However, in some cases,
713 participants who provided unsupported declarations were not identified as dumbfounded,
714 based on subsequent responses. A follow up analysis of the interview data revealed that 23
715 participants provided an unsupported declaration and proceeded to provide reasons for at
716 least one of their judgements; a further six participants provided an unsupported declaration
717 and proceeded to revise their judgement at least once. A stricter measure of dumbfounding,

718 one by which participants are required to explicitly acknowledge a state of dumbfoundedness
719 is necessary to address the issues with the selecting of an unsupported declaration that may
720 have led to the unusually high rates of dumbfounding observed in Study 2.

721 **Study 3a: Revised Computerised Task – College sample**

722 Study 3a was designed in response to the unexpectedly high rates of observed
723 dumbfounding in Study 2. Four limitations of the use of the unsupported declaration
724 selection as a measure of dumbfounding were identified. It was hypothesised that replacing
725 the unsupported declaration with an explicit admission of not having reasons would address
726 each of these limitations, and bring the option selection more in line with conversational
727 logic, making participants less willing to casually select the dumbfounded response. Making
728 participants explicitly acknowledge the absence of reasons for their judgement means that
729 their selecting of a dumbfounded response cannot be attributed to a mere misunderstanding
730 and thus, might provide a truer measure of dumbfounding.

731 **Method**

732 **Participants and design.** Study 3a was a frequency based, modified replication.
733 The aim was to identify if dumbfounded responding could be evoked. All participants were
734 presented with the same four moral vignettes. Results are primarily descriptive. Further
735 analysis tested for differences in responding depending on the vignette, or type of vignette,
736 presented.

737 A sample of 72 participants (46 female, 26 male; $M_{\text{age}} = 21.80$, min = 18, max = 46,
738 $SD = 3.91$) took part in this study. Participants were undergraduate students and
739 postgraduate students from MIC. Participation was voluntary and participants were not
740 reimbursed for their participation.

741 **Procedure and materials.** The materials in this study were almost the same as in
742 Study 2 with a change to the “dumbfounded” response option on the critical slide. Extra
743 questions were included following each of the counter-arguments. On the critical slide, the

744 unsupported declaration option was replaced with an admission of not having reasons (“It’s
745 wrong but I can’t think of a reason”). Following each counter-argument, participants were
746 asked if they (still) thought the behaviour was wrong, and if they had a reason for their
747 judgement. There was also a revision to the question on the post-discussion questionnaire
748 asking if participants had changed their judgements was changed: “did your judgement
749 change?” with a binary “yes/no” response option reverted back to “how much did your
750 judgement change?” with a seven point Likert scale response (as in Study 1). The same four
751 dilemmas *Heinz*, *Incest*, *Cannibal* and *Trolley* (Appendix A) along with the same prepared
752 counter arguments (Appendix B) as in Study 2 were used in Study 3a. Both the MLQ
753 (Steger et al., 2008); and CRSi7 (S. Huber & Huber, 2012) were also used. This study was
754 conducted in a designated psychology computer lab in MIC and was administered entirely on
755 individual computers using OpenSesame (Mathôt et al., 2012).

756 Participants were seated, given instructions, and allowed to begin the computer task.
757 The four vignettes from Study 1 *Heinz*, *Incest*, *Cannibal* and *Trolley* (Appendix A) along
758 with the same pre-prepared counter arguments (Appendix B) were used. Dumbfounding was
759 measured using the critical slide. The updated critical slide contained a statement defending
760 the behaviour and a question as to how the behaviour could be wrong (e.g., “Julie and
761 Mark’s behaviour did not harm anyone, how can there be anything wrong with what they
762 did?”) with three possible response options: (a) “There is nothing wrong”; (b) “It’s wrong,
763 but I can’t think of a reason”; (c) “It’s wrong and I can provide a valid reason”. The order of
764 these response options was randomised. Participants who selected (c) were required to
765 provide a reason. The selecting of option (b), the admission of not having reasons, was taken
766 to be a dumbfounded response. When participants had completed all questions relating to
767 all four dilemmas they completed the same longer questionnaire as in Studies 1 and 2
768 containing the MLQ (Steger et al., 2008), the Centrality of Religiosity Scale (S. Huber &
769 Huber, 2012), and some questions relating to demographics. The entire study lasted
770 approximately fifteen to twenty minutes.

771 Results and Discussion

772 Participants who selected the admission of not having reasons on the critical slide
773 (option b) were identified as dumbfounded. Forty of the 72 participants (56%) selected the
774 admission of not having reasons at least once. Table 1 shows the ratings of the behaviours
775 across each scenario. Table 2 and Figure 1 show the percentage of participants displaying
776 dumbfounded responses for each dilemma. Table 3 shows the responses to the questionnaires
777 presented between dilemmas. Again there was no statistically significant difference in
778 responses to the critical slide depending on the order of scenario presentation, $\chi^2(6, N =$
779 $288) = 0.61, p = .996$. There was no difference in responses to the critical slide depending on
780 scenario, $\chi^2(6, N = 288) = 9.60, p = .142$, or, type of scenario (32 instances for intuition
781 scenarios, 27 instances for reasoning scenarios), $\chi^2(2, N = 288) = 4.53, p = .104$. Thirteen
782 participants (18.06%) selected the admission of having no reasons for *Heinz*. Fourteen
783 participants (19.44%) selected the admission of not having reasons for *Cannibal* and *Trolley*.
784 Eighteen participants (25%) selected the admission of not having reasons for *Incest*.

785 The replacing of an unsupported declaration with an admission of having no reasons
786 led to substantially lower rates of dumbfounding than observed in Study 2. As such, it
787 appears that the issues associated with the selecting of an unsupported declaration have
788 been addressed in Study 3a. However, the rates of dumbfounding observed for *Incest* and
789 *Cannibal* in Study 3a were considerably lower than those observed in Study 1. This suggests
790 the revised measure may be too strict, measuring only open admissions of not having
791 reasons, but not accounting for a failure to provide reasons. As in the first computerised
792 task, participants who selected “It’s wrong and I can provide a valid reason” were then
793 required to provide a reason. In order to provide a measure of a failure to provide reasons,
794 these responses were analysed and coded, by the primary researcher. Those containing
795 unsupported declarations were taken as evidence for a failure to provide a reason and
796 identified as dumbfounded responses.

797 During the coding, another class of dumbfounded response was identified. Participants

798 occasionally provided undefended tautological responses as justification for their judgements,
799 whereby they simply named or described the behaviour in the scenario as justification for
800 their judgement (e.g., “They are related”, “Because it is cannibalism” [typographical error in
801 response]). These responses may be viewed as largely equivalent to unsupported declarations
802 (e.g., Mallon & Nichols, 2011). In Study 1, they were not identified as dumbfounded
803 responses, because when provided in an interview setting, they were always followed by
804 further questioning. This further questioning could lead to two possible responses: (a) a
805 dumbfounded response (unsupported declaration or an admission of not having reasons) or
806 (b) an alternative reason. A computerised task does not allow for a follow-up probe to
807 encourage participants to elaborate on such responses. Participants were not placed under
808 time pressure and could articulate and review their typed reason at their own pace. It is
809 reasonable to expect then, that, if participants did have a valid reason for their judgement,
810 they would have provided it along with, or instead of, the undefended tautological response.
811 As such, an undefended tautological reason appears to be evidence of a failure to identify
812 reasons . For this reason, these undefended tautological reasons were also coded as
813 dumbfounded responses, along with the unsupported declarations.

814 Table 2 and Figure 2 show the number and percentage of dumbfounded responses when
815 the coded string responses are included in the analysis. When the coded string responses are
816 included in the analysis, the number of participants displaying a dumbfounded response at
817 least once increased from 40 (56%) to 57 (79%). Observed rates of dumbfounding increased
818 for each scenario when the coded open-ended responses were included, with 19 participants
819 (26.39%) appearing to be dumbfounded by *Heinz*, 21 (29.17%) by *Cannibal*, 31 (43.06%) by
820 *Incest*, and 22 (30.56%) apparently dumbfounded by *Trolley*. Still, rates of dumbfounded
821 responding did not vary with type of moral scenario (52 instances for intuition scenarios, 41
822 instances for reasoning scenarios) being discussed, $\chi^2(1, N = 288) = 1.59, p = .208$. There
823 was no association between number of times dumbfounded and score on either measure on
824 the Meaning and Life questionnaire; Presence $r(72) = 0.82, p = .413$, or Search, $r(72) =$

825 0.07, $p = .945$, or the Centrality of Religiosity Scale $r(72) = 1.29$, $p = .201$.

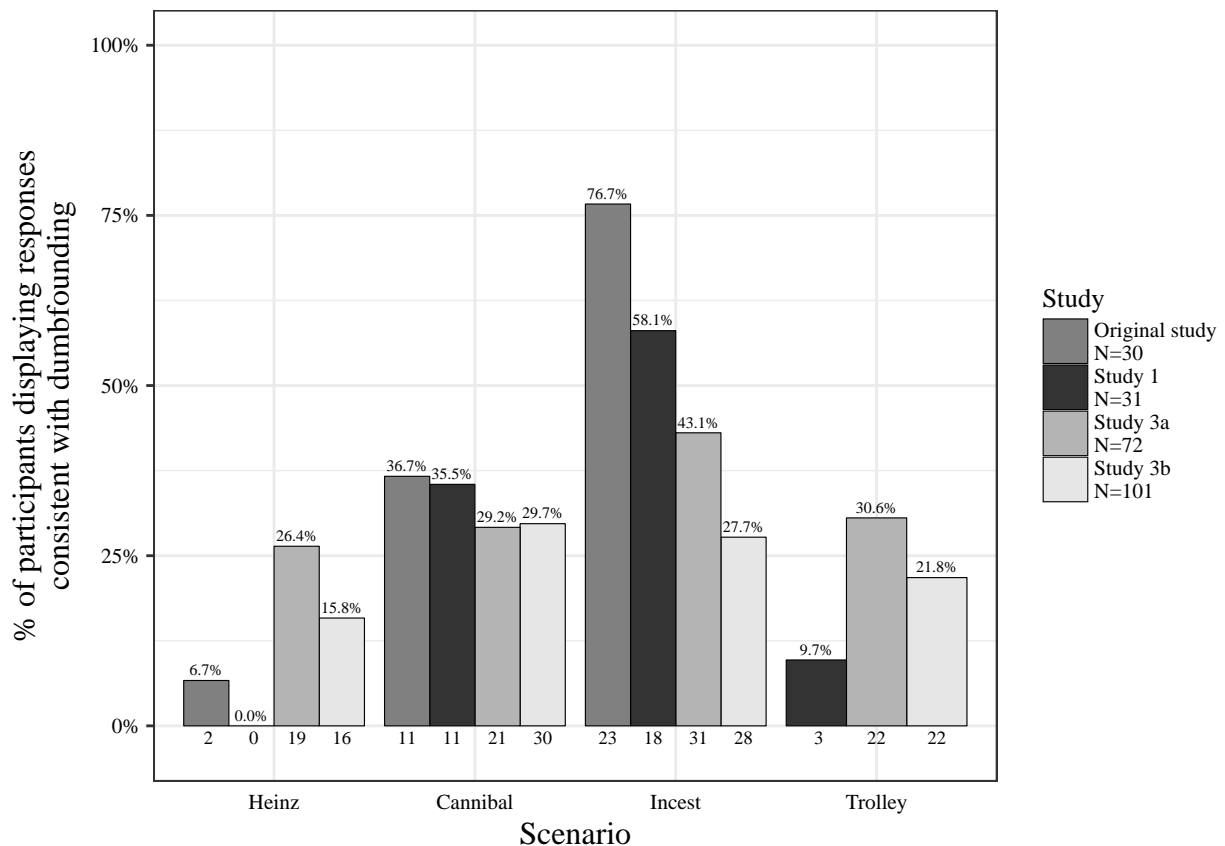


Figure 2. Rates of observed dumbfounding for each scenario across each study, including coded string responses.

826 When the coded open-ended responses were included in the analysis, the proportion of
 827 participants displaying a dumbfounded response at least once in Study 3a (79%) was much
 828 closer to that observed in the interview in Study 1 (74%) than before the open-ended
 829 responses were included (56%). The variation in observed rates of dumbfounding between
 830 dilemmas that was observed in the interview was not present in the computerised task. As
 831 such there remains a difference between the dumbfounding elicited during an interview and
 832 that elicited as part of a computerised task. However, it is clear that dumbfounded responses
 833 can be elicited as part of a computerised task. The participants in Studies 1, 2, and 3a were
 834 all college students (largely from the same institution) and as such, the following study
 835 investigated the phenomenon in a more diverse sample.

Study 3b: Revised Computerised Task – MTurk

836

837 Having successfully elicited dumbfounded responses in a college sample using a
838 computerised task in Study 3a, Study 3b was conducted in an attempt to replicate Study 3a
839 using more diverse sample using online recruiting through MTurk (Amazon Web Services
840 Inc., 2016).

841 Method

842 **Participants and design.** Study 3b was a frequency based, modified replication.
843 The aim was to identify if dumbfounded responding could be evoked. All participants were
844 presented with the same four moral vignettes. Results are primarily descriptive. Further
845 analysis tested for differences in responding depending on the vignette, or type of vignette,
846 presented.

847 A sample of 101 participants (53 female, 47 male; $M_{\text{age}} = 36.58$, min = 18, max = 69,
848 $SD = 12.45$) took part in this study. Participants were recruited online through MTurk
849 (Amazon Web Services Inc., 2016). Participation was voluntary and participants were paid
850 0.70 US dollars for their participation. Participants were recruited from English speaking
851 countries or from countries where residents generally have a high level of English (e.g., The
852 Netherlands, Denmark, Sweden). Location data for individual participants was not recorded,
853 however, based on other studies, using the same selection criteria, it is likely that 90% of the
854 sample was from the United States.

855 **Procedure and materials.** The materials in this study were almost the same as in
856 Study 3a, however, a different software package was used to present the materials and collect
857 the responses. OpenSesame (Mathôt et al., 2012) was replaced with Questback (Unipark,
858 2013) in order to facilitate online data collection. This meant that the recording of responses
859 changed from keyboard input to mouse input. It also allowed for multiple questions to be
860 displayed on the screen at the same time. Other than these changes, the materials were the
861 same as in Study 3a.

862 The computer task in Study 3b was much the same as Study 3a. The four vignettes
863 from Study 1: *Heinz*, *Incest*, *Cannibal*, and *Trolley* (Appendix A) along with the same
864 pre-prepared counter arguments (Appendix B). Dumbfounding was measured using the
865 critical slide.

866 The critical slide contained a statement defending the behaviour and a question as to
867 how the behaviour could be wrong, with three possible response options: (a) “There is
868 nothing wrong”; (b) “It’s wrong but I can’t think of a reason”; (c) “It’s wrong and I can
869 provide a valid reason”. Participants who selected (c) were required to provide a reason. The
870 order of these response options was randomised. When participants had completed all
871 questions relating to all four dilemmas they completed the same longer questionnaire as in
872 Studies 1 and 2 containing the Meaning and Life questionnaire (Steger et al., 2008), the
873 Centrality of Religiosity Scale (S. Huber & Huber, 2012), and some questions relating to
874 demographics. The entire study lasted approximately fifteen to twenty minutes.

875 **Results and Discussion**

876 Participants who selected the admission of not having reasons on the critical slide
877 (option b) were identified as dumbfounded. Table 1 shows the ratings of the behaviours
878 across each scenario. Table 2 and Figure 1 show the percentage of participants displaying
879 dumbfounded responses for each scenario. Table 3 shows the responses to the questionnaires
880 presented between scenario. On this occasion there was a statistically significant difference in
881 responses to the critical slide depending on the order of scenario presentation, $\chi^2(6, N =$
882 $404) = 14.77, p = .022$. The observed rates of dumbfounded responses were higher for the
883 third scenario, however they went down again for the fourth scenario along with rates of
884 selecting “nothing wrong”, meaning that the rates of participants providing reasons went up
885 again for the fourth scenario. The higher rates of providing reasons observed for the fourth
886 scenario presented means that this fluctuation is unlikely to be due to experimental fatigue,
887 which was the primary reason for testing for order effects. There was also a difference in

888 responses to the critical slide depending on scenario, $\chi^2(6, N = 404) = 15.18, p = .019$ with
889 more people selecting “nothing wrong” for *Incest* and fewer people selecting “nothing wrong”
890 for *Cannibal*. When dumbfounded responses are isolated and contrasted against other
891 responses this difference is no longer present, $\chi^2(3, N = 404) = 1.86, p = .602$. Forty four
892 participants (44%) selected the admission of not having reasons at least once. Twelve
893 participants (11.88%) selected the admission of having no reasons for *Heinz*. Sixteen
894 participants (15.84%) selected the admission of not having reasons for *Incest* and *Trolley*.
895 Nineteen participants (18.81%) selected the admission of not having reasons for *Cannibal*.

896 As in Study 3a, participants who selected option (c) “It’s wrong and I can provide a
897 valid reason”, were there then required to provide a reason through open-ended response.
898 These open-ended responses were coded, by the primary researcher, for dumbfounded
899 responses, again, identified as unsupported declarations or as undefended tautological
900 responses. Table 2 and Figure 2 show the rates of observed dumbfounding when the coded
901 open-ended responses were included in the analysis. As expected, the number of participants
902 displaying a dumbfounded response at least once increased, from 44 (44%) to 57 (56%).
903 Observed rates of dumbfounding increased for each scenario when the coded reasons were
904 included with 16 participants (15.84%) appearing to be dumbfounded by *Heinz*, 30 (29.70%)
905 by *Cannibal*, 28 (27.72%) by *Incest*, and 22 (21.78%) apparently dumbfounded by *Trolley*.
906 Taking these revised rates of dumbfounding there was a no significant difference in rates of
907 dumbfounded responding depending on scenario, $\chi^2(3, N = 404) = 6.56, p = .087$. There
908 was however, significantly more dumbfounded responding for the intuition scenarios (58
909 instances) than for the reasoning scenarios (38 instances), $\chi^2(1, N = 404) = 4.93, p = .026$.

910 There was no association between number of times dumbfounded and score on either
911 measure on the Meaning and Life questionnaire; Presence $r(101) = -0.78, p = .436$, or
912 Search, $r(101) = 0.63, p = .532$, or the Centrality of Religiosity Scale $r(101) = 0.44, p =$
913 $.662$. This is consistent with Studies 1, 2, and 3a. It appears that susceptibility to
914 dumbfounding is not related to either measure.

Combined Results and Discussion

915

Evaluating each Measure of Dumbfounding

916

917 The current research identifies moral dumbfounding as a rare demonstration of a
918 separation between intuitions and reasons for these intuitions (Barsalou, 2003, e.g., 2008,
919 2009; Crockett, 2013; Cushman, 2013). Two ways in which this separation may manifest
920 were identified. Firstly participants may acknowledge that they do not have reasons for their
921 judgements, admitting to not having reasons. Secondly, participants may fail to provide
922 reasons when asked, providing responses that fail to answer the question they were asked.
923 Two such responses were identified, unsupported declarations and tautological responses.

924 Measuring dumbfounding according to an admission of not having reasons only, in
925 Studies 1, 3a and 3b ($N = 204$), 100 participants (49%) were identified as dumbfounded at
926 least once. When a failure to provide reasons (taken as the providing of unsupported
927 declarations in Study 1, and, unsupported declarations and tautological responses in Study
928 3) was included as a dumbfounded response, 136 participants (67%) were identified as
929 dumbfounded at least once. When the selecting of an unsupported declaration (Study 2, N
930 = 72) was included ($N = 276$), 204 participants, (74%) were identified as dumbfounded at
931 least once.

932 The disparity in results between Study 2 and the other studies suggests that the
933 selection of an unsupported declaration does not provide a good measure of moral
934 dumbfounding. Participants in Studies 1, 3a, and 3b, recognised the illegitimacy
935 unsupported declarations as justifications for their judgement, with the majority of
936 participants avoided resorting to this type of response at all. The vast majority of
937 participants appeared to be willing to ignore the illegitimacy of the response, with large
938 numbers of participants selecting the unsupported declaration. While Study 2 did not
939 identify a means to measure dumbfounding, these results are interesting, and may provide an
940 insight into the cognitive processes that lead to dumbfounding.

941 Providing an unsupported declaration is clearly different to selecting one from a list of

942 possible responses. One possible explanation, is that dumbfounding is an aversive state,
943 similar to experiencing a threat to meaning (Heine et al., 2006; Proulx & Inzlicht, 2012), or
944 cognitive dissonance (Cooper, 2007; Festinger, 1957; E. Harmon-Jones & Harmon-Jones,
945 2007). The selecting of an unsupported declaration without deliberation allows participants
946 to avoid or minimise the impact of this aversive state and move on. Providing an
947 unsupported declaration involves more deliberation, making the illegitimacy of it more
948 salient, reducing its effectiveness in avoiding the aversive state of dumbfoundedness.
949 Furthermore, the relative attractiveness of these different responses to participants may be
950 linked to social desirability (Chung & Monroe, 2003; Latif, 2000; Morris & McDonald, 2013).
951 Follow-up work could investigate these questions directly.

952 The explicit acknowledgement of an absence of reasons can be measured systematically
953 by the selection of an admission of having no reasons. This is an unambiguous measure of
954 moral dumbfounding, does not account for participants who fail to provide reasons.
955 Measuring a failure to provide reasons, however, is more problematic. What is termed as a
956 valid reason is subjective. The providing of unsupported declarations and tautological
957 responses has been identified here as an indicator of a failure to provide reasons. This is
958 grounded in discussions of dumbfounding in the wider literature (Haidt, 2001; Mallon &
959 Nichols, 2011; Prinz, 2005), and the theoretical framework adopted here. Evidence for
960 equivalence of unsupported declarations and admissions of not having reasons was also found
961 in Study 1 whereby both measures displayed similar variability in non-verbal behaviours
962 when contrasted against participants who provided reasons, and participants who did not
963 rate the behaviour as wrong. However, caution is advised in taking unsupported declarations
964 as evidence for dumbfounding, particularly given the pattern of responses in Study 2, and
965 that a number of participants in Study 1 who provided an unsupported declaration
966 proceeded to provide reasons, or a revised judgement.

967 The current research identified two measures of dumbfounding. Limitations are
968 associated with each. Relying on admissions of having no reasons only, provides an overly

969 strict measure whereby a failure to provide reasons is not measured. Taking unsupported
970 declarations (and tautological reasons) as a measure of dumbfounding may provide too broad
971 a measure, risks identifying lazy or inattentive participants as dumbfounded. The providing
972 of a type-written response as part of a computerised task requires effort, and the majority of
973 participants avoid the use of unsupported declarations as justifications for their judgements.
974 This suggests that those who provided unsupported declarations did so because they failed
975 to identify alternative reason. It appears that the most practicable means to measure
976 dumbfounding accurately requires each of the responses: providing/selecting admissions of
977 not having reasons, and the providing of an unsupported declaration, to be accounted for.
978 Participants providing either of these responses may be identified as dumbfounded.

979 **Differences between Scenarios**

980 In Study 1 we found that rates of dumbfounded responding varied depending on the
981 scenario presented. Study 2 recorded high rates of dumbfounded responses for all scenarios.
982 In Studies 3a and 3b, we observed low rates of dumbfounded responding for all scenarios. In
983 Study 1 and Study 3b, we observed varying rates of dumbfounded responses depending on
984 scenario type. When Studies 3a and 3b are analysed together this variation is still observed,
985 with significantly more dumbfounded responses recorded for the intuition scenarios (110
986 instances) than for the reasoning scenarios (79 instances), $\chi^2(1, N = 288) = 6.55, p = .010$.
987 However, this combined analysis may be skewed in favour of Study 3b, due to the larger
988 sample size, 101 participants; Study 3a had only 72 participants. Further research and
989 continued replication is needed to confirm the reliability of this finding. When the
990 open-ended responses coded as tautological were included in the analysis of Studies 3a and
991 3b, the rates of dumbfounding appeared to be closer to those observed in Study 1.

992 Table 2 and Figure 1 show the initial observed rates of dumbfounding for each study.
993 Table 2 and Figure 2 show the revised rates of observed dumbfound responding in each study
994 once the open-ended coded responses from Studies 3a and 3b are included. Rates of

995 dumbfounding reported by Haidt et al. (2000) are also included for comparison. Study 2 was
996 a primarily a pilot study, and, as discussed, the observed rates of dumbfounding do not
997 appear to be representative of the phenomenon being studied, as such Study 2 is not
998 included in Figure 2.

999 **Differences between the Samples**

1000 The trend in observed rates of dumbfounded responses, across the dilemmas, identified
1001 by Haidt et al. (2000) appears to also be present in Study 1 (Interview). There does not
1002 appear to be a difference between scenarios in the computerised tasks. When the open-ended
1003 responses are included, the rates of observed dumbfounding for *Cannibal* appear to be
1004 similar across all the studies included in Figure 2 (two interviews and two computerised
1005 tasks). The computerised tasks appear to have higher rates of dumbfounding for both *Heinz*
1006 and *Trolley* than the interviews. There is a large degree of variation in the observed rate of
1007 dumbfounding for *Incest* between the four studies.

1008 *Incest* recorded higher rates of dumbfounding than the other scenarios in both
1009 interview studies (Study 1 and Haidt et al., 2000) and, to some degree, in Study 3a, the
1010 computer task with a college sample. The rate of dumbfounding observed for *Incest* with the
1011 online sample, in Study 3b, is lower than that observed with the college sample in Study 3a
1012 and is also slightly lower than that observed for *Cannibal* in the online sample. This is
1013 surprising, in that, the *Incest* dilemma is the most commonly cited example (e.g., Haidt,
1014 2001; Prinz, 2005; Royzman et al., 2015), and, in Studies 1, 2, and 3a, is the most reliable for
1015 eliciting dumbfounding, consistently eliciting higher rates than the other dilemmas. Looking
1016 at the ratings of the behaviours in each dilemma for each study may provide some clue as to
1017 where this variation comes from. The online sample were less inclined to rate the behaviour
1018 in *Incest* as wrong relative to the participants in the other studies. The percentage of
1019 participants initially rating *Incest* as wrong for each study are as follows: Study 1: 83.87%;
1020 Study 2: 87.50%; Study 3a: 84.72%; Study 3b: 70.30%. Furthermore, on the critical slide,

Table 4

Percentage of participants dumbfounded excluding participants who selected nothing wrong

	Heinz		Cannibal		Incest		Trolley	
	N	percent	N	percent	N	percent	N	percent
Study 1 (N = 31)	0/25	0%	11/23	47.83%	18/20	90%	3/23	13.04%
Study 2 (N = 72)	45/64	70.31%	46/68	67.65%	54/70	77.14%	46/62	74.19%
Study 3a (N = 72)	19/58	32.76%	21/68	30.88%	31/60	51.67%	22/57	38.6%
Study 3b (N = 101)	16/80	20%	30/91	32.97%	28/70	40%	22/77	28.57%

1021 the proportion of participants who selected “nothing wrong” for *Incest* for Study 3b (30.69%;
 1022 31 participants) was nearly double the proportion that selected “nothing wrong” for *Incest*
 1023 for Study 3a (16.67; 12 participants). When these participants are excluded from the
 1024 analysis of Study 3b (see Table 4 and Figure 3), the percentage of participants appearing to
 1025 be dumbfounded by *Incest* (22.86%; 16 participants; or 40%; 28 participants when
 1026 open-ended responses are included; $N = 70$) exceeds the percentage of participants
 1027 appearing to be dumbfounded by *Cannibal* (20.88%; 19 participants; or 32.97%; 30
 1028 participants when open-ended responses are included; $N = 91$). As such, it appears that the
 1029 apparent uncharacteristically low rates of observed dumbfounding for *Incest* in Study 3b,
 1030 when compared to *Cannibal*, may be due to the online sample being less inclined to rate the
 1031 behaviour as morally wrong rather than a difference in this sample’s ability to provide
 1032 justifications for their judgements to the two scenarios.

1033 It has been argued that moral dumbfounding occurs as a result of social pressure to
 1034 conform to conversational norms (Royzman et al., 2015). The findings presented by
 1035 Royzman et al. (2015) do not fully support this claim, however, they demonstrate that
 1036 incidences of moral dumbfounding are sensitive to social pressure. Studies 2 and 3, aimed to
 1037 reduce the influence of social pressure by testing dumbfounding as part of a computerised

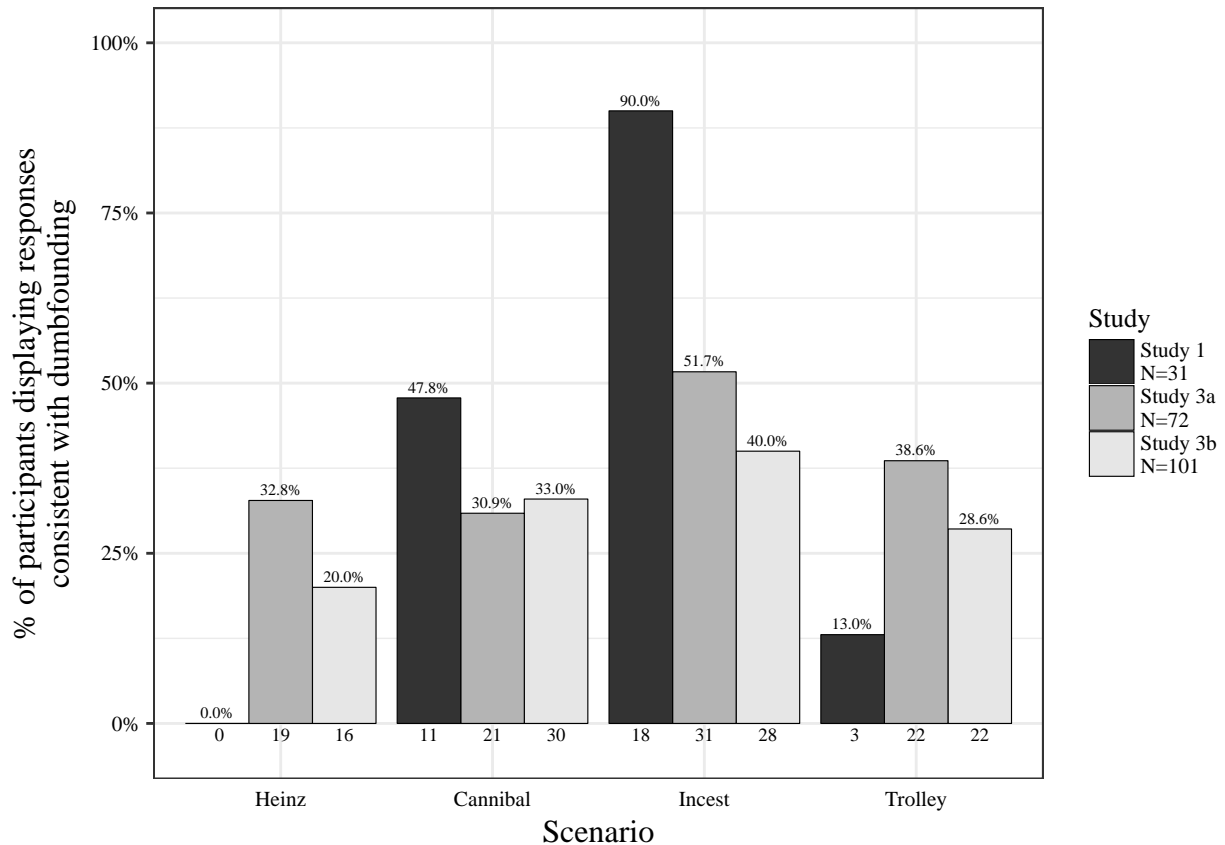


Figure 3. Percentage of dumbfounded responses when “nothing wrong” is excluded.

1038 task, as opposed to in an interview setting. The varying rates of dumbfounding depending
 1039 on task type indicate that the computerised task is different from the interview.

1040 Evidence that social pressure is reduced in the computerised task can be found by
 1041 examining the degree to which participants changed their minds, as measured in the
 1042 self-report response, and by comparing the initial judgements and revised judgements. The
 1043 self-report responses for Study 2 were of a binary yes/no form, whereas the responses in the
 1044 other studies were provided on a 1-7 Likert scale. As such the self-report data from Study 2
 1045 is not included in the analysis that follows.

1046 The mean responses for the self-report question “How much did you change your
 1047 mind?” are as follows: Study 1, $M = 2.88$, $SD = 1.59$; Study 3a, $M = 2.01$, $SD = 1.46$;
 1048 Study 3b, $M = 1.69$, $SD = 1.27$. A one-way ANOVA revealed significant differences in
 1049 responses to this question between the different studies, $F(2, 809) = 33.81$, $p < .001$, partial

1050 $\eta^2 = .077$. Tukey's post-hoc pairwise comparison revealed that responses in Study 1 were
1051 significantly higher than both Study 3a, $p < .001$, and Study 3b, $p < .001$. The responses in
1052 Study 3a were also significantly higher than the responses in Study 3b, $p = .008$.

1053 The initial judgements and revised judgements in the computer tasks were binned for
1054 comparison with the interview. "Wrong" judgements were assigned a value of "-1", "Right"
1055 judgements were assigned a value of "+1", "neutral" judgements were assigned a value of 0.
1056 The values for the revised judgements were subtracted from values for the initial judgements
1057 to create a new variable containing positive values ranging from -2 to +2. Negative values
1058 represent a change in judgement towards a more favourable judgement, and positive values
1059 represent a change in judgement towards condemning the actions. Higher values represent a
1060 greater swing in judgement. In the interview, there was only one incidence of a participant
1061 changing their judgement from favourable to condemnation, whereas 11 participants changed
1062 their judgement towards a more favourable judgement. In the computerised tasks, the
1063 numbers of participants changing their judgement in each direction is more balanced (see
1064 Table 1). There was a significant association between type of study and whether or not
1065 participants changed their mind in a given direction, $\chi^2(12, N = 1104) = 37.18, p < .001$.
1066 When Study 1 was removed this association disappeared, $\chi^2(8, N = 980) = 10.11, p = .258$.
1067 This pattern of results suggests that participants reacted differently in the interview than in
1068 the computerised tasks.

1069

General Discussion

1070 The goal of this research was to examine the replicability of dumbfounded responding
1071 following a moral judgement task, and identify specific measurable responses that may be
1072 viewed as indicators of moral dumbfounding. Four studies, with a combined total sample of
1073 $N = 276$, were conducted in an attempt to replicate and extend the original demonstration
1074 ($N = 30$) of moral dumbfounding by Haidt et al. (2000). We predicted that dumbfounded
1075 responses would be evoked when participants were required to provide justification for their

1076 moral judgements, when their basic intuitive justifications had been refuted. Two measures
1077 of moral dumbfounding were taken, an explicit acknowledgement of the absence of reasons,
1078 and a failure to provide reasons when pushed. Rates of observed dumbfounding vary
1079 depending on which measure is being employed.

1080 **Intuition versus Reasoning**

1081 Haidt et al. (2000) attribute the observed trend in dumbfounded responding to
1082 differences in type of scenario. They argue that *Heinz* is a “reasoning” scenario while
1083 *Cannibal* and *Incest* are “intuition” scenarios. Prinz (2005) suggests that these “intuition”
1084 scenarios have an emotional component, specifically that they elicit disgust, which leads to
1085 the judgement. Prinz argues that judgements grounded in disgust are more difficult to
1086 justify because they are grounded in emotion rather than reason. The variability between
1087 scenarios may be evidence for Haidt et al. prediction that judgements on the “intuition”
1088 scenarios would be more difficult to justify than the “reasoning” scenarios.

1089 Study 1, the interview, was the only study to produce robust differences between the
1090 scenarios.⁵ The results of the computerised tasks may indicate that there is no difference
1091 between the reasoning scenarios and the intuition scenarios. Alternatively, this may have
1092 highlighted a difference between an interview and a computerised task that influences the
1093 way people make moral judgements.

1094 It is possible that there exists a social influence in an interview setting that changes
1095 the way participants respond (e.g., Asch, 1956; Sabini, 1995; Staub, 2013) and, that the
1096 interviewer may be seen as a person in authority, demanding justifications for judgements
1097 made (e.g., Milgram, 1974). This may motivate participants to identify reasons to justify
1098 their judgements, leading to the suppression of dumbfounded responses. On the other hand,
1099 it may also motivate participants to heed the counter-arguments offered by the experimenter.
1100 This may lead to an interaction between scenario difficulty and social pressure to emerge,
1101 with the social pressure leading to fewer dumbfounded responses to the easier “reasoning”

1102 scenarios, but leading to more dumbfounded responses to the more difficult “intuition”
1103 scenarios. It may be the case that the rates of dumbfounding found in the computer tasks
1104 provide something of a crude baseline measure of participants’ initial perception of their own
1105 ability to justify their judgement of the scenario, having read the scenario and a number of
1106 counter-arguments. In the interview, these initial responses to the scenarios are distilled by
1107 the discussion with the experimenter to reflect the variation in difficulty between the
1108 scenarios.

1109 **Implications**

1110 The existence of moral dumbfounding has informed various theories of moral
1111 judgement either directly (e.g., Cushman et al., 2010; Haidt, 2001; Hauser et al., 2008; Prinz,
1112 2005) or indirectly (Crockett, 2013; Cushman, 2013; Greene, 2008, 2013). The original
1113 demonstration of moral dumbfounding remains unpublished in peer reviewed form (Haidt et
1114 al., 2000) and has not been directly replicated. The studies presented here aimed to replicate
1115 and extend this original moral dumbfounding study (Haidt et al., 2000) and thus, assess the
1116 notion that moral dumbfounding is in fact a psychological phenomenon that can be
1117 consistently observed. Study 1 successfully replicated the original study. Study 2 piloted the
1118 use of a computer task and recorded unexpectedly high rates of dumbfounded responding.
1119 Possible reasons for this were identified and addressed in Studies 3a and 3b. Study 3a and 3b
1120 recorded more moderate rates of dumbfounding with two different samples. All three studies
1121 successfully elicited dumbfounded responding identified as (a) admissions of not having
1122 reasons; (b) use of unsupported declarations as justification of a judgement; or (c) use of
1123 undefended tautological response as justification for a judgement; however, differences
1124 remain between the interview in Study 1 and the computerised task in Studies 3a and 3b.
1125 Taking these responses to be indicators of a state of dumbfoundedness, it appears that moral
1126 dumbfounding can be evoked in face-to-face and online contexts. As such, the research
1127 presented here may be seen as more support for the existence of intuitionist theories of moral

1128 judgement (e.g., Cushman et al., 2010; Greene, 2008; Haidt, 2001; Hauser et al., 2008; Prinz,
1129 2005) over rationalist theories (e.g., Kohlberg, 1971; Topolski et al., 2013).

1130 **Responding to Criticisms**

1131 The present research did not directly address the questions raised by Royzman et al.
1132 (2015). Those researchers suggest that there are two main factors that lead participants to
1133 produce responses that appear to be indicative of dumbfounding. Firstly, they argue that
1134 dumbfounded responding occurs as a result of social pressure to avoid appearing
1135 “uncooperative” (Royzman et al., 2015, p. 299), “inattentive” or “stubborn” (Royzman et al.,
1136 2015, p. 310). However, recall that the original definition of dumbfounding, which Royzman
1137 et al., employ, refers to the “stubborn” maintenance of a judgement. This creates a
1138 paradoxical situation whereby presenting as stubborn (as part of a dumbfounded response)
1139 occurs as a result of an attempt to avoid appearing stubborn. Secondly, they claim that
1140 participants’ judgements can be attributed to either norm-based reasons, or reason of
1141 potential harm. This claim is tested by presenting participants with questions relating to
1142 norm-based reasons and harm-based reasons, and excluding participants from analysis, based
1143 on their responses to these questions. They showed that almost all participants who rated
1144 the behaviour as wrong also endorsed at least one of these reasons. When controlling for the
1145 endorsing of these reasons Royzman et al. report a dumbfounding estimate of 1/53 which
1146 they report to be “not significantly greater than 0/53 ($z = 1.00, p = .32$)” (Royzman et al.,
1147 2015, p. 309) leading to the conclusion that, when controlling for norm-based reasons or
1148 harm-based reasons, moral dumbfounding does not occur. There are three main issues with
1149 the way this conclusion is reached.

1150 Firstly, the initial estimate of incidences of dumbfounding was 4/53 (7.55%). Based on
1151 the same calculations used by Royzman et al. (2015), this estimate of 4/53 is significantly
1152 greater than 0/53, $z = 2.04, p = .041$. These four participants were then interviewed further,
1153 during which, the “inconsistencies” in participants’ “responses were pointed out directly”

1154 (Royzman et al., 2015, p. 308). Following this interview, Royzman et al. were left with a
1155 dumbfounding estimate of 1/53 (which they claim is not significantly greater than 0/53).

1156 It is surprising that, having made the claim that dumbfounding arises as a result of
1157 social pressure, providing convincing evidence for this claim required a follow up interview,
1158 in which participants are exposed to social pressure. Using the same logic employed by
1159 Royzman et al. it would not be surprising if participants revised their responses after being
1160 “advised to carefully review and, if appropriate, revise” their responses (Royzman et al.,
1161 2015, p. 308). From this, it appears that incidences of dumbfounding can be reduced by
1162 changing the demands of the social situation. In effect, Royzman et al. (2015) have shown
1163 that moral dumbfounding is sensitive to social pressure. Demanding consistency between
1164 judgement and the endorsing of principles that may be relevant for a judgement reduces
1165 incidences of dumbfounding, whereas demanding consistency between a judgement and
1166 information contained in the vignette leads to increased dumbfounding. This is not the same
1167 as their claim that moral dumbfounding is caused by social pressure. Furthermore, the role
1168 of social pressure in the reduced incidences of dumbfounding observed is not acknowledged.

1169 Secondly, following this interview, Royzman et al. (2015) are still left with one
1170 participant who, by their own criteria, can be identified as dumbfounded (Royzman et al.,
1171 2015, p. 308). No explanation for the responding of this participant is offered, and cannot be
1172 explained by the theoretical position adopted in the conclusion. It is argued that one
1173 participant from a sample of 53, is not significantly greater than 0/53, $z = 1.00$, $p = .32$.
1174 Disregarding this estimate of moral dumbfounding as not statistically significant, $p = .32$,
1175 avoids offering an explanation for a response that is inconsistent with the argument made in
1176 the paper.

1177 Thirdly, and most importantly, the current research identifies dumbfounding as a rare
1178 demonstration of the separation between intuitions and reasons for these intuitions.
1179 Practical challenges to demonstrating this separation have already been identified: (a)
1180 post-hoc rationalisation and identification of reasons that are consistent with a judgement;

1181 (b) the possibility that the intuition emerged as a result of a well-rehearsed reasoned
1182 response. The work presented by Royzman et al. (2015) may be viewed as a practical
1183 demonstration of this first challenge; helping participants identify reasons that are consistent
1184 with their judgement and providing an opportunity them to endorse these reasons. As
1185 previously noted, the endorsing of a reason does not imply that the reason contributed to the
1186 judgement. This view of moral dumbfounding presents two methodological considerations
1187 that need to be addressed before accepting the claim that judgements in the dumbfounding
1188 paradigm can be attributed to either norm-based reasons or harm-based reasons. The first
1189 relates to participants' ability to articulate either harm-based or norm-based reasons. The
1190 second relates to the consistency with which these reasons guide judgements.

1191 Firstly, the final study reported by Royzman et al. (2015) does not report whether or
1192 not participants who endorsed either norm-based reasons or harm-based reasons also
1193 articulated the same reason. The mere endorsing of a principle or reason does not provide
1194 evidence that this principle guided the making of a judgement. To illustrate this point,
1195 consider the following scenario:

1196 Two friends (John and Pat) are bored one afternoon and trying to think of
1197 something to do. John suggests they go for a swim. Pat declines stating that it's
1198 too much effort - to get changed, and then to get dried and then washed and
1199 dried again after; he says he'd rather do something that requires less effort. John
1200 agrees and adds "Oh yeah, and there's that surfing competition on today so the
1201 place will be mobbed". To which Pat replies "Yeah exactly!".

1202 When John mentioned the surfing competition Pat immediately adopted it as another
1203 reason not to go for a swim however it is clear that this reason played no part in Pat's
1204 original judgement. It is possible that in identifying other reasons that are consistent with a
1205 particular judgement researchers may falsely attribute the judgement made to these reasons.
1206 The studies described by Royzman et al. (2015) do not sufficiently guard against the
1207 possibility of falsely attributing judgements to reasons endorsed, allowing for the possibility

1208 that some participants were falsely excluded from analysis. One way to avoid the false
1209 exclusion of participants would be to include an open-ended string response option
1210 immediately after the presenting of the vignette, in which participants are invited to provide
1211 the reason(s) for their judgement. Participants are then only excluded from analysis if they
1212 both articulated and endorsed a given principle.

1213 Secondly, consider the harm-based reasons, or the application of the harm principle.
1214 Royzman et al. (2015) argue that if participants do not believe that no harm came from the
1215 actions of Julie and Mark then concerns of harm may be considered a legitimate reason for
1216 judging the behaviour as wrong. Essentially, they have identified the harm principle as “it is
1217 wrong for two people to engage in an activity whereby harm may occur”. Royzman et al.
1218 (2015) argue that the application of this principle provides participants with a legitimate
1219 reason for their judgements. If this principle is guiding the judgements of participants, then
1220 this principle should be applied consistently across differing contexts. Royzman do not
1221 demonstrate that the participants in their sample consistently apply this principle across
1222 differing contexts (e.g., contact sports/boxing).

1223 Two indicators, measuring dumbfounding by differing standards, have been identified
1224 here: admissions of not having reasons, demonstrating an explicit acknowledgement of the
1225 absence of reasons; and unsupported declarations, demonstrating a failure to provide reasons
1226 when asked. The materials and measures developed here can be used in follow-up work in
1227 order address the methodological issues identified in the work of Royzman et al. (2015) and
1228 assess the strength of the concerns they identified in a more rigorous manner.

1229 **Limitations and Future Directions**

1230 The current research recorded variability between the different studies that remains
1231 unexplained. The interview recorded variation in responses between the different scenarios
1232 that was not observed in the computerised tasks. Possible explanations for this difference
1233 between computer task and interview have been offered here, however these are merely

1234 speculative and should be investigated further.

1235 The studies presented here are exploratory in design. The aim was to identify whether
1236 or not the phenomenon of moral dumbfounding could be elicited in a robust fashion. There
1237 was no experimental manipulation and analyses were primarily descriptive. These studies
1238 raise significant questions about the mechanisms underlying dumbfounded responses to
1239 moral judgement tasks, but clearly indicate that such dumbfounded responses can be reliably
1240 elicited, and demonstrate interesting variability. Future research is needed to identify specific
1241 variables that may moderate dumbfounding; examples may include meaning maintenance
1242 and meaning threat (Heine et al., 2006; Proulx & Inzlicht, 2012), need for closure
1243 (Kruglanski, 2013; Kruglanski & Webster, 1996), or zeal (McGregor, 2006a, 2006b;
1244 McGregor, Nash, & Prentice, 2012; McGregor, Zanna, Holmes, & Spencer, 2001).

1245

Conclusion

1246 The primary aim of the current studies was to examine the reliability of dumbfounded
1247 responding in moral judgements, and identify specific measurable indicators of moral
1248 dumbfounding. This is of particular interest considering the extent to which moral
1249 dumbfounding exists as a known phenomenon in the morality literature and its existence
1250 appears to inform theories of moral judgement. Two indicators of dumbfounding were taken:
1251 an admission of not having reasons and a failure to provide reasons when requested
1252 (measured by the providing of unsupported declarations/tautological responses). Four studies
1253 revealed varying rates of moral dumbfounding as recorded by these indicators depending on
1254 the type of task and on which indicator is being used. While further work is necessary to
1255 identify the specific variables that may moderate this variability, the research presented here
1256 demonstrated that two types of dumbfounded responding can be reliably elicited. In other
1257 words, we found that people are not always able to justify their moral judgements; they
1258 maintain their judgements in the absence of supporting reasons, in some cases they resort to
1259 unsupported declarations as justifications for judgements, in others admit that they do not

¹²⁶⁰ have reasons for their judgement. Further research is required to establish why this occurs.

References

1261

- 1262 Amazon Web Services Inc. (2016). Amazon Mechanical Turk.
- 1263 Asch, S. E. (1956). Studies of independence and submission to group pressures.
1264 *Psychological Monographs*, 70, 416.
- 1265 Aust, F. (2016). *Citr: 'RStudio' Add-in to Insert Markdown Citations*.
- 1266 Aust, F., & Barth, M. (2017). *Papaja: Create APA manuscripts with R Markdown*.
- 1267 Barsalou, L. W. (2003). Situated simulation in the human conceptual system. *Language and*
1268 *Cognitive Processes*, 18(5-6), 513–562. doi:[10.1080/01690960344000026](https://doi.org/10.1080/01690960344000026)
- 1269 Barsalou, L. W. (2008). Grounded Cognition. *Annual Review of Psychology*, 59(1), 617–645.
1270 doi:[10.1146/annurev.psych.59.103006.093639](https://doi.org/10.1146/annurev.psych.59.103006.093639)
- 1271 Barsalou, L. W. (2009). Simulation, situated conceptualization, and prediction.
1272 *Philosophical Transactions of the Royal Society B: Biological Sciences*, 364(1521),
1273 1281–1289. doi:[10.1098/rstb.2008.0319](https://doi.org/10.1098/rstb.2008.0319)
- 1274 Bates, D., & Maechler, M. (2017). *Matrix: Sparse and Dense Matrix Classes and Methods*.
- 1275 Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models
1276 Using lme4. *Journal of Statistical Software*, 67(1), 1–48. doi:[10.18637/jss.v067.i01](https://doi.org/10.18637/jss.v067.i01)
- 1277 Bellin, Z. (2012). The quest to capture personal meaning in psychology. *International*
1278 *Journal of Existential Psychology and Psychotherapy*, 4(1), 27.
- 1279 Berry, D., & Dienes, Z. P. (1993). *Implicit Learning: Theoretical and Empirical Issues*.
1280 Psychology Press.
- 1281 Cameron, C. D., Payne, B. K., & Doris, J. M. (2013). Morality in high definition: Emotion
1282 differentiation calibrates the influence of incidental disgust on moral judgments.
1283 *Journal of Experimental Social Psychology*, 49(4), 719–725.
1284 doi:[10.1016/j.jesp.2013.02.014](https://doi.org/10.1016/j.jesp.2013.02.014)
- 1285 Case, D. O., Andrews, J. E., Johnson, J. D., & Allard, S. L. (2005). Avoiding Versus Seeking:
1286 The Relationship of Information Seeking to Avoidance, Blunting, Coping, Dissonance,
1287 and Related Concepts. *Journal of the Medical Library Association : JMLA*, 93(3),

- 1288 353–362.
- 1289 Chaiken, S. (1980). Heuristic versus systematic information processing and the use of source
1290 versus message cues in persuasion. *Journal of Personality and Social Psychology*,
1291 *39*(5), 752–766. doi:[10.1037/0022-3514.39.5.752](https://doi.org/10.1037/0022-3514.39.5.752)
- 1292 Chaiken, S., & Trope, Y. (1999). *Dual-process Theories in Social Psychology*. Guilford Press.
- 1293 Chang, W. (2014). *Extrafont: Tools for using fonts*.
- 1294 Chung, J., & Monroe, G. S. (2003). Exploring Social Desirability Bias. *Journal of Business*
1295 *Ethics*, *44*(4), 291–302. doi:[10.1023/A:1023648703356](https://doi.org/10.1023/A:1023648703356)
- 1296 Cooper, J. (2007). *Cognitive dissonance: Fifty years of a classic theory* (Vol. xi). Thousand
1297 Oaks, CA: Sage Publications Ltd.
- 1298 Crockett, M. J. (2013). Models of morality. *Trends in Cognitive Sciences*, *17*(8), 363–366.
1299 doi:[10.1016/j.tics.2013.06.005](https://doi.org/10.1016/j.tics.2013.06.005)
- 1300 Cushman, F. A. (2013). Action, Outcome, and Value A Dual-System Framework for
1301 Morality. *Personality and Social Psychology Review*, *17*(3), 273–292.
1302 doi:[10.1177/1088868313495594](https://doi.org/10.1177/1088868313495594)
- 1303 Cushman, F. A., Young, L., & Greene, J. D. (2010). Multi-system Moral Psychology. In J.
1304 M. Doris & F. A. Cushman (Eds.), *The Moral Psychology Handbook* (pp. 47–71).
1305 Oxford; New York: Oxford University Press.
- 1306 Cushman, F. A., Young, L., & Hauser, M. D. (2006). The Role of Conscious Reasoning and
1307 Intuition in Moral Judgment Testing Three Principles of Harm. *Psychological Science*,
1308 *17*(12), 1082–1089. doi:[10.1111/j.1467-9280.2006.01834.x](https://doi.org/10.1111/j.1467-9280.2006.01834.x)
- 1309 Dreyfus, H. L., & Dreyfus, S. E. (1990). What is moral maturity? A phenomenological
1310 account of the development of ethical expertise. *Universalism Vs.*
1311 *Communitarianism*, 237–264.
- 1312 Dwyer, S. (2009). Moral Dumbfounding and the Linguistic Analogy: Methodological
1313 Implications for the Study of Moral Judgment. *Mind & Language*, *24*(3), 274–296.

- 1314 doi:[10.1111/j.1468-0017.2009.01363.x](https://doi.org/10.1111/j.1468-0017.2009.01363.x)
- 1315 Epstein, S. (1994). Integration of the cognitive and the psychodynamic unconscious.
1316 *American Psychologist*, *49*(8), 709–724. doi:[10.1037/0003-066X.49.8.709](https://doi.org/10.1037/0003-066X.49.8.709)
- 1317 Evans, J. S. B. T. (2003). In two minds: Dual-process accounts of reasoning. *Trends in*
1318 *Cognitive Sciences*, *7*(10), 454–459. doi:[10.1016/j.tics.2003.08.012](https://doi.org/10.1016/j.tics.2003.08.012)
- 1319 Evans, J. S. B. T., & Wason, P. C. (1976). Rationalization in a Reasoning Task. *British*
1320 *Journal of Psychology*, *67*(4), 479–486. doi:[10.1111/j.2044-8295.1976.tb01536.x](https://doi.org/10.1111/j.2044-8295.1976.tb01536.x)
- 1321 Festinger, L. (1957). *A theory of cognitive dissonance*. Stanford CA: Stanford University
1322 Press.
- 1323 Fox, J., & Weisberg, S. (2011). *An R Companion to Applied Regression* (Second.). Thousand
1324 Oaks CA: Sage.
- 1325 Friard, O., & Gamba, M. (2015, December). BORIS - Behavioral Observation Research
1326 Interactive Software. Italy.
- 1327 Gazzaniga, M. S., & LeDoux, J. E. (2013). *The Integrated Mind*. Springer Science &
1328 Business Media.
- 1329 Gray, K., Schein, C., & Ward, A. F. (2014). The myth of harmless wrongs in moral
1330 cognition: Automatic dyadic completion from sin to suffering. *Journal of*
1331 *Experimental Psychology: General*, *143*(4), 1600–1615. doi:[10.1037/a0036149](https://doi.org/10.1037/a0036149)
- 1332 Greene, J. D. (2008). The Secret Joke of Kant’s Soul. In *Moral Psychology Volume 3: The*
1333 *neurosciences of morality: Emotion, brain disorders, and development* (pp. 35–79).
1334 Cambridge (Mass.): the MIT press.
- 1335 Greene, J. D. (2013). *Moral tribes: Emotion, reason, and the gap between us and them*.
- 1336 Greene, J. D., Sommerville, R. B., Nystrom, L. E., Darley, J. M., & Cohen, J. D. (2001). An
1337 fMRI investigation of emotional engagement in moral judgment. *Science (New York,*
1338 *N.Y.)*, *293*(5537), 2105–2108. doi:[10.1126/science.1062872](https://doi.org/10.1126/science.1062872)
- 1339 Haidt, J. (2001). The emotional dog and its rational tail: A social intuitionist approach to
1340 moral judgment. *Psychological Review*, *108*(4), 814–834.

- 1341 doi:[10.1037/0033-295X.108.4.814](https://doi.org/10.1037/0033-295X.108.4.814)
- 1342 Haidt, J. (2007). The New Synthesis in Moral Psychology. *Science*, *316*(5827), 998–1002.
1343 doi:[10.1126/science.1137651](https://doi.org/10.1126/science.1137651)
- 1344 Haidt, J., & Björklund, F. (2008). Social Intuitionists Answer Six Questions about Moral
1345 Psychology. In W. Sinnott-Armstrong (Ed.), *Moral psychology Volume 2, The*
1346 *cognitive science of morality: Intuition and diversity* (pp. 181–217). London: MIT.
- 1347 Haidt, J., & Hersh, M. A. (2001). Sexual Morality: The Cultures and Emotions of
1348 Conservatives and Liberals. *Journal of Applied Social Psychology*, *31*(1), 191–221.
1349 doi:[10.1111/j.1559-1816.2001.tb02489.x](https://doi.org/10.1111/j.1559-1816.2001.tb02489.x)
- 1350 Haidt, J., Björklund, F., & Murphy, S. (2000). Moral dumbfounding: When intuition finds
1351 no reason. *Unpublished Manuscript, University of Virginia*.
- 1352 Haidt, J., Koller, S. H., & Dias, M. G. (1993). Affect, culture, and morality, or is it wrong to
1353 eat your dog? *Journal of Personality and Social Psychology*, *65*(4), 613–628.
1354 doi:[10.1037/0022-3514.65.4.613](https://doi.org/10.1037/0022-3514.65.4.613)
- 1355 Harmon-Jones, E., & Harmon-Jones, C. (2007). Cognitive Dissonance Theory After 50 Years
1356 of Development. *Zeitschrift Für Sozialpsychologie*, *38*(1), 7–16.
1357 doi:[10.1024/0044-3514.38.1.7](https://doi.org/10.1024/0044-3514.38.1.7)
- 1358 Hauser, M. D., Cushman, F. A., Young, L., Kang-Xing Jin, R., & Mikhail, J. (2007). A
1359 Dissociation Between Moral Judgments and Justifications. *Mind & Language*, *22*(1),
1360 1–21. doi:[10.1111/j.1468-0017.2006.00297.x](https://doi.org/10.1111/j.1468-0017.2006.00297.x)
- 1361 Hauser, M. D., Young, L., & Cushman, F. A. (2008). Reviving Rawls’s Linguistic Analogy:
1362 Operative Principles and the Causal Structure of Moral Actions. In W.
1363 Sinnott-Armstrong (Ed.), *Moral psychology Volume 2, The cognitive science of*
1364 *morality: Intuition and diversity* (pp. 107–155). London: MIT.
- 1365 Heine, S. J., Proulx, T., & Vohs, K. D. (2006). The Meaning Maintenance Model: On the
1366 Coherence of Social Motivations. *Personality and Social Psychology Review*, *10*(2),

- 1367 88–110. doi:[10.1207/s15327957pspr1002_1](https://doi.org/10.1207/s15327957pspr1002_1)
- 1368 Huber, S., & Huber, O. W. (2012). The Centrality of Religiosity Scale (CRS). *Religions*,
1369 3(3), 710–724. doi:[10.3390/rel3030710](https://doi.org/10.3390/rel3030710)
- 1370 IBM Corp. (2015). SPSS. Armonk, NY: IBM Corp.
- 1371 Jacobson, D. (2012). Moral Dumbfounding and Moral Stupefaction. In *Oxford studies in*
1372 *normative ethics* (Vol. 2, p. 289).
- 1373 Jacoby, L. L. (1983). Remembering the data: Analyzing interactive processes in reading.
1374 *Journal of Verbal Learning and Verbal Behavior*, 22(5), 485–508.
- 1375 Johansson, P., Hall, L., Sikström, S., & Olsson, A. (2005). Failure to Detect Mismatches
1376 Between Intention and Outcome in a Simple Decision Task. *Science*, 310(5745),
1377 116–119. doi:[10.1126/science.1111709](https://doi.org/10.1126/science.1111709)
- 1378 Kahneman, D. (2011). *Thinking, fast and slow*. London: Allen Lane.
- 1379 Kohlberg, L. (1971). *From is to Ought: How to Commit the Naturalistic Fallacy and Get*
1380 *Away with it in the Study of Moral Development*.
- 1381 Kruglanski, A. W. (2013). *The Psychology of Closed Mindedness*. Psychology Press.
- 1382 Kruglanski, A. W., & Webster, D. M. (1996). Motivated closing of the mind: “Seizing” and
1383 “freezing.” *Psychological Review*, 103(2), 263–283. doi:[10.1037/0033-295X.103.2.263](https://doi.org/10.1037/0033-295X.103.2.263)
- 1384 Latif, D. A. (2000). The Link Between Moral Reasoning Scores, Social Desirability, and
1385 Patient Care Performance Scores: Empirical Evidence from the Retail Pharmacy
1386 Setting. *Journal of Business Ethics*, 25(3), 255–269. doi:[10.1023/A:1006049605298](https://doi.org/10.1023/A:1006049605298)
- 1387 Lenth, R. (2016). *Estimability: Tools for Assessing Estimability of Linear Predictions*.
- 1388 Lenth, R. V. (2016). Least-Squares Means: The R Package lsmeans. *Journal of Statistical*
1389 *Software*, 69(1), 1–33. doi:[10.18637/jss.v069.i01](https://doi.org/10.18637/jss.v069.i01)
- 1390 Mallon, R., & Nichols, S. (2011). Dual Processes and Moral Rules. *Emotion Review*, 3(3),
1391 284–285. doi:[10.1177/1754073911402376](https://doi.org/10.1177/1754073911402376)
- 1392 Marwick, B. (n.d.). *Wordcountaddin: Word counts and readability statistics in R markdown*

- 1393 *documents.*
- 1394 Mathôt, S., Schreij, D., & Theeuwes, J. (2012). OpenSesame: An open-source, graphical
1395 experiment builder for the social sciences. *Behavior Research Methods*, *44*(2),
1396 314–324. doi:[10.3758/s13428-011-0168-7](https://doi.org/10.3758/s13428-011-0168-7)
- 1397 McGregor, I. (2006a). Offensive Defensiveness: Toward an Integrative Neuroscience of
1398 Compensatory Zeal After Mortality Salience, Personal Uncertainty, and Other
1399 Poignant Self-Threats. *Psychological Inquiry*, *17*(4), 299–308.
1400 doi:[10.1080/10478400701366977](https://doi.org/10.1080/10478400701366977)
- 1401 McGregor, I. (2006b). Zeal Appeal: The Allure of Moral Extremes. *Basic and Applied Social*
1402 *Psychology*, *28*(4), 343–348. doi:[10.1207/s15324834basp2804_7](https://doi.org/10.1207/s15324834basp2804_7)
- 1403 McGregor, I., Nash, K. A., & Prentice, M. (2012). Religious zeal after goal frustration. In M.
1404 A. Hogg & D. L. Baylock (Eds.), *Extremism and the Psychology of Uncertainty* (pp.
1405 147–164). Hoboken NJ: Wiley-Blackwell.
- 1406 McGregor, I., Zanna, M. P., Holmes, J. G., & Spencer, S. J. (2001). Compensatory
1407 conviction in the face of personal uncertainty: Going to extremes and being oneself.
1408 *Journal of Personality and Social Psychology*, *80*(3), 472–488.
1409 doi:[10.1037/0022-3514.80.3.472](https://doi.org/10.1037/0022-3514.80.3.472)
- 1410 McHugh, C. (2017). *Desnum: Creates some useful functions.*
- 1411 Milgram, S. (1974). *Obedience to Authority: An Experimental View*. New York: Harper and
1412 Row.
- 1413 Morris, S. A., & McDonald, R. A. (2013). The Role of Moral Intensity in Moral Judgments:
1414 An Empirical Investigation. In A. C. Michalos & D. C. Poff (Eds.), *Citation Classics*
1415 *from the Journal of Business Ethics* (pp. 463–479). Springer Netherlands.
1416 doi:[10.1007/978-94-007-4126-3_23](https://doi.org/10.1007/978-94-007-4126-3_23)
- 1417 Narvaez, D. (2005). The neo-Kohlbergian tradition and beyond: Schemas, expertise, and
1418 character. In G. Carlo & C. Pope-Edwards (Eds.), *Nebraska symposium on*

- 1419 *motivation* (Vol. 51, p. 119).
- 1420 Nisbett, R. E., & Wilson, T. D. (1977). Telling more than we can know: Verbal reports on
1421 mental processes. *Psychological Review*, *84*(3), 231.
- 1422 Pizarro, D. A., & Bloom, P. (2003). The intelligence of the moral intuitions: A comment on
1423 Haidt (2001). *Psychological Review*, *110*(1), 193–196.
1424 doi:[10.1037/0033-295X.110.1.193](https://doi.org/10.1037/0033-295X.110.1.193)
- 1425 Prinz, J. J. (2005). Passionate Thoughts: The Emotional Embodiment of Moral Concepts.
1426 In D. Pecher & R. A. Zwaan (Eds.), *Grounding Cognition: The Role of Perception*
1427 *and Action in Memory, Language, and Thinking* (pp. 93–114). Cambridge University
1428 Press.
- 1429 Proulx, T., & Inzlicht, M. (2012). The Five “A”s of Meaning Maintenance: Finding Meaning
1430 in the Theories of Sense-Making. *Psychological Inquiry*, *23*(4), 317–335.
1431 doi:[10.1080/1047840X.2012.702372](https://doi.org/10.1080/1047840X.2012.702372)
- 1432 R Core Team. (2017a). *Foreign: Read Data Stored by Minitab, S, SAS, SPSS, Stata, Systat,*
1433 *Weka, dBase, .*
- 1434 R Core Team. (2017b). *R: A Language and Environment for Statistical Computing*. Vienna,
1435 Austria: R Foundation for Statistical Computing.
- 1436 Reber, A. S. (1989). Implicit learning and tacit knowledge. *Journal of Experimental*
1437 *Psychology: General*, *118*(3), 219–235. doi:[10.1037/0096-3445.118.3.219](https://doi.org/10.1037/0096-3445.118.3.219)
- 1438 Royzman, E. B., Kim, K., & Leeman, R. F. (2015). The curious tale of Julie and Mark:
1439 Unraveling the moral dumbfounding effect. *Judgment and Decision Making*, *10*(4),
1440 296–313.
- 1441 Sabini, J. (1995). *Social psychology*. New York; London: Norton.
- 1442 Schnell, T. (2011). Individual differences in meaning-making: Considering the variety of
1443 sources of meaning, their density and diversity. *Personality and Individual*

- 1444 *Differences*, 51(5), 667–673. doi:[10.1016/j.paid.2011.06.006](https://doi.org/10.1016/j.paid.2011.06.006)
- 1445 Singmann, H., Bolker, B., & Westfall, J. (2015). *Afex: Analysis of Factorial Experiments*.
- 1446 Sneddon, A. (2007). A Social Model of Moral Dumbfounding: Implications for Studying
1447 Moral Reasoning and Moral Judgment. *Philosophical Psychology*, 20(6), 731–748.
1448 doi:[10.1080/09515080701694110](https://doi.org/10.1080/09515080701694110)
- 1449 Staub, E. (2013). *Positive Social Behavior and Morality: Social and Personal Influences*.
1450 Elsevier.
- 1451 Steger, M. F., Kashdan, T. B., Sullivan, B. A., & Lorentz, D. (2008). Understanding the
1452 Search for Meaning in Life: Personality, Cognitive Style, and the Dynamic Between
1453 Seeking and Experiencing Meaning. *Journal of Personality*, 76(2), 199–228.
1454 doi:[10.1111/j.1467-6494.2007.00484.x](https://doi.org/10.1111/j.1467-6494.2007.00484.x)
- 1455 Sun, R., Slusarz, P., & Terry, C. (2005). The Interaction of the Explicit and the Implicit in
1456 Skill Learning: A Dual-Process Approach. *Psychological Review*, 112(1), 159–192.
1457 doi:[10.1037/0033-295X.112.1.159](https://doi.org/10.1037/0033-295X.112.1.159)
- 1458 Topolski, R., Weaver, J. N., Martin, Z., & McCoy, J. (2013). Choosing between the
1459 emotional dog and the rational pal: A moral dilemma with a tail. *Anthrozoös*, 26(2),
1460 253–263. doi:[10.2752/175303713X13636846944321](https://doi.org/10.2752/175303713X13636846944321)
- 1461 Unipark, Q. (2013). QuestBack Unipark.(2013).
- 1462 Wickham, H. (2007). Reshaping Data with the reshape Package. *Journal of Statistical
1463 Software*, 21(12), 1–20.
- 1464 Wickham, H. (2009). *Ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New
1465 York.
- 1466 Wickham, H. (2011). The Split-Apply-Combine Strategy for Data Analysis. *Journal of
1467 Statistical Software*, 40(1), 1–29.
- 1468 Wickham, H. (2016). *Scales: Scale Functions for Visualization*.
- 1469 Wickham, H., & Chang, W. (2017). *Devtools: Tools to Make Developing R Packages Easier*.
- 1470 Wielenberg, E. J. (2014). *Robust Ethics: The Metaphysics and Epistemology of Godless*

- 1471 *Normative Realism*. OUP Oxford.
- 1472 Wilson, T. D., & Bar-Anan, Y. (2008). The unseen mind. *Science*, *321*(5892), 1046–1047.
- 1473 Zajonc, R. B. (1980). Feeling and thinking: Preferences need no inferences. *American*
- 1474 *Psychologist*, *35*(2), 151–175. doi:[10.1037/0003-066X.35.2.151](https://doi.org/10.1037/0003-066X.35.2.151)

Table captions

1475

- 1476 *Table 1.* Ratings of each scenario for each study
- 1477 *Table 2.* Observed frequency and percentage of each of the responses: dumb-
1478 founded, nothing wrong, and reasons provided
- 1479 *Table 3.* Responses to post-discussion questionnaire questions
- 1480 *Table 4.* Percentage of participants dumbfounded excluding participants who se-
1481 lected nothing wrong

1482

Figure captions

- 1483 *Figure 1.* Rates of observed dumbfounding for each scenario across each study.
- 1484 *Figure 2.* Rates of observed dumbfounding for each scenario across each study,
1485 including coded string responses.
- 1486 *Figure 3.* Percentage of dumbfounded responses when “nothing wrong” is excluded.

Footnotes

1487

1488 ¹In the present paper we will follow the practice of the majority of authors discussing dumbfounding in
1490 focusing on the unpublished Haidt et al. Manuscript, as it is freely available to download from the University
1491 of Virginia.

1492 ²Recent work by Royzman, Kim, and Leeman (2015) includes a demonstration of dumbfounding using the
1493 incest scenario. This work is an attempt to identify possible reasons that may be guiding the judgement of
1494 participants and in limiting its focus to a single scenario (Incest), it is not classed here as a direct replication
1495 of the original work by Haidt et al. (2000).

1496 ³These are largely theoretical arguments offering explanations of dumbfounding that are consistent with a
1497 rationalist perspective (e.g., Kohlberg, 1971; Topolski, Weaver, Martin, & McCoy, 2013). However Royzman,
1498 Kim, and Leeman (2015) present some empirical evidence in support of this position. This is discussed in
1499 more detail below.

1500 ⁴R (3.4.1, R Core Team, 2017b) and the R-packages *afex* (0.15.2, Singmann, Bolker, & Westfall, 2015), *car*
1501 (2.1.5, Fox & Weisberg, 2011), *citr* (0.2.0.9047, Aust, 2016), *desnum* (0.1.1, McHugh, 2017), *devtools* (1.13.2,
1502 Wickham & Chang, 2017), *estimability* (1.2, R. Lenth, 2016), *extrafont* (0.17, Chang, 2014), *foreign* (0.8.69, R
1503 Core Team, 2017a), *ggplot2* (2.2.1, Wickham, 2009), *lme4* (1.1.13, Bates, Mächler, Bolker, & Walker, 2015),
1504 *lsmeans* (2.27.2, R. V. Lenth, 2016), *Matrix* (1.2.11, Bates & Maechler, 2017), *papaja* (0.1.0.9492, Aust &
1505 Barth, 2017), *plyr* (1.8.4, Wickham, 2011), *reshape2* (1.4.2, Wickham, 2007), *scales* (0.5.0, Wickham, 2016),
1506 and *wordcountaddin* (0.2.0, Marwick, n.d.)

1507 ⁵Some differences were observed in Study 3b, however these existed only when scenarios were grouped by
1508 type, this inter-scenario variation in rates of dumbfounding is not equivalent to that observed in Study 1.

Appendices

Appendix A: Moral Scenarios

Heinz

In Europe, a woman was near death from a very bad disease, a special kind of cancer. There was one drug that the doctors thought might save her. It was a form of radium for which a druggist was charging ten times what the drug cost him to make. The sick woman's husband, Heinz, went to everyone he knew to borrow the money, but he could only get together about half of what it cost. He told the druggist that his wife was dying, and asked him to sell it cheaper or let him pay later. But the druggist said, "No, I discovered the drug and I'm going to make money from it." So, Heinz got desperate and broke into the man's store to steal the drug for his wife. The druggist had Heinz arrested and charged (Haidt et al., 2000).

Trolley

A Trolley is hurtling down a track towards five people. It will kill them all on impact. Paul is on a bridge under which it will pass. He can stop it by putting something very heavy in front of it. As it happens, there is a very fat man next to him – Paul's only way to stop the trolley is to push him over the bridge and onto the track, killing him to save five. Paul decides to push the man (adapted from Greene et al., 2001).

Cannibal (original)

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

1534 waste perfectly edible meat. So she cut off a piece of flesh, and took it home and cooked it.
1535 The person had died recently of a heart attack, and she cooked the meat thoroughly, so there
1536 was no risk of disease (Haidt et al., 2000).

1537 **Cannibal (revised)**

1538 Jennifer works in a medical school pathology lab as a research assistant. The lab
1539 prepares human cadavers that are used to teach medical students about anatomy. The
1540 cadavers come from people who had donated their body for the general use of the researchers
1541 in the lab. The bodies are normally cremated, however, severed cuts may be disposed of at
1542 the discretion of lab researchers, One night Jennifer is leaving the lab when she sees a body
1543 that is going to be discarded the next day. Jennifer was a vegetarian, for moral reasons. She
1544 thought it was wrong to kill animals for food. But then, when she saw a body about to be
1545 cremated, she thought it was irrational to waste perfectly edible meat. So she cut off a piece
1546 of flesh, and took it home and cooked it. The person had died recently of a heart attack, and
1547 she cooked the meat thoroughly, so there was no risk of disease

1548 **Incest**

1549 Julie and Mark, who are brother and sister, are travelling together in France. They
1550 are both on summer vacation from college. One night they are staying alone in a cabin near
1551 the beach. They decide that it would be interesting and fun if they tried making love. At
1552 very least it would be a new experience for each of them. Julie was already taking birth
1553 control pills, but Mark uses a condom too, just to be safe. They both enjoy it, but they
1554 decide not to do it again. They keep that night as a special secret between them, which
1555 makes them feel even closer to each other (Haidt et al., 2000).

Appendix B: Sample Statements to Challenge Judgements

1556

Heinz/Druggist

1557

- 1558 • Do you agree that the druggist has to make a living?
- 1559 • And do you accept that Heinz broke into the druggist's store?
- 1560 • And do you accept that he stole from him?

1561

-
- 1562 • Do think that Heinz should try to save his wife's life?
 - 1563 • And do you agree that he tried to get the money together
 - 1564 • And do you accept that Heinz tried to negotiate with the druggist

Trolley

1565

- 1566 • Do you accept that five people would have died if Paul didn't push the man?
- 1567 • And this man is the only way available to stop the trolley? (Paul does not weigh
- 1568 enough)
- 1569 • Do you agree that in stopping the trolley Paul saved the lives of five people?

Cannibal

1570

- 1571 • The body had been donated for research, it was to be discarded the next day. You
- 1572 must agree then that it had obviously fulfilled its purpose?
- 1573 • Do you accept that the body was already dead?
- 1574 • And do you accept that there was no risk of disease?

Incest

1575

- 1576 • Do you not agree that any concerns regarding reproductive complications are eased by
- 1577 their using of two forms of contraception?
- 1578 • And do you accept that they are both consenting adults, and that they both consented
- 1579 and enjoyed it?

1580

- And do you concede that nobody else was affected by their actions?

Contributions:

1605

- 1606 • Substantial contributions to conception and design: CMH, MMG, ERI, ELK,
- 1607 • Acquisition of data: CMH
- 1608 • Analysis and interpretation of data: CMH, MMG, ERI, ELK,
- 1609 • Drafting the article or revising it critically for important intellectual content: CMH,
- 1610 MMG, ERI, ELK,
- 1611 • Final approval of the version to be published: CMH, MMG, ERI, ELK,

1612

Competing Interests:

1613

1614 The authors declare that there are no competing interests.

1615

Data accessibility statement:

1616

1617 All participant data, and analysis scripts can be found on this paper's project page on
1618 the Open Science Framework at <https://osf.io/wm6vc/>