

The Moral Behavior of Ethicists and the Rationalist Delusion

Joshua Rust
Department of Philosophy
Stetson University
421 North Woodland Boulevard
DeLand, Florida 32723

Eric Schwitzgebel
Department of Philosophy
University of California at Riverside
Riverside, CA 92521-0201

January 2, 2013

The Moral Behavior of Ethicists and the Rationalist Delusion

Professional ethicists behave no morally better, on average, than do other professors. At least that's what we have found in a series of empirical studies that we will summarize below. Our results create a prima facie challenge for a certain picture of the relationship between intellectual reasoning and moral behavior – a picture that Jonathan Haidt (2012) has called “the rationalist delusion”. The rationalist delusion, according to Haidt, is the view that intellectual forms of moral reasoning substantially influence the moral attitudes and moral behavior of the reasoner. Haidt seeks to replace this rationalistic picture with an “intuitionist” model of moral psychology, in which the core of morality is a suite of pro-social emotions and arational intuitions. Haidt argues that our empirical results favor his view. After all, if not even professional ethicists are moved by intellectual moral reasoning, who else would be? While we agree with Haidt that our results support intuitionism over some rationalistic rivals, we believe that other models of moral psychology are also consistent with our findings, and some of these models reserve an important role for reasoning in shaping behavior and attitudes. Part One summarizes our empirical findings. Part Two explores five different theoretical models more or less consistent with those findings.

Part One: Our Empirical Studies

Missing library books. Our first study (Schwitzgebel 2009) examined the rates at which ethics books were missing from 32 leading academic libraries, compared to other philosophy books, according to those libraries' online catalogs. The primary analysis was confined to relatively obscure books we thought likely to be borrowed primarily by specialists in the field –

275 books reviewed in *Philosophical Review* between 1990 and 2001, excluding titles cited five or more times in the *Stanford Encyclopedia of Philosophy*. Among these books, we found ethics books somewhat *more* likely to be missing than non-ethics books: 8.5% of the ethics books that were off the shelf were listed as missing or as more than one year overdue, compared to 5.7% of the non-ethics philosophy books that were off the shelf (χ^2 test, $p = .03$). This result holds despite a similar total number of copies of ethics and non-ethics books held, similar total overall checkout rates of ethics and non-ethics books, and a similar average publication date of the books. Similarly, classic pre-20th-century ethics texts appear to go missing more often than do comparable non-ethics texts.

Peer ratings. Our second study examined peer opinion about the moral behavior of professional ethicists (Schwitzgebel and Rust 2009). We set up a table in a central location at the 2007 Pacific Division meeting of the American Philosophical Association and offered passersby gourmet chocolate in exchange for taking a “5-minute philosophical-scientific questionnaire”, which they completed on the spot. One version of the questionnaire asked respondents their opinion about the moral behavior of ethicists in general, compared to other philosophers and compared to non-academics of similar social background (with parallel questions about the moral behavior of specialists in metaphysics and epistemology). Opinion was divided: 36% of respondents rated ethicists morally better behaved on average than other philosophers, 44% rated them about the same, and 19% rated them worse. When ethicists’ behavior was compared to that of non-academics, opinion was split 50%-32%-18% between better, same, and worse. Another version of the questionnaire asked respondents to rate the moral behavior of the individual ethicist in their department whose last name comes next in alphabetical order, looping back from Z to A if necessary, with a comparison question about the moral behavior of a similarly

alphabetically chosen specialist in metaphysics and epistemology. Opinion was again split: 44% of respondents rated the arbitrarily selected ethics specialist better than they rated the arbitrarily selected M&E specialist, 26% rated the ethicist the same, and 30% rated the ethicist worse. In both versions of the questionnaire, the skew favoring the ethicists was driven primarily by respondents reporting a specialization or competence in ethics, who tended to avoid rating ethicists worse than others. Non-ethicist philosophers tended to split about evenly between rating the ethicists better, same, or worse.

Voting rates. We assume that regular participation in public elections is a moral duty, or at least that it is morally better than non-participation (though see Brennan 2011). In an opinion survey to be described below, we found that over 80% of sampled U.S. professors share that view. Accordingly, we examined publicly available voter participation records from five U.S. states, looking for name matches between voter rolls and online lists of professors in nearby universities, excluding common and multiply-appearing names (Schwitzgebel and Rust 2010). In this way, we estimated the voting participation rates of four groups of professors: philosophical ethicists, philosophers not specializing in ethics, political scientists, and professors in departments other than philosophy and political science. We found that all four groups of professors voted at approximately the same rates, except for the political science professors, who voted about 10-15% more often than did the other groups. This result survived examination for confounds due to gender, age, political party, and affiliation with a research-oriented vs. teaching-oriented university.

Courtesy at philosophy conferences. While some rules of etiquette can be morally indifferent or even pernicious, we follow Confucius (5th c. BCE/2003), Karen Stohr (2012), and others in seeing polite, respectful daily behavior as an important component of morality. With

this in mind, we examined courteous and discourteous behavior at meetings of the American Philosophical Association, comparing ethics sessions with non-ethics sessions (Schwitzgebel, Rust, Huang, Moore, and Coates 2012). We used three measures of courtesy – talking audibly during the formal presentation, allowing the door to slam when entering or exiting mid-session, and leaving behind litter at one’s seat – across 2800 audience-hours of sessions at four different APA meetings. None of the three measures revealed any statistically detectable differences in courtesy. Audible talking (excluding brief, polite remarks like “thank you” for a handout) was rare: .010 instances per audience hour in the ethics sessions vs. .009 instances per audience hour in the non-ethics sessions (two-proportion z test, $p = .77$). The median rate of door-slamming per session (compared to mid-session entries and exits in which the audience member attempted to shut the door quietly) was 18.2% for the ethics sessions and 15.4% for the non-ethics sessions (Mann-Whitney test, $p = .95$). Finally, ethicists were not detectably less likely than non-ethicists to leave behind cups (16.8% vs. 17.8% per audience member, two-proportion z test, $p = .48$) or trash (11.6% vs. 11.8%, two-proportion z test, $p = .87$). The latter result survives examination for confounds due to session size, time of day, and whether paper handouts were provided. However, we did find that the audience members in *environmental ethics* sessions in particular left behind less trash than did the audience in all other sessions combined (3.0% vs. 11.9%, Fisher’s exact test, $p = .02$).

APA free riding. We assume a prima facie duty for program participants in philosophy conferences to pay the modest registration fees that the organizers of those conferences typically charge. However, until recently the American Philosophical Association had no mechanism to enforce conference registration, which resulted in a substantial free-riding problem. With this in mind, we examined the Pacific Division APA programs from 2006-2008, classifying sessions

into ethics, non-ethics, or excluded. We then examined the registration compliance of program participants in ethics sessions vs. program participants in non-ethics sessions by comparing anonymously encrypted lists of participants in those sessions (participants with common names excluded) to similarly encrypted lists of people who had paid their registration fees (Schwitzgebel forthcoming). (Although the APA Pacific Division generously supplied the encrypted data, this research was neither solicited by nor conducted on behalf of the APA or the Pacific Division.) During the period under study, ethicists appear to have paid their conference registration fees at about the same rate as did non-ethicist philosophers (74% vs. 76%, two-proportion z test, $p = .43$). This result survives examination for confounds due to gender, institutional prestige, program role, year, and status as a faculty member vs. graduate student.

Responsiveness to student emails. Yet another study examined the rates at which ethicists responded to brief email messages designed to look as though written by undergraduates (Rust and Schwitzgebel forthcoming). We sent three email messages – one asking about office hours, one asking for the name of the undergraduate advisor, and one inquiring about an upcoming course – to ethicists, non-ethicist philosophers, and a comparison group of professors in other departments, drawing from online faculty lists at universities across several U.S. states. All messages addressed the faculty member by name, and some included additional specific information such as the name of the department or the name of an upcoming course the professor was scheduled to teach. The messages were checked against several spam filters, and we had direct confirmation through various means that over 90% of the target email addresses were actively checked. Overall, ethicists responded to 62% of our messages, compared to a 59% response rate for non-ethicist philosophers and 58% for non-philosophers – a difference that

doesn't approach statistical significance despite (we're somewhat embarrassed to confess) 3,109 total trials (χ^2 test, $p = .18$).

Self-reported attitudes and behavior. Our most recent study examined ethicists', non-ethicist philosophers', and non-philosophers' self-reported attitudes and behavior on a number of issues including membership in disciplinary societies, voting, staying in touch with one's mother, vegetarianism, organ and blood donation, responsiveness to student emails, charity, and honesty in responding to survey questionnaires. The survey was sent to about a thousand professors in five different U.S. states, with an overall response rate of 58% or about 200 respondents in each of the three groups. Identifying information was encrypted for participants' privacy. On some issues – voting, email responsiveness, charitable donation, societal membership, and survey response honesty – we also had direct, similarly encrypted, observational measures of behavior that we could compare with self report. Aggregating across the various measures, we found no difference among the groups in overall self-reported moral behavior, in the accuracy of the self-reports for those measures where we had direct observational evidence, or in the correlation between expressed normative attitude and either self-reported or directly observed behavior. The one systematic difference we did find was this: Across several measures – voting, vegetarianism, charitable donation, and organ and blood donation – ethicists appeared to embrace more stringent moral views than did non-philosophers, while non-ethicist philosophers held views of intermediate stringency. However, this increased stringency of attitude was not unequivocally reflected in ethicists' behavior.

This last point is best seen by examining the two measures on which we had the best antecedent hope that ethicists would show moral differences from non-ethicists: vegetarianism and charitable donation. Both issues are widely discussed among ethicists, who tend to have

comparatively sophisticated philosophical opinions about these matters, and professors appear to exhibit large differences in personal rates of charitable donation and in meat consumption. Furthermore, ethicists' stances on these issues are directly connected to specific, concrete behaviors that they can either explicitly implement or not (e.g., to donate 10% annually to famine relief; to refrain from eating the meat of such-and-such animals). This contrasts with exhortations like "be a kinder person" that are difficult to straightforwardly implement or to know if one has implemented. Looking, then, in more detail at our findings on vegetarianism and charitable donation:

Self-reported attitude and behavior: eating meat. We solicited normative attitude about eating meat by asking respondents to rate "regularly eating the meat of mammals such as beef or pork" on a nine-point scale from "very morally bad" to "very morally good" with the mid-point marked "morally neutral". On this normative question, there were large differences among the groups: 60% of ethicist respondents rated meat-eating somewhere on the bad side of the scale, compared to 45% of non-ethicist philosophers and only 19% of professors from other departments (χ^2 test, $p < .001$). Later in the survey we posed two behavioral questions. First, we asked "During about how many meals or snacks per week do you eat the meat of mammals such as beef or pork?" Next, we asked "Think back on your last evening meal, not including snacks. Did you eat the meat of a mammal during that meal?" On the meals-per-week question, we found a modest difference among the groups: Ethicists reported a mean of 4.1 meals per week, compared to 4.6 for non-ethicist philosophers and 5.3 for non-philosophers (ANOVA, square-root transformed, $p = .006$). We also found 27% of ethicists to report no meat consumption (zero meat meals per week), compared to 20% of non-ethicist philosophers and 13% of non-philosophers (χ^2 test, $p = .01$). However, statistical evidence suggested that respondents were

fudging their meals-per-week answers: Self-reported meals per week was not mathematically consistent with what one would expect given the numbers reporting having eaten meat at the previous evening meal. And when asked about their previous evening meal, the groups' self-reports differed only marginally, with ethicists in the intermediate group: 37% of ethicists reported having eaten the meat of a mammal at their previous evening meal, compared to 33% of non-ethicist philosophers and 45% of non-philosophers (χ^2 test, $p = .06$).

Self-reported attitude and behavior: charity. We solicited normative opinion about charity in two ways. First, we asked respondents to rate “donating 10% of one’s income to charity” on the same nine-point scale we used for the question about eating meat. Ethicists expressed the most approval, with 89% rating it as good and a mean rating of 7.5 of the scale, vs. 85% and 7.4 for non-ethicist philosophers and 73% and 7.1 for non-philosophers (χ^2 test, $p < .001$; ANOVA, $p = .01$). Second, we asked what percentage of income the typical professor should donate to charity (instructing participants to enter “0” if they think it’s not the case that the typical professor should donate to charity). 9% of ethicists entered “0”, vs. 24% of non-ethicist philosophers and 25% of non-philosophers (χ^2 test, $p < .001$). Among those not entering “0”, the geometric mean was 5.9% for the ethicists vs. 4.8% for both of the other groups (ANOVA, $p = .03$). Later in the survey, we asked participants what percentage of their income they personally had donated to charity in the previous calendar year. Non-ethicist philosophers reported having donated the least, but there was no statistically detectable difference between the self-reported donation rates of the ethicists and the non-philosophers. (Reporting zero: 4% of ethicists vs. 10% of non-ethicist philosophers and 6% of non-philosophers, χ^2 test, $p = .052$; geometric mean of the non-0’s 3.7% vs. 2.6% vs. 3.6%, ANOVA, $p = .004$.) However, we also had one direct measure of charitable behavior: Half of the survey recipients were given a charity

incentive to return the survey – \$10 to be donated to their selection from among Oxfam America, World Wildlife Fund, CARE, Make-a-Wish Foundation, Doctors Without Borders, or American Red Cross. By this measure, the non-ethicist philosophers showed up as the *most* charitable, and in fact were the only group who responded at detectably higher rates when given the charity incentive (67% vs. 59%, two-proportion z test, one-tailed, $p = .048$; compared to 59% on both versions for ethicists and 55% vs. 52% for non-philosophers). While we doubt that this is a dependably valid measure of charitable behavior overall, we are also somewhat suspicious of the self-report measures. We judge the overall behavioral results to be equivocal, and certainly not to decisively favor the ethicists over both of the two other groups.

Conclusion. Across a wide variety of measures, it appears that ethicists, despite expressing more stringent normative attitudes on some issues, behave not much differently than do other professors. However, we did find some evidence that philosophers litter less in environmental ethics sessions than in other APA sessions, and we found some equivocal evidence that might suggest slightly higher rates of charitable giving and slightly lower rates of meat-eating among ethicists than among some other subsets of professors. On one measure – the return of library books – it appears that ethicists might behave morally worse.

Part Two: Possible Explanations.

The rationalist delusion. Haidt (2012) has accused many philosophers, back at least to Plato, of subscribing to what he calls “the rationalist delusion”: They assume that moral reasoning leads people to moral truth, enabling them to behave morally better. Kant (1785/1998), for example, says that human virtue must find its source in practical reasoning, since inclinations toward self-love otherwise threaten to lead us astray. Haidt, in contrast,

compares reasoning and intuition to a rider on an elephant, with the rider, reasoning, largely powerless to control the direction of the elephant. The rationalist delusion is then the erroneous conviction that the rider can control the elephant. Haidt also compares the role of reasoning to that of a lawyer rather than a judge: The lawyer does her best to advocate for the positions given to her by her clients – in this case the intuitions – producing whatever ideas and arguments are convenient for the pre-determined conclusion. The rationalist delusion is then the mistaken impression that reason serves as a neutral judge over moral arguments rather than a paid-off advocate compelled to plump for one side. Haidt cites our work as evidence for this view (e.g., Haidt 2012, p. 89), and we're inclined to agree that it fits nicely with his view and so in that way lends support. We have also recently found evidence that philosophers, perhaps especially ethics PhDs, may be especially good at or especially prone toward embracing moral principles as a form of post-hoc rationalization of covertly manipulated moral judgments (Schwitzgebel and Cushman 2012).

However, Haidt's metaphors, while radical in a way, are conservative in another way, since they preserve and embrace the tradition – also going back to Plato – of sharply distinguishing between reason and other means of arriving at judgments. We find the distinction problematic: Emotional responses, intuitions, spontaneous judgments, gut feelings can have, too, a kind of sensitivity to reasons about them, which conscious, explicit reasoning sometimes fails to know. The elephant might sometimes be more rational than the rider even if the reasons cannot be made explicit (e.g., Damasio 1994; Arpaly 2003).

Furthermore, even if we accept the traditional distinction, Haidt's picture might be misleading if the elephant frequently collaborates with the rider and the lawyer frequently listens to the judge. Rawls's (1971) picture of philosophical method as involving “reflective

equilibrium” between intuitive assessments of particular cases and rationally appealing general principles is one model of how this might occur. The idea is that just as one sometimes adjusts one’s general principles to match one’s intuitions about particular cases, one also sometimes rejects one’s intuitions about particular cases in light of one’s general principles.

Thus we imagine – and we find it hard to believe that it is not sometimes the case – that explicit moral reasoning can sometimes lead one to shift one’s intuitive judgments. Perhaps one comes to believe, in college, that there is nothing wrong with homosexuality; perhaps one comes to see something wrong with certain sorts of sexist or racist behavior one previously approved of; perhaps one comes to see the merits of vegetarianism or donation to famine relief or reducing carbon emissions or unconventional marital arrangements or limiting certain well-meaning forms of business regulation. And then one might shift one’s behavior to some extent in the direction of one’s new views. We would not deny that emotional appeal and the application of social pressure tend to greatly enhance the likelihood of such conversions of attitude; but it would be a depressing determinism indeed if such forces are *all* there is, if philosophical thinking and explicit reasoning can only be the servant and never the collaborator and partial guide of these other forces. An absolutely extreme version of the rationalist delusion picture – and we’re not sure how extreme Haidt is willing to be (he does insert qualifications at important points) – seems both humanistically unattractive and antecedently empirically implausible.

Therefore, we think it worth considering other possible explanations of the evidence. We focus on our own evidence, but we recognize that a plausible interpretation of it must be contextualized with other sorts of evidence from recent moral psychology that seem to support the idea that people suffer from a rationalist delusion – including Haidt’s own dumbfounding evidence (summarized in his 2012); evidence that we have poor knowledge of the principles

driving our moral judgments about puzzle cases (e.g., Cushman, Young, and Hauser 2006; Mikhail 2011; Ditto and Liu 2012); and evidence about the diverse factors influencing moral judgment (e.g., Hauser 2006; Greene 2008; Schnall et al. 2008).

Narrow principles. Professional ethicists might have two different forms of expertise. One might concern the most general principles and unusually clean hypothetical cases – the kinds of principles and cases at stake when ethicists argue about deontological vs. consequentialist ethics using examples of runaway trolleys and surgeons who can choose secretly to carve up healthy people to harvest their organs. Expertise of that sort might have little influence on one's day-to-day behavior. A second form of expertise might be much more concretely practical but concern only narrow principles – principles like whether it's okay to eat meat and under what conditions, whether one should donate to famine relief and how much, whether one has a duty to vote in public elections. An ethicist can devote serious, professional-quality attention to only a limited number of such practical principles; and once she does so, her behavior might be altered favorably as a result. But such reflection would only alter the ethicist's behavior in those few domains that are the subject of professional focus.

If philosophical moral reasoning tends to lead toward improved moral behavior only in specifically selected narrow domains, we might predict that the sorts of domains that ethicists tend to select for special professional focus would be the ones where we would tend to see overall better behavior on average. Those who select environmental ethics for a career focus might consequently pollute and litter less than they otherwise would, for example, in accord with our results. (Though it is also possible, of course, that people who tend to litter less are more likely to be attracted to environmental ethics in the first place.) Ethicists specializing in issues of gender or racial equality might succeed in mitigating sexist and racist behavior. Perhaps, too, we

will see ethicists donating more to famine relief and being more likely to embrace vegetarianism – issues that have received wide attention in recent Anglophone ethics and on which we found some equivocal evidence of ethicists’ better behavior.

Common topics of professional focus tend also to be interestingly difficult and nuanced. So, it might be, contra Haidt, that intellectual forms of ethical reflection do make a large difference in one’s personal behavior, but only in hard cases, where our pre-reflective intuitions fail to be reliable guides: The reason why ethicists are no more likely than non-ethicists to call their mothers or answer student emails might be because the moral status of these action is not, for them, an intuitively nonobvious, attractive subject of philosophical analysis and they take no public stand on it.

Depending on other facts about moral psychology, the Narrow Principles hypothesis might even predict – as we seem to find for the vegetarianism and charity data – that attitude differences will tend to be larger than behavioral differences, since the attitude must shift before the behavior and since behavioral change requires further exertion beyond attitude change. Note that, in contrast, a view on which people embrace attitudes wholly to rationalize their existing behaviors or behavioral inclinations would probably not predict that ethicists would show highly stringent attitudes where their behavior is unexceptional.

The Narrow Principles model, then, holds that professional focus on narrow principles can make a behavioral difference. In their limited professional domains, ethicists might then behave a bit more morally than they otherwise would. Whether they also therefore behave morally better overall might then turn on whether the attention dedicated to one moral issue results in moral backsliding on other issues, for example due to moral licensing (the phenomenon in which acting well in one way seems to license people to act worse in others; Merritt, Effron,

and Monin 2010) or ego depletion (the phenomenon according to which dedicating self-control in one matter leaves fewer resources to cope with temptation in other matters; Mead, Alquist, and Baumeister 2010).

Reasoning might lead one to behave more permissibly but no better. Much everyday practical moral reasoning seems to be dedicated not to figuring out what is morally the best course – often we know perfectly well what would be morally ideal, or think we do – but rather toward figuring out whether something that is less than morally ideal is still permissible. Consider for example, sitting on the couch relaxing while one’s spouse does the dishes. A very typical occasion of moral reflection for some of us! One knows perfectly well that it would be *morally better* to get up and help. The topic of reflection is not that, but instead whether, despite not being morally ideal, it is still *permissible* not to help: Did one have a longer, harder day? Has one been doing one’s fair share overall? And so forth. It might be the case that explicit moral reasoning can help one see one’s way through these issues. And it might furthermore be the case that explicit moral reasoning generates two different results approximately equally often: the result that what one might have thought was morally permissible is not in fact permissible (thus motivating one to avoid it, e.g., to get off the couch) and the result that what one might have thought was morally impermissible is in fact permissible (thus licensing one not to do the morally ideal thing, e.g., to stay on the couch). If reasoning does generate these two results about equally often, people who tend to engage in lots of moral reflection of this sort might be well calibrated to permissibility and impermissibility, and thus behave *more permissibly* overall than do other people, despite not acting *morally better* overall. The rationalist picture might work reasonably well for permissibility even if not for goodness and badness. Imagine someone who tends to fall well short of the moral ideal but who hardly ever

does anything that would really qualify as morally *impermissible*, contrasted with a sometimes-sinner sometimes-saint.

This model, if correct, could be straightforwardly reconciled with our data as long as the issues we have studied – except insofar as they reveal ethicists behaving differently – allow for cross-cutting patterns of permissibility, e.g., if it is often but not always permissible not to vote. It would also be empirically convenient for this view if it were more often permissible to steal library books than non-ethicists are generally inclined to think and ethical reflection tends to lead people to discover that fact.

Compensation for deficient intuitions. Our empirical research can support the conclusion that philosophical moral reflection is not morally improving only given several background assumptions, such as (i.) that ethicists do in fact engage in more philosophical moral reflection than do otherwise socially similar non-ethicists and (ii.) that ethicists do not start out morally worse and then use their philosophical reflection to bring themselves up to average. We might plausibly deny the latter assumption. Here's one way such a story might go. Maybe some people, from the time of early childhood or at least adolescence, tend to have powerful moral intuitions and emotions across a wide range of cases while other people have less powerful or less broad-ranging moral intuitions and emotions. Maybe some of the people in the latter group tend to be drawn to intellectual and academic thought; and maybe those people then use that intellectual and academic thought to compensate for their deficient moral intuitions and emotions. And maybe those people, then, are disproportionately drawn into philosophical ethics. More or less, they are trying to figure out intellectually what the rest of us are gifted with effortlessly. These people have basically made a career out of asking “What is this crazy ethics thing, anyway, that everyone seems so passionate about?” and “Everyone else seems to have

strong opinions about donating to charity or not, and when to do so and how much, but they don't seem able to defend those opinions very well and I don't find myself with that same confidence; so let's try to figure it out." It needn't be the case that most ethicists are like this, as long as there are enough to balance out whatever positive force moral reflection delivers to the group as a whole.

If this were the case, one might find ethicists, even though no morally better behaved overall, more morally well behaved than they would have been without the crutch of intellectual reflection, and perhaps also morally better behaved than non-ethicists are in cases where the ordinary intuitions of the majority of people are in error. Conversely one might find ethicists morally worse behaved in cases where the ordinary intuitions of the majority of people are a firmer guide than abstract principle. We hesitate to conjecture about what issues might fit this profile but *if*, for example, ordinary intuition is a poorer guide than abstract principle about issues like vegetarianism, charity, and environmentalism and a better guide about the etiquette of day-to-day social interactions with one's peers, then one would expect ethicists to behave better than average on the issues of the former sort and worse on issues of the latter sort.

Rationally driven moral improvement plus toxic rationalization in equal measure. A final possibility is this: Perhaps the view that Haidt is criticizing as "the rationalist delusion" is entirely right some substantial proportion of the time, but also a substantial proportion of the time explicit rational reflection is actually toxic, leading one to behave worse; and these two tendencies approximately cancel out in the long run. So perhaps we sometimes care about morality for its own sake, think things through reasonably well, and then act on the moral truths we thereby discover. And maybe the tools and habits of professional ethics are of great service in this enterprise. You might stop to think about whether it would be morally good to refrain

from eating a second cookie from the batch left in the mailroom for people to share, decide on intellectual grounds that it would be good to refrain, and thus do refrain, acting morally better as a result of your reflection than you otherwise would have acted. Call this view the rationally-driven moral improvement view. But then maybe also, in equal measure, things go just as badly wrong: When we stop to reflect, what we do is rationalize immoral impulses that we would otherwise *not* have acted on, generating a superficially plausible patina of argument that licenses viciousness that we would otherwise have avoided. Robespierre convinces himself that forming the Committee of Public Safety really is for the best, and consequently does evil that he would otherwise have avoided. Much less momentously, I concoct a superficial consequentialist or deontological story on which stealing that library book really is just fine, and so do it. And the tools of moral philosophy aid me all the more in this noxious reasoning.

If this bivalent view of moral reflection is correct, we might expect moral reflection to produce movement away from the moral truth and toward one's inclinations where common opinion is in the right and our inclinations are vicious but not usually acted on, and movement toward the moral truth where common opinion and our inclinations and behavior are all in the wrong. When widely held norms frustrate our desires, the temptation toward toxic rationalization can arise acutely and professional ethicists might be especially skilled in such rationalization. But this misuse of reason might be counterbalanced by a genuine noetic desire, which – perhaps especially with the right training – sometimes steers us right when otherwise we would have steered wrong. In the midst of widespread moral misunderstanding that accords with people's pretheoretic intuitions and inclinations, there might be few tools that allow us to escape error besides the tools of moral philosophy.

Again, one might make conditional predictions, depending on what one takes to be the moral truth. For example, if vegetarianism is not common opinion, and if it is contrary to our inclinations and yet morally good, one might predict more ethicist vegetarians. If stealing library books is widely frowned upon and not usually done, though tempting, we might expect ethicists to do so at higher rates.

Conclusion. We decline to choose among these five models. There might be truth in all of them; and still other views are available too. Maybe ethicists find themselves increasingly disillusioned about the value of morality at the same time they improve their knowledge of what morality in fact requires. Or maybe ethicists learn to shield their personal behavior from the influence of their professional reflections, as a kind of self-defense against the apparent unfairness of being held to higher standards because of their choice of profession. Or.... We believe the empirical evidence is insufficient to justify even tentative conclusions. We recommend the issues for further empirical study and for further armchair reflection.

References:

- Arpaly, Nomy (2003). *Unprincipled virtue*. Oxford: Oxford.
- Brennan, Jason (2011). *The ethics of voting*. Princeton: Princeton.
- Confucius (5th c. BCE/2003). *The analects*, trans. E. Slingerland. Indianapolis: Hackett.
- Damasio, Antonio (1994). *Descartes' error*. New York: Penguin.
- Ditto, Peter H., and Brittany Liu (2012). Deontological dissonance and the consequentialist crutch. *The social psychology of morality*, ed. M. Mikulincer and P.R. Shaver. Washington, DC: American Psychological Association.
- Green, Joshua D. (2008). The secret joke of Kant's soul. In *Moral psychology, vol. 3*, ed. W. Sinnott-Armstrong. Cambridge, MA: MIT.
- Haidt, Jonathan (2012). *The righteous mind*. New York: Random.
- Hauser, Marc D. (2006). *Moral minds*. New York: HarperCollins.
- Kant, Immanuel (1785/2003). *Groundwork of the metaphysics of morals*, trans. M. Gregor. Cambridge: Cambridge.
- Mead, Nicole L., Jessica L. Alquist, and Roy F. Baumeister (2010). Ego depletion and the limited resource model of self-control. In *Self-control in society, mind, and brain*, ed. R. Hassin, K. Ochsner, and Y. Trope. Oxford: Oxford.
- Merritt, Anna C., Daniel A. Effron, and Benoît Monin (2010). Moral self-licensing: When being good frees us to be bad. *Social and Personality Psychology Compass*, 4, 344-357.
- Mikhail, John (2011). *Elements of moral cognition*. Cambridge: Cambridge.
- Rawls, John (1971). *A theory of justice*. Harvard: Harvard.

- Rust, Joshua, and Eric Schwitzgebel (forthcoming). Ethicists' and non-ethicists' responsiveness to student emails: Relationships among self-reported behavior, expressed normative attitude, and directly observed behavior. *Metaphilosophy*.
- Schnall, Simone, Jonathan Haidt, Gerald L. Clore, and Alexander H. Jordan (2008). Disgust as embodied moral judgment. *Personality and Social Psychology Bulletin*, 34, 1096-1109.
- Schwitzgebel, Eric (2009). Do ethicists steal more books? *Philosophical Psychology*, 22, 711-725.
- Schwitzgebel, Eric (forthcoming). Are ethicists any more likely to pay their registration fees at professional meetings? *Economics & Philosophy*.
- Schwitzgebel, Eric, and Fiery Cushman (2012). Expertise in moral reasoning? Order effects on moral judgment in professional philosophers and non-philosophers. *Mind & Language*, 27, 135-153.
- Schwitzgebel, Eric, and Joshua Rust (2009). The moral behaviour of ethicists: Peer opinion. *Mind*, 118, 1043-1059.
- Schwitzgebel, Eric, and Joshua Rust (2010). Do ethicists and political philosophers vote more often than other professors? *Review of Philosophy and Psychology*, 1, 189-199.
- Schwitzgebel, Eric, and Joshua Rust (forthcoming). The moral behavior of ethics professors: Relationships among expressed normative attitude, self-described behavior, and directly observed behavior. *Philosophical Psychology*.
- Schwitzgebel, Eric, Joshua Rust, Linus T. Huang, Alan T. Moore, and Justin Coates (2012). Ethicists' courtesy at philosophy conferences. *Philosophical Psychology*, 35, 331-340.
- Stohr, Karen (2012). *On manners*. New York: Routledge.