Running head: **MORAL BAYESIANISM**

**Empirical Evidence for Moral Bayesianism**

Haim Cohen,1,2\* Ittay Nissan-Rozen,3\* Anat Maril1,4

**1** The Hebrew University of Jerusalem,

Department of Cognitive Science,

Mt. Scopus, Jerusalem, 91905 Israel.

2 The Federmann Center for the Study of Rationality,

Givat Ram, Jerusalem, 91904 Israel.

3 The Hebrew University of Jerusalem,

Department of Philosophy,

Mt. Scopus, Jerusalem, 9190501 Israel.

4 The Hebrew University of Jerusalem,

Department of Psychology,

Mt. Scopus, Jerusalem, 91905 Israel.

\* Equal contribution

Corresponding Author:

Haim Cohen

haim.cohen3@mail.huji.ac.il

Phone 972-546- 359838

**Abstract**

Many philosophers in the field of meta-ethics believe that rational degrees of confidence in moral judgements should have a probabilistic structure, in the same way as do rational degrees of belief. The current paper examines this position, termed “moral Bayesianism,” from an empirical point of view. To this end, we assessed the extent to which degrees of moral judgments obey the third axiom of the probability calculus, , known as finite additivity, as compared to degrees of beliefs on the one hand and degrees of desires on the other. Results generally converged to show that degrees of moral judgment are more similar to degrees of belief than to degrees of desire in this respect. This supports the adoption of a Bayesian approach to the study of moral judgements. To further support moral Bayesianism, we also demonstrated its predictive power. Finally, we discuss the relevancy of our results to the meta-ethical debate between moral cognitivists and moral non-cognitivists.

*Keywords:* *moral judgements, meta-ethics, degrees of belief, Moral Bayesianism, desire as belief thesis, moral cognitivism, moral non-cognitivism*

1. **Introduction**

Intuitively, moral judgements come in degrees of confidence. For example, one can judge a certain act to be right or wrong with more or less confidence. These degrees of confidence of moral judgements differ, however, from the degrees of wrongness or rightness attributed to the acts, or the states of affair, that are the objects of the moral judgements. For example, many people are equally confident in their judgements that it is wrong to lie to a friend for personal gain and that it is wrong to kill an innocent person, but still attach a higher degree of wrongness to the latter act.

Following Michael Smith (2002), we will call the former type of degree “certitude,” and the latter “importance.”[[1]](#endnote-1) Smith, famously, used the above observation to argue against moral non-cognitivism, the thesis that moral judgements do not express beliefs or belief-like attitudes, but rather a different type of attitude.[[2]](#endnote-2) Moral cognitivists take moral judgements to express beliefs in propositions with moral content, and so, straightforwardly, can understand certitude as degrees of beliefs in such propositions and importance as degrees of wrongness/rightness of the acts, or states of affair, which are the subject matter of these propositions. Smith presented some evidence in favor of the claim that only moral cognitivists have the conceptual resources to account for the two different types of degrees associated with moral judgements.

In response, some philosophers (for example Lenman 2003; Ridge, 2007; Sepielli, 2012; Eriksson & Olinder, 2016; Makins, 2021) suggested alternative non-cognitivist ways to account for the two types of degrees, without understanding any of them as degrees of belief in propositions with moral content. Some philosophers (such as Ridge, 2007; Eriksson & Olinder, 2016) suggested that degrees of certitude of moral judgements should be understood as degrees of beliefs in claims about features of either the non-cognitivist attitude (which is the moral judgement) itself or about non-moral features of the acts or states of affair which are its objects. Eriksson & Olinder (2016), for example, suggested that degrees of certitudes of moral judgements are degrees of beliefs in propositions that describe classifications of the acts which are the objects of the moral judgement, when these classifications by themselves are understood to be completely descriptive. We can call this type of non-cognitivist responses to Smith’s argument, type 1 non-cognitivist responses.

Other philosophers (for example Sepielli, 2012; Makins, 2021) suggested that degrees of certitude should be understood not as degrees of beliefs (in any type of propositions) but rather as degrees of some feature of the non-cognitivist attitude (which is the moral judgement) which is not its strength. Makins (2021), for example, suggested that degrees of certitude of moral judgements are the degrees of the ambivalence of the agent’s moral desire (e.g., the degree in which the agent’s motivation in favor of performing a certain act conflicts with the agent’s motivation in favor of avoiding to perform the same act). We can call this type of non-cognitivist responses to Smith’s argument, type 2 non-cognitivist responses.

Both types of non-cognitivists responses faced, in turn, cognitivist criticisms (see for example Bykvist & Olson 2011 and 2017). In any case, it is clear that in order for any attempt to account for certitude in non-cognitivist terms to succeed, it must capture key features of the role certitude plays in moral reasoning and decision-making. This is true, we believe, both on a normative level and on a descriptive level. That is, the non-cognitivist must explain how certitude, understood in non-cognitivist terms, can play (at least most of the time) the central roles certitude actually plays, both when it comes to an ideal morally-motivated rational agent and when it comes to real people.

The latter task stands at the heart of this paper. It is not enough, for the non-cognitivist, to show that the feature that is supposed to play the role of certitude in their account *can* do so for an ideal agent. It is necessary to show also that this feature typically plays this role for actual agents. Without such a demonstration, the non-cognitivist account will be vulnerable to the objection that it fails to identify the relevant type of attitude as moral judgements, as attitudes are, at least partly, individuated by the key roles that they play.[[3]](#endnote-3)

One central feature that many philosophers have, implicitly or explicitly, assumed moral judgements’ degrees of certitude have, and should have, is a probabilistic structure.[[4]](#endnote-4) In accordance with Nissan-Rozen (2017), we call the position according to which degrees of certitude of moral judgments have a probabilistic structure “moral Bayesianism.”

Little attention has been given in the literature to the task of directly justifying moral Bayesianism. This is unsurprising from a cognitivist point of view. For cognitivists, moral judgements are just beliefs in moral propositions and there is a long tradition in philosophy that tries to justify probabilism (the thesis that *rational* degrees of beliefs have a probabilistic structure) and a long tradition in psychology that tries to show that everyday degrees of beliefs are approximately probabilistic.

Similarly, non-cognitivists of type 1 (who takes degrees of certitudes to be degrees of beliefs in some types of non-moral propositions) also do not have any problem explaining this feature. However, non-cognitivists that deny that degrees of certitudes are degrees of belief cannot turn to these lines of literature to justify moral Bayesianism and should look for justification elsewhere.

Staffel (2019) discussed non-cognitivist moral Bayesianism from a normative point of view. She examined how traditional arguments for probabilism translate into two leading non-cognitivist accounts (of type 2) of the certitude of moral judgements. However, to the best of our knowledge, there have been no previous attempts to examine moral Bayesianism from a descriptive point of view. In the relevant psychological literature, participants in experiments are usually asked to accept or reject various moral claims, not to report their degree of confidence in the truth of these claims (see, for example, Cushman, Young, & Hauser, 2006; Greene et al., 2009; Greene, Sommerville, Nystrom, Darley, & Cohen, 2001; Hauser, Cushman, Young, Kang‐Xing Jin, & Mikhail, 2007; McGuire, Langdon, Coltheart, & Mackenzie, 2009; Moore, Clark, & Kane, 2008; Moore, Lee, Clark, & Conway, 2011; Valdesolo & DeSteno, 2006).[[5]](#endnote-5) We do not know of any previous attempt to directly examine whether degrees of moral judgements have a probabilistic structure.

This paper examines whether moral Bayesianism holds true as a descriptive thesis. Informally described, our main finding is that degrees of certitude in moral judgement tend to have a structure that is probabilistic in the same way as the structure of non-moral beliefs (with respect to a plausible measure). Importantly, this approximately probabilistic structure of degrees of belief and of moral judgements is not shared by degrees of desire. However, relying on the assumption that people often desire a state of affairs to be true to the extent that they find it morally worthy, we were able to use the approximately probabilistic structure of degrees of certitude of moral judgements to predict a change in the structure of degrees of desire when moving from a non-moral to a moral context. As we explain in section 6, this last finding poses a special challenge not only to non-cognitivists of type 2 but also to non-cognitivists of type 1.

Together, these findings provide strong evidence for moral Bayesianism (as a descriptive thesis) and pave the way for further research employing the thesis to study different aspects of moral judgements. As the Bayesian perspective has been employed successfully in research on the nature and dynamics of non-moral beliefs (e.g., in Griffiths & Tenenbaum, 2006), the prospect of applying it in the context of moral judgements is exciting.

The paper contributes to the meta-ethical debate between moral cognitivists and moral non-cognitivists by pointing to two features of moral judgements that any plausible non-cognitivist position should account for, namely that their degrees of certitude have a probabilistic structure, and that these degrees relate to degrees of desires in a certain way. Importantly, while these two features are unsurprising from a cognitivist point of view, they are challenging to account for from a non-cognitivist point of view (the first feature is challenging to account for only from the point of view of non-cognitivists of type 2).

The structure of the rest of the paper is as follows. In Section 2, we set the stage by providing necessary philosophical background and introducing the idea behind our experimental design, which uses conformity to the axiom of Finite Additivity (FA) as a measure for the extent to which degrees of a given attitude can be taken to have a probabilistic structure. In section 3, we comment on the relation between our study and other studies in the psychology of meta-ethics. In section 4, we present our first study, which establishes that there is a significant difference between the extent to which degrees of belief on the one hand, and degrees of desire on the other, obey FA. We also show that this difference is robust across different levels of importance. These findings allow us to use, in studies 2 and 3, the level of conformity to FA of degrees of belief and degrees of desire as benchmarks to which degrees of certitude of moral judgement can be compared. In Section 5, we present our second study, which establishes that, in terms of conformity to FA, degrees of certitude of moral judgment are more similar to degrees of belief than to degrees of desire. In Section 6 we present our third study, which shows that when moving from a non-moral context to a moral context the level of conformity to FA of degrees of desire increases significantly. We also show that degrees of desire and degrees of certitude of moral judgments are significantly correlated in moral contexts, in line with the natural assumption that people tend to desire a proposition to be true to the extent that they judge it to be morally right. In the philosophical literature this assumption is discussed under the title “the desire as belief thesis” (see for example Lewis, 1988). Section 7 concludes.

1. **Philosophical background**

Let be a function from a finite set of propositions,  that has the structure of a Boolean algebra to the set of real numbers. represents a measure of the degrees of a given attitude (belief, desire, moral judgement) directed to different propositions in the set. We say that is probabilistic if (1) it assigns to each proposition in the set a non-negative number, (2) it assigns the number 1 to the disjunction of all the propositions in the set, and (3) it obeys the axiom of FA:

.

The first two axioms can be understood as expressing no more than a choice of scale for measuring the degrees of the attitude in question (that is, the unit interval scale). However, the FA axiom puts a substantial restriction on the structure of these degrees. While it seems reasonable to demand that a measure of rational degrees of belief will obey FA (there is a long tradition of presenting theoretical arguments for this claim; for an overview, see Leitgeb and Pettigrew, 2010a, 2010b), there seems to be no reason to make the same assumption with respect to rational desires.

Suppose, for example, you are told that hot drinks are going to be served soon and that you are certain (that is, believe to degree 1) that “hot drinks” stands for “either coffee or tea.” While rationality demands that if you believe that it will be tea to degree x, you must believe that it will be coffee to degree 1-x (assuming your degrees of belief are measured on a unit interval scale), rationality does not demand that if you want it to be tea to degree x, you must want it to be coffee to degree 1-x (even when degrees of desire are measured on a unit interval scale). If anything, rationality puts a different restriction on degrees of desire: if you want it to be tea to degree x and want it to be coffee to degree y, you must want it to be “either coffee or tea” to a degree which equals a weighted average of x and y (i.e., not equal to the sum of x and y, as in the case of rational degrees of belief).

This intuitive difference between the structure of *rational* degrees of desire and rational degrees of belief (a difference formally represented in any version of Bayesian decision theory) suggests that it might be possible to use the tendency to conform to FA as a way to distinguish between *actual* degrees of desire and actual degrees of belief. If it turns out that this is the case, we can use the tendencies to conform to FA of the two attitude types as benchmarks with respect to which the tendency to conform to FA of degrees of certitude of moral judgements can be evaluated. To the extent that degrees of certitude of moral judgement are shown to be similar to degrees of belief and different from degrees of desire with respect to their tendency to conform to FA, adopting a Bayesian approach when studying the psychology of moral judgement seems justifiable (even if moral judgements are not beliefs).

This is the basic idea that stands at the basis of our experimental design. Of course, the ultimate test for any methodological approach is its predictive success. Thus, after establishing that the structure of degrees of certitudes of moral judgement is more similar to that of degrees of belief than to that of degrees of desire, we demonstrate (in Section 5) how it is possible to employ this finding to make a novel prediction.

Several studies have shown that, given a binary dilemma, degrees of belief approximately satisfy FA (Tversky & Fox, 1995; Tversky & Koehler, 1994; Wallsten, Budescu, & Zwick, 1993). However, to the best of our knowledge, none of these studies employed a control/comparison group and it is thus unknown whether the measurements distinguished between degrees of belief and degrees of other attitudes, such as desire. In addition, to the best of our knowledge, no study has assessed whether degrees of desire and degrees of certitude of moral judgement approximately satisfy FA. Our goal in the first two studies was to answer these questions.

1. **Psychological background**

Although we are unaware of any study that has examined moral Bayesianism as a descriptive thesis, there is a vast psychological literature on meta-ethical questions (e.g., Beebe & Sackris, 2016; Goodwin & Darley, 2008, 2012; Nichols, 2004; Nichols & Folds-Bennett, 2003; Sarkissian, Park, Tien, Wright, & Knobe, 2011; Theriault, Waytz, Heiphetz, & Young, 2017; Wainryb, Shaw, Langley, Cottam, & Lewis, 2004; Wright, Grandjean, & McWhite, 2013). The current study differs from others in this literature not only in its subject matter (i.e., moral Bayesianism), but also in its methodological approach.

Most studies in the psychology of meta-ethics asked subjects to *explicitly* report what they believed to be the correct meta-ethical position. While providing important insights, these methods also have limitations that philosophers (Pölzler, 2017; Sinnott‐Armstrong, 2009) and psychologists (Beebe & Sackris, 2016) have pointed out. Many empirical studies (e.g., Greenwald & Farnham, 2000; Greenwald, Poehlman, Uhlmann, & Banaji, 2009) have confirmed that there is often a gap between explicit reports of an internal state and implicit measures of that state. For instance, implicit measures (e.g., the Implicit Association Test) have revealed racial biases in individuals who explicitly report that they do not hold such biases. This suggests that explicit reports *do not* *accurately* reflect participant attitudes. It cannot be taken for granted, then, that the *explicit* meta-ethical reports of participants accurately reflect their attitudes.[[6]](#endnote-6)

Our study, on the other hand, does not rely on explicit reports of the subjects regarding the conformity to FA of their moral judgments. As we specify below, we used an inter-subjective design in which, with respect to each moral claim, X, each subject reported only their degrees of confidence in either X or not-X. Subjects were not asked to report both their degrees of confidence in X and in not-X. Thus, our results point to a pattern in the subjects’ moral judgements of which they may not even be aware.

1. **Study 1: Setting the benchmarks**

In our first study, we showed that degrees of belief and degrees of desire can serve as benchmarks against which degrees of certitude of moral judgement can be evaluated, in terms of their tendency to conform to FA. In addition, we tested whether the difference we found between the level of conformity to FA of degrees of belief and degrees of desire was stable across different contexts. Specifically, we examined whether the difference would occur in dilemmas that were relatively more important to the participants (e.g., choosing a profession) as well as dilemmas that were less important to them (e.g., choosing a type of drink). We found that the level of importance of the dilemma did not affect the difference in level of conformity to FA between the degrees of the two types of attitudes. We conducted this examination to rule out the possibility that our results (Study 2) could be attributed to differences between the level of importance of the matters addressed by moral judgments and that addressed by desires.

* 1. **Participants**

The 389 participants (ages 18-35 years, *M* = 25.61, *SD* = 3.19; 52% female) in this study completed the moral dilemma questionnaire (designed using Qualtrics; see description below) on their personal computers. Participants were recruited through the website [www.panel4all.co.il](http://www.panel4all.co.il), and were compensated with gift certificates from the panel4all organization. The ethics committee of the Department of Psychology at the Hebrew University of Jerusalem approved all three experiments described in this paper.

* 1. **Materials**

Participants were randomly assigned to one of three (a, b, c) online moral dilemma questionnaires, all with similar structures. Each questionnaire contained two main parts, which participants completed at least four days apart. In addition, each questionnaire had a male version and a female version.

The first part contained 76 dilemmas, which were randomly divided among the three questionnaires.[[7]](#endnote-7) Each questionnaire was randomly divided into two blocks, a block of dilemmas (N = 13), in which participants were asked about their degree of desire, and a separate block of dilemmas (N = 13 or 11) in which they were asked about their degree of belief. Two of the questionnaires comprised 26 dilemmas and the remaining questionnaire comprised 24 dilemmas. For this part, we constructed two types of statement: beliefs (*To what extent do you believe that x?*) and desires (How much would you like to x*?*). Critically, each statement had two complementary versions in which the possible outcome was either A or B, and the conjunction of the two was impossible.

In the second part, the same dilemmas from the first part were presented once again. This time, for each dilemma, participants were asked: “How significant is the dilemma in your opinion?”

* 1. **Experimental design**

Participants received an email with a link to one of the three questionnaires. The number of participants, the age range, mean, standard deviation and percentage of females in each of the three questionnaire groups were as follows: a – *N* = 134; age range = 18-34 years, *M* = 25.8, *SD* = 2.9; 40% female; b – *N* = 122; age range = 19-35 years, *M* = 25.6, *SD* = 3.5; 57% female; c – *N* = 132; age range = 18-34 years, *M* = 25.3, *SD* = 3.01; 58% female. For each dilemma in the first part, participants were randomly assigned to report either degrees of belief or degrees of desire with respect to statement A or to statement B, but not to both. Take for example the dilemma, “You are preparing to take a big trip, and wondering whether to travel to India or South America.” For this dilemma, the corresponding statements were as follows: (A) “To what extent do you believe that you will go to India? / To what extent do you desire to go to India?” or (B) “To what extent do you believe that you will go to South America? / To what extent do you desire to go to South America?” For a complete list of all dilemmas, see Supplementary Materials. Participants reported their degree of desire or belief with respect to A or B using a non-numeric and continuous scale on which they had to slide a marker to the desired point, with only two extreme anchor points, “Not at all” and “Very much.” The slider was initially set in the middle of the scale. Responses were translated to numeric values ranging from 0 to 100.

In the second part, the dilemmas presented to participants in the first part were presented to them again. This time, for each dilemma participants had to answer the question: “How significant is the dilemma in your opinion?” These questions were answered by placing a marker on a scale between 0 and 100.

In each questionnaire, two catch trials were presented in each part, in which participants were asked to place the marker at the right or left of the scale. These questions were designed to examine whether participants were alert and reading the instructions carefully. Participants who were wrong on one or more catch trials were not included in the analysis of the results (3.4%).

* 1. **Measurements**

The experiment employed a between-subjects design, with two independent variables: attitude (two levels: belief and desire; manipulated) and mean level of importance of each dilemma over all participants (continuous variable; measured). The dependent variables werethe level of additivity of the components (A, B), which were evaluated independently. For dilemma i in attitude j, we calculated the level of conformity to FA as follows:. Unlike previous studies (e.g.,Tversky & Koehler, 1994) that measured additivity using only the sum of the averages, we calculated the absolute value of the sum of the averages to avoid cases in which subadditivity overrides super-additivity, and vice versa. For example, if question x was sub-additive and question y was super-additive, the total of the two would be close to 100. Thus, calculation of the absolute distance from 100 would lead to the conclusion that x and y were non-additive.

* 1. **Results**
		1. **Frequentist analysis**

The results were analyzed with multiple linear regression using R (Team, 2015) to assess the effect of attitude (two levels: belief and desire) and level of importance (continuous) on LoC. Our model for this analysis was as follows:, where was the dependent variable LoC for dilemma i after aggregating data between participants, represented the fixed effect attitude, represented the fixed effect level of importance, and represented the residuals. The results of the regression indicated that the two predictors explained 27.9% of the variance [R2 = .27, F(3,72) = 9.31, p < 0.001] (see Table 1). There was a significant main effect [0001] of attitude (belief: ; desire: . There was no main effect of importance [67], nor was there an interaction effect [45] (see Table 2). Thus, conformity to FA could serve as a test for whether moral judgment is a belief-like or a desire-like attitude (see Figure 1). We verified that the sample size was acceptable by running a Bayesian t-test and reporting the Bayes factor, which indicated substantial evidence ] for the hypothesis that there is a difference between the LoC of desire and belief rather than there is not. This finding indicates that one can stop collecting data since the evidence is conclusive (Rouder, 2014).

 **Figure 1.** The X axis represents level of importance and the Y axis represents LoC. There was no difference between beliefs and desires in level of importance (represented by the two distributions on the X axis); however, they did differ in LoC (represented by the two distributions on the Y axis).

Table 1. *Regression results using LoC as the criterion*

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Predictor | *B* | *B*95% CI[LL, UL] | *sr2* | *sr2*95% CI[LL, UL] | Fit |
| (Intercept) | -0.53\*\* | [-0.82, -0.25] |  |  |  |
| desire  | 1.02\*\* | [0.62, 1.42] | .26 | [.09, .43] |  |
| importance | -0.02 | [-0.28, 0.24] | .00 | [-.01, .01] |  |
| desire: importance | 0.16 | [-0.25, 0.57] | .01 | [-.02, .03] |  |
|  |  |  |  |  | *R2*  = .280\*\* |
|  |  |  |  |  | 95% CI[.10,.41] |
|  |  |  |  |  |  |

*Note.* A significant *b*-weight indicates that the semi-partial correlation is also significant. *b* represents unstandardized regression weight. *sr2* represents the semi-partial correlation squared. *LL* and *UL* indicate the lower and upper limits of a confidence interval, respectively. \**p* < .05. \*\**p* < .01.

* 1. **Discussion**

A key requirement for our analysis was that the levels of conformity to FA of desires and of beliefs would be significantly different from one another.Study 1 found this to be the case, such that levels of conformity to FA of desires and of beliefs can be used as benchmarks for evaluating degrees of certitude of moral judgements.[[8]](#endnote-8) In addition, we found no evidence that level of importance moderates the effect between level of conformity to FA of beliefs and of desires.

1. **Study 2**

This experiment attempted to replicate the significant difference between degrees of belief and degrees of desire in level of conformity to FA, and – more importantly – to determine whether degrees of certitude of moral judgements were more similar to the former or the latter.

* 1. **Participants**

The 355 participants (ages 18-58 years, *M* = 25.36, *SD* = 3.75; 53% female) in this study completed the moral dilemma questionnaire (designed using Qualtrics) on their personal computers. Participants were recruited from the internet site www.panel4all.co.il and received gift certificates from the panel4all organization.

* 1. **Materials**

As in the previous experiment, each condition (belief, desire, and moral judgment) had different dilemmas. However, we made three changes. To begin with, while in the first experiment each participant responded to both the belief and the desire dilemmas, in this experiment the participants were randomly divided into three groups, such that each responded to either the belief, the desire, or the moral judgment dilemmas. The goal was to separate the probabilistic (belief-like) and the non-probabilistic (desire-like) ways of thinking from the moral judgment condition. Second, we added 20 moral dilemmas that were classified as such by participants in a pilot study.[[9]](#endnote-9) Finally, we randomly selected 20 questions characterized as having high importance from the benchmark conditions (belief and desire). Note that in the previous experiment, we did not find a significant effect of level of importance. However, since a pilot study verified that all the moral dilemmas were characterized by a high level of importance, we selected the belief and desire dilemmas that were also characterized by a high level of importance.[[10]](#endnote-10) (For a complete list of all dilemmas, see Supplementary Materials.)

Participants were randomly assigned to one of three online questionnaires: moral judgment, desire, or belief (M, D, or B, respectively). We constructed three types of statement: beliefs (*To what extent do you believe that x?*), desires (How much would you like to *x?*), and moral judgments (*To what extent do you deem it appropriate to act x?*).[[11]](#endnote-11) Critically, each statement had two complementary versions in which the possible outcome was either A or B. The conjunction of the two was impossible, enabling us to measure responses to the same dilemma.

* 1. **Experimental design**

Three groups of subjects were tested under three conditions in a between-subject design. Each participant was randomly assigned to one of the three conditions. Participants reported responses to a possible dilemma in the belief condition (*To what extent do you believe that A/B?* *N* = 113; age range = 18-39 years, *M* = 24.94, *SD* = 2.87; 60% female), in the desire condition (How much would you like to *A/B?* *N* = 122; age range = 18-58 years, *M* = 26.15, *SD* = 5.52; 48% female), or in the moral judgment condition (*To what extent do you deem it appropriate to act A/B?* *N* = 117; age range = 18-34 years, *M* = 24.99; *SD* = 2.87; 52% female). For each dilemma in each condition, participants were randomly assigned to report their degrees of belief/desire/moral judgment with respect to statement A or to statement B, but not to both. Take for example the dilemma, “Refugees from Africa seek to enter your country.” For this dilemma, the corresponding statements were as follows: (A) “Your country will allow them to enter” or (B) “Your country will send them back.” As in Study 1, participants indicated their degrees of belief/desire/moral judgment with respect to A or B using a non-numerical, continuous scale with only two extreme anchor points “not at all” and “very much,” using a slider that was initially set to the middle of the scale. For each dilemma, participants had to slide the marker to their desired point.

Reaction time (RT) data were collected automatically by Qualtrics using additional code to correct for deviations in measurement stemming from different participants using different devices. RT results were transformed using the log transformation. First, we excluded responses for which the RT was less than 1 sec (167 answers, 2.4%). In addition, for each dilemma, we excluded responses for which the RT was lower than the mean minus two standard deviations (287 answers, 4.2%). In total, we excluded 454 out of 6849 responses (6.6%).

* 1. **Measurements**

The experiment had a between-subjects design with the independent variable attitude (three levels: belief, desire, and moral judgment). The dependent variables werethe levels of conformity to FA of the components (A, B), which were evaluated independently. For dilemma i in attitude j, we calculated the LoC, as before.

* 1. **Results**
		1. **Frequentist analysis**

The results were analyzed with multiple linear regression using R (Team, 2015) to assess the effect of attitude (three levels: belief, desire, and moral judgment) on LoC. Our model for this analysis was as follows:, where was the dependent variable LoC for dilemma i after aggregating data between participants, represented the fixed effect attitude, and represented the residuals. The results of the regression analysis indicated that the main effect of attitude explained 19% of the variance [R2 = .19, F(2,56) = 6.6, p = 0.002]. To examine whether level of conformity to FA of moral judgments behaved like that of beliefs or that of desires, we began by comparing desires (M = 16.62 ,SD = 10.88) and beliefs (M = 7.04, SD = 6.14), and found a significant difference, such that the LoC of desires was higher than that of beliefs [. We then examined whether moral judgments (M = 10.75, SD = 7.36) behaved like beliefs or like desires. The LoC of moral judgments was significantly smaller than that of desires , while there was no significant difference between moral judgments and beliefs [.

* + 1. **Bayesian analysis**

We used a Bayesian model to estimate whether moral judgments were similar to beliefs or to desires. To clarify our motivation to use Bayesian analysis, we explain how Bayesian parameter estimation allowed us to test whether moral judgments were more similar to desires or to beliefs. Bayesian parameter estimation assesses null values by setting a range of values, including the null value. The result of this analysis is a posterior distribution, or a distribution showing the relative plausibility of each possible level of conformity to FA of each attitude, conditional on the data, the priors, and the model. This analysis enables us to test the *similarity* between attitudes and does not limit us to the differences, as does the frequentist analysis. In particular, by comparing the posterior distributions of moral judgments to those of beliefs and to those of desires, we could determine the (dis)similarity between them.

The first analysis was a Bayesian estimation model that focused on the posterior high density interval (HDI) to reflect the most likely dense value of the parameter and of the differences between each of the two attitudes. If the value 0 is not included in the HDI, we conclude that the attitudes are different; otherwise, we conclude that the result is inconclusive. In addition, we calculated the effect size via ROPE analysis with bayestestR package (Makowski, Ben-Shachar, & Lüdecke, 2019). Practically speaking, a ROPE is the range of parameters near value 0. If 95% of the HDI falls inside the defined ROPE [-.1,.1], then one can accept the null hypothesis; otherwise, the results are inconclusive (Kruschke, 2018).

Parameters of the Bayesian estimation model were analyzed using JAGS (Depaoli, Clifton, & Cobb, 2016) and the rjags package in R (Plummer, 2016) to run MCMC samples. In all estimations, we ran three separate chains, each consisting of 10,000 samples (after 1000 burn-in iterations and thin = 10). Next, we assessed convergence by computing the Gelman-Rubin statistic (Gelman & Rubin, 1992), with values below 1.01 indicating successful convergence. In this model, was the dependent variable LoC for dilemma i after aggregating data between participants for attitude j and standardizing it. We estimated by a t distribution . The first parameter is , which is equal to. represents the intercept and represents the fixed effect of attitude (three levels: belief, desire, and moral judgment). Each attitude has its own scale parameter, . The scale parameter of each group comes from a gamma distribution that has mode ω and standard deviation , so that each group’s scale is estimated separately and is vague on the scale of the data (Kruschke, 2015). The third parameter of the t distribution is ν. When v is more than 30, the t distribution is normal; thus, we require that the prior of v give equal opportunity to small values of ν (less than 30) and larger values of ν (greater than 30). To capture this, the prior of v was exponentially distributed from 1 to infinity with a mean of 30 (Kruschke, 2013; 2015). Note that since this is a novel approach to the study of moral judgements, no previous studies assess the parameters. Therefore, we chose uninformative parameters aligned with Kruschke's suggestions (2015). The description of the parameters of the model are provided in Figure 2.

**Figure 2.** Description of the parameters of the Bayesian estimation model

In testing the robustness of the frequentist analysis results from the first study, we expected that the difference between the posterior distributions of beliefs and desires would yield a central tendency with a significant positive number (see Figure 3 for visualization of the posterior distribution). When we subtracted the posterior distribution of desire from the posterior distribution of belief, the results revealed that 95% of the HDI were between 0.21 and 1.52, and that the mode was 0.85, indicating that the LoC of beliefs was significantly lower than that of desires (inside ROPE = 0%). Similarly, when we subtracted the posterior distribution of desire from the posterior distribution of moral judgment, 95% of the HDI were between 0.04 and 1.34, and the median was 0.66, indicating that the LoC of moral judgments was significantly lower than that of desire (inside ROPE = 1.5%). Note, that the frequentist, the HDI, and the ROPE analyses all indicate a significant difference between belief-desire. For moral judgments-desire, the frequentist and the HDI both indicate a significant difference, but the ROPE analysis result was not zero but 1.5%, indicating a somewhat less conclusive result.

Finally, when we subtracted the posterior distribution of beliefs from the posterior distribution of moral judgments, 95% of the HDI were between -0.66 and 0.29, and the median was 0.12, indicating that the LoC of moral judgments was not significantly different from that of beliefs (inside ROPE = 27.7%).

We ran two analyses to test whether degrees of certitude of moral judgments were more similar to degrees of desires or to degrees of beliefs in terms of their conformity to FA. The first analysis compared the overlap between the posterior distributions of beliefs and of moral judgments and the overlap between the posterior distributions of desires and moral judgments. Regarding the posterior distributions of desires and moral judgments, the Cohen’s d was 1.08, and 86% of the LoC of desires would be above the mean of the level of conformity to FA of moral judgments (Cohen’s U3). In addition, there was a 59% overlap between the two groups, and a 78% chance that a LoC picked at random from the desire group would have a higher score than a LoC picked at random from the moral judgment group (probability of superiority). Regarding the posterior distributions of beliefs and moral judgments, the Cohen’s d was 0.26, and 60% of the LoC of moral judgments would be above the mean of the LoC of beliefs (Cohen’s U3). In addition, there was 90% overlap between the two groups, and a 57% chance that a LoC picked at random from the moral judgment group would have a higher score than a LoC picked at random from the belief group (probability of superiority). Thus, it appears that the posterior distribution of moral judgments is much more similar to the posterior distribution of beliefs than to that of desires.

**Figure 3.** Posterior distributions for beliefs, desires, and moral judgments, obtained using the MCMC procedure



*Note*. LoC denotes the level of conformity to finite additivity

The second analysis addressed the research question using a model comparison analysis. The first model compared the likelihood of the hypothesis that the level of conformity to FA of moral judgments is similar to that of beliefs to the hypothesis that the level of conformity to FA of moral judgments is not similar to that of beliefs. The second model compared the likelihood of the hypothesis that the level of conformity to FA of moral judgments is similar to that of desires to the hypothesis that the level of conformity to FA of moral judgments is not similar to that of desires.

The Bayes factor of each model represents the ratio of the likelihoods of the data given the first hypothesis versus the likelihoods of the data given the second hypothesis. By comparing the Bayes factors of the models, we can estimate whether the data is more likely if the level of conformity to FA of moral judgments is similar to that of beliefs or if the level of conformity to FA of moral judgments is similar to that of desires.

Regarding Bayesian model selection, we ran a model that compared the two Bayes factors (see Figure 4 for a description of the parameters of the model). In the model, the effect size of each model (denoted by d1 and d2) was estimated by a Cauchy distribution and , which is a t distribution with equal to 0 and equal to 0.707(Wagenmakers et al., 2018). The prior of v, which is the third parameter, was exponentially distributed and the range was from 1 to infinity with a mean of 30 (Kruschke, 2013; 2015). DB (desire and belief) and DM (desire and moral judgment) are the Bernoulli likelihoods of hypothesis 1 (the attitudes are similar) and hypothesis 2 (the attitudes are not similar) of each model, when the odds are even (0.5). Thus, hypothesis 2 puts all prior mass for d1 and d2 on a single point, that is, H2: d1/d2 = 0, whereas hypothesis 1 assumes that d1/d2 has a Cauchy distribution, with H1: d1/d2 ∼ Cauchy(0, 1). The observed data yM (moral judgment), yD (desires), and yB (beliefs) follow a t distribution . The μ of the t distributions of beliefs and of moral judgments had an uninformative prior N(0, 0.001). The μ of desires in the first model (desire and belief) was given by = . For the second model (desire and moral judgment), = . The of the t distribution of each attitude had a uniform prior, . The prior of v, which is the third parameter of the t distribution, was exponentially distributed and the range was from 1 to infinity with a mean of 30. To assess convergence, we ran three chains, each consisting of 10000 samples. We assessed convergence by computing the Gelman-Rubin statistic, with values below 1.01 indicating successful convergence.

Regarding the first model, in which we tested whether beliefs and moral judgments were similar, there was no evidence supporting either hypothesis 1 or hypothesis 2 ( ; 95%, CI: [-0.44, 0.79], mean = 0.33). Regarding the second model, in which we tested whether desires and moral judgments were similar, there was evidence that desires and moral judgments were significantly different, as assumed in hypothesis 1 ( ; 95%, CI: [-1.76, 0.02], mean = 0.82). Finally, we computed the Bayes factor of the first model (LoC of moral judgments is similar to that of beliefs/LoC of moral judgments is not similar to that of beliefs) divided by the Bayes factor of the second model (LoC of moral judgments is similar to that of desires/LoC of moral judgments is not similar to that of desires). The results indicated that the first model predicted the observed data approximately 9.58 times better than did the second model.

**Figure 4.** Description of the parameters of the Bayesian model selection

* 1. **Discussion**

In this study, levels of conformity to FA were used as an indirect test to examine the question: do degrees of certitude of moral judgement behave more like degrees of desire or more like degrees of belief? A key requirement for our analysis was that the level of conformity to FA of desires would be significantly different from that of beliefs. This difference was found in Study 1 and replicated in Study 2. Study 2 also established that degrees of certitude of moral judgement are more similar to degrees of belief than to degrees of desire in terms of their tendency to conform to FA.[[12]](#endnote-12)

As explained in the introduction, this finding contributes to the meta-ethical debate between moral non-cognitivists and moral cognitivists by placing further demand on any plausible non-cognitivist account of moral certitude: such accounts should not only point to a feature of moral judgments that can play the role of certitude, but also make sure that this feature comes in degrees that – for most people – have a probabilistic structure, to the same extent that degrees of belief do.

It also supports the adoption of a Bayesian approach to studying the nature of moral judgments, their dynamic, and interaction with other attitudes. The next study (Study 3) aimed to further support this claim by demonstrating that the Bayesian approach has a predictive power in this context. Another important contribution of study 3 is that it gives evidence for another demand (that is, other than the demand that certitude has a probabilistic structure) that non-cognitivist accounts of moral certitude must respect. As explained in the introduction, this last demand (unlike the former demand) applies to both type 1 and type 2 non-cognitivists accounts.

To this end, we examined whether the level of conformity to FA of degrees of desire in a *moral context* is similar to the level of conformity to FA of degrees of belief and degrees of moral judgement. The hypothesis that it would indeed be so is based on the results of the two previous studies and on one additional natural assumption (that makes sense from a cognitivist point of view): that in moral contexts, people tend to desire a proposition to be true approximately to the extent to which they judge it to be morally worthy.

1. **Study 3**

In the philosophical literature, the assumption that people desire a proposition to be true to the extent to which they judge it to be morally worthy is discussed under the title “the desire as belief thesis” (DBT). David Lewis (1988, 1996) famously showed that, given several minimal assumptions about the structure of rational beliefs and desires (except in trivial cases),[[13]](#endnote-13) even if an agent obeys the DBT at one point in time, they can always learn additional information (to which they assign a positive probability before learning it) that will rationally lead them to stop obeying the DBT.

Thus, Lewis concluded, the DBT cannot be taken as a necessary normative demand for rational agents. Several philosophers (e.g., Bradley & List, 2009; Hájek & Pettit, 2004; Russell & Hawthorne, 2016; Nissan-Rozen, 2015) have discussed Lewis’ result and its normative significance and suggested alternative formulations of the thesis. Here, however, we are interested in the DBT not as a normative constraint, but rather as an empirical thesis, which – even if violated in exceptional cases – usually holds. We are unaware of any previous attempt to test the DBT empirically, but on an intuitive level we find it natural. Indeed, in the philosophical discussion, it is generally taken for granted that most people obey the DBT most of the time.

If the DBT indeed holds, then in light of the results of Studies 1 and 2, we can expect that, specifically in moral contexts, degrees of desire will conform to FA approximately as well as to degrees of moral judgments. This is exactly what we found in Study 3. Thus, by adopting a Bayesian approach regarding moral judgements, we were able to give empirical validation to the DBT (as an empirical thesis).

Notice that the DBT, as originally formulated, explicitly refers to the relation between moral beliefs and desires. It is natural, then, to take it as expressing what can be regarded as a cognitivist rationale, according to which moral judgements are beliefs in moral propositions and morally-motivated rational agents ought to desire a proposition to be true to the degree that they believe it is morally right for it to be true. On a descriptive level, the hypothesis would be that real people, even if they sometimes fail to respect this restriction, usually obey it.[[14]](#endnote-14)

It is more challenging, however, to find a rationale for the DBT from a non-cognitivist point of view. Importantly, this is true with respect to both non-cognitivist accounts of type 1 and of type 2. Whatever the feature, the degrees of which are taken to be degrees of certitude according to a given non-cognitivist account, is, it is not a belief in a moral proposition. Thus, the main motivation for adopting the DBT, namely that it seems to capture the motivating role of moral judgements, does not apply from the point of view of moral non-cognitivism.

For example, on the face of it, there seems to be no reason to think that a morally motivated rational agent’s degree of desire for performing a certain act would be equal to the agent’s degree of belief that the act should be classified as belonging to some (non-moral) category, as someone who accepts Eriksson’s and Olinder’s (2016) account *and the DBT* would have to argue.

We do not rule out the possibility of a non-cognitivist account that can vindicate the DBT while interpreting degrees of certitude of moral judgements, not as degrees of beliefs in moral propositions, but rather in another, non-cognitivist way. Our results show, however, that any plausible non-cognitivist account should be able to do so. The results of Study 3, in other words, ground a further restriction on non-cognitivist accounts: not only do they have to point to a feature of moral judgements that can be interpreted as coming in degrees with a probabilistic structure, but these degrees should also restrict degrees of desire in a specific way (at least most of the time). This demand applies equally well, of course, to moral cognitivist accounts. However, from a cognitivist point of view it is, as explained above, completely unsurprising.

We use, then, the DBT as a psychological contraction that raises the following hypotheses: (a) degree of desires and degree of moral judgments are correlated; and (b) the level of conformity to FA of desires in a moral context is significantly different from the level of conformity to FA of desires in a non-moral context.

* 1. **Participants**

The 717 participants (ages 18-40 years, *M* = 26.1, *SD* = 3.8; 54% female) in this study completed the moral dilemma questionnaire (designed using Qualtrics) on their personal computers. Participants were recruited from the Hebrew University of Jerusalem (HU: *N* = 210) and from the internet site www.panel4all.co.il (PA: *N* = 507). Participants from HU received 25 shekels or academic credit and those from PA received gift certificates from the panel4all organization.

* 1. **Materials**

Thequestionnaire in this study contained two parts. The first and main part contained 21 moral dilemmas, which had been classified as such by participants in a pilot study.[[15]](#endnote-15) For this part, we constructed three types of statements in which the root element was identical in three conditions: beliefs (*To what extent do you believe that x?*), desires (How much would you like to *x?*), and moral judgments (*To what extent do you deem it appropriate to act x?*). Critically, each statement had two complementary versions in which the possible outcome was either A or B. The conjunction of the two was impossible, enabling us to measure participant responses to the same dilemma.

The second part cosntained six measures examining possible individual differences. None of these yielded significant results and they will not be discussed further. The questionnaire included additional parts, the results of which are not analyzed and discussed in this paper. Overall, then, for all three conditions, the structure of the questionnaire was as follows: (1) personal questions (age, gender, and profession); (2)general instructions for the next parts; (3) the main part: 21 moral dilemmas; (4)three moral dilemmas in which the possible responses included three options (instead of two options); (5) political attitude questionnaires; (6) Bayesian version of the trolley problem; (7) personal information used to assess individual differences in the Ellsberg paradox, a mathematical question, need for closure, additional mathematical questions, and Rational-Experiential Inventory questionnaires.

* 1. **Experimental design and measurement**

Six groups of participants were tested in a between-subjects design. Participants reported either beliefs, desires, or moral judgments with respect to the 21 binary dilemmas presented in one of the complementary versions. In each condition, the 21 dilemmas were presented in a different random order to each participant. Participants indicated their degrees of belief (*To what extent do you believe A/B?* *N* = 229: HU = 70, PA = 159; age range 19-40 years, *M* = 26.36, *SD* = 3.80; 41% female); degrees of desire (*To what extent do you desire A/B?* *N* = 243: *HU* = 60, *PA* = 183; age range 18-35 years, *M* = 25.61, *SD* = 3.59; 60% female); degrees of moral judgment (*To what extent do you deem* *it appropriate to act A/B?* *N* =245: *HU* = 80, *PA* = 165; age range 18-40 years, *M* = 26.23, *SD* = 4.01; 61% female). For each dilemma in each condition, participants were randomly assigned to report their degrees of belief/desire/moral judgment with respect to statement A or to statement B, but not to both.

* 1. **Measurements**

The independent variable was attitude (three levels: belief, desire, and moral judgment). The dependent variables werethe levels of conformity to FA of the components (A, B), which were evaluated independently. For dilemma i in attitude j, we calculated the level of conformity to FA, as before.

* 1. **Results**
		1. **Are degree of moral judgment and degree of desire correlated?**

The relationships between degrees of moral judgment, degrees of desire, and degrees of belief were tested using correlations between the average responses of desires and of moral judgments for each statement (A or B) for each dilemma. There was a significant correlation between degrees of desire and degrees of moral judgment []. No significant correlations were found between degrees of belief and degrees of moral judgment [], nor between degrees of belief and degrees of desire []. In addition, we ran a Bayesian correlation which indicated substantial evidence ] for the hypothesis that there was a correlation between degrees of desire and degrees of moral judgments. This finding indicates that one can stop collecting data since the evidence is conclusive (Rouder, 2014).

* + 1. **Is the level of conformity to FA of desires in a moral context significantly different from that of desires in a non-moral context?**

To assess whether the level of conformity to FA of desires in a moral context was significantly different from the level of conformity to FA of desires in a non-moral context, we tested whether the level of conformity to FA of desires in the third study (moral context) was significantly lower than the level of conformity to FA of desires in the first and second studies (non-moral context). The results were analyzed using a one-way between subjects ANOVA to compare the effect of context (three levels: Study 1, Study 2, and Study 3) on the level of conformity to FA of desires. There was a significant main effect of context on level of conformity to FA []. In addition, post hoc comparisons using the Tukey HSD test indicated that the mean score for the moral context (Study 3) was significantly different () than the mean score for the non-moral context (Study 1: ). Moral context (Study 3) was also significantly different () from the second non-moral context (Study 2: ). Meanwhile, there was no significant difference between the two non-moral context conditions (Study 1 and Study 2). Thus, the level of conformity to FA of desires was significantly lower in a moral context than it was in non-moral contexts.

* 1. **Discussion**

In Study 3, we showed that the tendency of degrees of certitude of moral judgments to obey FA (in a similar way to degrees of beliefs) has a predictive power.[[16]](#endnote-16) As the DBT thesis predicts, degrees of desire and degrees of moral judgment were significantly correlated and the level of conformity to FA of desires was significantly higher in a moral context than in a non-moral context.

As explained, these results do not only provide further support to moral Bayesianism as a descriptive thesis, but also point to an additional restriction that non-cognitivist accounts of certitudes of moral judgements must explain.

It is worth mentioning that, while it is very intuitive, the DBT is non-trivial. To see why (and in what sense) this is so, recall the distinction between degrees of certitude in moral judgments (the type of degrees we study in this paper) and the degrees of importance of moral judgements (i.e., the degrees of wrongness or rightness of the behaviors that are the objects of moral judgements). As noted, these two types of degrees seem independent.

Consider now a case in which you are uncertain whether a given behavior is morally permissible and assign different degrees of rightness/wrongness (let us call these different degrees of “moral value”) to the behavior in case it is permissible and in case it is impermissible. For example, suppose you are uncertain whether eating meat is permissible, and believe that in case it is permissible, it is not a particularly good thing to do (and not a particularly bad thing to avoid), while in case it is not permissible, it is a very bad thing to do. Many people find this type of attitude very natural.

Some philosophers argue (see for example, Sepielli, 2013) that in such a case, the degree of desirability a morally-motivated rational agent should have toward the prospect of eating meat must be equal to the expected degree of moral value of the prospect. In other words, it should be equal to a weighted average of the moral values she assigns to eating meat in case it is not permissible to eat meat and the moral value she assigns to eating meat in case it is permissible, when the weights are the probabilities that she attaches to each of the two hypotheses regarding the moral value of the act.

If, for example, she assigns a degree of moral value of 10 to eating meat, in case it is permissible to eat meat, assigns a moral value of -100 to eating meat in case it is not permissible to eat meat, and assigns an equal probability to the two possibilities, her degree of desire for the prospect of eating meat must be (10\*0.5) + (-100\*0.5) = -45. Of course, the expectation can be calculated relative to more than two hypotheses regarding the moral value of the act. Let us call this the “desire as expected moral value thesis.”

This “desire as expected moral value thesis” is based on the assumption that there exists a numerical scale for measuring degrees of moral value. The important point, in the context of our discussion, is that the DBT (as originally formulated by Lewis) is consistent with the “desire as expected moral value thesis” only when the scale used for measuring degrees of moral value is the (quite trivial) binary scale in which there are only two possible degrees, 0 (not permissible) and 1 (permissible).

In other words, the DBT is a private case of the “desire as expected moral value thesis,” the private case in which moral value is measured on a trivial binary 0/1 scale. Since it is at least plausible that the “desire as expected moral value thesis,” understood as an empirical thesis, is approximately true, and that a trivial binary 0/1 scale cannot adequately represent the way our subjects evaluated the moral status of the scenarios presented to them, our results – which clearly support the DBT – are surprising.

1. **Conclusion**

Our studies show that, in terms of one intuitive measure, degrees of certitudes of moral judgements tend to have an approximately probabilistic structure. We additionally showed that these degrees are related to degrees of desires in the way predicted by the DBT.

Both these findings are unsurprising from the point of view of moral cognitivists, but are more challenging to explain when adopting non-cognitivist positions. We do not argue that no (existing or merely possible) non-cognitivist position can meet this challenge. Whether such a claim is true is an open question. The paper establishes, however, that constructing such an account is a challenge moral non-cognitivists should face.

In light of the well-known successes of the Bayesian approach in the context of non-moral beliefs and in light of the predictive success of moral Bayesianism demonstrated in section 3, our findings also support the possibility that adopting a Bayesian approach with respect to other questions in moral psychology might be a fruitful exercise. We hope that others will follow our steps and employ a Bayesian approach to such questions.

**Acknowledgements**

We would like to thank Itzhak Aharon, Ryan Doody, David Enoch, Jonathan Najenson, Thomas Polzler, Brian Weatherson, and three anonymous reviewers for their thoughtful comments and suggestions on this article. Earlier versions of this paper were presented in the annual meeting of the European Society for Philosophy and Psychology, Hatfield, England (2017) and in Tel hai Seminar, Department of Philosophy, Tel-Hai College, Israel (2018). We thank the participants of these events for useful discussions. This work was funded by The Israel Science Foundation grants (grants 1471/20) to Anat Maril, by The Israel Science Foundation grant (grants 1042/13) to Ittay Nissan-Rozen, and by a fellowship from the Sidney M. Edelstein Center for History and Philosophy of Technology and Medicine at the Hebrew University of Jerusalem to Haim Cohen.

**References**

Beebe, J. R., & Sackris, D. (2016). Moral objectivism across the lifespan. Philosophical Psychology, 29(6), 912–929. <https://doi.org/10.1080/09515089.2016.1174843>

 Bradley, R., & List, C. (2009). Desire-as-belief revisited. Analysis, 69(1), 31–37. <https://doi>. org/10.1093/analys/ann005

Bykvist, K., & Olson, J. (2012). Against the being for account of normative certitude. Journal of Ethics and Social Philosophy, 6(2), 1–8. [http://dx.doi.org/10.26556/jesp.v6i2.63 . PHILOSOPHICAL PSYCHOLOGY 27](http://dx.doi.org/10.26556/jesp.v6i2.63%20.%20%09PHILOSOPHICAL%20PSYCHOLOGY%2027)

Bykvist, K., & Olson, J. (2017). Non-cognitivism and fundamental moral certitude: Reply to Eriksson and Francén Olinder. Australasian Journal of Philosophy, 95(4), 794–799. <https://doi.org/10.1080/00048402.2016.1269352>

Cushman, F., Young, L., & Hauser, M. (2006). The role of conscious reasoning and intuition in moral judgment: Testing three principles of harm. Psychological Science, 17(12), 1082– 1089. <https://doi.org/10.1111/j.1467-9280.2006.01834.x>

 Depaoli, S., Clifton, J. P., & Cobb, P. R. (2016). Just another Gibbs sampler (JAGS) flexible software for MCMC implementation. Journal of Educational and Behavioral Statistics, 41 (6), 628–649. <https://doi.org/10.3102/1076998616664876>

Eriksson, J., & Olinder, R. F. (2016). Non-cognitivism and the classification account of moral uncertainty. Australasian Journal of Philosophy, 94(4), 719–735. <http://dx.doi.org/> 10.1080/00048402.2016.1149199

Gelman, A., & Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. Statistical Science, 7(4), 457–472. <https://doi.org/10.1214/ss/1177011136>

Goodwin, G. P., & Darley, J. M. (2008). The psychology of meta-ethics: Exploring objectivism. Cognition, 106(3), 1339–1366. <https://doi.org/10.1016/j.cognition.2007.06.007>

Goodwin, G. P., & Darley, J. M. (2012). Why are some moral beliefs perceived to be more objective than others? Journal of Experimental Social Psychology, 48(1), 250–256. https:// doi.org/10.1016/j.jesp.2011.08.006

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, 293(5537), 2105– 2108. <https://doi.org/10.1126/science.1062872>

Greene, J. D., Cushman, F. A., Stewart, L. E., Lowenberg, K., Nystrom, L. E., & Cohen, J. D. (2009). Pushing moral buttons: The interaction between personal force and intention in moral judgment. Cognition, 111(3), 364–371. <https://doi.org/10.1016/j.cognition.2009.02.001>

Greenwald, A. G., & Farnham, S. D. (2000). Using the implicit association test to measure self- esteem and self-concept. Journal of Personality and Social Psychology, 79(6), 1022. <https://doi.org/10.1037/0022-3514.79.6.1022>

Greenwald, A. G., Poehlman, T. A., Uhlmann, E. L., & Banaji, M. R. (2009). Understanding and using the Implicit Association Test: III. Meta-analysis of predictive validity. Journal of Personality and Social Psychology, 97(1), 17. <https://doi.org/10.1037/a0015575>

Griffiths, T. L., & Tenenbaum, J. B. (2006). Optimal predictions in everyday cognition. Psychological Science, 17(9), 767–773. <https://doi.org/10.1111/j.1467-9280.2006.01780.x>

 Hájek, A., & Pettit, P. (2004). Desire beyond belief. Australasian Journal of Philosophy, 82(1), 77–92. <https://doi.org/10.1080/713659805>

Hauser, M., Cushman, F., Young, L., Kang-Xing Jin, R., & Mikhail, J. (2007). A dissociation between moral judgments and justifications. Mind & Language, 22(1), 1–21. <https://doi>. org/10.1111/j.1468-0017.2006.00297.x

Kruschke, J. K. (2013). Bayesian estimation supersedes the t test. Journal of Experimental Psychology. General, 142(2), 573. <https://doi.org/10.1037/a0029146>

Kruschke, J. K. (2015). Doing bayesian data analysis, Second Edition: A Tutorial with R, JAGS, and Stan. Academic Press.

Kruschke, J. K. (2018). Rejecting or accepting parameter values in Bayesian estimation. Advances in Methods and Practices in Psychological Science, 1(2), 270–280. <https://doi>. org/10.1177/2515245918771304

Leitgeb, H., & Pettigrew, R. (2010a). An objective justification of Bayesianism I: Measuring inaccuracy. Philosophy of Science, 77(2), 201–235. https://doi.org/10.1086/651317 28

Leitgeb, H., & Pettigrew, R. (2010b). An objective justification of Bayesianism II: The consequences of minimizing inaccuracy. Philosophy of Science, 77(2), 236–272. https:// doi.org/10.1086/651318

Lenman, J. (2003). Non-cognitivism and the dimensions of evaluative judgment. Brown Electronic Article Review Service, 15. <http://www.brown.edu/Departments/Philosophy/> bears/homepage.html

Lewis, D. (1988). Desire as belief. Mind, XCVII(387), 323–332. <https://doi.org/10.1093/> mind/XCVII.387.323

Lewis, D. (1996). Desire as belief II. Mind, 105(418), 303–313. <https://doi.org/10.1093/mind/> 105.418.303

Lockhart, T. (2000). Moral uncertainty and its consequences. Oxford University Press.

MacAskill, M., Bykvist, K., & Ord, T. (2020). Moral uncertainty. Oxford University Press..

Makins, N. (2021). Attitudinal ambivalence: Moral uncertainty for non-cognitivists. Australasian Journal of Philosophy, 1–15. <https://doi.org/10.1080/00048402.2021.1908380>

Makowski, D., Ben-Shachar, M., & Lüdecke, D. (2019). bayestestR: describing effects and their uncertainty, existence and significance within the bayesian framework. Journal of Open Source Software, 4(40), 1541. <https://doi.org/10.21105/joss.01541>

McGuire, J., Langdon, R., Coltheart, M., & Mackenzie, C. (2009). A reanalysis of the personal/impersonal distinction in moral psychology research. Journal of Experimental Social Psychology, 45(3), 577–580. <https://doi.org/10.1016/j.jesp.2009.01.002>

Moore, A. B., Clark, B. A., & Kane, M. J. (2008). Who shalt not kill? Individual differences in working memory capacity, executive control, and moral judgment. Psychological Science, 19(6), 549–557. <https://doi.org/10.1111/j.1467-9280.2008.02122.x>

Moore, A. B., Lee, N. L., Clark, B. A., & Conway, A. R. (2011). In defense of the personal/ impersonal distinction in moral psychology research: Cross-cultural validation of the dual process model of moral judgment. Judgment and Decision Making, 6(3), 186.

Nichols, S., & Folds-Bennett, T. (2003). Are children moral objectivists? ’Children’s judgments about moral and response-dependent properties. Cognition, 90(2), B23–B32. [https://doi.org/10.1016/S0010-0277(03)00160-4](https://doi.org/10.1016/S0010-0277%2803%2900160-4)

Nichols, S. (2004). After objectivity: An empirical study of moral judgment. Philosophical Psychology, 17(1), 3–26. <https://doi.org/10.1080/0951508042000202354>

Nissan-Rozen, I. (2015). A triviality result for the “Desire by Necessity” thesis. Synthese, 192 (8), 2535–2556. <https://doi.org/10.1007/s11229-015-0666-8>

Nissan-Rozen, I. (2017). Reasoning with comparative moral judgements: An Argument for Moral Bayesianism. In: Urbaniak, R., Payette, G. (eds). Applications of Formal Philosophy. Logic, Argumentation & Reasoning, vol 14. Springer, Cham. [https://doi.org/10.1007/978- 3-319-58507-9\_6](https://doi.org/10.1007/978-%20%093-319-58507-9_6)

Plummer, M. (2012). JAGS Version 3.3. 0 user manual. International agency for research on cancer. http://sourceforge.net/projects/mcmc-jags/files/ .

Pölzler, T. (2017). Revisiting folk moral realism. Review of Philosophy and Psychology, 8(2), 455–476. <https://doi.org/10.1007/s13164-016-0300-9>

R, Core Team. (2016). R: A language and environment for statistical computing. Foundation for Statistical Computing. <https://www.R-project.org/>

Ridge, M. (2007). Ecumenical expressivism: The best of both worlds? In R. Shafer-Landau (Ed.), Oxford Studies in Metaethics (Vol. 2). New York: Oxford.

Riedener, S. (2021). Uncertain Values: An Axiomatic Approach to Axiological Uncertainty. Berlin, Boston: De Gruyter. <https://doi.org/10.1515/9783110736199>.

Rouder, J. N. (2014). Optional stopping: No problem for Bayesians. Psychonomic Bulletin & Review, 21(2), 301–308. <https://doi.org/10.3758/s13423-014-0595-4>

Russell, J. S., & Hawthorne, J. (2016). General dynamic triviality theorems. Philosophical Review, 125(3), 307–339. <https://doi.org/10.1215/00318108-3516936>

Sarkissian, H., Park, J., Tien, D., Wright, J. C., & Knobe, J. (2011). Folk moral relativism. Mind & Language, 26(4), 482–505. <https://doi.org/10.1111/j.1468-0017.2011.01428.x>

Sepielli, A. (2012). Normative uncertainty for non-cognitivists. Philosophical Studies, 160(2), 191–207. <https://doi.org/10.1007/s11098-011-9713-6>

Sepielli, A. (2013). Moral uncertainty and the principle of equity among moral theories. Philosophy and Phenomenological Research, 86(3), 580–589. <https://doi.org/10.1111/j>. 1933-1592.2011.00554.x

Sinnott-Armstrong, W. (2009). Mixed-Up meta-ethics. Philosophical Issues, 19, 235–256. <https://doi.org/10.1111/j.1533-6077.2009.00168.x>

Smith, M. (2002). Evaluation, uncertainty and motivation. Ethical Theory and Moral Practice, 5(3), 305–320. [https://doi.org/10.1023/A:1019675327207](https://doi.org/10.1023/A%3A1019675327207)

Staffel, J. (2019). Expressivism, normative uncertainty, and arguments for probabilism. W: T.S. Gendler & J. Hawthorne (red.), Oxford Studies in Epistemology (t. 6). Oxford: Oxford University Press, 161–189. <http://dx.doi.org/10.1093/oso/9780198833314.003.0007>

Theriault, J., Waytz, A., Heiphetz, L., & Young, L. (2017). Examining overlap in behavioral and neural representations of morals, facts, and preferences. Journal of Experimental Psychology. General, 146(11), 1586. <https://doi.org/10.1037/xge0000350>

Tversky, A., & Koehler, D. J. (1994). Support theory: A nonextensional representation of subjective probability. Psychological Review, 101(4), 547. [https://doi.org/10.1037/0033- 295X.101.4.547](https://doi.org/10.1037/0033-%20%09295X.101.4.547)

Tversky, A., & Fox, C. R. (1995). Weighing risk and uncertainty. Psychological Review, 102 (2), 269. <https://doi.org/10.1037/0033-295X.102.2.269>

Valdesolo, P., & DeSteno, D. (2006). Manipulations of emotional context shape moral judgment. PSYCHOLOGICAL SCIENCE-CAMBRIDGE-, 17(6), 476. <https://doi.org/10>. 1111/j.1467-9280.2006.01731.x

van Roojen, M. (2018). Motivation, Recommendation, Non-Cognitivism, and the Naturalistic Fallacy. In N. Sinclair (Ed.), The Naturalistic Fallacy (Classic Philosophical Arguments, pp. 117-137). Cambridge: Cambridge University Press. [http://dx.doi.org/10. 1017/9781316717578.008](http://dx.doi.org/10.%20%091017/9781316717578.008)

Wagenmakers, E. J., Love, J., Marsman, M., Jamil, T., Ly, A., Verhagen, J., . . . Morey, R. D. (2018). Bayesian inference for psychology. Part II: Example applications with JASP. Psychonomic Bulletin & Review, 25(1), 58–76. [https://doi.org/10.3758/s13423-017-1323- 7](https://doi.org/10.3758/s13423-017-1323-%097)

Wainryb, C., Shaw, L. A., Langley, M., Cottam, K., & Lewis, R. (2004). ’Children’s thinking about diversity of belief in the early school years: Judgments of relativism, tolerance, and disagreeing persons. Child Development, 75(3), 687–703. <https://doi.org/10.1111/j.1467-> 8624.2004.00701.x

Wallsten, T. S., Budescu, D. V., & Zwick, R. (1993). Comparing the calibration and coherence of numerical and verbal probability judgments. Management Science, 39(2), 176–190. <https://doi.org/10.1287/mnsc.39.2.176>

Wright, J. C., Grandjean, P. T., & McWhite, C. B. (2013). The meta-ethical grounding of our moral beliefs: Evidence for meta-ethical pluralism. Philosophical Psychology, 26(3), 336–361. https://doi.org/10.1080/09515089.2011.633751

1. Smith referred also to a third type of degree associated with moral judgements, their degree of stability over time. Smith called this type of degree “robustness.” This third type of degree will not play a role in this paper. [↑](#endnote-ref-1)
2. Different non-cognitivist accounts point to different types of attitudes. See van Roojen (2018) for a good review. [↑](#endnote-ref-2)
3. Alternatively, the non-cognitivist can argue that most people, most of the time, fail to obey the normative ideal. Although in some contexts this is certainly the right move to take, we believe it is uncontroversial that, other things being equal, one should aim to avoid such conclusions. [↑](#endnote-ref-3)
4. This is an implicit assumption, for example, in almost all discussions of decision-making under normative uncertainty (see for example Lochhart, 2000; macAskill, Bykvist & Ord, 2020; Reidener. 2021). [↑](#endnote-ref-4)
5. As an anonymous referee commented, however, it seems unlikely that most moral psychologists would deny that people can be more or less confident in their moral judgements. A more plausible view would be that – much as many philosophers believe is the case with respect to non-moral beliefs – moral judgement may take either a binary (accept/reject) or a graded form. However, accepting that moral judgements may come in degrees of confidence does not imply accepting that these degrees have a probabilistic structure. This latter claim is what we examine in this paper. [↑](#endnote-ref-5)
6. The persistent disagreement among philosophers regarding the status of moral judgements may be taken to suggest that holding incorrect meta-ethical views regarding one's own meta-ethics (as reflected in one's behavior) is a very common phenomenon. [↑](#endnote-ref-6)
7. To limit the duration of the online questionnaires to 10 minutes, we limited each to a maximum of 26 questions. [↑](#endnote-ref-7)
8. An anonymous referee has raised the worry that our results might reflect responses that are trivial in some sense. The reviewer suggested that this might be the case with respect to some of the desire reports, which might reflect complete indifference with respect to the dilemmas (thus, leading the participants to report degrees very close to 50%). A related worry might be that, when it comes to beliefs, participants might report degrees close to 100% and 0%. To address these worries we checked the percentages of participants who reported degrees between 45% to 55% (6% for beliefs and 2% for desires) and the percentages of participants who reported degrees of either 100% or 0% (9% for beliefs and 15% for desires) and established that, even if the analysis is limited to the range of values that do not fall into one of these two categories, we obtain the same pattern of result as reported above. [↑](#endnote-ref-8)
9. Participants classified dilemmas as either moral or non-moral dilemmas. They were asked to rate each dilemma on a 7-point scale ranging from 1 (does not require a moral judgment) to 7 (requires a moral judgment). Finally, we defined a dilemma as belonging to the moral domain if more than 70% of participants scored it 5 or higher. [↑](#endnote-ref-9)
10. Level of importance was defined as follows: "High" for responses greater than 60, "Low" for responses lower than 40, and "Neutral" for responses between 40 and 60. [↑](#endnote-ref-10)
11. An anonymous referee has raised the worry that participants may have interpreted the word “appropriate” in a non-moral way (e.g., as referring to social conventions, or legal status). We do not believe this is the case, for three reasons. First, the questions were posed to the participants in Hebrew. The exact formulation was "עד כמה ראוי בעיניך ש...?". The Hebrew word "ראוי" bears a distinctive moral connotation lacking in the English “appropriate” (we debated whether to translate it here as “morally appropriate” but chose to stick with a verbatim translation). Second, in the Hebrew formulation there is an emphasis on one’s personal assessment. The verbatim translation of the Hebrew question is: “how much appropriate, in your eyes is…?” Third, as explained above, the moral dilemmas were classified as such by participants in a pilot experiment. [↑](#endnote-ref-11)
12. Following Note 8, here too, we have checked for “trivial” responses (i.e., responses around 50%, 100%, and 0%). Responses around 0% or 100%, were observed for beliefs in 8% of the responses, for desires in 13%, and moral judgments in 19%. Responses around 50% were observed for beliefs in 5% of the responses, for desires in 5% and for moral judgments in 5%. Even if the analysis is limited to the range of values that do not fall into one of these categories, we get the same result-pattern as reported above. [↑](#endnote-ref-12)
13. Cases in which the degree of belief in either a given proposition, x, or the proposition "x is morally appropriate," are equal to 1 or 0. [↑](#endnote-ref-13)
14. Importantly, both internalists and externalists about moral motivation can, should and usually do accept this claim. The difference between internalists and externalists about moral motivation is that the former, but not the latter, take the connection between moral judgments and moral motivation to be necessary. However, externalists too may, and usually do, accept that typically (but not necessarily) moral judgements restrict moral motivation. In the context of the DBT, both internalists and externalists should accept it as a descriptive thesis that usually holds, but only internalists should be troubled by Lewis’ result according to which it cannot always hold for a rational agent. [↑](#endnote-ref-14)
15. Participants classified dilemmas as either moral dilemmas or non-moral dilemmas. They were asked to rate each dilemma on a 7-point scale ranging from 1 (does not require a moral judgment) to 7 (requires a moral judgment). Finally, we defined a dilemma as belonging to the moral domain if more than 70% of participants scored it between 5 and 7. [↑](#endnote-ref-15)
16. Following Note 8, here too we have checked for “trivial” responses (i.e., responses around 50%, 100% and 0%). Responses around 0% or 100% were observed for beliefs in 8% of the responses, for desires in 17% and for moral judgments in 16%. Responses around 50%, observed for beliefs in 3% of the responses, for desires in 6% and for moral judgments in 6%. Even if the analysis is limited to the range of values that do not fall into one of these categories, we get the same result pattern as reported above. [↑](#endnote-ref-16)