UNIVERSITY of CALIFORNIA PRESS | Collabra

McHugh, C., et al. (2017). Searching for Moral Dumbfounding: Identifying Measurable Indicators of Moral Dumbfounding. *Collabra: Psychology,* 3(1): 23, DOI: https://doi.org/10.1525/collabra.79

ORIGINAL RESEARCH REPORT

Searching for Moral Dumbfounding: Identifying Measurable Indicators of Moral Dumbfounding

Cillian McHugh*, Marek McGann*, Eric R. Igou† and Elaine L. Kinsella†

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

Keywords: Morality; Dumbfounding; Judgement; Intuitions; Reasoning

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

Defining and Measuring Moral Dumbfounding

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

It is apparent from the literature that there is no single, agreed definition of moral dumbfounding. That said, an absence of reasons for, or an inability to justify or defend, a moral judgement, is consistently identified across definitions. However, even despite this apparent consistency, there remains considerable variation in the language used to describe this "failure to provide reasons for a moral judgement". Indeed, the lack of definitional specificity has led to differing interpretations of moral dumbfounding. It also allows for the possibility of disagreement relating to the implications, both theoretical and practical, of moral dumbfounding.

* Mary Immaculate College, University of Limerick, IE

† University of Limerick, IE

Corresponding author: Cillian McHugh (cillian.mchugh@mic.ul.ie)

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

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

This view of dumbfounding includes the elements of the phenomenon that are mentioned most frequently within the wider literature. It is also consistent with the way dumbfounding is described in the original study by Haidt et al. (2000). They report interesting variation in a number of non-verbal behaviours that may be linked with stubbornness or puzzlement, but beyond these, they do not offer a specific indication of how stubbornness and puzzlement are operationalised. Furthermore, other than appearing in the introductory definition for dumbfounding, in the abstract, (Haidt et al., 2000, p. 2), the terms "stubborn" and "puzzled" do not appear again for the remainder of the paper, suggesting that they are not core elements of the phenomenon.

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

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

(The Short) History of Moral Dumbfounding

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

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

The following year, Haidt and Hersh (2001) investigated differences between conservatives and liberals, across a range of responses to moral issues, and found that conservatives produced more dumbfounded type responses (e.g., stuttering, stating "I don't know", admitting

they could not explain their answers (Haidt & Hersh, 2001, p. 200)), than liberals when discussing particular issues. Although this study did not investigate dumbfounding directly, the findings indicate that there may be individual differences that drive moral judgements which have not yet been fully investigated.

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

Moral Dumbfounding and Moral Intuitions

Moral dumbfounding is used as supporting evidence for a range of "intuitionist" theories of moral judgement (e.g. Cushman et al., 2010; Haidt, 2001; Prinz, 2005). According to these intuitionist theories, our moral judgements are grounded in an emotional or intuitive automatic response rather than slow deliberate reasoning (Cameron, Payne, & Doris, 2013; Crockett, 2013; Cushman, 2013; Cushman et al., 2010; Greene, 2008; Haidt, 2001; Prinz, 2005). Two of the most influential such theories of moral judgement have been Haidt's social intuitionist model (Haidt, 2001; Haidt, & Björklund, 2008) and Greene's dual processes model (Greene, 2008, 2013; Greene, Sommerville, Nystrom, Darley, & Cohen, 2001). Haidt (2001) in his social intuitionist model likens the distinction between fast moral intuitions and slow moral reasoning to the distinction between fast and slow thinking that appears in dual systems theories of cognition (Chaiken, 1980; see also Chaiken, & Trope, 1999; Epstein, 1994; Haidt, 2001; Kahneman, 2011; Zajonc, 1980). In introducing and defending this model, Haidt makes specific reference to one of the dumbfounding scenarios, and the findings from the unpublished manuscript relating to this dilemma (Haidt, 2001; see also Haidt, & Björklund, 2008; Haidt, & Hersh, 2001). Greene draws heavily on Haidt's work in defending his dual-process model of moral judgement (Greene, 2008). In more recent years, Cushman (2013; Cushman et al., 2010) and Crockett (2013), building on the work of Haidt and Greene have continued the development intuitionist/dual-process theories of moral judgement (Crockett, 2013; Cushman, 2013; Greene, 2008; Haidt, 2001).

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

However, the claim that reasons for intuitions are learned independently of the intuition does not necessarily imply that there are no reasons for a given intuition. This leads to two difficulties in demonstrating this separation between intuitions and reasons for the intuition. Firstly, in many circumstances, it is possible to trace the emergence of a given social or moral norm to particular reasons. Pizarro and Bloom (2003) defend the claim that moral intuitions may be rational, and informed by prior reasoning or deliberation. A related, more general claim is that deliberative (model-based) responses can, over time, become automatic or habitual (e.g., Barsalou, 2003; Cushman, 2013; H. L. Dreyfus, & Dreyfus, 1990). Secondly, in many cases, after an intuitive judgement is made, reasons that are consistent with the judgement may be identified through post-hoc rationalisation (e.g., Cushman et al., 2006). This means that, although there is a clear theoretical case for a separation between intuitions and reasons for these intuitions, demonstrating this separation is problematic.

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

The existence of moral dumbfounding, therefore, is compelling evidence for intuitionist theories of moral judgement. These theories are supported by a large body of other empirical evidence, however, they are also either directly (e.g., Cushman et al., 2010; Haidt, 2001; Hauser et al., 2008; Prinz, 2005) or indirectly (e.g., Crockett, 2013; Cushman, 2013; Greene, 2008, 2013) grounded in the assumption that moral dumbfounding is a real phenomenon. The present research aims, to test the validity of the claim that moral dumbfounding is a real

phenomenon through an attempted replication of the widely-cited unpublished study by Haidt et al. (2000). This will also test the strength of existing moral theories grounded in its existence. In addition to this, we aim to identify specific, measurable indicators of dumbfounding and develop practicable methods for eliciting and measuring dumbfounding in larger samples. These may be used to explore the phenomenon in greater depth, informing the further development of moral theory.

Challenges to Moral Dumbfounding

In recent years moral dumbfounding has been challenged by a number of authors (e.g., Gray et al., 2014; Jacobson, 2012; Sneddon, 2007; Wielenberg, 2014), arguing, in line with rationalist theories of moral judgement (Kohlberg, 1971; Narvaez, 2005; Topolski, Weaver, Martin, & McCoy, 2013), that moral judgements are grounded in reasons. Recent work by Royzman, Kim, and Leeman (2015), involving a series of studies focusing on the Incest dilemma, identified two reasons that may be guiding participants' judgements. The reasons identified were: (a) potential harm – where participants believed that harm could arise as a result of the actions of the characters in the scenario despite the vignette stating that no harm arose; and (b) normativity – where citing a moral norm is seen as sufficient justification for making a judgement consistent with that norm. They found, that, when participants who endorsed either of these reasons were excluded from analysis, there were only four participants (from a sample of fifty-three) who rated the behaviour as wrong without offering a reason. Following a subsequent interview, two of these participants changed their judgement, and one changed her response to the question relating to normative reasons. This left just one participant who maintained that the behaviour was wrong without valid reason and, in their view, could be truly identified as dumbfounded. Consequently, they argue that dumbfounding is not as prevalent a phenomenon as portrayed by (Haidt et al. 2000; Royzman et al., 2015, p. 310). In identifying reasons that appear to be guiding people's judgements, they claim to have found evidence for rationalist theories of moral judgement (Royzman et al., 2015, p. 311) over intuitionist theories. They argue that the dumbfounded behaviours observed by Haidt et al. (2000) can be attributed to social pressure that exists in an interview setting, whereby participants accept the counter-arguments offered by the interviewer, even if they disagree, in order to appear cooperative (Royzman et al., 2015, p. 299).

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

The inability of people to articulate principles that are consistent with, and therefore may arguably be guiding moral judgements has been documented in a study by Cushman et al. (2006). They identified three distinct principles that appear to guide moral judgements; these are: (a) harm caused by action is worse than harm caused by omission; (b) harm intended is worse than harm foreseen; (c) harm involving physical contact is worse than harm without physical contact. They conducted a series of studies in which participants' judgements were largely consistent with these principles. Interestingly, however, when questioned afterwards, participants were only reliably able to articulate two of these principles (a) and (c). Principle (b), while consistent with the judgements made, was not well articulated by participants. It appears that, making judgements consistent with a principle does not imply that participants can articulate this principle. It is this inability to articulate principles or reasons for a moral judgement that is the hallmark of moral dumbfounding and is of key interest in the current research.

The Current Research

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

Study 1: Interview

The primary aim of Study 1 was to replicate the original dumbfounding study (Haidt et al., 2000). Four moral judgement vignettes were used (Appendix A). Three of

these vignettes (*Heinz*, *Incest*, and *Cannibal*) were taken from Haidt et al. (2000). A fourth vignette (*Trolley*) was adapted from Greene et al. (2001). Haidt et al. (2000) contrasted *Heinz*, a so-called reasoning scenario, against *Cannibal* and *Incest*, so-called intuition scenarios. Their study also included two tasks that did not have any moral content. For the purposes of consistency and balance, the non-moral tasks were omitted from the present study, and a second moral reasoning vignette was included in their stead, such that two reasoning vignettes (*Heinz* and *Trolley*) were contrasted against two intuition vignettes (*Incest* and *Cannibal*). We hypothesised that dumbfounding would be elicited and that rates of dumbfounded responding would vary depending on the content of the dilemma, with the intuition scenarios eliciting more dumbfounded responses than the reasoning scenarios. Two measures of dumbfounding were taken reflecting the two distinct ways in which absence of reasons may present: admissions of not having reasons (self-reported dumbfounding), and the use of an unsupported declaration (it's just wrong) as a justification for a judgement, with a failure to provide any alternative reason when the unsupported declaration was questioned (a failure to provide reasons). As in the original study (Haidt et al., 2000), various non-verbal measures were also recorded in an attempt to account for stubbornness and puzzlement.

Method

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

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

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

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

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

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

Sample counter-arguments were prepared for each scenario. To ensure that participants were only pushed to defend a judgement of "morally wrong" these counter-arguments exclusively defended the potentially questionable behaviour of the characters. A list of prepared counter-arguments can be seen in Appendix B. A post-discussion questionnaire, taken from Haidt et al. (2000) was administered after discussion of each scenario (Appendix C).

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

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

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

Results and Discussion

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

Twenty two of the 31 participants (70.97%) produced a dumbfounded response (admission of having no reasons; or the use of an unsupported declaration as a justification for a judgement, with a failure to provide any alternative reason when the unsupported declaration was questioned) at least once. Examples of such responses included "It just seems wrong and I cannot explain why, I don't know", "because I just think it's wrong, oh God, I don't know why, it's just [pause] wrong". **Table 2** shows the number, and percentage, of participants who displayed dumbfounded responses and non-dumbfounded responses for each

Table 1: Ratings of each scenario for each study.

		Heinz		Cannibal		Incest		Trolley	
Study	**Judgement**	**N**	**percent**	**N**	**percent**	**N**	**percent**	**N**	**percent**
Study 1	Initial: Wrong	27	87.10%	25	80.65%	26	83.87%	23	74.19%
	Initial: Neutral	0	0%	0	0%	0	0%	0	0%
	Initial: OK	4	12.90%	6	19.35%	5	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	73.61%	68	94.44%	63	87.5%	50	69.44%
	Initial: Neutral	9	12.50%	3	4.17%	3	4.17%	6	8.33%
	Initial: OK	10	13.89%	1	1.39%	6	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	75%	67	93.06%	61	84.72%	48	66.67%
	Initial: Neutral	6	8.33%	3	4.17%	7	9.72%	10	13.89%
	Initial: OK	12	16.67%	2	2.78%	4	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	80.20%	85	84.16%	71	70.3%	66	65.35%
	Initial: Neutral	9	8.91%	13	12.87%	20	19.8%	14	13.86%
	Initial: OK	11	10.89%	3	2.97%	10	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%

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	Nothing wrong	14	19.44%	4	5.56%	12	16.67%	15	20.83%
(critical slide)	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	Nothing wrong	14	19.44%	4	5.56%	12	16.67%	15	20.83%
(coded)	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	Nothing wrong	21	20.79%	10	9.9%	31	30.69%	24	23.76%
(critical slide)	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	Nothing wrong	21	20.79%	10	9.9%	31	30.69%	24	23.76%
(coded)	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%

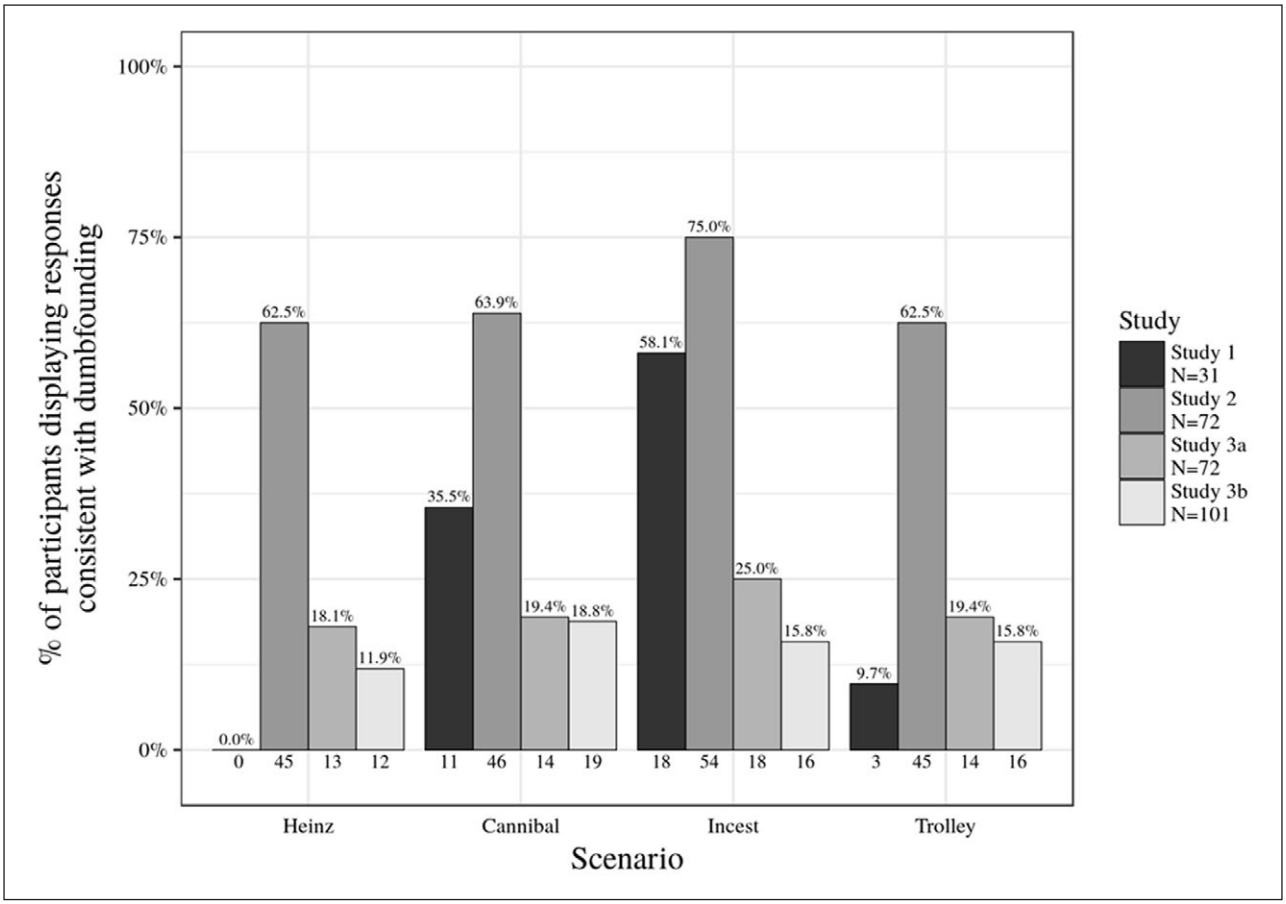

Figure 1: Rates of observed dumbfounding for each scenario across each study.

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

dilemma. The rates of each type of dumbfounded response are also displayed. **Figure 1** shows the percentage of participants displaying dumbfounded responses for each dilemma. **Table 3** shows the responses to the questionnaires presented between dilemmas.

In line with the original study (Haidt et al., 2000), the videos were also coded by the primary researcher across a range of measures. Haidt et al. (2000) report differences, between intuition and reasoning scenarios. They do not, however, report comparisons between participants identified as dumbfounded and participants not identified as dumbfounded. The current research, aiming to identify measurable indicators of dumbfounding, categorised participants as dumbfounded according to the two types of verbal responses (admissions and unsupported declaration) and compared these groups with participants who were not identified as dumbfounded, across a range of measures. There were two stages in this analysis. Firstly, all participants identified as dumbfounded were compared against participants who provided reasons only. Secondly, participants identified as dumbfounded were grouped according to type of dumbfounded response, and participants who did not rate the behaviour as wrong were also included in the analysis.

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

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

Paralinguistic variables reported by Haidt et al. (2000) include frequency (per minute) of: "ums, uhs, hmms", "turns with laughter", "turns with face touch", "doubt faces", and "turns with pen fiddle". As with the argument variables, the coding of the non-verbal/paralinguistic responses also varies slightly from what was reported by Haidt et al. (2000). We coded for both verbal hesitations ("um/em/uh") and non-verbal hesitations/stuttering. "Turns" was coded independently of other behaviours as changing position. Laughter was coded for independently of changing position. The coding of hands touching the self was not limited to the face. Participants did not have pens to fiddle with, however we coded for generic fidgeting. The term "doubt faces" presented as problematic to code for rigorously across different individuals. As such,

two distinctive and opposing facial expressions were coded for: smiling and frowning.

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

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

A one-way ANOVA revealed a significant difference in number of times per minute participants laughed between the dumbfounded group, (M = 2.81, SD = 2.84) and the group who provided reasons (M = 1.18, SD = 1.25), F (1, 89) = 14.35, p < .001, partial η^2 = .139. Similarly, a one-way ANOVA revealed a significant difference in relative amount of time spent smiling (as a proportion of the total time spent on the given scenario) between the dumbfounded group, (M = .32, SD = .15) and the group who provided reasons (M = .16, SD = .14), F (1, 89) = 25.24, p < .001, partial η^2 = .221. Consistent with the results reported by Haidt et al. (2000), a series of one-way ANOVAs revealed no differences in verbal hesitations, F (1, 89) = 2.35, p = .129, partial η^2 = .026, non-verbal hesitations, p = .074, changing posture, p = .485, hands on the self, p = .864, frowning, p = .958, and fidgeting, F (1, 89) = 1.66, p = .201, partial η^2 = .018. A one-way ANOVA revealed a significant difference relative amount of time spent in silence (as a proportion of the total time spent on the given scenario) between the dumbfounded group, (M = .14, SD = .08) and the group who provided reasons (M = .09, SD = .06), F (1, 89) = 9.72, p = .002, partial η^2 = .098.

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

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

A one-way ANOVA revealed a significant difference in relative frequency of utterances whereby participants were working towards a reason depending on overall reaction to scenarios, F (3, 120) = 7.54, p < .001, partial η^2 = .159. Tukey's post-hoc pairwise comparison revealed that participants who provided reasons were identified as working towards a reason significantly more frequently (M = 2.70, SD = 1.53) than participants who did not rate the behaviour as wrong (M = 1.76, SD = 1.48), p = .021, and more frequently than participants who provided unsupported declarations as justifications (M = .64, SD = .72), p < .001. There was no difference between participants who admitted to not having reasons (M = 1.90, SD = 1.56) and any of the other groups. A one-way ANOVA revealed no significant difference in relative frequency of expressions of doubt depending on overall reaction to scenarios, F (3, 120) = 2.17, p = .096, partial η^2 = .051.

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

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

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

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

Further analyses. An exploratory analysis revealed no association between number of times dumbfounded and score on either measures from the MLQ: Presence, r (31) = 0.74, p = .466, or Search, r (31) = 1.38, p = .179, or the Centrality of Religiosity Scale r (31) = 0.35, p = .726. There was no difference in observed rates of dumbfounded responses depending on the order of scenario presentation, χ^2(6, N = 124) = 4.01, p = .676. Rates of dumbfounded responses varied depending on which moral dilemma was being discussed, χ^2(6, N = 124) = 46.82, p < .001. The highest rate of dumbfounding was recorded for *Incest*, with 18 of the 31 (58.06%) participants displaying dumbfounded responses. Eleven participants (35.48%) displayed dumbfounded responses for *Cannibal* and three participants (9.68%) displayed dumbfounded responses for *Trolley*. The lowest recorded rate of dumbfounded response was for the Heinz dilemma, with no participants resorting to unsupported declarations as justification or admitting to not having reasons for their judgement. This trend is generally consistent with that which emerged in the original study (with the exception of *Trolley*, which was not used in the original study). Furthermore, rates of dumbfounded responding varied depending on which type of moral scenario was being discussed. *Heinz* and *Trolley*, identified as reasoning scenarios, were contrasted against the intuition scenarios *Incest* and *Cannibal*. There was significantly more dumbfounded responding for the intuition scenarios (29 instances) than for the reasoning scenarios (3 instances), χ^2(2, N = 124) = 38.17, p < .001.

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

Study 1 provides evidence supporting the view that moral dumbfounding is a genuine phenomenon and can be elicited in an interview setting when participants are pressed to justify their judgements of particular moral scenarios. Two key limitations have been identified as a result of conducting studies in an interview setting. Firstly, conducting video-recorded interviews, and the accompanying analyses, is particularly labour intensive, which leads to a smaller sample size. The aims of the present research were to examine the replicability of dumbfounding, and to identify specific measurable indicators of dumbfounding. A sample size of thirty-one is not sufficient in fulfilling the first aim. Secondly, an interview setting introduces a social context that may influence the responses of participants, in that, participants may feel a social pressure to behave in a particular way (Royzman et al., 2015). Alternative methods are required to examine dumbfounding with a larger sample, and whether it still occurs in the absence of the social pressure that is present in an interview setting. Two responses have been identified as indicators of dumbfounding. The degree to which each of these responses can be elicited in a setting other than an interview is investigated in Studies 2 and 3.

Study 2: Initial Computerised Task

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

the critical slide, or if they provided an unsupported declaration as a reason.

Method

Participants and design. Study 2 was a frequency-based, conceptual replication of Study 1. The aim was to identify if dumbfounded responding could be evoked via a computer-based task. All participants were presented with the same four moral vignettes. Results are primarily descriptive. Further analysis tested for differences in responding depending on the vignette, or type of vignette, presented.

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

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

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

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

Results and Discussion

Participants who selected the unsupported declaration on the critical slide were identified as dumbfounded. **Table 1** shows the ratings of the behaviours across each scenario. **Table 2** shows the number, and percentage, of participants who displayed "dumbfounded" responses (identified as the selecting of an unsupported declaration) and non-dumbfounded responses for each dilemma. **Figure 1** shows the percentage of participants displaying dumbfounded responses for each dilemma. **Table 3** shows the responses to the questionnaires presented between dilemmas. The open-ended responses provided by participants who selected option (c) "It's wrong and I can provide a valid reason" were analysed and coded, by the primary researcher, and unsupported declarations provided here were also identified as dumbfounded responses. Following this coding, one additional participant was identified as dumbfounded for *Trolley*. Sixty eight of the 72 participants (94%) selected the unsupported declaration at least once. There was no statistically significant difference in responses to the critical slide depending on the order of scenario presentation, $\chi^2(6, N = 288) = 4.13$, $p = .659$. There was no statistically significant difference in responses to the critical slide depending on scenario presented, $\chi^2(6, N = 288) = 9.00$, $p = .173$. Rates of dumbfounded responding did not vary with type of moral scenario (100 instances for intuition scenarios, 90 instances for reasoning scenarios) being discussed, $\chi^2(2, N = 288) = 6.58$, $p = .037$. Forty five participants (62.50%) selected the unsupported for *Heinz*. Forty six participants (63.89%) selected (or provided) the unsupported declaration for *Cannibal* and *Trolley*. Fifty four participants (75%) selected the unsupported declaration for *Incest*. There was no association between number of times dumbfounded and score on either measure on the Meaning and Life questionnaire; Presence $r(72) = -0.44$, $p = .662$, or Search, $r(72) = 1.12$, $p = .268$, or the Centrality of Religiosity Scale $r(72) = 1.24$, $p = .220$.

The most striking result from this study was the willingness of participants to select the unsupported declaration in response to a challenge to their judgement. This is inconsistent with what was found in in both Study 1 and in the original study by Haidt et al. (2000). In these

studies, participants did not readily offer an unsupported declaration as justification for their judgement, rather it was a last resort following extensive cross-examining. The exceptionally high rates of dumbfounding observed in Study 2 do not appear to be representative of the phenomenon more generally. There is, therefore, clearly a difference between offering an unsupported declaration as a justification for a judgement during an interview and selecting an unsupported declaration from a list of possible response options during a computerised task. It is possible that, during the interview, participants experienced a social pressure to successfully justify their judgement. This social pressure may also have made participants were more aware of the illegitimacy of using an unsupported declaration as a justification for their judgement. It is also possible that, seeing it written down as a possible answer legitimises selecting it as a justification for the judgement. The unsupported declaration does not provide an acceptable answer to the question on the critical slide, however, its presence in the list of possible response options may imply to participants that it is an acceptable answer, particularly if they do not put too much thought into it. By selecting the unsupported declaration participants can move quickly along to the next stage in the study without necessarily acknowledging any inconsistency in their reasoning, avoiding potentially dissonant cognitions (e.g., Case, Andrews, Johnson, & Allard, 2005; E. Harmon-Jones & Harmon-Jones, 2007; see also Heine, Proulx, & Vohs, 2006). Selecting the unsupported declaration may also allow the participant to proceed without expending effort trying to think of reasons for their judgement beyond the intuitive justifications that had already been de-bunked.

Rates of dumbfounded responding in Study 2 were higher than expected. Possible reasons for this could be (a) reduced social pressure to appear to have reasons for judgements; (b) a failure of participants to comprehend that the unsupported declaration does not provide a logically justifiable response to the question asked in the critical slide; (c) the apparent legitimising of the unsupported declaration by its inclusion in the list of possible response options; or (d) the selecting by participants of an "easy way out" option without thinking about it fully (through carelessness/laziness/eagerness to move on to a less taxing task). It appears that the selecting of unsupported declarations is not an accurate measure of dumbfounding. In Study 1, participants were only identified as dumbfounded based on the providing of an unsupported declaration if they subsequently failed to provide further reasons when the unsupported declaration was questioned. However, in some cases, participants who provided unsupported declarations were not identified as dumbfounded, based on subsequent responses. A follow up analysis of the interview data revealed that 23 participants provided an unsupported declaration and proceeded to provide reasons for at least one of their judgements; a further six participants provided an unsupported declaration and proceeded to revise their judgement at least once. A stricter measure of dumbfounding, one by which participants are required to explicitly acknowledge a state of dumbfoundedness is necessary to address the issues with the selecting of an unsupported declaration that may have led to the unusually high rates of dumbfounding observed in Study 2.

Study 3a: Revised Computerised Task – College sample

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

Method

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

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

Procedure and materials. The materials in this study were almost the same as in Study 2 with a change to the "dumbfounded" response option on the critical slide. Extra questions were included following each of the counter-arguments. On the critical slide, the unsupported declaration option was replaced with an admission of not having reasons ("It's wrong but I can't think of a reason"). Following each counter-argument, participants were asked if they (still) thought the behaviour was wrong, and if they had a reason for their judgement. There was also a revision to the question on the post-discussion questionnaire asking if participants had changed their judgements was changed: "did your judgement change?" with a binary "yes/no" response option reverted back to "how much did your judgement change?" with a seven point Likert scale response (as in Study 1). The same four dilemmas *Heinz*, *Incest*, *Cannibal* and *Trolley* (Appendix A) along with the same prepared counter-arguments (Appendix B) as in Study 2 were used in Study 3a. Both the MLQ (Steger et al., 2008); and CRSi7 (S. Huber & Huber, 2012) were also used. This study was conducted

in a designated psychology computer lab in MIC and was administered entirely on individual computers using OpenSesame (Mathôt et al., 2012).

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

Results and Discussion

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

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

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

Table 2 and **Figure 2** show the number and percentage of dumbfounded responses when the coded string responses are included in the analysis. When the coded string responses are included in the analysis, the number of participants displaying a dumbfounded response at least once increased from 40 (56%) to 57 (79%). Observed rates of dumbfounding increased for each scenario when the coded open-ended responses were included, with 19 participants (26.39%) appearing to be dumbfounded by *Heinz*, 21 (29.17%) by *Cannibal*, 31 (43.06%) by *Incest*, and 22 (30.56%) apparently dumbfounded by *Trolley*. Still, rates of dumbfounded responding did not vary with type of moral scenario (52 instances for intuition scenarios, 41 instances for reasoning scenarios) being discussed, $\chi^2(1, N = 288) = 1.59$, $p = .208$. There was no association between number of times dumbfounded and score on either measure on the Meaning and Life questionnaire; Presence $r(72) = 0.82$, $p = .413$, or Search, $r(72) = 0.07$, $p = .945$, or the Centrality of Religiosity Scale $r(72) = 1.29$, $p = .201$.

When the coded open-ended responses were included in the analysis, the proportion of participants displaying a dumbfounded response at least once in Study 3a (79%) was much closer to that observed in the interview in Study 1 (74%) than before the open-ended responses

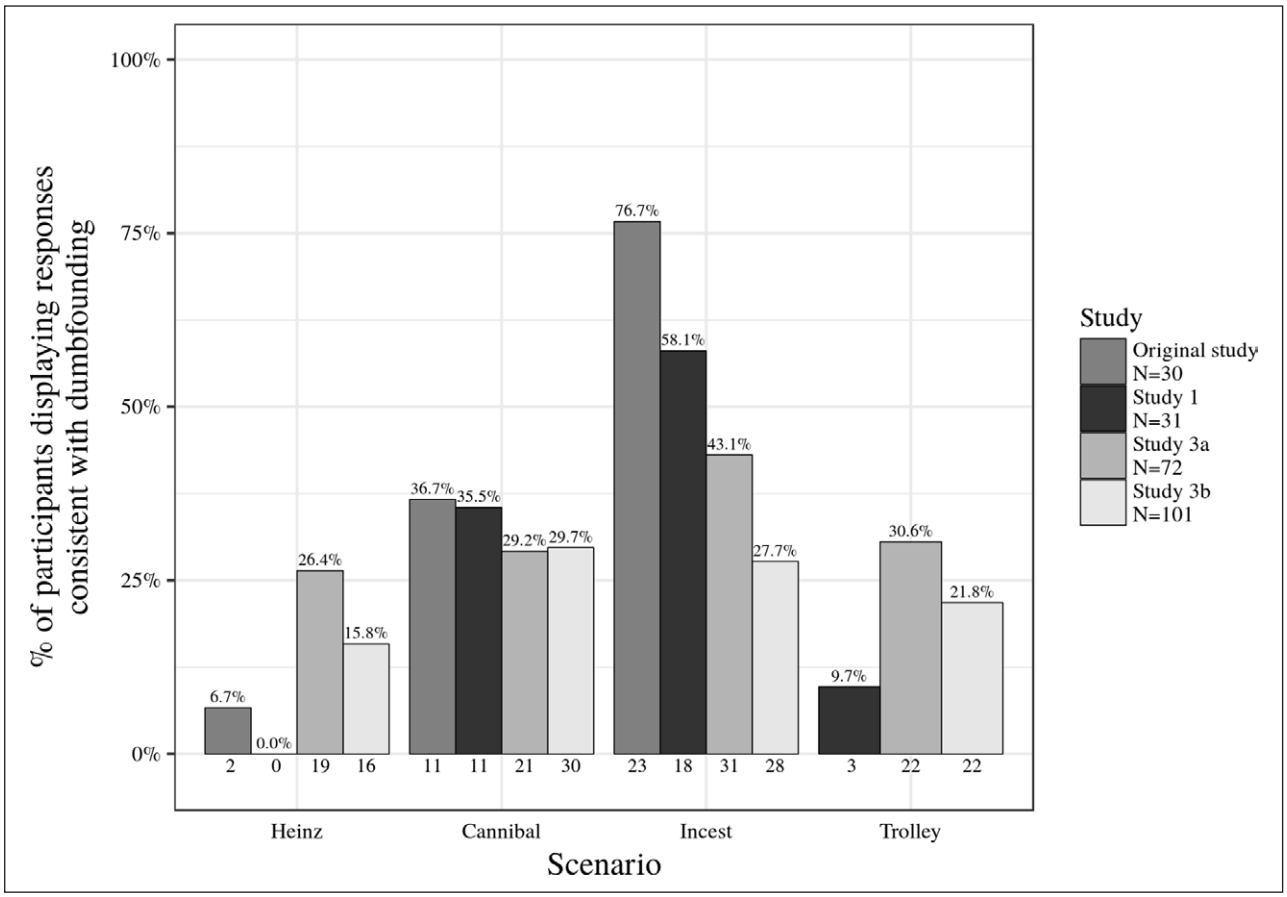

Figure 2: Rates of observed dumbfounding for each scenario across each study, including coded string responses.

were included (56%). The variation in observed rates of dumbfounding between dilemmas that was observed in the interview was not present in the computerised task. As such there remains a difference between the dumbfounding elicited during an interview and that elicited as part of a computerised task. However, it is clear that dumbfounded responses can be elicited as part of a computerised task. The participants in Studies 1, 2, and 3a were all college students (largely from the same institution) and as such, the following study investigated the phenomenon in a more diverse sample.

Study 3b: Revised Computerised Task – MTurk

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

Method

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

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

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

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

The critical slide contained a statement defending the behaviour and a question as to how the behaviour could be wrong, with three possible response options: (a) "There is nothing wrong"; (b) "It's wrong but I can't

think of a reason"; (c) "It's wrong and I can provide a valid reason". Participants who selected (c) were required to provide a reason. The order of these response options was randomised. When participants had completed all questions relating to all four dilemmas they completed the same longer questionnaire as in Studies 1 and 2 containing the Meaning and Life questionnaire (Steger et al., 2008), the Centrality of Religiosity Scale (S. Huber & Huber, 2012), and some questions relating to demographics. The entire study lasted approximately fifteen to twenty minutes.

Results and Discussion

Participants who selected the admission of not having reasons on the critical slide (option b) were identified as dumbfounded. **Table 1** shows the ratings of the behaviours across each scenario. **Table 2** and **Figure 1** show the percentage of participants displaying dumbfounded responses for each scenario. **Table 3** shows the responses to the questionnaires presented between scenario. On this occasion there was a statistically significant difference in responses to the critical slide depending on the order of scenario presentation, $\chi^2(6, N = 404) = 14.77$, $p = .022$. The observed rates of dumbfounded responses were higher for the third scenario, however they went down again for the fourth scenario along with rates of selecting "nothing wrong", meaning that the rates of participants providing reasons went up again for the fourth scenario. The higher rates of providing reasons observed for the fourth scenario presented means that this fluctuation is unlikely to be due to experimental fatigue, which was the primary reason for testing for order effects. There was also a difference in responses to the critical slide depending on scenario, $\chi^2(6, N = 404) = 15.18$, $p = .019$ with more people selecting "nothing wrong" for *Incest* and fewer people selecting "nothing wrong" for *Cannibal*. When dumbfounded responses are isolated and contrasted against other responses this difference is no longer present, $\chi^2(3, N = 404) = 1.86$, $p = .602$. Forty four participants (44%) selected the admission of not having reasons at least once. Twelve participants (11.88%) selected the admission of having no reasons for *Heinz*. Sixteen participants (15.84%) selected the admission of not having reasons for *Incest* and *Trolley*. Nineteen participants (18.81%) selected the admission of not having reasons for *Cannibal*.

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

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

Combined Results and Discussion

Evaluating each Measure of Dumbfounding

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

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

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

Providing an unsupported declaration is clearly different to selecting one from a list of possible responses. One possible explanation, is that dumbfounding is an aversive state, similar to experiencing a threat to meaning (Heine et al., 2006; Proulx & Inzlicht, 2012), or cognitive dissonance

(Cooper, 2007; Festinger, 1957; E. Harmon-Jones, & Harmon-Jones, 2007). The selecting of an unsupported declaration without deliberation allows participants to avoid or minimise the impact of this aversive state and move on. Providing an unsupported declaration involves more deliberation, making the illegitimacy of it more salient, reducing its effectiveness in avoiding the aversive state of dumbfoundedness. Furthermore, the relative attractiveness of these different responses to participants may be linked to social desirability (Chung, & Monroe, 2003; Latif, 2000; Morris, & McDonald, 2013). Follow-up work could investigate these questions directly.

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

The current research identified two measures of dumbfounding. Limitations are associated with each. Relying on admissions of having no reasons only, provides an overly strict measure whereby a failure to provide reasons is not measured. Taking unsupported declarations (and tautological reasons) as a measure of dumbfounding may provide too broad a measure, risks identifying lazy or inattentive participants as dumbfounded. The providing of a type-written response as part of a computerised task requires effort, and the majority of participants avoid the use of unsupported declarations as justifications for their judgements. This suggests that those who provided unsupported declarations did so because they failed to identify alternative reason. It appears that the most practicable means to measure dumbfounding accurately requires each of the responses: providing/selecting admissions of not having reasons, and the providing of an unsupported declaration, to be accounted for. Participants providing either of these responses may be identified as dumbfounded.

Differences between Scenarios

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

Table 2 and **Figure 1** show the initial observed rates of dumbfounding for each study. **Table 2** and **Figure 2** show the revised rates of observed dumbfound responding in each study once the open-ended coded responses from Studies 3a and 3b are included. Rates of dumbfounding reported by Haidt et al. (2000) are also included for comparison. Study 2 was a primarily a pilot study, and, as discussed, the observed rates of dumbfounding do not appear to be representative of the phenomenon being studied, as such Study 2 is not included in **Figure 2**.

Differences between the Samples

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

Incest recorded higher rates of dumbfounding than the other scenarios in both interview studies (Study 1 and Haidt et al., 2000) and, to some degree, in Study 3a, the computer task with a college sample. The rate of dumbfounding observed for *Incest* with the online sample, in Study 3b, is lower than that observed with the college sample in Study 3a and is also slightly lower than that observed for *Cannibal* in the online sample. This is surprising, in that, the *Incest* dilemma is the most commonly cited example (e.g., Haidt, 2001; Prinz, 2005; Royzman et al., 2015), and, in Studies 1, 2, and 3a, is the most reliable for eliciting dumbfounding, consistently eliciting higher rates than the other dilemmas. Looking at the ratings of the behaviours in each dilemma for each study may provide some clue as to where this variation comes from. The online sample were less inclined to rate the behaviour in *Incest* as wrong relative to the participants in the other studies. The percentage of participants initially

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%

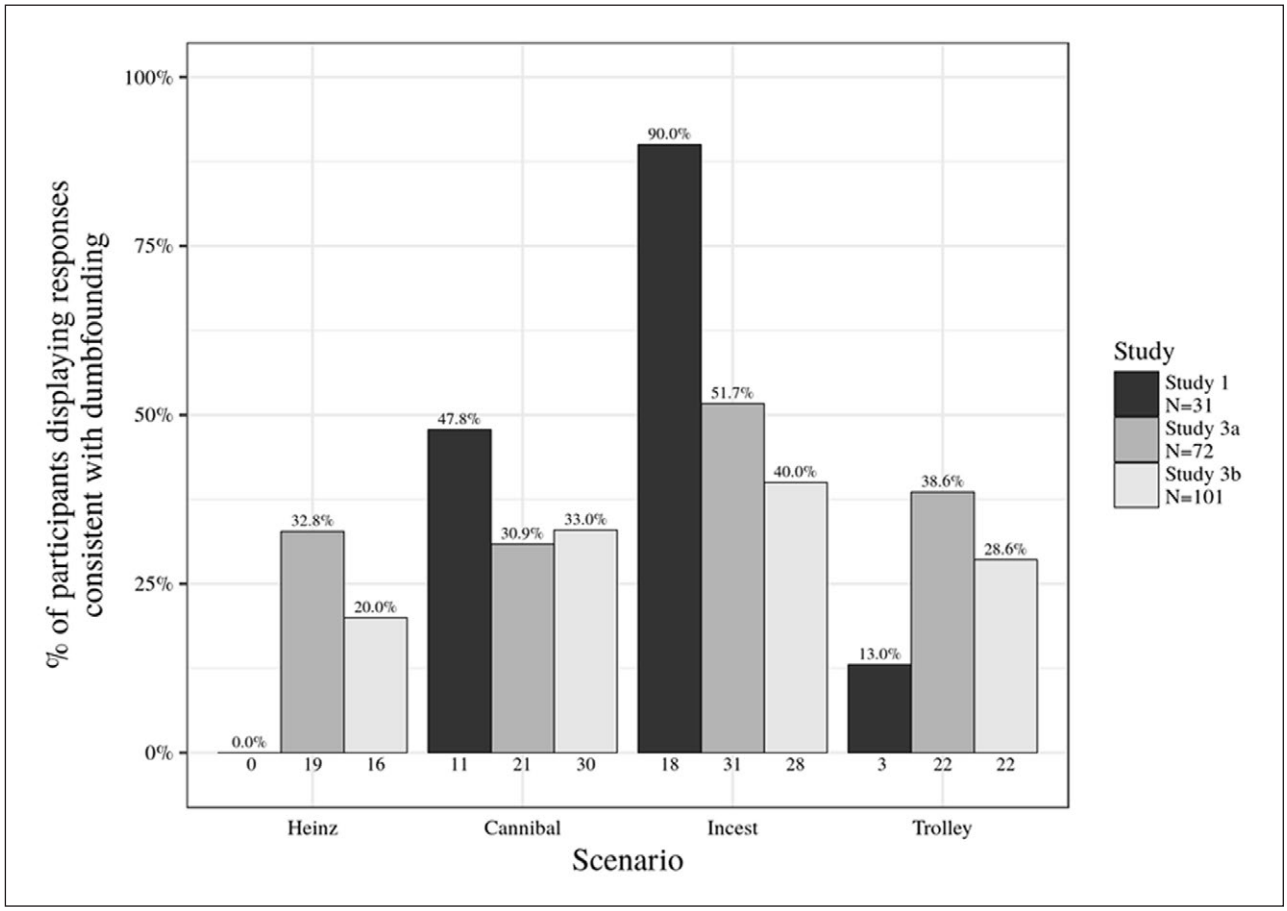

Figure 3: Percentage of dumbfounded responses when "nothing wrong" is excluded.

rating *Incest* as wrong for each study are as follows: Study 1: 83.87%; Study 2: 87.50%; Study 3a: 84.72%; Study 3b: 70.30%. Furthermore, on the critical slide, the proportion of participants who selected "nothing wrong" for *Incest* for Study 3b (30.69%; 31 participants) was nearly double the proportion that selected "nothing wrong" for *Incest* for Study 3a (16.67; 12 participants). When these participants are excluded from the analysis of Study 3b (see **Table 4** and **Figure 3**), the percentage of participants appearing to be dumbfounded by *Incest* (22.86%; 16 participants; or 40%; 28 participants when open-ended responses are included; $N = 70$) exceeds the percentage of participants appearing to be dumbfounded by *Cannibal* (20.88%; 19 participants; or 32.97%; 30 participants when open-ended responses are included; $N = 91$). As such, it appears that the apparent uncharacteristically low rates of observed dumbfounding for *Incest* in Study 3b, when compared to *Cannibal*, may be due to the online sample being less inclined to rate the behaviour as morally wrong rather than a difference in this sample's ability to provide justifications for their judgements to the two scenarios.

It has been argued that moral dumbfounding occurs as a result of social pressure to conform to conversational norms (Royzman et al., 2015). The findings presented by Royzman et al. (2015) do not fully support this claim, however, they demonstrate that incidences of moral dumbfounding are sensitive to social pressure. Studies 2 and 3, aimed to reduce the influence of social pressure by testing dumbfounding as part of a computerised task, as opposed to in an interview setting. The varying rates of dumbfounding depending on task type indicate that the computerised task is different from the interview.

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

The mean responses for the self-report question "How much did you change your mind?" are as follows: Study 1, $M = 2.88$, $SD = 1.59$; Study 3a, $M = 2.01$, $SD = 1.46$; Study 3b, $M = 1.69$, $SD = 1.27$. A one-way ANOVA revealed significant differences in responses to this question between the different studies, F (2, 809) = 33.81, $p < .001$, partial $\eta^2 = .077$. Tukey's post-hoc pairwise comparison revealed that responses in Study 1 were significantly higher than both Study 3a, $p < .001$, and Study 3b, $p < .001$. The responses in Study 3a were also significantly higher than the responses in Study 3b, $p = .008$.

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

General Discussion

The goal of this research was to examine the replicability of dumbfounded responding following a moral judgement task, and identify specific measurable responses that may be viewed as indicators of moral dumbfounding. Four studies, with a combined total sample of $N = 276$, were conducted in an attempt to replicate and extend the original demonstration ($N = 30$) of moral dumbfounding by Haidt et al. (2000). We predicted that dumbfounded responses would be evoked when participants were required to provide justification for their moral judgements, when their basic intuitive justifications had been refuted. Two measures of moral dumbfounding were taken, an explicit acknowledgement of the absence of reasons, and a failure to provide reasons when pushed. Rates of observed dumbfounding vary depending on which measure is being employed.

Intuition versus Reasoning

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

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

It is possible that there exists a social influence in an interview setting that changes the way participants respond (e.g., Asch, 1956; Sabini, 1995; Staub, 2013) and, that the interviewer may be seen as a person in authority, demanding justifications for judgements made (e.g., Milgram, 1974). This may motivate participants to identify reasons to justify their judgements, leading to the suppression of dumbfounded responses. On the other hand, it may also motivate participants to heed the counter-arguments offered by the experimenter. This may lead to an interaction between scenario difficulty and social pressure to emerge, with the social pressure leading to fewer dumbfounded responses to the easier "reasoning" scenarios, but leading to more dumbfounded responses to the more difficult "intuition" scenarios. It may be the case that the rates of dumbfounding found in the computer tasks provide something of a crude baseline measure of participants' initial perception of their own ability to justify their judgement of the scenario, having read the scenario and a number of counter-arguments. In the interview, these initial responses to the scenarios are distilled by the discussion with the experimenter to reflect the variation in difficulty between the scenarios.

Implications

The existence of moral dumbfounding has informed various theories of moral judgement either directly (e.g., Cushman et al., 2010; Haidt, 2001; Hauser et al., 2008; Prinz, 2005) or indirectly (Crockett, 2013; Cushman, 2013; Greene, 2008, 2013). The original demonstration of moral dumbfounding remains unpublished in peer reviewed form (Haidt et al., 2000) and has not been directly replicated. The studies presented here aimed to replicate and extend this original moral dumbfounding study

(Haidt et al., 2000) and thus, assess the notion that moral dumbfounding is in fact a psychological phenomenon that can be consistently observed. Study 1 successfully replicated the original study. Study 2 piloted the use of a computer task and recorded unexpectedly high rates of dumbfounded responding. Possible reasons for this were identified and addressed in Studies 3a and 3b. Study 3a and 3b recorded more moderate rates of dumbfounding with two different samples. All three studies successfully elicited dumbfounded responding identified as (a) admissions of not having reasons; (b) use of unsupported declarations as justification of a judgement; or (c) use of undefended tautological response as justification for a judgement; however, differences remain between the interview in Study 1 and the computerised task in Studies 3a and 3b. Taking these responses to be indicators of a state of dumbfoundedness, it appears that moral dumbfounding can be evoked in face-to-face and online contexts. As such, the research presented here may be seen as more support for the existence of intuitionist theories of moral judgement (e.g., Cushman et al., 2010; Greene, 2008; Haidt, 2001; Hauser et al., 2008; Prinz, 2005) over rationalist theories (e.g., Kohlberg, 1971; Topolski et al., 2013).

Responding to Criticisms

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

Firstly, the initial estimate of incidences of dumbfounding was 4/53 (7.55%). Based on the same calculations used by Royzman et al. (2015), this estimate of 4/53 is significantly greater than 0/53, $z = 2.04$, $p = .041$. These four participants were then interviewed further, during which, the "inconsistencies" in participants' "responses were pointed out directly" (Royzman et al., 2015, p. 308). Following this interview, Royzman et al. were left with a dumbfounding estimate of 1/53 (which they claim is not significantly greater than 0/53).

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

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

Thirdly, and most importantly, the current research identifies dumbfounding as a rare demonstration of the separation between intuitions and reasons for these intuitions. Practical challenges to demonstrating this separation have already been identified: (a) post-hoc rationalisation and identification of reasons that are consistent with a judgement; (b) the possibility that the intuition emerged as a result of a well-rehearsed reasoned response. The work presented by Royzman et al. (2015) may be viewed as a practical demonstration of this first challenge; helping participants identify reasons that are consistent with their judgement and providing an opportunity them to endorse these reasons.

As previously noted, the endorsing of a reason does not imply that the reason contributed to the judgement. This view of moral dumbfounding presents two methodological considerations that need to be addressed before accepting the claim that judgements in the dumbfounding paradigm can be attributed to either norm-based reasons or harm-based reasons. The

first relates to participants' ability to articulate either harm-based or norm-based reasons. The second relates to the consistency with which these reasons guide judgements.

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

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

When John mentioned the surfing competition Pat immediately adopted it as another reason not to go for a swim however it is clear that this reason played no part in Pat's original judgement. It is possible that in identifying other reasons that are consistent with a particular judgement researchers may falsely attribute the judgement made to these reasons. The studies described by Royzman et al. (2015) do not sufficiently guard against the possibility of falsely attributing judgements to reasons endorsed, allowing for the possibility that some participants were falsely excluded from analysis. One way to avoid the false exclusion of participants would be to include an open-ended string response option immediately after the presenting of the vignette, in which participants are invited to provide the reason(s) for their judgement. Participants are then only excluded from analysis if they both articulated and endorsed a given principle.

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

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

Limitations and Future Directions

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

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

Conclusion

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

Data accessibility statement

All participant data, and analysis scripts can be found on

this paper's project page on the Open Science Framework at https://osf.io/wm6vc/.

Additional Files

The Additional files for this article can be found as follows:

- **Appendix A.** Moral Scenarios. DOI: https://doi.org/10.1525/collabra.79.s1
- **Appendix B.** Sample Statements to Challenge Judgements. DOI: https://doi.org/10.1525/collabra.79.s2
- **Appendix C.** Post Discussion Duestionnaire. DOI: https://doi.org/10.1525/collabra.79.s3

Notes

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

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

[3] These are largely theoretical arguments offering explanations of dumbfounding that are consistent with a rationalist perspective (e.g., Kohlberg, 1971; Topolski, Weaver, Martin, & McCoy, 2013). However Royzman, Kim, and Leeman (2015) present some empirical evidence in support of this position. This is examined, in detail, in the Introduction and in the Discussion.

[4] R (3.4.1, R Core Team, 2017b) and the R-packages *afex* (0.15.2, Singmann, Bolker, & Westfall, 2015), *car* (2.1.5, Fox, & Weisberg, 2011), *citr* (0.2.0.9047, Aust, 2016), *desnum* (0.1.1, McHugh, 2017), *devtools* (1.13.2, Wickham, & Chang, 2017), *estimability* (1.2, R. Lenth, 2016), *extrafont* (0.17, Chang, 2014), *foreign* (0.8.69, R Core Team, 2017a), *ggplot2* (2.2.1, Wickham, 2009), *lme4* (1.1.13, Bates, Mächler, Bolker, & Walker, 2015), *lsmeans* (2.26.3, R. V. Lenth, 2016), *Matrix* (1.2.10, Bates, & Maechler, 2017), *papaja* (0.1.0.9492, Aust, & Barth, 2017), *plyr* (1.8.4, Wickham, 2011), *reshape2* (1.4.2, Wickham, 2007), *scales* (0.4.1, Wickham, 2016), *shiny* (1.0.3, Chang, Cheng, Allaire, Xie, & McPherson, 2017), and *wordcountaddin* (0.2.0, Marwick, n.d.).

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

Competing Interests

The authors have no competing interests to declare.

Author contributions

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

Author Information

All procedures performed in studies involving human participants were approved by institutional research ethics committee and conducted in accordance with the Code of Professional Ethics of the Psychological Society of Ireland, and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards. Informed consent was obtained from all individual participants included in the study.

References

Amazon Web Services Inc. (2016). Amazon Mechanical Turk.

Asch, S. E. (1956). Studies of independence and submission to group pressures. *Psychological Monographs*, *70*, 416.

Aust, F. (2016). *Citr: 'RStudio' Add-in to Insert Markdown Citations*.

Aust, F., & **Barth, M.** (2017). *Papaja: Create APA manuscripts with R Markdown*.

Barsalou, L. W. (2003). Situated simulation in the human conceptual system. *Language and Cognitive Processes*, *18*(5–6), 513–562. DOI: https://doi.org/10.1080/01690960344000026

Barsalou, L. W. (2008). Grounded Cognition. *Annual Review of Psychology*, *59*(1), 617–645. DOI: https://doi.org/10.1146/annurev.psych.59.103006.093639

Barsalou, L. W. (2009). Simulation, situated conceptualization, and prediction. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *364*(1521), 1281–1289. DOI: https://doi.org/10.1098/rstb.2008.0319

Bates, D., & **Maechler, M.** (2017). *Matrix: Sparse and Dense Matrix Classes and Methods*.

Bates, D., Mächler, M., Bolker, B., & **Walker, S.** (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, *67*(1), 1–48. DOI: https://doi.org/10.18637/jss.v067.i01

Bellin, Z. (2012). The quest to capture personal meaning in psychology. *International Journal of Existential Psychology and Psychotherapy*, *4*(1), 27.

Berry, D., & **Dienes, Z. P.** (1993). *Implicit Learning: Theoretical and Empirical Issues*. Psychology Press.

Cameron, C. D., Payne, B. K., & **Doris, J. M.** (2013). Morality in high definition: Emotion differentiation calibrates the influence of incidental disgust on moral judgments. *Journal of Experimental Social Psychology*, *49*(4), 719–725. DOI: https://doi.org/10.1016/j.jesp.2013.02.014

Case, D. O., Andrews, J. E., Johnson, J. D., & **Allard, S. L.** (2005). Avoiding Versus Seeking: The Relationship of Information Seeking to Avoidance, Blunting, Coping, Dissonance, and Related Concepts. *Journal of the Medical Library Association: JMLA*, *93*(3), 353–362.

Chaiken, S. (1980). Heuristic versus systematic information processing and the use of source versus message cues in persuasion. *Journal of Personality and Social Psychology*, *39*(5), 752–766. DOI: https://doi.org/10.1037/0022-3514.39.5.752

Chaiken, S., & **Trope, Y.** (1999). *Dual-process Theories in Social Psychology*. Guilford Press.

Chang, W. (2014). *Extrafont: Tools for using fonts*.

Chang, W., Cheng, J., Allaire, J. J., Xie, Y., & **McPherson, J.** (2017). *Shiny: Web Application Framework for R*.

Chung, J., & **Monroe, G. S.** (2003). Exploring Social Desirability Bias. *Journal of Business Ethics*, *44* (4), 291–302. DOI: https://doi.org/10.1023/A:1023648703356

Cooper, J. (2007). *Cognitive dissonance: Fifty years of a classic theory*, *xi*. Thousand Oaks, CA: Sage Publications Ltd.

Crockett, M. J. (2013). Models of morality. *Trends in Cognitive Sciences*, *17*(8), 363–366. DOI: https://doi.org/10.1016/j.tics.2013.06.005

Cushman, F. A. (2013). Action, Outcome, and Value A Dual-System Framework for Morality. *Personality and Social Psychology Review*, *17*(3), 273–292. DOI: https://doi.org/10.1177/1088868313495594

Cushman, F. A., Young, L., & **Greene, J. D.** (2010). Multi-system Moral Psychology. In: Doris, J. M., & Cushman, F. A. (Eds.), *The Moral Psychology Handbook*, 47–71. Oxford; New York: Oxford University Press. DOI: https://doi.org/10.1093/acprof:oso/9780199582143.003.0003

Cushman, F. A., Young, L., & **Hauser, M. D.** (2006). The Role of Conscious Reasoning and Intuition in Moral Judgment Testing Three Principles of Harm. *Psychological Science*, *17*(12), 1082–1089. DOI: https://doi.org/10.1111/j.1467-9280.2006.01834.x

Dreyfus, H. L., & **Dreyfus, S. E.** (1990). What is moral maturity? A phenomenological account of the development of ethical expertise. *Universalism Vs. Communitarianism*, 237–264.

Dwyer, S. (2009). Moral Dumbfounding and the Linguistic Analogy: Methodological Implications for the Study of Moral Judgment. *Mind & Language*, *24*(3), 274–296. DOI: https://doi.org/10.1111/j.1468-0017.2009.01363.x

Epstein, S. (1994). Integration of the cognitive and the psychodynamic unconscious. *American Psychologist*, *49*(8), 709–724. DOI: https://doi.org/10.1037/0003-066X.49.8.709

Evans, J. S. B. T. (2003). In two minds: Dual-process accounts of reasoning. *Trends in Cognitive Sciences*, *7*(10), 454–459. DOI: https://doi.org/10.1016/j.tics.2003.08.012

Evans, J. S. B. T., & **Wason, P. C.** (1976). Rationalization in a Reasoning Task. *British Journal of Psychology*, *67*(4), 479–486. DOI: https://doi.org/10.1111/j.2044-8295.1976.tb01536.x

Festinger, L. (1957). *A theory of cognitive dissonance*. Stanford CA: Stanford University Press.

Fox, J., & **Weisberg, S.** (2011). *An R Companion to Applied Regression* (Second). Thousand Oaks CA: Sage.

Friard, O., & **Gamba, M.** (2015, December). BORIS – Behavioral Observation Research Interactive Software. Italy.

Gazzaniga, M. S., & **LeDoux, J. E.** (2013). *The Integrated Mind*. Springer Science & Business Media.

Gray, K., Schein, C., & **Ward, A. F.** (2014). The myth of harmless wrongs in moral cognition: Automatic dyadic completion from sin to suffering. *Journal of Experimental Psychology: General*, *143*(4), 1600–1615. DOI: https://doi.org/10.1037/a0036149

Greene, J. D. (2008). The Secret Joke of Kant's Soul. In: *Moral Psychology: The neurosciences of morality: Emotion, brain disorders, and development, 3*, 35–79. Cambridge (Mass.): the MIT press.

Greene, J. D. (2013). *Moral tribes: Emotion, reason, and the gap between us and them*.

Greene, J. D., Sommerville, R. B., Nystrom, L. E., Darley, J. M., & **Cohen, J. D.** (2001). An fMRI investigation of emotional engagement in moral judgment. *Science (New York, N.Y.)*, *293*(5537), 2105–2108. DOI: https://doi.org/10.1126/science.1062872

Haidt, J. (2001). The emotional dog and its rational tail: A social intuitionist approach to moral judgment. *Psychological Review*, *108*(4), 814–834. DOI: https://doi.org/10.1037/0033-295X.108.4.814

Haidt, J. (2007). The New Synthesis in Moral Psychology. *Science*, *316*(5827), 998–1002. DOI: https://doi.org/10.1126/science.1137651

Haidt, J., & **Björklund, F.** (2008). Social Intuitionists Answer Six Questions about Moral Psychology. In: Sinnott-Armstrong, W. (Ed.), *Moral psychology Volume 2: The cognitive science of morality: Intuition and diversity*, 181–217. London: MIT.

Haidt, J., Björklund, F., & **Murphy, S.** (2000). Moral dumbfounding: When intuition finds no reason. *Unpublished Manuscript*. University of Virginia.

Haidt, J., & **Hersh, M. A.** (2001). Sexual Morality: The Cultures and Emotions of Conservatives and Liberals. *Journal of Applied Social Psychology*, *31*(1), 191–221. DOI: https://doi.org/10.1111/j.1559-1816.2001.tb02489.x

Haidt, J., Koller, S. H., & **Dias, M. G.** (1993). Affect, culture, and morality, or is it wrong to eat your dog? *Journal of Personality and Social Psychology*, *65*(4), 613–628. DOI: https://doi.org/10.1037/0022-3514.65.4.613

Harmon-Jones, E., & **Harmon-Jones, C.** (2007). Cognitive Dissonance Theory After 50 Years of Development. *Zeitschrift Für Sozialpsychologie*, *38*(1), 7–16. DOI: https://doi.org/10.1024/0044-3514.38.1.7

Hauser, M. D., Cushman, F. A., Young, L., Kang-Xing Jin, R., & **Mikhail, J.** (2007). A Dissociation Between Moral Judgments and Justifications. *Mind & Language*, *22*(1), 1–21. DOI: https://doi.org/10.1111/j.1468-0017.2006.00297.x

Hauser, M. D., Young, L., & **Cushman, F. A.** (2008). Reviving Rawls's Linguistic Analogy: Operative Principles and the Causal Structure of Moral Actions. In: Sinnott-Armstrong, W. (Ed.), *Moral psychology Volume 2: The cognitive science of morality: Intuition and diversity,* 107–155. London: MIT.

Heine, S. J., Proulx, T., & **Vohs, K. D.** (2006). The Meaning Maintenance Model: On the Coherence of Social Motivations. *Personality and Social Psychology Review*, *10*(2), 88–110. DOI: https://doi.org/10.1207/s15327957pspr1002_1

Huber, S., & **Huber, O. W.** (2012). The Centrality of Religiosity Scale (CRS). *Religions*, *3*(3), 710–724. DOI: https://doi.org/10.3390/rel3030710

IBM Corp. (2015). SPSS. Armonk, NY: IBM Corp.

Jacobson, D. (2012). Moral Dumbfounding and Moral Stupefaction. In: *Oxford studies in normative ethics*, *2*, 289. DOI: https://doi.org/10.1093/acprof:oso/9780199662951.003.0012

Jacoby, L. L. (1983). Remembering the data: Analyzing interactive processes in reading. *Journal of Verbal Learning and Verbal Behavior*, *22*(5), 485–508. DOI: https://doi.org/10.1016/S0022-5371(83)90301-8

Johansson, P., Hall, L., Sikström, S., & **Olsson, A.** (2005). Failure to Detect Mismatches Between Intention and Outcome in a Simple Decision Task. *Science*, *310*(5745), 116–119. DOI: https://doi.org/10.1126/science.1111709

Kahneman, D. (2011). *Thinking, fast and slow*. London: Allen Lane.

Kohlberg, L. (1971). *From is to Ought: How to Commit the Naturalistic Fallacy and Get Away with it in the Study of Moral Development*.

Kruglanski, A. W. (2013). *The Psychology of Closed Mindedness*. Psychology Press.

Kruglanski, A. W., & **Webster, D. M.** (1996). Motivated closing of the mind: "Seizing" and "freezing." *Psychological Review*, *103*(2), 263–283. DOI: https://doi.org/10.1037/0033-295X.103.2.263

Latif, D. A. (2000). The Link Between Moral Reasoning Scores, Social Desirability, and Patient Care Performance Scores: Empirical Evidence from the Retail Pharmacy Setting. *Journal of Business Ethics*, *25*(3), 255–269. DOI: https://doi.org/10.1023/A:1006049605298

Lenth, R. (2016). *Estimability: Tools for Assessing Estimability of Linear Predictions*.

Lenth, R. V. (2016). Least-Squares Means: The R Package lsmeans. *Journal of Statistical Software*, *69*(1), 1–33. DOI: https://doi.org/10.18637/jss.v069.i01

Mallon, R., & **Nichols, S.** (2011). Dual Processes and Moral Rules. *Emotion Review*, *3*(3), 284–285. DOI: https://doi.org/10.1177/1754073911402376

Marwick, B. (n.d.). *Wordcountaddin: Word counts and readability statistics in R markdown documents*.

Mathôt, S., Schreij, D., & **Theeuwes, J.** (2012). OpenSesame: An open-source, graphical experiment builder for the social sciences. *Behavior Research Methods*, *44*(2), 314–324. DOI: https://doi.org/10.3758/s13428-011-0168-7

McGregor, I. (2006a). Offensive Defensiveness: Toward an Integrative Neuroscience of Compensatory Zeal After Mortality Salience, Personal Uncertainty, and Other Poignant Self-Threats. *Psychological Inquiry*, *17*(4), 299–308. DOI: https://doi.org/10.1080/10478400701366977

McGregor, I. (2006b). Zeal Appeal: The Allure of Moral Extremes. *Basic and Applied Social Psychology*, *28*(4), 343–348. DOI: https://doi.org/10.1207/s15324834basp2804_7

McGregor, I., Nash, K. A., & **Prentice, M.** (2012). Religious zeal after goal frustration. In: Hogg, M. A., & Baylock, D. L. (Eds.), *Extremism and the Psychology of Uncertainty*, 147–164. Hoboken NJ: Wiley-Blackwell.

McGregor, I., Zanna, M. P., Holmes, J. G., & **Spencer, S. J.** (2001). Compensatory conviction in the face of personal uncertainty: Going to extremes and being oneself. *Journal of Personality and Social Psychology*, *80*(3), 472–488. DOI: https://doi.org/10.1037/0022-3514.80.3.472

McHugh, C. (2017). *Desnum: Creates some useful functions*.

Milgram, S. (1974). *Obedience to Authority: An Experimental View*. New York: Harper and Row.

Morris, S. A., & **McDonald, R. A.** (2013). The Role of Moral Intensity in Moral Judgments: An Empirical Investigation. In: Michalos, A. C., & Poff, D. C. (Eds.), *Citation Classics from the Journal of Business Ethics*, 463–479. Springer Netherlands. DOI: https://doi.org/10.1007/978-94-007-4126-3_23

Narvaez, D. (2005). The neo-Kohlbergian tradition and beyond: Schemas, expertise, and character. In: Carlo, G. & Pope-Edwards, C. (Eds.), *Nebraska symposium on motivation*, *51*, 119.

Nisbett, R. E., & **Wilson, T. D.** (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review*, *84*(3), 231. DOI: https://doi.org/10.1037/0033-295X.84.3.231

Pizarro, D. A., & **Bloom, P.** (2003). The intelligence of the moral intuitions: A comment on Haidt (2001). *Psychological Review*, *110*(1), 193–196. DOI: https://doi.org/10.1037/0033-295X.110.1.193

Prinz, J. J. (2005). Passionate Thoughts: The Emotional Embodiment of Moral Concepts. In: Pecher, D., & Zwaan, R. A. (Eds.), *Grounding Cognition: The Role of Perception and Action in Memory, Language, and Thinking*, 93–114. Cambridge University Press. DOI: https://doi.org/10.1017/CBO9780511499968.005

Proulx, T., & **Inzlicht, M.** (2012). The Five "A"s of Meaning Maintenance: Finding Meaning in the Theories of Sense-Making. *Psychological Inquiry*, *23*(4), 317–335. DOI: https://doi.org/10.1080/1047840X.2012.702372

R Core Team. (2017a). *Foreign: Read Data Stored by Minitab, S, SAS, SPSS, Stata, Systat, Weka, dBase.*

R Core Team. (2017b). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.

Reber, A. S. (1989). Implicit learning and tacit knowledge. *Journal of Experimental Psychology: General*, *118*(3), 219–235. DOI: https://doi.org/10.1037/0096-3445.118.3.219

Royzman, E. B., Kim, K., & **Leeman, R. F.** (2015). The curious tale of Julie and Mark: Unraveling the moral dumbfounding effect. *Judgment and Decision Making*, *10*(4), 296–313.

Sabini, J. (1995). *Social psychology*. New York; London: Norton.

Schnell, T. (2011). Individual differences in meaning-making: Considering the variety of sources of meaning, their density and diversity. *Personality and Individual Differences*, *51*(5), 667–673. DOI: https://doi.org/10.1016/j.paid.2011.06.006

Singmann, H., Bolker, B., & **Westfall, J.** (2015). *Afex: Analysis of Factorial Experiments*.

Sneddon, A. (2007). A Social Model of Moral Dumbfounding: Implications for Studying Moral Reasoning and Moral Judgment. *Philosophical Psychology*, *20*(6), 731–748. DOI: https://doi.org/10.1080/09515080701694110

Staub, E. (2013). *Positive Social Behavior and Morality: Social and Personal Influences*. Elsevier.

Steger, M. F., Kashdan, T. B., Sullivan, B. A., & **Lorentz, D.** (2008). Understanding the Search for Meaning in Life: Personality, Cognitive Style, and the Dynamic Between Seeking and Experiencing Meaning. *Journal of Personality*, *76*(2), 199–228. DOI: https://doi.org/10.1111/j.1467-6494.2007.00484.x

Sun, R., Slusarz, P., & **Terry, C.** (2005). The Interaction of the Explicit and the Implicit in Skill Learning: A Dual-Process Approach. *Psychological Review*, *112*(1), 159–192. DOI: https://doi.org/10.1037/0033-295X.112.1.159

Topolski, R., Weaver, J. N., Martin, Z., & **McCoy, J.** (2013). Choosing between the emotional dog and the rational pal: A moral dilemma with a tail. *Anthrozoös*, *26*(2), 253–263. DOI: https://doi.org/10.2752/175303713X13636846944321

Unipark, Q. (2013). QuestBack Unipark.

Wickham, H. (2007). Reshaping Data with the reshape Package. *Journal of Statistical Software*, *21*(12), 1–20. DOI: https://doi.org/10.18637/jss.v021.i12

Wickham, H. (2009). *Ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. DOI: https://doi.org/10.1007/978-0-387-98141-3

Wickham, H. (2011). The Split-Apply-Combine Strategy for Data Analysis. *Journal of Statistical Software*, *40*(1), 1–29. DOI: https://doi.org/10.18637/jss.v040.i01

Wickham, H. (2016). *Scales: Scale Functions for Visualization*.

Wickham, H., & **Chang, W.** (2017). *Devtools: Tools to Make Developing R Packages Easier*.

Wielenberg, E. J. (2014). *Robust Ethics: The Metaphysics and Epistemology of Godless Normative Realism*. OUP Oxford. DOI: https://doi.org/10.1093/acprof:oso/9780198714323.001.0001

Wilson, T. D., & **Bar-Anan, Y.** (2008). The unseen mind. *Science*, *321*(5892), 1046–1047. DOI: https://doi.org/10.1126/science.1163029

Zajonc, R. B. (1980). Feeling and thinking: Preferences need no inferences. *American Psychologist*, *35*(2), 151–175. DOI: https://doi.org/10.1037/0003-066X.35.2.151

Peer review comments
The author(s) of this paper chose the Open Review option, and the peer review comments are available at: http://doi.org/10.1525/collabra.79.pr

How to cite this article: McHugh, C., McGann, M., Igou, E. R. and Kinsella, E. L. (2017). Searching for Moral Dumbfounding: Identifying Measurable Indicators of Moral Dumbfounding. *Collabra: Psychology,* 3(1): 23, DOI: https://doi.org/10.1525/collabra.79

Submitted: 02 February 2017 **Accepted:** 31 July 2017 **Published:** 04 October 2017