

ARTICLE



## The moral behavior of ethics professors: A replication-extension in German-speaking countries

Philipp Schönegger <sup>a</sup> and Johannes Wagner<sup>a</sup>

<sup>a</sup>Department of Philosophy, University of St Andrews

### ABSTRACT

What is the relation between ethical reflection and moral behavior? Does professional reflection on ethical issues positively impact moral behaviors? To address these questions, Schwitzgebel and Rust empirically investigated if philosophy professors engaged with ethics on a professional basis behave any morally better or, at least, more consistently with their expressed values than do non-ethicist professors. Findings from their original US-based sample indicated that neither is the case, suggesting that there is no positive influence of ethical reflection on moral action. In the study at hand, we attempted to cross-validate this pattern of results in the German-speaking countries and surveyed 417 professors using a replication-extension research design. Our results indicate a successful replication of the original effect that ethicists do not behave any morally better compared to other academics across the vast majority of normative issues. Yet, unlike the original study, we found mixed results on normative attitudes generally. On some issues, ethicists and philosophers even expressed more lenient attitudes. However, one issue on which ethicists not only held stronger normative attitudes but also reported better corresponding moral behaviors was vegetarianism.

### ARTICLE HISTORY

Received 20 June 2018  
Accepted 26 August 2018

### KEYWORDS

Experimental philosophy;  
replication-extension; moral  
attitudes; moral behavior

## 1. Introduction

It is an open but highly relevant question how one can properly conceive the relation between ethical reflection and moral action. Max Scheler, when asked about the disparity between the ethical formalism he advocated and his own personal life famously remarked that “signposts do not walk in the direction they point to”<sup>1</sup> (Böhme, 2011, p. 47), thereby suggesting a gap between insights into moral truths and their embodiment in action.

**CONTACT** Philipp Schönegger  [ps234@st-andrews.ac.uk](mailto:ps234@st-andrews.ac.uk)  Eschensiedlung 100, 8530 Deutschlandsberg, Austria

We are especially thankful for the continued assistance of Eric Schwitzgebel, whose guidance has contributed greatly to this project in many ways.

We are also grateful for helpful comments, suggestions, and generous support from Florian Cova, Ferenc Kemény, Edouard Machery, Lukas Meyer, Norbert Paulo, Thomas Pölzler, Joshua Rust, Peter Singer, Markus Seethaler, and Pascale Willemsen.

 Supplemental data for this article can be accessed [here](#).

© 2019 Informa UK Limited, trading as Taylor & Francis Group

The fields of experimental philosophy and moral psychology have substantially contributed to our general understanding of moral intuitions and judgments (Haidt, 2001; Helion & Pizarro, 2015; Knobe & Nichols, 2008; Sinnott-Armstrong, 2008). However, few empirical studies have directly addressed the relation between ethical reflection and moral action. Research in this area struggles to derive clearly testable hypotheses from the underlying philosophical questions. An obvious hindrance for empirical testing is that serious ethical reflection hardly seems to be experimentally inducible.

One way to circumvent this issue is to investigate the attitudes and behaviors of ethics professors, as ethicists are assumed to engage with well-above-average intensity in ethical reflection in virtue of their professional occupation. Yet, most existing studies on this topic have been primarily concerned with lay people (e.g. Gold, Pulford, & Colman, 2015), which is surprising, as the element of professional engagement has been successfully used in other research contexts to operationalize constructs that are hard to experimentally manipulate (e.g. Horvath & Wiegemann, 2016). Hence, comparing professional ethicists' moral attitudes and behaviors to appropriate reference groups remains a promising way to empirically test aspects of the relationship between ethical reflection and moral action.

In a series of studies by Eric Schwitzgebel, co-authored with Joshua Rust (2009, 2010, 2011, 2013, 2014) and Fiery Cushman (2012, 2015), the empirical relations between the normative attitudes and moral behaviors of professional ethicists have been investigated systematically. Their research covered a variety of methodologies and topics like evaluations of peer opinion concerning ethicists' moral behavior, research on order effects concerning ethical intuitions in trolley cases, and ethicists' voting behavior. In their most well-known study (2014), Schwitzgebel and Rust compared the self-reported and directly observed moral behaviors of professional ethicists with their espoused normative views to determine their consistency. As their findings proved to be both empirically informative and highly relevant to how one thinks about the relation between ethical reflection and action, this underscores the value of investigating ethics professors to understand the nature and corollaries of ethical reflection. In order to contribute to and validate this pioneering work, we herewith conducted a replication attempt of Schwitzgebel and Rust's seminal study in German-speaking countries.

A few points about the basic assumptions of this kind of expertise-based research are in place. According to Schwitzgebel and Rust, this research format rests on two plausible but "empirically open" (2011, p. 6) foundational presuppositions: First, that professional ethicists are more likely to engage in philosophical moral reflection or that their reflection is of superior quality compared to academics from other disciplines,

and second, that there are no substantial differences between ethicists and other professors except the intensity or quality of moral reflection. These two assumptions are central to build the case that a different intensity of ethical reflection is, in fact, the main factor in potential group differences. If these assumptions are plausible, then it can be argued that investigating the moral behavior of ethicists might offer insights into the relationship between ethical reflection and behavior.

Furthermore, being empirical in character, this type of research does not make any claim for or against the existence, possibility, or nature of moral truths, nor does it make a claim on how or to what degree such truths are accessible to human insight. One major task is thus to clarify the models one intends to put to the test and to clearly separate their testable aspects from their untestable ones. In order to separate the competing views and to determine which of their aspects can actually be subjected to empirical testing, Schwitzgebel and Rust (2014, p. 295–296) formulated four broad models of the relation between ethical reflection and behavior. They think that research on the moral behavior of ethics professors could potentially provide evidence for or against these models.

First, the *booster view* holds that philosophical reflection leads to the discovery of moral truths and that these moral truths have a significant impact on behavior. This view has been defended by a large array of philosophers, from the Greek and Roman stoics (Cooper, 2012) to most of the western (Duignan, 2011) and eastern (Ivanhoe, 2000)<sup>2</sup> philosophical traditions.

Second, the *rationalization view* claims that ethical reflection does not alter a person's behavior but, rather, leads the person to selectively accept those ethical views that match their already existing habits and behaviors. This view has empirical basis in well-established psychological effects of motivated reasoning (Ditto, Pizarro, & Tannenbaum, 2009; Uhlmann, Pizarro, Tannenbaum, & Ditto, 2009) and in the phenomenon of moral dumbfounding (Björklung, Haidt, & Scott, 2000; McHugh, McGann, Igou, & Kinsella, 2017), which suggest that moral judgments and behaviors display resilient tendencies when confronted with evidence sufficient to warrant adaptation.

Third, according to the *inert discovery view*, ethical reflection leads to the discovery of moral truths, but it does not lead to a corresponding change in behavior. In other words, despite the attainment of moral insights through reflection, no change in behavior occurs.

Fourth and last, the *epiphenomenalist view* contends that philosophical reflection generally has no influence on either moral attitudes or behavior. This is to say, insights into moral truths would not even affect the beliefs of a person, let alone their behaviors.

The question is this: Which aspects of these models can be empirically investigated? While it is not possible to test the existence of moral truths or

whether they are discoverable by human reflection, it is possible to inquire whether people themselves believe in moral truths and inquire about their relation to ethical reflection (as we did in our extension). Furthermore, if ethical reflection actually leads to some moral insights, professional ethicists might be assumed to have more of such insights on average, because of either the intensity or the frequency of ethical reflection they engage in. Moreover, if ethical reflection has any determinable effects on persons, this should be reflected in the normative attitudes that they hold, which are empirically leviably. Thus, comparing ethicists with appropriate reference groups might allow one to draw conclusions about the effects of extensive moral reflection. On the other hand, the morality of actions can be investigated either by observing or by asking for self-reports of behaviors that are normatively relevant. Normative relevance can be defined in three ways: (1) On behalf of a particular moral outlook, (2) on the basis of common-sense plausibility, or (3) relative to participants' own normative attitudes. Following Schwitzgebel and Rust (2014), we relied on the two latter options in order to operationalize the normativity of behaviors in a theory-independent way. Finally, the consistency between normative attitudes and moral behaviors can be empirically tested in a similar way.

In their original 2014 study, Schwitzgebel and Rust investigated whether the moral attitudes and behaviors of ethics professors offer support for any of the four models outlined above. To this end, they reached out to 980 tenured, tenure-track, and emeritus professors from universities in five US states,<sup>3</sup> resulting in 198 ethicists, 208 non-ethicist philosophers, and 167 non-philosophy professors taking the survey. In the first part of the questionnaire, participants were asked for their normative evaluations of a wide array of potentially normatively charged issues, ranging from voting and blood donation to response rates to student e-mails, vegetarianism, and charity. In the second part, participants were asked to self-report their behavior on most of these topics.

In our replication attempt, we aimed to reproduce three major types of effects that Schwitzgebel and Rust (2014) observed: First, that ethicists expressed “somewhat more stringent normative attitudes on some issues” (p. 293) compared with the non-ethicist philosopher and non-philosopher groups; second, that ethicists did not show “unequivocally better behavior” on any of the specified normative issues in comparison to the two reference groups (p. 293); and, third, the mixed effects in terms of attitude-behavior relationship, with ethicists showing the strongest relationship on the topic of “voting” but the weakest on “charitable donation” (p. 293).

Though Schwitzgebel and Rust (2014) took their results to be rather inconclusive with regard to their proposed four explanatory models, they did feel sufficiently confident to “reject the view that ethicists behave, on average, morally better than do non-ethicists” (p. 320). This assertion was

based on their main finding that ethics professors expressed more pronounced normative attitudes whilst not correspondingly standing out on any of the behavioral measures, both in comparison to the two reference groups. In light of these findings, the authors issued three possible explanations: First, that there might be no impact of moral reflection on moral behavior; second, that moral reflection might have facilitating as well as inhibitory influences on manifesting corresponding behaviors so that the respective effects do not manifest on average; and, third, they conceded that “one of the background assumptions of our research [might be] false” (p. 320).

Our replication-extension study intended to shed further light on these possible explanations by (1) attempting to replicate Schwitzgebel and Rust’s (2014) primary findings and (2) extending the survey with focused questions that might help explain the pattern of results. Hence, while we implemented the original research design as accurately as possible, we extended the original questionnaire by three additional items. In this extension, we asked participants how they themselves conceive of the relation between ethical reflection, moral truth, and behavior. We also asked them to what degree they agree that “the belief that something is wrong is sufficient to act accordingly,” as we wanted to locate participants on the internalism-externalism spectrum of moral motivation.

In the main body of the survey, we attempted to cross-validate Schwitzgebel and Rust’s pattern of results in German-speaking countries. A successful replication in a different cultural context would strengthen the evidence base of the alleged effect. This is because of the large number of differing factors such as language, culture, institutions, history, tradition, and time, given that this replication takes place nearly a decade after the original survey was conducted.<sup>4</sup>

Given Schwitzgebel and Rust’s three major findings, we attempt to replicate the findings that ethicists, on average, have the most demanding normative attitudes, that ethicists, on average, do not behave better, and that attitude-behavior consistency measures are mixed and do not offer a clear directional picture. Against this backdrop, we derived the following null hypotheses for our replication attempt: (1) Ethicists, non-ethicist philosophers, and non-philosophers do not differ in the stringency of their normative assessments in the normative issues specified; (2) ethicists, non-ethicist philosophers, and non-philosophers do not differ in their moral behaviors on any of the normative issues specified; and (3) ethicists, non-ethicist philosophers, and non-philosophers do not differ in their overall attitude-behavior consistency.

## 2. Methods

This study constitutes a replication attempt of Schwitzgebel and Rust’s study (2014), together with an extension. First and foremost, we aim to

replicate the effects of the original study in a different context, while staying as close as possible to the original methods and design. By replicating the survey, part of the study, and following Brandt et al. (2014), this part of the replication could be aptly located towards the “close replication” (Brandt et al., 2014, p. 218) part of the spectrum ranging from close to conceptual replication. However, while Schwitzgebel and Rust (2014) had direct observational data of participants’ behavior on issues such as voting (through voting records) or e-mail responsiveness (through a different experiment (Schwitzgebel & Rust, 2013)), our present study has not gathered any observational data. In this sense, the project as a whole would be better classified as a conceptual replication, though we adopted procedures as close to those of the original study as possible. Furthermore, in adding an extension, we aimed to “extend the results of the prior stud[y]” (Bonett, 2012, p. 409) in order to explore potential explanations of the results found in the original study.

For our study, we reached out to 418 ethics professors, 527 non-ethicist philosophy professors, and 521 non-philosophy professors from 62 German-speaking universities located in Germany, Austria, and Switzerland. The term “professor” was defined as any person holding a PhD and occupying an academic position in their respective university. This was done to ensure cross-country comparability, as different systems of titles are present in the countries surveyed and in the United States, where the original study was conducted.

We contacted potential participants selectively based on their areas of expertise, as evidenced by information available on their public departmental websites. Where the information was insufficient to classify a professor as either ethicist or non-ethicist philosopher, we did not contact this person. We took the listing of “ethics,” “applied ethics,” “normative theory,” “ethics of X,”<sup>5</sup> or similar terms at least once on a philosopher’s page as both sufficient and necessary to classify him or her as an ethicist. If no such terms were present in a philosopher’s stated areas of specialization or expertise, he or she was classified as a non-ethicist philosopher. For the non-philosopher comparison group, we contacted linguists, philologists, historians, cultural scientists, archaeologists, and people from related fields at the same universities that the ethicists and non-ethicist philosophers were sampled from, in order to counterbalance possible differences in salary, social status of disciplines, and other possible confounds based on locality. To grant full anonymity to all participants, we generated three different survey links to assign people to groups according to our CV-based classification. We did not consider information pertaining to age, gender, or exact academic position.

We sent out two e-mails with several days in between. In the first e-mail, we wanted to establish personal contact by inviting recipients to participate

in the study and sketching the topic of the survey. In the second e-mail, we repeated relevant facts of the first but, this time, embedded the actual link in the online survey.<sup>6</sup> All e-mails were sent out manually, as we wanted to address each participant personally. This procedure was chosen to establish better personal contact with potential participants, thus making participation in the survey more inviting. Unlike the original study, we did not reach out to participants via hard copy letters in addition to e-mail. Even in the original study, which was conducted nine years ago, the hard copy modality only accounted for around 20% of the responses (E. Schwitzgebel, personal communication, 21 February 2018). Given the almost exclusive use of e-mails in academia in 2018, we considered utilizing e-mail exclusively to be a minor limiting factor.

We did not send our survey to 59 persons that were on our initial mailing list, as they either had malfunctional e-mail-addresses, made explicit statements that they did not want to participate in response to the first mail, or gained direct awareness of our project and its goals. We effectively sent our survey link to 1415 professors, and we received sufficiently completed surveys from 151 ethics professors, 133 non-ethicist philosophy professors, and 133 non-philosophy professors (a response rate of 29.5%). The original study achieved an overall response rate of 58%, a difference that we think is due to a number of reasons: First, both of the original authors were already professors; second, our study did not include a ten-dollar charity incentive; and, third, we did not send out printed versions of the survey.

With the exception of the extension items, the items of the survey were virtually identical to those of the original study, other than that we translated them into German. If direct translation of the item wording did not make sense due to cultural or institutional reasons (like differing election systems or varying legal structures), then the questions were changed slightly to neutralize the differences. For instance, whereas the original study asked whether one's driver license included a "statement or symbol indicating willingness to be an organ donor in the event of death" (Schwitzgebel & Rust, 2014, p. 308), our version of the item only asked for the willingness to be an organ donor, as the respective legal systems of Austria, Germany, and Switzerland encompass both opt-in and opt-out systems.

The main body of the survey consisted of three parts: The first part asked participants for their normative assessments of a wide array of normative issues. These issues were theft, membership in an academic society, voting in public elections, talking to one's mother, eating meat,<sup>7</sup> willingness to donate one's organs in the case of death, donating blood, answering student e-mails, giving 10% of one's income to charity, and honesty in surveys. The second part required participants to self-report their behavior on these very same

topics. Examples included (but were not limited to) questions such as how much one donated to charity in 2017, how many meals a week included the meat of mammals, or how often the person had voted in the course of the last ten nation-wide elections. The third part was presented to ethicists and non-ethicist philosophers only, and they were asked on which level(s) of abstraction they deliberate on ethical problems: metaethics, normative ethics, or applied ethics (with the fourth option being “ethics is not among my specializations”). Furthermore, participants were asked to state which of the following was their preferred normative theory: deontology, consequentialism, virtue ethics, skepticism, or “no settled position.”

Like the original study, we used a 9-point Likert scale ranging from 1 (very morally bad/strongly disagree)<sup>8</sup> through 5 (morally neutral/neither agree nor disagree) to 9 (very morally good/strongly agree) to assess normative attitudes and our extension questions. For the question of how much one ought to give to charity and the behavioral measures, we used a variety of item formats, but we strictly followed the original study when choosing which specific format to use for which question. Response formats ranged from adjusting a controller such that a specific percentage value was assigned, a blank field for a simple integer to enter, and multiple or single choice format.<sup>9</sup>

At the beginning of part 1, we also cautioned participants in the fashion of the original by stating that we were aware that ranking moral badness and goodness along a scale may be difficult, that there may be other ways to speak of this subject matter, such as through concepts of permissibility or duty, and that the normative status of actions might be highly context-dependent. However, we encouraged participants to interpret the questions as straightforwardly as possible and set aside such considerations as far as possible.

In addition to the main body of our survey, we extended the original survey by three additional items on attitudes pertaining to (1) reasons (internalism/externalism) for moral motivation, (2) moral truth, and (3) moral behavior. All three items were phrased as statements without philosophical jargon, and participants were asked to indicate if they agreed or disagreed with the three statements on a 9-point Likert scale. The first statement claimed that the belief that some act is wrong is sufficient to act accordingly, the second that philosophical moral reflection leads to moral truths, and the third that moral reflection has a positive impact on moral action. These items were included in order to provide some possible explanation for a possible gap between normative attitudes and behavior. We also hoped to receive some indication of how participants themselves conceive of the relationship between normative attitudes and behavior with regards to the four models introduced above.

Generally speaking, all replication studies are subject to a certain number of limitations. This replication was conducted in German-speaking countries rather than in the United States, thereby testing the “study’s assumptions in new contexts” (Brandt et al., 2014, p. 222), which is a hallmark of replication studies. As a consequence, there were a number of necessarily differing factors present in our research that could have impacted our results. This is always the case with “close replication attempts” (Brandt et al., 2014, p. 218) as there is “no such thing as an exact replication” (Schmidt, 2009, p. 92). We do, however, believe that none of these factors had a distorting influence on our study or its results.<sup>10</sup>

### 3. Results

#### 3.1. Normative attitudes

Generally speaking, in our study we found mixed results on the question of ethicists’ normative attitudes. On five normative topics, no significant differences emerged between ethicists, non-ethicist philosophers, and non-philosophers. Unlike the original study, we did not find ethicists unequivocally expressing stronger normative attitudes when there were significant differences between groups. Rather, we found a mix of ethicists being both on the stringent but also on the lenient side of significant group differences in normative assessments across various issues. In terms of summary statistics of normative attitudes, we found that it wasn’t the ethicists but the non-philosophers who expressed significantly more stringent attitudes when compared to non-ethicist philosophers. Furthermore, when significant differences did emerge, ethicists and non-ethicist philosophers often concurred in their normative assessments bidirectionally, that is, in both the directions of stringency and leniency. In terms of the individual normative topics, ethicists were more stringent on how much one ought to donate to charity compared to the two reference groups, and they were more lenient on staying in touch with one’s mother compared to non-philosophers. Together with non-ethicist philosophers, they expressed a more stringent attitude towards vegetarianism, and they were more lenient on the issues of theft, voting, and paying membership fees to one’s academic society. Generally, our data failed to show significant differences in how many questions were rated as neutral: 1.9 (ethicists) versus 2.1 (non-ethicist philosophers) versus 2.4 (non-philosophers), as opposed to the findings of the original study: 1.8 versus 2.1 versus 2.7 for ethicists, non-ethicist philosophers, and non-philosophers respectively.

In concordance with the original study, we utilized proportional analysis<sup>11</sup> and found significant differences on the normative topics “theft,” “paying fees to one’s academic society,” and “vegetarianism.” On

group mean comparisons, we found significant differences on these same topics, but additional differences emerged on the issues of “voting,” “staying in touch with one’s mother,” and “charity in percentage” (what percentage of income an average professor ought to give to charity). On the normative issues of “organ donation,” “blood donation,” “responding to student e-mails,” “giving 10% of one’s income to charity,” and “honesty in this survey,” no differences between groups were found on either measure. For a summary of group-mean comparisons across all normative issues, see Table 1.

In our sample, all normative-attitude variables as well as all continuous-behavior variables and extension variables deviated significantly from normal distribution (Kolmogorow-Smirnow tests for normal distribution, all  $p < .01$ ). Like the original study, we conducted our analyses even if we found violations of normal distribution. Despite normal distribution being a precondition for analysis of variance, simulation studies have shown that one-factor ANOVAs are relatively robust against violations of normality (Blanca, Alarcón, Arnau, Bono, & Bendayan, 2017) if the sample size is large enough (Ghasemi & Zahediasl, 2012 p. 486; Pallant, 2007, p. 179–200), which is the case in our study.

### 3.1.1. Theft

When asked how good or bad it was to steal 1000 euros from a friend’s house, for example, 87.5% of non-philosophers rated stealing 1000 euros at the extreme end of the scale (1: very morally bad). Ethicists and non-ethicist philosophers were much less likely to do so, with 66.2% and 69.2% respectively rating it as very morally bad,  $\chi^2 = 16.7$ ,  $p < .001$ . The same pattern

**Table 1.** Normative attitudes.

*Normative Attitudes: Mean differences between groups (ethicists, non-ethicist philosophers, and non-philosophers) with regards to normative attitudes on different normative issues*

	Evaluating statements on a 9-point Likert-scale From 1 (“Very morally bad”) to 9 (“Very morally good”)				
	ANOVA p (sig.)	Ethicists	Non-Ethicist Philosophers	Non- Philosophers	
1. Committing theft of 1000€	.002	8.61 <sup>a</sup>	8.61 <sup>a</sup>	8.84 <sup>b</sup>	reverse
2. Paying academic membership fees	.000	6.00 <sup>ab</sup>	5.92 <sup>a</sup>	6.63 <sup>b</sup>	
3. Regularly voting in public elections	.000	6.92 <sup>a</sup>	6.63 <sup>a</sup>	7.52 <sup>b</sup>	
4. Not regularly talking to one’s mum	.023	6.35 <sup>a</sup>	6.44 <sup>ab</sup>	6.75 <sup>b</sup>	reverse
5. Regularly eating meat of mammals	.000	6.43 <sup>a</sup>	6.35 <sup>a</sup>	5.80 <sup>b</sup>	reverse
6. Donate organs after death	.685	7.18	7.34	7.28	
7. Regularly donating blood	.873	7.09	7.18	7.10	
8. Not answering student e-mails	.837	6.30	6.38	6.28	reverse
9a. Donating 10% of income to charity	.623	7.50	7.32	7.35	
10. Not being honest in this survey	.459	6.55	6.56	6.74	reverse
9b. How much % of income should a professor donate to charity per year?	.019	6.92% <sup>a</sup>	5.14% <sup>ab</sup>	4.69% <sup>b</sup>	

**abc** indicating sig. group differences (Tukey).

emerged via variance analysis,  $F(2, 407) = 6.97, p = .019, \eta^2 = .033$ ; significant post-hoc comparisons were as follows: ethicists – non-philosophers  $p = .003, d = .44$ , non-ethicist philosophers – non-philosophers  $p = .004, d = .41$ .

### *3.1.2. Paying academic membership fees to one's academic society and voting in public elections*

Proportional analyses showed that 65% of non-philosophers rated membership in one academic society as morally good, while 56% of ethicists and only 48% of non-ethicist philosophers did so. Differences between non-ethicist philosophers and non-philosophers were significant,  $\chi^2 = 7.927, p < .001$ . Group mean comparisons revealed that non-philosophers judged paying academic membership fees to be more morally good than both ethicists and non-ethicist philosophers did,  $F(2, 406) = 11.10, p = .001, \eta^2 = .052$ ; significant post-hoc comparisons: ethicists – non-philosophers  $p = .001, d = .47$ , non-ethicist philosophers – non-philosophers  $p = .001, d = .52$ , a finding that deviates from the results in the original study.

Non-philosophers also rated voting as significantly more morally good than ethicists and non-ethicist philosophers did,  $F(2, 408) = 11.74, p < .001, \eta^2 = .054$ ; significant post-hoc comparisons: ethicists – non-philosophers  $p = .003, d = .41$ , non-ethicist philosophers – non-philosophers  $p = .001, d = .56$ . Again, no such effect had manifested in the original study. Interestingly, as in the case of theft, it was the non-philosophers who differed from both of the other groups. No significant differences emerged in the corresponding proportional analysis.

### *3.1.3. Not regularly talking to one's mother*

Non-philosophers judged not keeping in contact with one's mother as worse than did ethicists, while non-ethicist philosophers did not differ significantly from any of the two,  $F(2, 401) = 3.75, p = .024, \eta^2 = .018$ ; significant post-hoc comparisons: ethicists – non-philosophers  $p = .021, d = .3$ . On proportional analysis, no differences were found with 65% of ethicists, 70% of non-ethicist philosophers, and 75% of non-philosophers rating not keeping regular contact with one's mother on the bad side of the scale,  $\chi^2 = 3.507, p = .173$ .

### *3.1.4. Percent of income that ought to be given to charity*

We effectively assessed the normative attitude on charity in two ways: First, we let people rate on the typical Likert scale the moral badness or goodness of giving 10% of one's income to charity, where we found no group differences. We also asked participants how much they think an average professor ought to give to charity. On this second issue, a group difference emerged, as ethicists ( $M = 6.9\%$ ,  $SD = 9.6$ ) differed from non-philosophers

( $M = 4.6\%$ ,  $SD = 4.0$ ) but not from non-ethicist philosophers ( $M = 5.1\%$ ,  $SD = 5.5$ ),  $F(2, 414) = 3.99$ ,  $p = .019$ ,  $\eta^2 = .019$ ; significant post-hoc comparisons: ethicists – non-philosophers  $p = .023$ ,  $d = .29$ . If one excludes statistical outliers of demands of 30% or more, the group difference still remains significant,  $F(2, 411) = 3.02$ ,  $p = .05$ ,  $\eta^2 = .014$ ; however, post-hoc tests fail to establish significant group differences. When excluding those who entered “0,” ethicists entered a mean of 9.1%, with non-ethicist philosophers entering 7.8% and non-philosophers 6.7%, a difference that is only marginally significant,  $F(2, 291) = 2.97$ ,  $p = .053$ ,  $\eta^2 = .020$ . Moreover, groups did not differ with regards to whether professors should not give to charity at all: 24% of ethicists, 34% of non-ethicist philosophers, and 30% of non-philosophers entered “0” on this question. This differs from the original result, which observed only 9% of ethicists entering zero, but with 24% of non-ethicist philosophers and 25% of non-philosophers doing so (Schwitzgebel & Rust, 2014, p. 312).

### 3.1.5. Vegetarianism

Most interesting, perhaps, were the findings on vegetarianism. Ethicists and non-ethicist philosophers judged eating meat to be morally worse compared to non-philosophers,  $F(2, 405) = 8.20$ ,  $p = .001$ ,  $\eta^2 = .039$ ; sig. post-hoc comparisons: ethicists – non-philosophers  $p = .001$ ,  $d = .45$ , non-ethicist philosophers – non-philosophers  $p = .005$ ,  $d = .39$ . On proportional analysis, 67% of ethicists and 63% of non-ethicist philosophers rated eating meat as morally bad, both significantly more than only 39% of non-philosophers,  $\chi^2 = 23.5$ ,  $p < .001$ . Again, ethicists and non-ethicist philosophers differed significantly from non-philosophers but not from each other. This pattern of results concurs with the findings of the original study.

### 3.2. Moral behavior

Like the original study, we found no consistent differences in moral behaviors between ethics professors and control groups, both for individual normative issues and on overall assessments, with the exception of vegetarianism.

No significant differences in self-reported behavior were found on the issue of membership in academic societies, where 89% of ethicists, 87% of non-ethicist philosophers, and 92% of non-philosophers reported being in one,  $\chi^2 = 1.592$ ,  $p = .451$ . Similarly, the groups did not differ in their reports of contact with their mothers: 87% of ethicists reported keeping in contact at least twice a month, compared to 81% of non-ethicist philosophers and 89% of non-philosophers,  $\chi^2 = 3.815$ ,  $p = .559$ . There was also no difference in whether or not one would be willing to be an organ donor,

$\chi^2 = 1.164$ ,  $p = .148$ , or whether one donated blood at least once a year,  $\chi^2 = 2.152$ ,  $p = .341$ . This failure to distinguish groups on these items was also present in the original study. Differing from this pattern, however, is the self-reported percentage of student e-mails that are answered, where non-ethicist philosophers reported answering less e-mails than non-philosophers,  $F(2, 398) = 4.65$ ,  $p = .01$ ,  $\eta^2 = .023$ ; significant post-hoc comparisons: non-ethicist philosophers – non-philosophers  $p = .009$ ,  $d = .34$  (for group comparison on continuous variables, see Table 2).

When asked how much one had given to charity in 2017, and without counting those responses that entered “0,” ethicists reported having donated 4.6%, compared to non-ethicist philosophers’ 4.6% and non-philosophers’ 4.4%,  $F(2, 316) = .060$ ,  $p = .942$ ,  $\eta^2 < .001$ . Even when including the “0,” differences remained non-significant,  $F(2, 398) = .66$ ,  $p = .518$ ,  $\eta^2 = .003$  (see Table 2).

One major deviation from the original study with regards to moral behaviors emerged with the issue of vegetarianism. Whereas the original findings on this item remained inconclusive, our data presents a clear picture. In our sample, only 24% of ethicists reported having eaten meat at the last dinner, compared to 40% of non-ethicist philosophers and 39% of non-philosophers,  $\chi^2 = 9.308$ ,  $p = .01$ , a difference that is both significant and relevantly different from the original. Moreover, ethicists also reported significantly less meals containing meat per week: 2.1, compared to non-ethicist philosophers’ 2.8 and non-philosophers’ 3.0,  $F(2, 398) = 3.44$ ,  $p = .034$ ,  $\eta^2 = .017$ . The difference between ethicists and non-philosophers proved to be significant, ethicists – non-philosophers  $p = .038$ ,  $d = .31$  (see Figure 1). The fact that both measures of self-reported behavior established significant differences between ethicists and non-philosophers supports confidence in the validity of this exceptional finding.

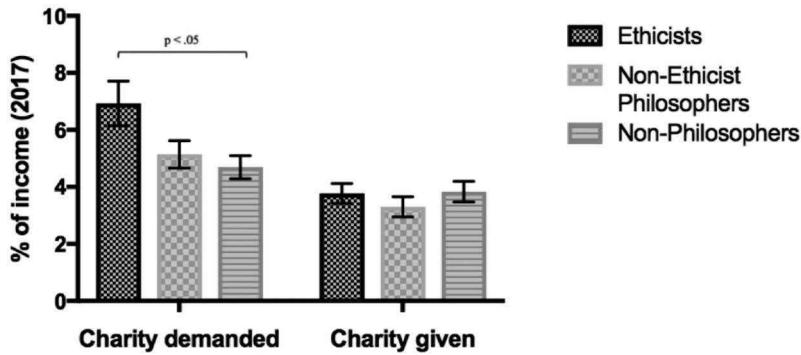
**Table 2.** Moral behaviors.

*Moral Behaviors: Mean differences between groups (ethicists, non-ethicist philosophers, and non-philosophers) with regards to normative attitudes on different normative issues*

	Self-report measures of moral behavior			
	ANOVA p (sig.)	Ethicists	Non-Ethicist Philosophers	Non- Philosophers
1. Number of votes cast during the last 10 elections	.128	8.77	8.63	9.21
2. Last contact with mum in days	.234	3.77	4.01	2.98
3. Number of meals containing meat per week	.034	2.16 <sup>a</sup>	2.81 <sup>ab</sup>	2.98 <sup>b</sup>
4. Percentage of student e-mails usually answered	.010	94.9 <sup>ab</sup>	90.8 <sup>a</sup>	96.7 <sup>b</sup>
5. Percent of income donated to charity in 2017	.518	3.77	3.30	3.83
6. Months since last blood donation	.595	140.1	140.4	116.3

□ Higher values indicate less moral behaviors.

**abc** indicating sig. group differences (Tukey).



**Figure 1.** Normative attitudes towards eating meat and meals containing meat per week.

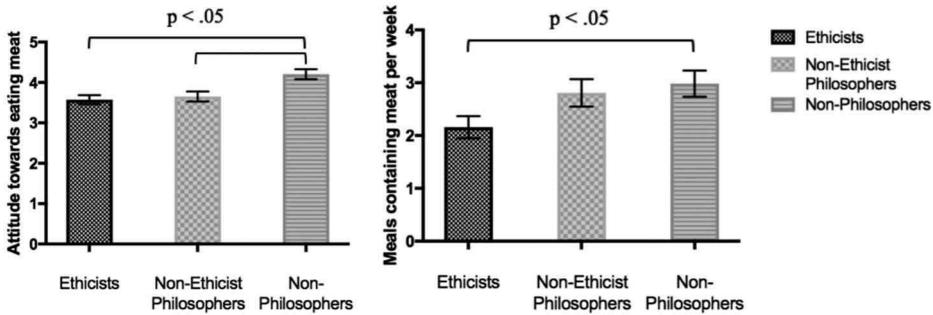
It is also of special interest to point to the difference in means of weekly meat consumption between our sample and the original one. Schwitzgebel and Rust (2014) report means of 4.1 ( $SD = 1.1$ ) meals for ethicists, 4.6 ( $SD = 1.1$ ) for non-ethicist philosophers, and 5.3 ( $SD = 1.2$ ) for non-philosophers. This differs greatly from our observed counterparts of 2.2 ( $SD = 2.52$ ), 2.8 ( $SD = 3.00$ ), and 3.0 ( $SD = 2.80$ ), respectively. Though this was not our actual research question, we found that all three groups ate significantly less meat than the groups of the original study,  $F(1, 941) = 64.85$ ,  $p = .001$ ,  $\eta^2 = .064$ , thereby suggesting that there are much more pronounced differences between cultures than between groups.

### 3.3. Attitude-behavior consistency

There are two broad ways of measuring attitude-behavior consistency: Through evaluating differences across groups in attitudes and behaviors, respectively, or through within-group correlational measures that can then be compared across groups in a further step.

In terms of group differences, we find considerable attitude-behavior consistency on vegetarianism for ethicists. Ethicists view eating meat as worse than do non-philosophers, and ethicists also eat less meat, again in comparison to non-philosophers (see Tables 1, 2; Figure 1). Conversely, no such consistency was found on the topic of charity, where ethicists demanded remarkably higher levels of charity than non-philosophers, while showing no corresponding difference in self-reported charitable giving (see Tables 1,2; Figure 2).

Secondly, we looked at correlational measures: We selectively report consistency on issues where relevant group differences emerged or issues that were mentioned by Schwitzgebel and Rust (2014). To start with, in terms of vegetarianism, we found significant correlation-based consistency within all three groups (ethicists,  $r = .347$ ,  $p = .001$ ; non-ethicist philosophers,  $r = .342$ ,  $p$



**Figure 2.** Charity demanded compared to charity given.

=.001; non-philosophers,  $r = .28$ ,  $p = .001$ ). On charity, ethicists showed a significant but weak relationship between how much one ought to give and how much one reported giving. This might seem to contradict the finding that ethicists are inconsistent on charity in terms of group comparisons, yet this has to be put into the context of the fact that both reference groups showed disproportionate correlations. Thus, while ethicists showed mild but significant correlational consistency, non-ethicist philosophers were clearly more consistent correlation-wise; what is more, both ethicists and non-ethicist philosophers are trumped by non-philosophers, who exhibited an extraordinarily high correlational consistency of .73 (ethicists,  $r = .17$ ,  $p = .048$ ; non-ethicist philosophers,  $r = .30$ ,  $p = .001$ ; non-philosophers,  $r = .73$ ,  $p = .001$ ). Hence, the relative inconsistency of ethicists on charity is indeed corroborated by this finding, not contradicted.

Despite no relevant group differences emerging on other issues in terms of attitude-behavior consistency, we also looked at the correlational consistency of voting and staying in touch with one's mother, for the sake of comparison with what was reported in the original study. On voting, we only found a significant attitude-behavior relationship for non-ethicist philosophers (ethicists,  $r = .09$ ,  $p = .264$ ; non-ethicist philosophers,  $r = .43$ ,  $p = .001$ ; non-philosophers,  $r = .03$ ;  $p = .722$ ). Interestingly, the original study found ethicists to be rather consistent on voting but the non-ethicist philosopher less consistent. With regards to contact with one's mother, our results, very much resembling Schwitzgebel and Rust's (2014), found considerable consistency in non-ethicist philosophers ( $r = .27$ ,  $p = .005$ ) but none in the other groups (ethicists,  $r = .13$ ,  $p = .175$ ; non-philosophers,  $r = .08$ ,  $p = .437$ ).

### 3.4. Aggregate measures of normative attitudes and moral behaviors

To gather evidence to illuminate whether ethicists generally stand out with respect to normative attitudes and moral behaviors compared to the

**Table 3.** Aggregate scores.

*Aggregate scores for normative attitude and moral behaviors: Group comparisons*

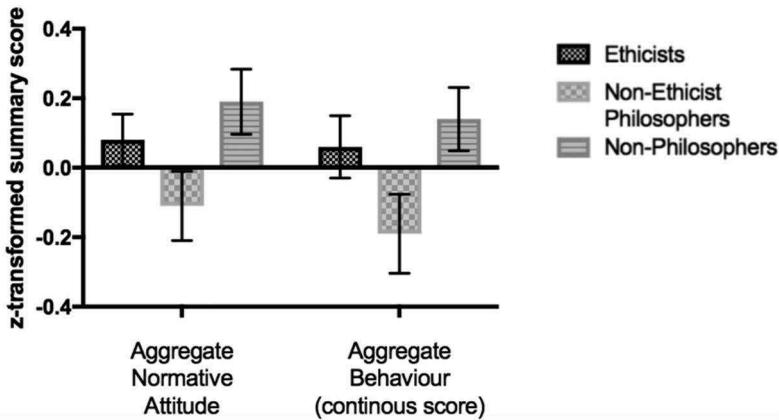
<i>Aggregate sum of z-transformed variables:</i>	<i>p (sig.)</i>	<i>Ethicists</i>	<i>Non-Ethicist Philosophers</i>	<i>Non-Philosophers</i>
A Stringency of Normative Attitudes	.031	0.08 <sup>ab</sup>	-0.11 <sup>a</sup>	0.19 <sup>b</sup>
A' without Vegetarianism	.001	-0.12 <sup>a</sup>	-0.14 <sup>a</sup>	0.27 <sup>b</sup>
B1 Moral Behaviors (continuous z-score summary score of all eight behavioral issues)	.048	0.06 <sup>ab</sup>	-0.19 <sup>a</sup>	0.14 <sup>b</sup>
B1 without Vegetarianism	.039	0.01 <sup>ab</sup>	-0.18 <sup>a</sup>	0.19 <sup>b</sup>
B2 Moral Behaviors (Yes/No summary score across all eight behavioral issues)	.182	-0.06	-0.06	0.13
B2 without Vegetarianism	.326	-0.01	-0.09	0.10

**abc** indicating sig. group differences (Tukey).

reference groups, we ran summary statistics that tested for potentially small but consistent effects across normative issues. As we lacked direct observational data of behavior, we had to deviate from the original study design in compiling the respective summary statistics. Furthermore, given the unusual pattern of vegetarianism reported, we decided to also analyze aggregate measures by singling out vegetarianism from all summary statistics as a contrast and secondary analysis (see Table 3).

In terms of normative attitudes, we first converted each normative attitude distribution into z-scores in order to standardize distributional parameters across normative issues that had been collected using a 9-point Likert scale. We then reversed distributions for question items with reverse polarity and summed up all the z-transformed normative issues except theft to compile an aggregate measure of normative stringency. We followed an analogous procedure to obtain aggregates for behaviors.<sup>12</sup> We did not collect behavioral data on theft; as we aimed for a symmetry between normative attitudes and self-reported behaviors on both the continuous and the summary score, the theft item was excluded. While the two aggregate measures of normative attitudes did not deviate significantly from normal distribution (Kolmogorow-Smirnow, both  $p = .200$ ), the distributions of our four behavior aggregates did deviate from normal distribution (Kolmogorow-Smirnow, all  $p < .01$ ).

In general, the aggregate scores show that, on normative attitudes, it is actually the non-philosophers who show more stringent attitudes than do non-ethicist philosophers, but not ethicists,  $F(2, 375) = 3.50$ ,  $p = .031$ ,  $\eta^2 = .018$ ; significant post-hoc comparisons: non-ethicist philosophers – non-philosophers  $p = .046$ ,  $d = .28$ . For moral behaviors, the same pattern is repeated in the continuous z-score summary,  $F(2, 298) = 3.06$ ,  $p = .048$ ,  $\eta^2 = .020$ ; significant post-hoc comparisons: non-ethicist philosophers – non-philosophers  $p = .050$ ,  $d = .33$ , but not in the yes-or-no score, where no group differences emerged (see Table 3). Hence, surprisingly, we found that the non-ethicist philosophers lag behind on one measure for aggregated moral behaviors compared to non-philosophers, while the ethicists



**Figure 3.** Aggregate scores.

are indistinguishable from both groups (see [Figure 3](#)). As the logic of our experimental design does not allow for a substantial interpretation of this finding, we do not want to speculate about its origin.

When excluding vegetarianism, however, ethicists and non-ethicist philosophers show concurrent patterns once again, as they both exhibit less stringent normative attitudes when compared to non-philosophers. Excluding vegetarianism on behavior, the only significant group difference that remained was between non-ethicist philosophers and non-philosophers (see [Table 3](#)).

### 3.5. Extension results

Our extension included three different statements that the participants were asked to agree or disagree with on a 9-point Likert scale. The first item pertained to the internalism-versus-externalism debate on moral motivation. We asked participants to evaluate the following statement: “The belief that an action is wrong is sufficient motivation to act according to this belief.” This item was chosen to offer a possible explanation for Schwitzgebel and Rust’s (2014) main finding that ethicists hold stronger normative views while being indistinguishable from the other two groups in their self-reported moral behavior. Internalism holds that the making of a judgment is linked to a minimal motivation to act according to it, whereas externalism posits that the motivation to act is contingent upon the “content of the motivational dispositions,” (Smith, 1994, p. 72), that is, content other than the judgment or belief itself. Strong leanings of ethicists toward externalism or at least a non-internalist result for the self-evaluation could give a possible explanation for the original effect found by Schwitzgebel and Rust (2014). If it is not

believed that a judgment about whether an act is right or wrong gives sufficient reason for action, then that would straightforwardly explain a lack of attitude-behavior consistency.

Judging on a Likert scale, with 1 signaling strong disagreement and 9 pertaining to full agreement, ethicists ( $M = 4.7$ ,  $SD = 3.0$ ) and non-ethicist philosophers ( $M = 4.9$ ,  $SD = 2.9$ ) were less inclined to agree with motivation internalism than non-philosophers were ( $M = 5.6$ ,  $SD = 2.5$ ),  $F(2, 367) = 3.56$ ,  $p = .029$ ,  $\eta^2 = .019$ ; significant post-hoc comparisons: ethicist – non-philosophers  $p = .029$ ,  $d = .14$ . Tukey post-hoc comparisons showed a significant difference between ethicists and non-philosophers,  $p = .029$ . This suggests that non-philosophers are more inclined to favor motivation internalism than ethicists and non-ethicist philosophers are. Potentially, philosophers (both ethicists and non-ethicist philosophers) might have a theoretically more focused understanding of “sufficiency.” The sample average as a whole did not differ significantly from 5 (neutral).

Second, we asked respondents to which extent they agree with the statement that philosophical moral reflection leads to the realization of moral truths, as we were interested in how participants themselves viewed this (empirically untestable) relation. We were unable to find any statistically significant group differences,  $F(2, 371) = .73$ ,  $p = .485$ ,  $\eta^2 = .004$ . It is noteworthy, however, that all three groups were slightly inclined to agree, on average, that ethical reflection yields moral truths: Sample mean = 5.8, compared to value: 5 (neutral). A one sample t-test produced this result:  $t = 6.520$ ,  $p = .001$ .

Lastly, we asked for evaluation of the explicit statement that philosophical moral reflection leads to improved moral behavior. The choice for this item was strongly motivated by the desire to illuminate how ethicists, non-ethicist philosophers, and non-philosophers themselves conceive of this relationship. We did not find any significant differences between the groups, non-ethicists philosophers ( $M = 5.2$ ,  $SD = 2.2$ ), ethicists ( $M = 5.5$ ,  $SD = 2.0$ ), and non-philosophers ( $M = 5.7$ ,  $SD = 2.0$ ),  $F(2, 371) = 1.08$ ,  $p = .339$ ,  $\eta^2 = .006$ , but we were able to determine mild overall agreement across groups that ethical reflection leads to better moral behavior, sample mean = 5.5, compared to value: 5 (neutral). A one sample t-test produced this result:  $t = 4.698$ ,  $p = .001$ .

### 3.6. Normative theory and level of abstraction

When asked which normative theory they find most attractive, philosophers (both ethicists and non-ethicists) showed a distribution that was close to that of the original study and diverged from other data on this matter. 25.7% chose virtue ethics and 20.7% preferred deontology, with

only 12.7% deciding for utilitarianism. Moreover, 6.8% claimed to adhere to skepticism, with the remaining 33.9% reporting having no settled position. Schwitzgebel and Rust (2014) similarly found virtue ethics and deontology to be the most popular options, whereas other surveys of professional philosophers found deontology and utilitarianism to be the two most popular theories, with virtue ethics trailing behind both (Bourget & Chalmers, 2014). Like the original study, we did not find systematic differences in normative attitudes or behavior among the categories of these items. However, we made use of the level of abstraction as an exploratory criterion to validate our differentiation criterion between ethicist and non-ethicist philosophers (see discussion).

#### 4. Discussion

Let us come back to Max Scheler's comment on the discrepancy between espoused ethical theory and a divergent private life, namely his proclamation that signposts do not walk in the direction they point towards. Schwitzgebel and Rust (2014) suggested that this might indeed characterize the overall relation between professionalized ethical reflection and moral action. This conclusion is at odds with a common belief about ethics, reflected in our study by the fact that across groups there was a mild agreement that ethical reflection indeed leads to better moral behavior: Ethicists ( $M = 5.5$ ,  $SD = 2.0$ ), non-ethicist philosophers ( $M = 5.2$ ,  $SD = 2.2$ ), and non-philosophers ( $M = 5.6$ ,  $SD = 2.1$ ). Based on our results, we might have to conclude that the picture is indeed complicated. We draw three general observations from our results.

First, our study was unable to provide ample evidence for the effect that ethicists unequivocally exhibit higher stringency in their normative attitudes. Rather, we found mixed results. When asked how much one ought to give to charity and how good or bad it is to eat the meat of mammals, ethicists indeed judged more stringently than non-philosophers did. Yet, on three other issues, they were on the lenient side in their normative assessments. Furthermore, our data failed to show significant differences in the amount of questions that were rated as neutral. Utilizing proportional analysis of how many persons judged on the "good" or "bad" side of the scale, we demonstrated that ethicists either did not differ, or, if they did, they almost always showed the same pattern as non-ethicist philosophers did, both differing from non-philosophers. Additionally, on aggregate scores of normative attitudes, ethicists also did not judge more stringently than any of the control groups did. Curiously, for the most part, ethicists and non-ethicist philosophers showed a concurring pattern in their normative assessments, in both the directions of stringency and leniency. We think that this finding in particular calls for further research.

Second, we were able to replicate the effect that ethicists do not behave better or closer to their expressed attitudes than do both other groups on the majority of issues. This was true for both individual normative issues and aggregate scores, where we did not detect any group differences between the ethicists and the control groups (although non-ethicist philosophers reported significantly fewer moral behaviors compared to non-philosophers, on aggregate). Hence, we understand our data to corroborate the basic finding of Schwitzgebel and Rust (2014), namely that ethicists do not show significantly different behavior on the vast majority of normative issues.

Third, an exceptional finding emerged with regards to vegetarianism. As reported above, ethicists and non-ethicist philosophers rated eating meat as worse than did non-philosophers, but it was only ethicists who stated that they eat less meat per week in comparison to non-philosophers and who were less likely than both other groups to report eating meat at the last dinner. This is a new finding that stands out from both our pattern of results and existing research on the moral behavior of ethics professors.

Given these three observations, the relation between ethical reflection and behavior does not appear to be straightforward. Do our data provide support for any one of the four particular models of this relation that were outlined in the introduction? As our results are rather heterogenous, we consider it informative to assess which model is evidenced issue by issue, rather than aiming only at a general assessment.

As mentioned, ethicists expressed more stringent assessments only on vegetarianism and how much one ought to give to charity. Yet, the relation to behavior was not uniform in these two cases: In the case of vegetarianism, ethicists reported behavior that corresponded to their more stringent normative attitudes. This might be taken as evidence for the booster view, namely, that ethical reflection leads to moral insight and is accompanied by corresponding moral behaviors. Hence, if it is indeed the case that eating as a vegetarian or giving to charity are morally good, then ethicists are closer to moral truth compared to control groups. In the case of charity, in contrast, ethicists did not stand out from the control groups with regard to the percentage they reported to have donated themselves, despite demanding a higher percentage of donation. This seems to correspond with the inert discovery view, namely, that ethical reflection leads to moral insight without correspondingly better behaviors.

For “blood donation,” “organ donation,” “donating 10% of one’s income,” and “honesty in this survey” no significant group differences emerged. This suggests the epiphenomenal view, namely, that professional ethical reflection does not have any noticeable impact on either normative attitudes or behavior, or, at least, not beyond the impact that the ethical reflection non-ethicists engage in has.<sup>13</sup>

Another question merits further attention: Is this heterogenous pattern across issues systematic or not, such that the ethicists' stronger or more lenient assessments of normative issues might plausibly be said to reflect a meaningful prioritization depending on the contents of the issue in question? Is there something in the content of normative issues that explains ethicists' relative stringency on vegetarianism and charity as well as their relative leniency on other issues?<sup>14</sup>

As a possible answer, we want to suggest that a divide between "properly moral" and "somewhat moral" might be illuminating with regard to the prioritization of charity and vegetarianism at the relative disadvantage of other normative topics. This divide is close to the distinction between "moral" and "conventional" established by Elliot Turiel (1983), which leads to similar predictions. For instance, the comparison between eating meat or not giving to charity and being more lenient on calling one's mother regularly and voting in elections would be an exemplification of such an effect, favoring some areas of conduct as properly moral over others as merely somewhat moral or conventional. Indeed, a number of comments by participants of all three groups remarked that some of our normative issues were "not moral" in the appropriate sense. While voting and calling one's mother received such comments, none of this kind were submitted for vegetarianism or charity.<sup>15</sup> This gives further credence to the claim that a distinction similar to the one suggested by Turiel (1983) is at work here.

Within professional ethics, it is relatively uncontroversial that vegetarianism and charity are properly moral domains which merit moral consideration. In other words, we believe that the prioritizing observed might reflect a disciplinary consensus in ethics. Indeed, the particular prioritizing pattern matches the agenda of popular movements in applied ethics such as "effective altruism," which has consistently advocated prioritizing some areas of moral urgency like poverty and animal suffering over smaller, more intermediate ones. This setting of priorities goes back to the foundational paper by Singer (1972), later supplanted by the growing animal rights movement (1990). Generally, the movement advocates that one achieves the most good by supporting the most effective charities and actions, such as abstaining from eating meat. This is directly in line with the emphasis on charity and vegetarianism over other issues like voting and keeping in touch with one's mother (MacAskill, 2015), which we observed for ethicists in our data.<sup>16</sup>

One might also explain this distribution of (properly) moral and conventional more broadly, namely, as a result of shared psychological tendencies of university professors. Following the work of Haidt, the vast majority of professors sampled here would count as Western, educated, industrial, rich, and democratic (Haidt, 2012), with ensuing converging beliefs on what counts as moral and what does not. There is an empirical

foundation for this claim in our data: We find generally high levels of vegetarianism among the groups studied compared to those of the general population (Gert, Barbosa, & Brettschneider, 2016).<sup>17</sup>

Coming back to the initial question whether professional ethical reflection leads to the discovery of moral truths, it might very well be that it leads to a more realistic assessment of what domains of moral conduct are tied up with the most serious impact. This might generate a comparatively framework-independent consensus on properly moral topics, based on the relative effectiveness of the behavioral change of individuals. There is one major trouble with conceiving this consensus as ethics-specific, however: In our data, ethicists and non-ethicist philosophers often showed concurrent trends, both in stringent and in lenient directions. Thus, despite some minor divergences, ethicists were largely synchronous with their non-ethicist philosopher colleagues in their normative assessments. Does this undermine our assumption that the observed priorities are due to ethical reflection specifically? Two broad possibilities suggest themselves: (1) There are substantial similarities between philosophers (both ethicist and non-ethicist) that systematically distinguish them from non-philosophers, and/or (2) the differentiation between ethicists and non-ethicist philosophers for group assignment is insufficient.

On the first possibility, it is certainly possible that there are psychological similarities between ethicists and non-ethicist philosophers, as well as similarities in the type of reflection they engage in. For instance, we hypothesize that professional philosophical reflection might result in an awareness of extreme and remote cases when contemplating assessment of normative issues. This could possibly explain why ethicists and non-ethicist philosophers judge theft of 1000 euros more leniently than do non-philosophers – they might just have worse atrocities in mind, such that they do not use the extreme end of the scale for an issue like theft.<sup>18</sup>

On the second possibility, it could be that we failed to distinguish between ethicists and non-ethicist philosophers. This could be because the CV-based group assignment does not yield a sufficiently discriminating criterion. To test this assumption exploratively, we run all our analyses by using a self-assigned, survey-internal criterion as well. Every philosophy professor who, in the survey, chose either “applied ethics” and/or “normative ethics” as the level of abstraction on which they consider ethical issues was identified as an ethicist; if not, we assigned the person to the non-ethicist philosopher group. Using this self-assignment criterion, we found no substantial differences from our original analyses, a result which underpins the robustness of our original criterion.

Furthermore, there were subtle but consistent differences between ethicists and non-ethicist philosophers. While both ethicists and non-

ethicist philosophers judged more stringently on vegetarianism, only the ethicist group behaved significantly better compared to non-philosophers. On how much one ought to give to charity, only the ethicists differed significantly from non-philosophers. On aggregate scores, although no significant difference emerged between ethicists and non-ethicist philosophers directly, only the non-ethicist philosophers judged more leniently and reported significantly fewer moral behaviors compared to non-philosophers, which might suggest that they are generally a bit more casual with regards to moral issues than ethicists and non-philosophers. Hence, though both ethicists and non-ethicist philosophers exhibited an almost uniform prioritizing pattern in terms of their normative attitudes, systematic differences manifested that justify treating ethicists as a distinctive group.

Hence, we think it is plausible that the prioritization in terms of normative attitudes mirrors a consensus in professional ethics, and the similarity of ethicists and non-ethicist philosophers (in combination with a shared psychological tendency to take into account the extreme and the remote) reflects a spill-over of this consensus to the discipline of philosophy as a whole. Ethics is, after all, one of the pillars of philosophy, and non-ethicist philosophers can be expected to be acquainted with an ethical consensus, if it indeed exists. One might thus hypothesize that the disciplinary consensus in ethics rubs off on non-ethicist philosophers who are not primarily occupied with ethics. The fact that the prioritization pattern is somewhat less pronounced for non-ethicist philosophers and that they are more casual in their moral behaviors supports this view, as normative attitudes are plausibly more easily impacted than moral behaviors.

One possibility is also that philosophers as a group are actually somewhat unconventional, and, as described in the work of Turiel, might have a different understanding of what is conventional and moral compared to non-philosophers. Arguably, philosophers might consider some issues conventional that non-philosophers consider moral.

Importantly, if there is indeed a consensus that is specific to the discipline of ethics (and potentially to the discipline of philosophy as a whole), it might almost, by definition, depart from “folk morality” in some respects. We think that this is a plausible explanation for the systematic differences between ethicists and non-philosophers that we observed. Indeed, one might interpret the concurrent leniency of ethicist and non-ethicist philosophers on “theft,” “voting,” and “talking to one’s mother” the other way around, namely, as a relatively stringent attitude of non-philosophers.<sup>19</sup> Indeed, non-philosophers showed significantly more stringent attitudes across issues compared to non-ethicist philosophers; if the exceptional case of vegetarianism was

excluded from the aggregate, non-philosophers judged more stringently compared to both ethicists and non-ethicist philosophers across issues. This might reflect a folk psychological tendency to overestimate the normative import of some issues, either by a naïve tendency of generally conflating what is conventional with what is moral or by a failure to relativize the question at hand with respect to other issues, including extreme and remote cases.

## 5. Conclusion

This replication-extension aimed to replicate the findings of Schwitzgebel and Rust (2014) which showed that, while ethicists showed stronger normative attitudes, they did not differ in their moral behavior or attitude-behavior consistency. Our research surveyed 417 professors from 62 universities in Austria, Germany, and Switzerland. The data gathered in this replication attempt offered mixed results on the topic of normative attitudes; generally, ethicists did not differ from both non-ethicist philosophers and non-philosophers on individual and on aggregate measures, with the exception of vegetarianism and charity. However, our study managed to replicate the finding that ethicists do not behave morally better or more in line with their expressed views than did either non-ethicist philosophers or non-philosophers. Given our results, we are inclined to agree with Schwitzgebel and Rust that the “psychological story is complicated” (Schwitzgebel & Rust, 2014, p. 320) but also that “ethicists [do not] behave, on average, morally better than do non-ethicists” (p. 320), with the notable exception of vegetarianism.

## Notes

1. Our translation.
2. We owe this citation and its place to Schwitzgebel and Rust (2014).
3. These five states were California, Florida, North Carolina, Minnesota, and Washington.
4. With our replication study we also aim to address the problem of the replication crisis in social science fields (Schmidt, 2009) like moral psychology and adjacent areas of enquiry such as experimental philosophy, even though it has recently been suggested that experimental philosophy replicates at a much higher rate than similar disciplines (Cova et al., 2018). Moreover, in order to contribute to this ideal of a cumulative science and adopt the “best-practices” (Cova et al., 2018, p. 10) suggested, we pre-registered this study in the Open Science Framework (Schönegger & Wagner, 2018) following the “replication recipe” (Brandt et al., 2014).
5. In German: “Ethik,” “Angewandte Ethik,” “Normative Theorie,” “Ethik der X,” and „X-Ethik“.
6. We used Polldaddy as survey platform.

7. Following Schwitzgebel and Rust, we decided to limit the “eating meat” item to eating the meat of mammals, as there might be some who draw an ethical distinction between mammals and fish. Focusing on mammals was a way to preempt confusion. This means that any further classification of “vegetarianism” could also include, at least in part, a certain number of pescatarians. We do recognize that not including poultry could have distorting impacts on the interpretability of the result. However, we did decide to stick with the original version for replication reasons. All further mention of “meat” will be taken to mean “meat of mammals.”
8. The first part of the survey utilized the “morally good – morally bad” dichotomy, whereas the extension made use of the “strongly disagree – strongly agree” variant of the same scale.
9. Like Schwitzgebel and Rust (2014), we gave the option of free text commentary at the end of each page to give participants the opportunity to clarify their answers or raise objections to the wording or methodology. We gave these comments a cautious reading. Most of them raised objections to specific wordings, pointed towards the importance of context in certain item descriptions, demanded more background information on certain questions, expressed disagreement with our choice of some items or possible answers, and commented more generally on whether or not they liked the survey or what they felt it missed. None of these comments suggested a fundamental flaw in the study design, similar to the responses to the original study (Schwitzgebel & Rust, 2014, p. 298–299).
10. According to the discipline’s standards, we put the alpha-level of significance at 0.05.
11. Like the original, we collapsed the values 1 through 4 into “morally bad” and values 6 through 9 into “morally good” for proportional analysis. When analyzing the good side of the scale, the bad side and the neutral values were collapsed and vice versa.
12. We aggregated yes and no questions on moral behavior, counting “yes” as 1, and “no” or “do not recall” as 0. On behavioral measures where we only had a continuous variable, we split the distribution at the median and assigned “1” to the upper and “0” to the lower 50% of the distribution. As there was thus exactly one point to get for moral behavior on every normative issue, participants could get between 0 and 7–8 points (depending on whether vegetarianism was excluded or not).
13. Non-ethicist philosophers fell behind on the issue of “answering students’ e-mails” as well as on one aggregate measure of moral behavior, while no such differences were observed in terms of normative attitudes. Though this effect does not allow for an inference with regard to the effects of professional ethical reflection, it speaks for a local inert discovery view as an explanation for non-ethicist philosophers’ moral insights, as non-ethicist philosophers did not live up to the equally strong attitudes they expressed in this instance compared to ethicists and non-philosophers.
14. It is very important to note, however, that in our talk of “relative stringency or leniency,” we are not talking about relativity among normative issues, but relativity among groups with regards to particular issues. The issue of theft, for instance, was unequivocally perceived as the most morally reprehensible by all groups. In fact, the selected normative issues are not straightforwardly comparable in their absolute numbers, as the responses to them might differ with the intensity of the example chosen – one could ask about a theft of 50 euros instead of 1000 euros, for example. Hence, in our talk about somewhat and properly moral, we are talking about relative differences among ethicists, non-ethicist philosophers, and non-philosophers, arguing that the issues which ethicists favor compared to non-philosophers might reflect properly moral issues.

15. Participants who commented on charity were mostly concerned with background assumptions of duty and the classification used for the question, resulting in a demand for a better focus on additional circumstances such as family situation and other obligations.
16. Notice how this provides a straightforward explanation of ethicists' more stringent attitudes and better behavior in terms of vegetarianism: It is highly recommended in terms of effectiveness, and in our contemporary society it's not too difficult to implement. It also deserves mention, however, that the difference between the original US sample and our European sample was much larger than the difference that was observed within cultures, that is, between groups. This indicates that culture is a more influential factor on eating meat than ethical reflection, which, however, does not thereby undermine ethical reflection as a significant factor. The difference between samples might reflect cultural differences or a general trend towards the reduction of meat consumption in the course of the last decade, at least among academic populations, as only 4.3% of Germans aged 18 to 79 (Gert et al., 2016) subscribe to a vegetarian diet. Interestingly, the original study also found 27% of ethicists (Schwitzgebel & Rust, 2014, p. 307) reporting strict vegetarianism, which is comparable to our number of 30%. On this picture, while the proportion of full-vegetarian ethicists remained roughly the same, we were able to observe a significant trend toward eating less meat across groups.
17. Following Haidt's understanding, one would expect a distribution favoring harm and fairness considerations, for example, in charity and vegetarianism, over others, such as calling one's mother. By contrast, one would expect this prioritization to change in other cultures. Indeed, one generally finds a varying pattern in other contexts (Graham, Meindl, Beall, Johnson, & Zhang, 2016) and, more specifically, with regard to different religions (Levine et al., 2018). To give an example, in Confucian cultures civility takes a central role in morality (Buchtel et al., 2015), which would suggest a different weighing, one which would favor calling one's mother substantially more. These cross-cultural analyses, however, can only illuminate the background of our research here, as they cannot explain the group differences studied.
18. A similar worry has been raised by the authors of the original study (Schwitzgebel & Rust, 2014, p. 299–300).
19. In addition, non-philosophers also judged more stringently than non-ethicist philosophers with regards to “paying academic fees.”

## Disclosure statement

No potential conflict of interest was reported by the authors.

## Notes on contributors

*Philipp Schönegger* is a graduate student in moral, political and legal philosophy at the University of St. Andrews and the University of Stirling.

*Johannes Wagner* is a graduate student in psychology at the University of Graz.

## ORCID

Philipp Schönegger  <http://orcid.org/0000-0001-9930-487X>

## References

- Björklund, F., Haidt, J., & Murphy, S. (2000). *Moral dumbfounding: When intuition finds no reason* (Lund Psychological Reports; Vol. 1 No 2). Lund: Department of Psychology, Lund University.
- Blanca, M., Alarcón, R., Arnau, J., Bono, R., & Bendayan, R. (2017). Non-normal data: is anova still a valid option?. *Psicothema*, 29(4), 552-557.
- Böhme, G. (2011). Geht denn der Wegweiser den Weg, den er zeigt?. In C. Anmann, B. Bleisch, & A. Goppel (Eds.), *Müssen Ethiker Moralisch Sein?* Frankfurt: Campus Verlag.
- Bonett, D. (2012). Replication Extension Studies. *Current Directions in Psychological Science*, 21(6), 409-412.
- Bourget, D., & Chalmers, D. J. (2014). What do philosophers believe? *Philosophical Studies*, 170(3), 465-500.
- Brandt, M., Jzerman, H., Dijksterhuis, A., Farach, F. J., Geller, J., Giner-Sorolla, R., & Van't Veer, A. (2014). The replication recipe: What makes for a convincing replication? *Journal of Experimental Social Psychology*, 50, 217-224.
- Buchtel, E. E., Guan, Y., Peng, Q., Su, Y., Sang, B., Chen, S. X., & Bond, M. H. (2015). Immorality East and West: Are immoral behaviors especially harmful, or especially uncivilized? *Personality and Social Psychology Bulletin*, 41(10), 1382-1394.
- Cooper, J. M. (2012). *Pursuits of wisdom: Six ways of life in ancient philosophy*. Princeton: Princeton University Press.
- Cova, F., Strickland, B., Abatista, A., Allard, A., Andow, J. A., & Zhou, X. (2018). Estimating the Reproducibility of Experimental Philosophy. doi:10.17605/OSF.IO/SXDAH
- Ditto, P. H., Pizarro, D. A., & Tannenbaum, D. (2009). Motivated moral reasoning. In B. H. Ross, D. M. Bartels, C. W. Bauman, L. J. Skitka, & D. L. Medin (Eds.), *Psychology of learning and motivation, vol. 50: Moral judgment and decision making*. San Diego, CA: Academic Press.
- Duignan, B. (ed.). (2011). *The history of western ethics*. New York, NY: Britannica.
- Gert, B. M., Barbosa, C. L., & Brettschneider, A. (2016). Prevalence of persons following a vegetarian diet in Germany. *Journal of Health Monitoring*, 1(2), 2-14.
- Ghasemi, A., & Zahediasl, S. (2012). Normality tests for statistical analysis: a guide for non-statisticians. *International Journal Of Endocrinology and Metabolism*, 10(2), 486. doi:10.5812/ijem
- Gold, N., Pulford, B. D., & Colman, A. M. (2015). Do as I say, don't do as i do: Differences in moral judgments do not translate into differences in decisions in real-life trolley problems. *Journal Of Economic Psychology*, 47, 50-61.
- Graham, J., Meindl, P., Beall, E., Johnson, K. M., & Zhang, L. (2016). Cultural differences in moral judgment and behavior, across and within societies. *Current Opinion in Psychology*, 8, 125-130.
- Haidt, J. (2001). The emotional dog and its rational tail: A social intuitionist approach to moral judgment. *Psychological Review*, 108, 814-818.
- Haidt, J. (2012). *The righteous mind: Why good people are divided by politics and religion*. New York, NY: Pantheon Books.
- Helion, C., & Pizarro, D. A. (2015). Beyond dual-processes: The interplay of reason and emotion in moral judgment. In N. Levy & J. Clausen (Eds.), *Springer handbook for neuroethics*. Dordrecht: Springer Netherlands.
- Horvath, J., & Wiegmann, A. (2016). Intuitive expertise and intuitions about knowledge. *Philosophical Studies*, 173(10), 2701-2726.

- Ivanhoe, P. (2000). *Confucian moral self-cultivation*. Indianapolis: Hackett.
- Knobe, J., & Nichols, S. (eds.). (2008). *Experimental philosophy*. Oxford: Oxford University Press.
- Levine, S., Rottmann, J., Davis, T., O'Neill, E., Stich, S., & Machery, E. (2018). *Religion's impact on conceptions of the moral domain*. Manuscript submitted for publication.
- MacAskill, W. (2015). *Doing good better: How effective altruism can help you make a difference*. New York, NY: Random House.
- McHugh, C., McGann, M., Igou, E. R., & Kinsella, E. L. (2017). Searching for moral dumbfounding: Identifying measurable indicators of moral dumbfounding. *Collabra: Psychology*, 3(1), 23.
- Pallant, J., & Manual, S. S. (2007). A step by step guide to data analysis using SPSS for windows version 15. SPSS Survival manual, 3.
- Schmidt, S. (2009). Shall we really do it again? The powerful concept of replication is neglected in the social sciences. *Review of General Psychology*, 13(12), 90–100.
- Schönegger, P., & Wagner, J. (2018). Replication-Extension of Schwitzgebel and Rust's "Relationships Among Self-Reported Behavior, Expressed Normative Attitude, and Directly Observed Behavior". Retrieved from [osf.io/wt2er](https://osf.io/wt2er)
- Schwitzgebel, E., & Cushman, F. (2012). Expertise in moral reasoning? Order effects on moral judgment in professional philosophers and non-philosophers. *Mind & Language*, 27, 135–153.
- Schwitzgebel, E., & Cushman, F. (2015). Philosopher's biased judgments persist despite training, expertise and reflection. *Cognition*, 141, 127–137.
- Schwitzgebel, E., & Rust, J. (2009). The moral behavior of ethicists: Peer opinion. *Mind*, 118, 1043–1059.
- Schwitzgebel, E., & Rust, J. (2010). Do ethicists and political philosophers vote more often than other professors? *Review of Philosophy and Psychology*, 1, 189–199.
- Schwitzgebel, E., & Rust, J. (2011). The self-reported moral behavior of ethics professors. Retrieved from <http://www.faculty.ucr.edu/~eschwitz/SchwitzAbs/EthSelfRep.htm>.
- Schwitzgebel, E., & Rust, J. (2013). Ethicists' and non-ethicists' responsiveness to student emails: Relationships among expressed normative attitude, self-described behavior, and experimentally observed behavior. *Metaphilosophy*, 44, 350–371.
- Schwitzgebel, E., & Rust, J. (2014). The moral behavior of ethics professors: Relationships among self-reported behavior, expressed normative attitude, and directly observed behavior. *Philosophical Psychology*, 27, 293–371.
- Singer, P. (1972). Famine, affluence and morality. *Philosophy and Public Affairs*, 1, 229–243.
- Singer, P. (1990). *Animal Liberation*. New York, NY: Random House.
- Sinnott-Armstrong, W. (ed.). (2008). *Moral psychology. Volume 2: The cognitive science of morality: Intuition and diversity*. Cambridge: MIT Press.
- Smith, M. (1994). *The moral problem*. Oxford: Basil Blackwell.
- Turiel, E. (1983). *The development of social knowledge: Morality and convention*. Cambridge University Press.
- Uhlmann, E. L., Pizarro, D. A., Tannenbaum, D., & Ditto, P. H. (2009). The motivated use of moral principles. *Judgment and Decision Making*, 4, 479–491.