

Philosophical Temperament

Jonathan Livengood, Justin Sytsma, Adam Feltz, Richard Scheines, and Edouard Machery

Abstract: Many philosophers have worried about what philosophy is. Often they have looked for answers by considering what it is that philosophers do. Given the diversity of topics and methods found in philosophy, however, we propose a different approach. In this article we consider the philosophical temperament, asking an alternative question: What are philosophers like? Our answer is that one important aspect of the philosophical temperament is that philosophers are especially reflective: They are less likely than their peers to embrace what seems obvious without questioning it. This claim is supported by a study of more than 4,000 philosophers and non-philosophers, the results of which indicate that even when we control for overall education level, philosophers tend to be significantly more reflective than their peers. We then illustrate this tendency by considering what we know about the philosophizing of a few prominent philosophers. Recognizing this aspect of the philosophical temperament, it is natural to wonder how philosophers came to be this way: Does philosophical training teach reflectivity or do more reflective people tend to gravitate to philosophy? We consider the limitations of our data with respect to this question and suggest that a longitudinal study be conducted.

1. Introduction

Many signs point to an identity crisis in contemporary philosophy.¹ As a group, we philosophers are puzzled and conflicted about what exactly philosophy is. Thus, in the preface of *The Philosophy of Philosophy*, Timothy Williamson writes, “This book grew out of a sense that contemporary philosophy lacks a self-image that does it justice” (2007, p. ix). The debate spurred by this identity crisis has by and large focused on methodology, and understandably so: If one wants to characterize the nature of philosophy, it seems natural to focus on what philosophers do and on what tools they employ. Unfortunately, this approach has so far been fruitless as there has been little convergence between philosophers defending different

1. We are certainly not the first generation of philosophers to wonder about our identity (e.g., Popper, 1952).

methodologies. In fact, this lack of convergence is plausibly not an accident since philosophers have always relied on a range of methods. The identity crisis is thus as urgent as ever.

In the present paper, we attack this identity crisis from a different angle. Rather than looking at what philosophers do, we examine who philosophers are, and rather than confronting philosophical methodology, we consider philosophical *temperament*. We offer evidence that philosophers are alike in being more *reflective* than their peers: They are less likely than their peers to embrace what seems obvious without questioning it, and they are disposed to submit to scrutiny their intuitive inclination to judge that something is the case. If this is an important facet of philosophical personality, as our data suggest, then one obviously wants to know where it comes from: Does it result from philosophical training as one might hope? Or, rather, is it the case that philosophers are more reflective because more reflective individuals are more likely to become philosophers?

Here is how we will proceed. In Section 2, we describe our study. In Section 3, we consider what our findings might tell us about philosophical temperament. In Section 4, we ask where philosophical temperament comes from, and we lament that our data do not answer this question, despite our best analytic efforts. Hence, in Section 5 we set out a proposal for further research into the origins of our reflective disposition.

2. Philosophical Training and Cognitive Reflectivity

In this section, we present the results of an extensive online study looking at cognitive reflectivity and how it varies with education in general and philosophical training in particular. Across more than four thousand participants, we found that philosophical training is positively

correlated with cognitive reflectivity even when overall education is taken into account. Our data suggest that thinking reflectively is part of philosophical temperament.

2.1 Cognitive Reflectivity

As we will use the phrase, *cognitive reflectivity* is a disposition to challenge one's own intuitions whenever presented with a novel problem, rather than simply relying on whatever first comes to mind. Being reflective in this sense is not necessarily equivalent to thinking hard or thinking deeply. One might think deeply about some problems when prompted to do so and yet not have any tendency to challenge one's intuitions in general. Or one might think hard about what follows from an intuition without ever stopping to ask whether or not that intuition is sound. To measure a person's cognitive reflectivity, we used the three-item Cognitive Reflection Test (CRT) developed by Shane Frederick (2005). Frederick notes that researchers often distinguish between two types of cognitive processes—what Keith Stanovich and Richard West (2000) call System-1 and System-2 processes. System-1 processes are executed quickly and spontaneously, typically with no conscious deliberation, and require little attention. By contrast, System-2 processes take more time, involve conscious deliberation, and require concentration. Frederick developed the CRT as a measure of how likely a person is to employ System-2 processes, thinking over a problem rather than simply relying on the intuitive answer.²

2. Frederick gives the following examples of tasks that rely on System-1 versus System-2 processes (2005, 26): "Recognizing that the face of the person entering the classroom belongs to your math teacher involves System-1 processes—it occurs instantly and effortlessly and is unaffected by intellect, alertness, motivation or the difficulty of the math problem being attempted at the time. Conversely, finding $\sqrt{19163}$ to two decimal places without a calculator involves System-2 processes—mental operations requiring effort, motivation, concentration, and the execution of learned rules." Note that the CRT remains an interesting measure of people's tendency to rely on their intuitions to solve various intellectual tasks even if one is skeptical of the distinction between System-1 and System-2 processes (for discussion of this distinction, see Gigerenzer and Regier, 1996; Machery, 2009, Chapter 5). Particularly, cognitive process tracing research indicates that CRT is associated with more elaborative and thorough reasoning during decision making (Cokely & Kelley, 2009).

The idea behind the CRT is to present participants with questions that have clear intuitive answers that are nonetheless incorrect. To arrive at the correct answer requires participants to move beyond the answer that initially comes to mind and to consciously reflect on the problem instead. The CRT consists of three questions of this sort (Frederick, 2005, p. 27):

- (1) A bat and a ball cost \$1.10 in total. The bat costs \$1.00 more than the ball. How much does the ball cost? _____ cents
- (2) If it takes 5 machines 5 minutes to make 5 widgets, how long would it take 100 machines to make 100 widgets? _____ minutes
- (3) In a lake, there is a patch of lily pads. Every day, the patch doubles in size. If it takes 48 days for the patch to cover the entire lake, how long would it take for the patch to cover half of the lake? _____ days

A participant's CRT score is simply the number of questions he or she answers correctly. Each CRT question has an intuitive but incorrect answer.³ However, when one stops to think about the problem, the intuitive answer is readily seen to be false. As Frederick notes, "Anyone who reflects upon it for even a moment would recognize that the difference between \$1.00 and 10 cents is only 90 cents, not \$1.00 as the problem stipulates. In this case, catching that error is tantamount to solving the problem" (2005, p. 26-27).

How should one expect philosophers to perform on the CRT? Anecdotal support can be given both for the prediction that philosophers will perform relatively poorly on the test and for

3. That these problems each have an intuitive answer is supported by our experience with the problems; even knowing the answers, the intuitive but incorrect answers jump out at us. Frederick offers additional support for this claim (2005, 27-28): "The proposition that the three CRT problems generate an incorrect 'intuitive' answer is supported by several facts. First, among all the possible wrong answers people could give, the posited intuitive answers (10, 100, 24) dominate. Second, even among those responding correctly, the wrong answer was often considered first, as is apparent from introspection, verbal reports and scribbles in the margin (for example, 10 cents was often crossed out next to 5 cents, but never the other way around). Third, when asked to judge problem difficulty (by estimating the proportion of other respondents who would correctly solve them), respondents who missed the problems thought they were easier than the respondents who solved them. For example, those who answered 10 cents to the 'bat and ball' problem estimated that 92 percent of people would correctly solve it, whereas those who answered '5 cents' estimated that 'only' 62 percent would. . . . Fourth, respondents do much better on analogous problems that invite more computation. For example, respondents miss the 'bat and ball' problem far more often than they miss the 'banana and bagel' problem: 'A banana and a bagel cost 37 cents. The banana costs 13 cents more than the bagel. How much does the bagel cost?'"

the prediction that they will perform relatively well. Recent discussions of philosophical methodology have been focused on the use of intuitions in philosophical theorizing.⁴ Accepting that many philosophers rely on their intuitions about specific cases in their philosophical theorizing, one might predict that since philosophers give high credence to their intuitions, they will tend to do poorly on the CRT. Alternatively, we might note that it is part of the stereotype of philosophers that they tend to think problems over, reflecting on them and considering them from various and sundry angles. Accepting the accuracy of this stereotype, one might predict that philosophers will do quite well on the CRT (Cokely & Feltz, 2009). To determine how philosophical training is actually related to CRT score we considered data we collected in connection with numerous other studies we conducted over the last year.

2.2 Cognitive Reflection Study

We collected data on 4,472 participants, 823 of whom reported having at least some philosophical training.⁵ Data was collected online through the Philosophical Personality website.⁶ In addition to the CRT, each participant was given a 10-item personality inventory, a philosophical probe,⁷ and short biographical and demographic questionnaires. Participants were 72.2% female, with a mean age of 35.8 years, and ranging in age from 18 to 100 years old.⁸

4. We explain what we take intuitions to be in Section 3.

5. These numbers exclude 1,041 participants who did not complete the survey, 1,326 participants who were underage and not eligible to participate, and 536 non-native English speakers owing to a suspected selection effect. Philosophical training was coded on a scale from 0 (no training) to 5 (PhD). Intermediate values represent (1) some undergraduate classes, (2) undergraduate degree, (3) some graduate classes, and (4) Master's degree.

6. <http://www.philosophicalpersonality.com>

7. CRT data was collected over the course of running a number of different studies. As such, subjects were given a wide range of different philosophical probes; none of these probes were specifically related to the CRT or cognitive reflection.

8. We have doubts about the participants who reported ages over 90 years. However, we note that when we exclude all those reporting ages greater than 90 (seven data points), the estimated models are nearly identical to those reported below using the complete dataset.

We found that the mean CRT score for the 823 participants with some philosophical training (0.98) was more than double the mean CRT score for those participants with no training in philosophy (0.44). Further, the mean CRT score for the 158 participants with some graduate training in philosophy (1.32) was triple the mean CRT score for those without training in philosophy. These results are shown graphically in Figure 1.

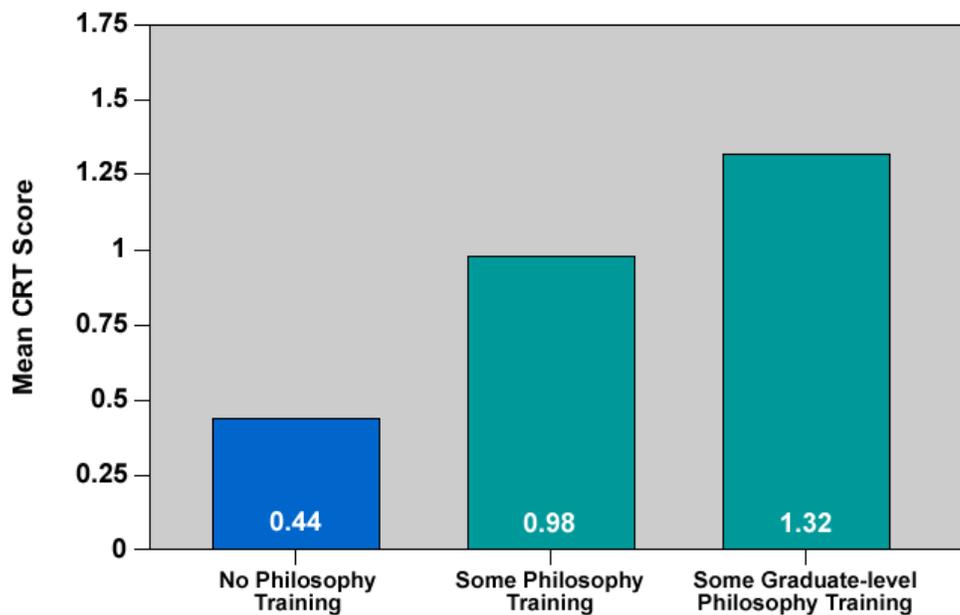


Figure 1: Mean CRT scores for participants with varying degrees of training in philosophy.

Of course, people with training in philosophy tend to be better educated than people with no training in philosophy, and it is likely that education is positively correlated with cognitive reflectivity. As such, it might be that the above numbers simply reflect the effect of education in general and not training in philosophy in particular.

What we actually find, however, is that training in philosophy positively correlates with CRT score even when controlling for education level. Thus, if we consider each level of

education⁹ separately, what we find is that with the exception of participants with a vocational or trade degree, the mean score for participants with some training in philosophy is higher than for those without training in philosophy. For example, 1201 participants reported that they had had some college education, and 189 of these also reported that they had had taken some philosophy classes. The mean CRT score for the 189 participants who had taken some philosophy courses (0.74) was nearly 70% greater than the 1012 who had not taken philosophy courses (0.43).

Similar effects held for other levels of college education, as can be seen in Figure 2.

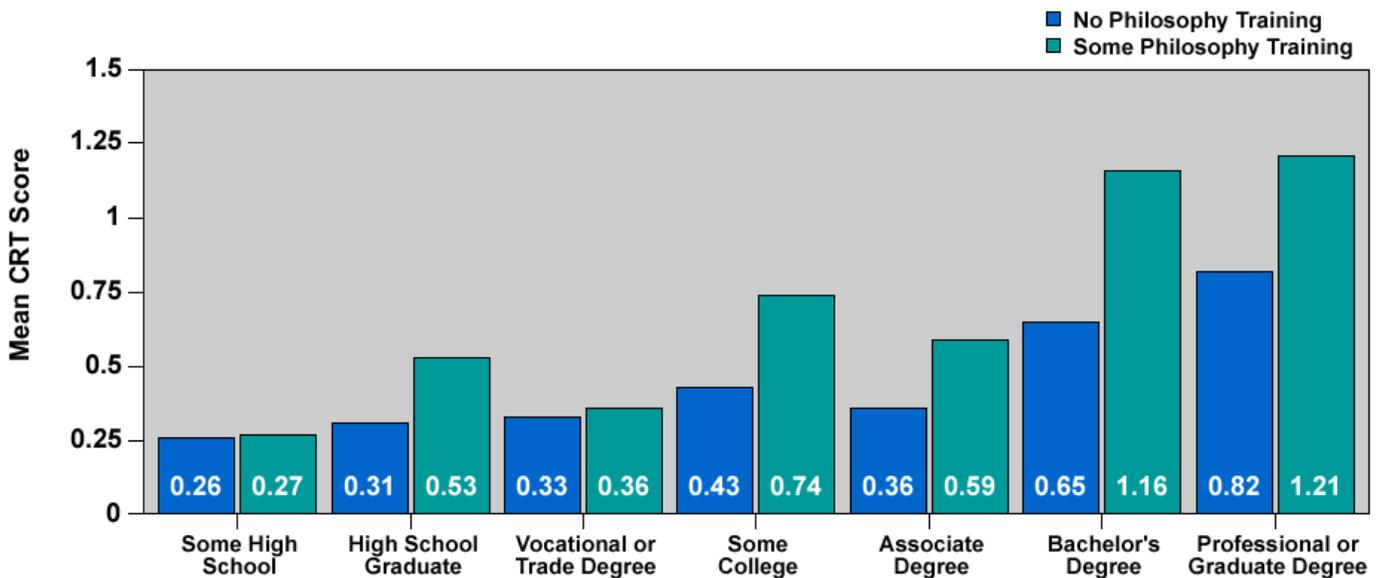


Figure 2: Mean CRT scores for participants with and without training in philosophy, broken down by level of education.

To put a finer point on it, we regressed CRT score on gender, education, philosophical training, and age according to the model:

$$\text{Score} = \beta_0 + \beta_1 \cdot \text{Gender} + \beta_2 \cdot \text{Philo} + \beta_3 \cdot \text{Edu} + \beta_4 \cdot \text{Age} + \varepsilon_S$$

9. Levels of general education were coded on a scale of 1 (some high school) to 7 (graduate or professional degree). Intermediate values represent (2) high school graduate or equivalent, (3) vocational or trade degree, (4) some college, (5) associate degree, and (6) bachelor's degree.

Philosophical training and education are both positively associated with CRT score and males have a significantly higher mean CRT score than females (with a model p-value of zero up to machine error). The intercept is not significantly different from zero. Point estimates for the other four parameters are $b_1 = 0.477$, $b_2 = 0.147$, $b_3 = 0.102$, and $b_4 = -0.003$. However, since the scales on the predictor variables are different (education has seven levels, philosophical training has five, gender has only two, and age is practically continuous), point estimates from ordinary regression may be misleading with respect to the relative importance of the predictors.

In order to clearly exhibit the relative importance of the predictors, we normalized the variables to obtain the following point estimates for the four non-zero parameters: $b_1 = 0.240$, $b_2 = 0.131$, $b_3 = 0.210$, and $b_4 = -0.042$. Philosophical training is better than half as good a predictor of CRT score as gender or overall education.¹⁰ To reiterate, philosophical training is predictive of CRT score even when one knows how much overall education an individual has. The more philosophical training one has, the higher one's CRT score is likely to be. In Section 4, we will consider whether our data support a causal interpretation of the regression model, but first, we consider what the mere association between philosophical training and CRT score might tell us about philosophical temperament.

3. Philosophical Temperament

Philosophers obviously form a disparate bunch: They engage with an impressively diverse range of issues (e.g. the reality of universals, the limits of knowledge, the nature of the levels of selection, the nature of just wars, etc.); they rely on very different methods (e.g. appeal to

10. When we added psychological training (measured on the same scale as philosophical training) to our model, we found that it was *negatively* associated with CRT score, with a normalized estimated coefficient of $b_5 = -0.067$. Also, it is worth noting that while the predictive value of gender might seem surprising, it was expected. Frederick (2005) found that on average men score significantly higher than women on the CRT (see Frederick, 2005, 37-38 for discussion).

intuitions, formalization, generalization from historical cases, or even experimentation); and they hold a wide range of opinions about specific problems (e.g. focusing merely on knowledge, foundationalism, coherentism, foundherentism, contextualism, conventionalism, reliabilism, pragmatism, etc.). In fact, the diversity of philosophers' interests, methods, and opinions is so great that looking for the nature of philosophy—philosophizing about philosophy, as it were—might strike some as a clear waste of time (even as others write books on the topic).

However, the data presented in Section 2 suggest that the variety of interests, views, and methods among philosophers hides some deep commonalities. Indeed, in light of that study, we hypothesize that philosophers tend to share a “philosophical temperament”—a cluster of dispositions that distinguishes philosophy from other intellectual endeavors. Our findings concerning philosophers' cognitive reflectivity cast light on a single, but important aspect of philosophical temperament. To describe this aspect properly, it will be useful if we first explain what we mean by “intuition”: An intuition is a spontaneous intellectual sensation: p seems to be true without being consciously inferred. In considering the first question of the CRT, for example, it intuitively seems that the answer must be 10 cents. Similarly, in the Gettier case, it intuitively seems that the agent does not have any knowledge, while in the bystander version of the Trolley case it intuitively seems morally permissible to push the lever. An intuitive seeming is analogous to a perceptual seeming: In both cases, something seems to be the case that is not consciously inferred from something else; in both cases too, people need not endorse what seems to be the case—i.e., people may refrain from judging that what *seems* to be the case really *is* the case.¹¹

11. Naturally, perceptual and intellectual seemings differ in that perception is involved in the former case, but not in the latter.

Now, some people tend to put more faith in their intellectual and perceptual seemings than others. That is, some people are disposed to believe that what seems to be the case really is the case, while others tend to “poke at their intuitions.” By “poking at intuitions,” we mean to cover a range of possible practices, which all have in common that they are meant to determine whether the intuition is trustworthy and should thus be endorsed.¹² To our knowledge, nobody has attempted to describe these practices systematically. Based on our own practices and what we can glean from philosophers’ (usually succinct) descriptions of their own philosophizing, these practices divide into at least three types.

First, some practices are meant to examine whether intuitions derive from processes known to be unreliable. This can be done in various ways. Philosophers may have acquired by experience the conviction that particular kinds of intuition are untrustworthy, and they might examine whether any particular intuition is of these kinds. Philosophers may also examine whether their intuitions are elicited by irrelevant aspects of the eliciting thought experiments. Finally, philosophers may appeal to psychological findings about the biases that affect judgments to cast doubts on some of their intuitions (as proposed, e.g., by Doris & Stich, 2006; Nichols & Knobe, 2007; Appiah, 2008; but see Machery, forthcoming, for discussion).

Second, philosophers often avoid endorsing an intuition before having considered a range of variants on the thought experiment eliciting the intuition. This way of poking at intuitions is well-described by Douglas Hofstadter in his critical discussion of Searle’s Chinese room thought experiment (Hofstadter & Dennett, 1981, p. 375):

In particular, we would like to show how Searle's setup is just one of a large family of related thought experiments, several of which are the topics of other selections of this book. Each member of this family of thought experiments is defined by a particular choice of “knob settings” on a thought-experiment generator. Its purpose is to create—in your mind’s eye—various sorts of imaginary simulations of human mental activity. Each

12. Which is not to say that such practices are reliable to meet this goal.

different thought experiment is an “intuition pump” (Dennett's term) that magnifies one facet or other of the issue, tending to push the reader toward certain conclusions.

More recently, Daniel Dennett (2005, p. 104) has noted that he embraces the method detailed by Hofstadter:

But is it a good intuition pump? How could we tell? Douglas Hofstadter's classic advice to philosophers confronted by a thought experiment is to treat it the way scientists treat a phenomenon of interest: vary it, turn it over, examine it from all angles, and in different settings and conditions, just to make sure you aren't taken in by illusions of causation. Turn all the knobs, he said, and see if the thing still pumps the same intuitions.

As with perception, Dennett suggests that we check the correctness of our intuitions by turning the object of intuition on all sides in order to get a good view of it.

Third, philosophers typically check whether intuitions and their implications are consistent with their beliefs, assuming that inconsistency would be a *prima facie* reason to question an intuition's trustworthiness. And we suspect that there might well be other intuition-poking practices employed by philosophers.

Two additional points are worth noting about intuition-poking. That philosophers employ such intuition-poking practices does not necessarily entail that they do so explicitly; rather, these practices may be skills that are implicit in expert philosophizing. In this regard, the situation may be similar to scientists' attitudes toward experimentation. Allan Franklin (2002, 2009) has convincingly argued that scientists embrace a class of strategies to determine the trustworthiness of experimental results in spite of having rarely identified these strategies explicitly.¹³

Furthermore, it is unclear whether the practices philosophers rely on to poke at intuitions really allow them to effectively sort trustworthy from untrustworthy intuitions: It could be that

13. But see, e.g., Cook and Campbell (1979) for a well-known attempt at systematizing these strategies in the behavioral sciences.

philosophers' usual practices are inefficient. Absent a systematic characterization of these practices, it is hard to say (for skepticism on this point, see Weinberg, 2007).

We propose that philosophers have a greater tendency to poke at their intuitions than do equally educated non-philosophers. That is, we propose that, by training or by selection (see Section 4), philosophers are less likely to blindly accept their intuitions and more likely to submit those intuitions to scrutiny. Philosophers ponder; they question their gut reactions, readily taking a skeptical eye toward how things seem to them. Philosophers' epistemic relation to their intellectual seemings can be usefully compared to the relation of airplane pilots to perceptual seemings during visually guided flights (e.g., the flights of non-commercial planes in the VFR ["visual flight rules"] corridors around Manhattan). Because vision through the cockpit may be distorted, it is part of pilots' expertise to be *less* likely to blindly follow their perceptual seemings.

The hypothesis that philosophers are more likely than non-philosophers to scrutinize their intuitions naturally explains our findings about philosophers' CRT scores: Because philosophers tend to poke at intuitions, they are more likely to identify those intuitions that are defective than equally educated people. This explains why they tend to do better than others on the CRT even when we control for education.

Importantly, to say that philosophers are particularly disposed to poke at intuitions is not to deny that philosophers often view the intuitiveness of a proposition as evidence for its truth. In fact, although some philosophers have challenged the claim that intuitiveness carries any epistemic weight (e.g., Hintikka, 1999; Williamson, 2007), a philosopher as influential as Saul Kripke writes (1972/1980, p. 42):

Of course, some philosophers think that something's having intuitive content is very inconclusive evidence in favor of it. I think it is very heavy evidence in favor of

anything, myself. I really don't know, in a way, what more conclusive evidence one can have about anything, ultimately speaking.

There is no inconsistency between being disposed to poke at one's intuitions and taking intuitiveness as evidence for the truth of propositions, for when the intuition survives an expert philosopher's critical scrutiny, the fact that a proposition was intuited might be taken to carry some particular weight with respect to the truth of this proposition.

However one feels about the evidential weight of intuitions, our claim is that philosophers have a distinctive tendency to question intuitions. Philosophers are typically more critical and more skeptical of what seem to be intuitively the case than non-philosophers. Our proposal about the philosophical temperament is well illustrated by what we know of the philosophizing of some prominent philosophers. David Lewis is a case in point (see also the quotation from Dennett above).¹⁴ Lewis characterizes his approach to philosophy succinctly at the beginning of *Philosophical Papers I*. He writes (1983, p. x):

Our "intuitions" are simply opinions; our philosophical theories are the same. Some are commonsensical, some are sophisticated; some are particular, some general; some are more firmly held, some less. But they are all opinions, and a reasonable goal for a philosopher is to bring them into equilibrium. Our common task is to find out what equilibria there are that can withstand examination, but it remains for each of us to come to rest at one or another of them.

On Lewis's view, philosophizing consists in probing one's and other's opinions—including one's philosophical beliefs and intuitions—through argument in order to test their coherence. One thereby determines the costs of holding various opinions, including the costs of endorsing one's intuitions. Continuing, Lewis tells us that philosophy is about measuring the price of maintaining specified collections of opinions (1983, p. x):

Philosophical theories are never refuted conclusively. (Or hardly ever. Gödel and Gettier may have done it.) The theory survives its refutation—at a price. Argle (a character in

14. Our description of Lewis' philosophizing is only based on our acquaintance with his philosophical writings.

[Lewis's] paper, *Holes*) has said what we accomplish in philosophical argument: we measure the price.

Surely, some will disagree with Lewis's view about the nature of philosophizing, and we do not mean to endorse or even recommend it here. Our point is merely that Lewis treats intuitions as opinions: They are things philosophers need to inspect on a par with philosophical claims and inspecting them consists in evaluating their costs—determining what views and what other intuitions would have to be rejected if one were to endorse them. Just like philosophical views and arguments, intuitions are thoroughly examined before being endorsed.

Anecdotes are also supportive of our characterization of the philosophical temperament. To give but one of many examples, in 2007, the Philosophy Talk radio program aired an episode titled, “If Truth is so valuable, why is there so much BS?”¹⁵ After a listener's interesting question near the end of the program, Ken Taylor (one of the hosts) asks Harry Frankfurt (the guest), “Harry, what do you think?” Frankfurt replies, “What do I think? ... Do I have *time* to think?” (emphasis in the original).

4. Education versus Selection

Where does the philosophical disposition to poke at intuitions come from? Is philosophical temperament something that can be instilled through training (as “educationists” would say) or are people with a skeptical temperament more likely to become philosophers, perhaps because they are pushed by philosophy professors to become philosophers (as “selectionists” would say)? Perhaps philosophical temperament is due in part to both learning and selection. The problem here is one of causal inference. Having seen that philosophical training is associated with CRT score, we want to know what the causal relationship is between philosophical training and

15. <http://www.philosophytalk.org/pastShows/BS.html>

cognitive reflectivity. In this section we apply graphical search and structural equation modeling to our data but are forced to conclude that they do not answer the causal question.

We would like to know whether philosophical training causes ordinary people to become more reflective or rather selects for more reflective people. Since CRT score is our only measure (or indicator) of reflectivity, in the context of the causal models below, we treat CRT as a proxy for reflectivity.¹⁶ The selectionist/educationist debate comes down to deciding whether Philo causes Reflect or whether Reflect causes Philo.¹⁷ We investigated whether our data could provide evidence on the direction of this causal relationship by using Tetrad IV¹⁸ to search for all the causal models (including models with unmeasured common causes) that are consistent with the sample covariances (Table 1) over Score, Philo, Edu, Age, and Gender.

	Score	Philo	Edu	Age	Gender
Score	0.802	0.172	0.417	-0.159	0.107
Philo	0.172	0.636	0.511	-0.154	0.056
Edu	0.417	0.511	3.422	5.591	0.017
Age	-0.159	-0.154	5.591	211.982	-0.317
Gender	0.107	0.056	0.017	-0.317	0.201

Table 1: Covariance matrix for the variables Score, Philo, Edu, Age, and Gender.

16. Since we are treating CRT score as equivalent to Reflectivity itself, all our models are observational. Identifying CRT score with the latent (unmeasured) variable Reflectivity is a strong assumption forced on us by the fact that we have only one measure of Reflectivity. Our data do not guarantee that CRT score is a good measure of Reflectivity, though such a guarantee is possible in principle. Note that we are not so daft as to claim that manipulating a participant’s CRT score might affect that participant’s level of philosophical training. Whenever a model asserts that CRT score causes philosophical training, it should be understood as asserting that cognitive reflectivity—a relatively persistent trait of an individual—causes philosophical training. Such models assert that if one could intervene to make an individual more reflective, then we would expect that individual to acquire more philosophical training with some unknown time lag.

18. If, as we suspect, philosophical training and reflectivity are connected by a feedback mechanism, the debate must change in character. However, as we shall see, many interesting issues remain if philosophical training and reflectivity are assumed to be mutual causes.

17. Tetrad is available at www.phil.cmu.edu/projects/tetrad

The FCI algorithm in Tetrad IV outputs the PAG (partial ancestral graph) shown in Figure 3. According to this PAG, in no causal model consistent with the data is Score a cause of any other variable except perhaps Philo.

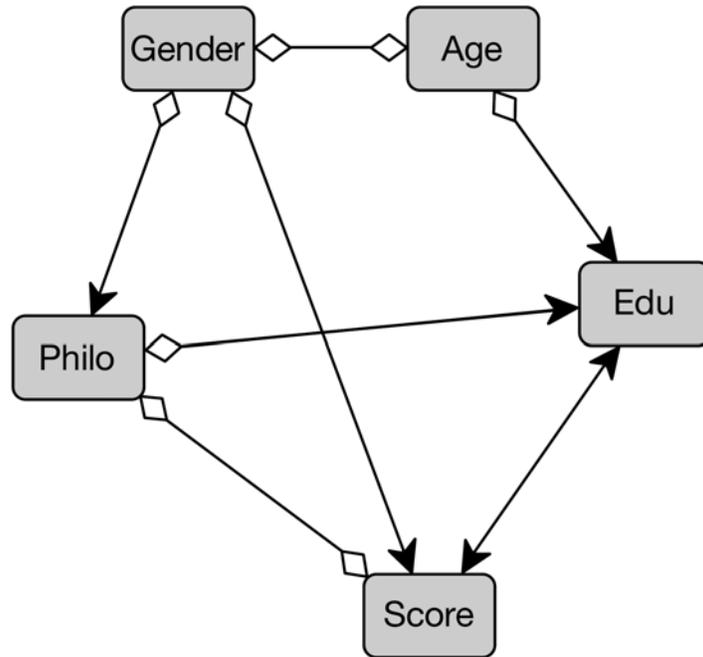


Figure 3: PAG output of an FCI search in Tetrad over the variables Score, Philo, Edu, Age, and Gender.

The unoriented edge between Gender and Age is an expected artifact—a selection effect due to the fact that women have longer life expectancy than men. The PAG entails many other causal constraints on the set of models that fit our data, but it does not orient the edge connecting Philo and Score, which means that there is not enough information in the data to determine the causal direction between them, and thus not enough information to choose between a selectionist or educationist model.

Although a PAG represents an equivalence class of causal models that all fit or don't fit the data equally well, it does not guarantee that any member of the class actually fits the data.

To answer this question, we chose two representatives of the equivalence class and two close variants that agreed on all causal relations except the one between Philo and Score. In all four models, Gender causes Age, Philo, and Score; Age causes Edu; and Philo causes Edu. The four possible relationships we estimated between Philo and Score are: (1) Philo causes Score, (2) Score causes Philo, (3) Philo and Score are effects of an unmeasured common cause (treated as an unexplained covariance), and (4) Philo and Score are mutual causes. Label the models (1) Educationist, (2) Selectionist, (3) Covariance, and (4) Feedback (see Figure 4). Models 1 and 3 are members of the equivalence class.

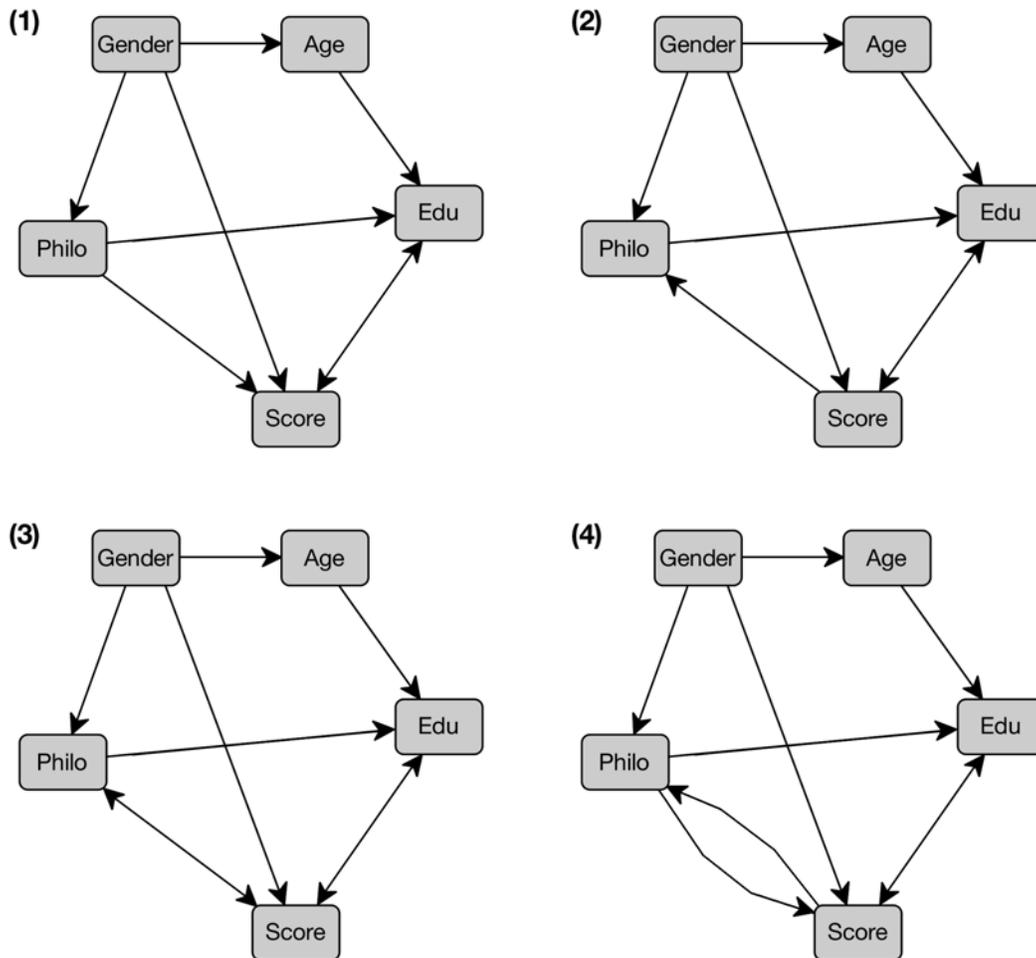


Figure 4: Four distinct causal models: (1) Educationist, (2) Selectionist, (3) Covariance, (4) Feedback.

Each graph can be statistically interpreted as a structural equation model, and its parameters and statistical fit estimated in Tetrad or with the sem package in R. Standardized estimates are given in Figure 5. In these models, the edge coefficients represent the expected change in an effect upon intervening to change the cause by one unit while holding fixed all other variables in the model. Since all variables are standardized to have unit variance and mean zero, estimates on multiple causes of a single variable represent the relative strengths of these causes.

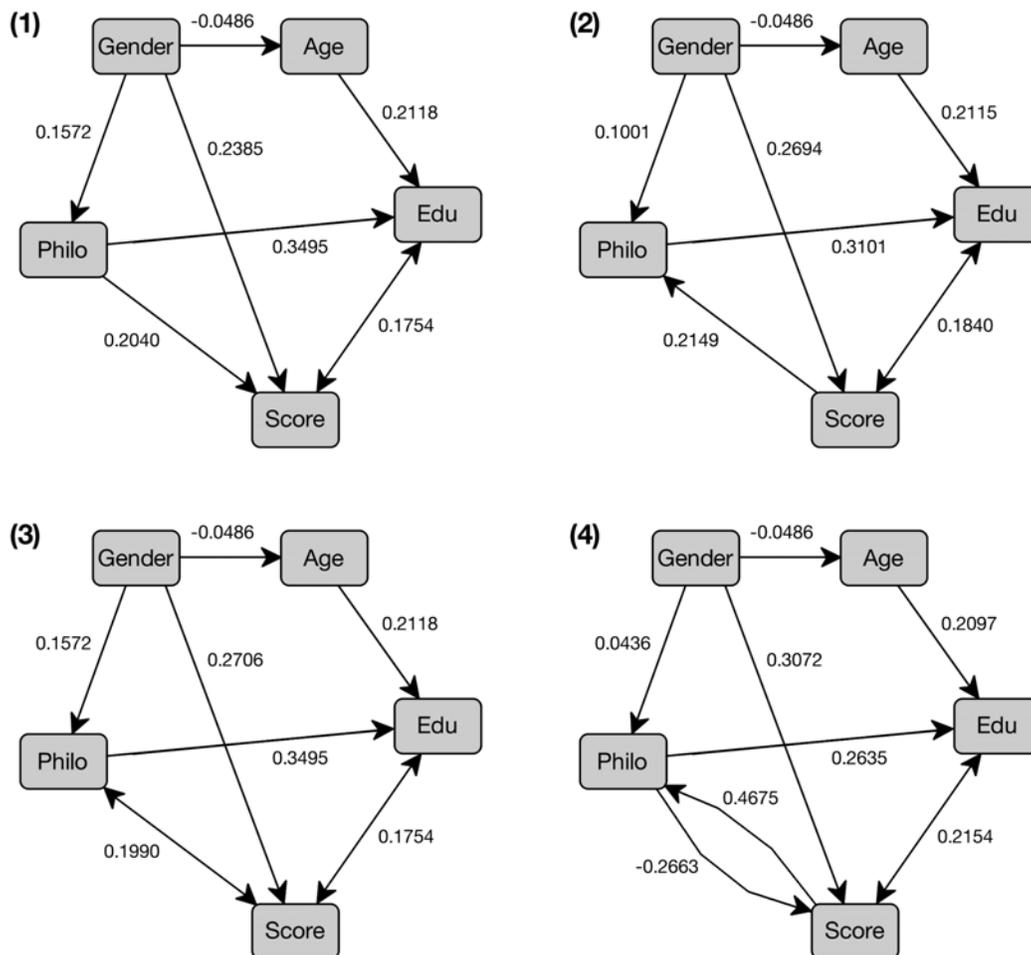


Figure 5: Standardized estimates of the relative strengths of the causal connections in each of the models shown in Figure 4.

None of these models may be rejected on the basis of our data, as can be seen from their values for selected indices of fit in Table 2.

	(1) Educationist	(2) Selectionist	(3) Covariance	(4) Feedback
Model Chi-Square	3.2845	1.8714	3.2845	2.1092
Degrees of Freedom	3	3	3	2
P-Value	0.3498	0.5995	0.34981	0.3483
Goodness of Fit	0.9997	0.9998	0.9997	0.9998
Adjusted GFI	0.9985	0.9992	0.9985	0.9986
Bentler-Bonett NFI	0.9980	0.9989	0.9980	0.9987
Bentler CFI	0.9998	1.0000	0.9998	0.9999
SRMR	0.0067	0.0052	0.0067	0.0061

Table 2: Selected indices of fit for the four models shown in Figure 4.

In fact, all four models fit the data extremely well.¹⁹ The fact that the data fit all of these models is bad news. Our data do not answer the interesting causal question: was it training or selection that made philosophers more reflective than their peers?

The careful reader may have noticed something strange about the parameter estimates for Model 4 in Figure 5. Specifically, R reports that the effect of reflectivity on philosophical training is positive but the effect of philosophical training on reflectivity is *negative*. We want to be *very* cautious about endorsing any specific set of parameter estimates of Model 4, because we

19. The null hypothesis that the chi-square is testing here is that the covariance matrix generated by the model is identical to the observed covariance matrix. Hence, unlike in ordinary hypothesis testing, we *do not* want to reject the null. Except for SRMR, the other measures indicate a good fit when they are close to one. SRMR indicates a good fit when it is close to zero.

suspect it is locally under-identified. That is, we suspect that the data do not determine a unique estimate for the coefficients in the feedback loop, though they may uniquely determine some of the other parameters in the model. Hence, we checked our results by estimating the parameters using a Bayesian technique (MCMC with non-informative priors) in AMOS 17.0. We found nearly identical estimates for Models 1-3 but not for Model 4. In Model 4, AMOS reversed the signs for the edges from Gender to Philo and from Philo to Edu. Moreover, AMOS gave different estimates for the coefficients in the feedback loop, though they had the same signs as the ones computed in R.

Were Model 4 (with the estimates computed in R or in AMOS) correct, it would indicate that while philosophy preferentially attracts and retains reflective people, philosophical training as such makes people *less* reflective than they would have been without their training. Anything said about this possibility is doubly speculative, since on the one hand, all four of the causal models we have considered fit the data, and on the other hand, Model 4 is itself under-identified. However, keeping those caveats in mind, we will venture a speculative story. Perhaps doing good philosophy requires a certain reluctance to endorse one's intuitions. Nevertheless, as one does more and more philosophy, one comes to trust one's intuitions more and more. At every stage of education, only the best philosophers are selected to go on, but those philosophers will incur a cost for their devotion—they become less reflective. This story is meant to be provocative, not assertive, since the data are not clear on this point.

Our data are surprisingly rich and they place significant constraints on possible causal models, including that gender and education are not directly connected and that age and philosophical training are not directly connected. Also they tell us something interesting conditional on which model is true; they tell us something *very* interesting conditional on the

Feedback model being true. Nonetheless, they do not provide an unconditional answer to the question, “Are philosophers more reflective than their peers because of learning or selection?” Both the educationist model and the selectionist model fit the data very well, and two other models that are not clearly educationist or selectionist also fit the data.

Overall, we find that our data do not tell us how philosophers came to be more reflective than their peers. And we cannot do any better than this without either adding theoretical constraints (which we have no defensible reason for doing) or collecting data in a way that entails further constraints (by collecting data generated from an experiment or a quasi-experiment). One simple way to do the latter is to conduct a longitudinal study and we consider how this might be done in the next section.

5. A Proposed Longitudinal Study

Our data do not speak clearly to the question of how we came to be more reflective than those around us. What are we to do? The question would certainly be answered by a controlled experiment in which we intervened to increase some participants’ level of philosophical training. But such an experiment is not feasible. Instead, we will recommend a longitudinal study.

Longitudinal studies are well-suited to deciding between causal models like the ones considered in Section 4 because data is gathered from participants at more than one point in time, allowing the researcher to see how the participants have changed over the observation period. If we were to test the cognitive reflectivity of a group of students entering college and again upon graduation, and we found that students who took philosophy courses showed a greater increase in cognitive reflectivity over that period of time than those students who did not, then we would have reason to believe that philosophical training teaches cognitive reflectivity. Alternatively, if

we found that students who end up receiving graduate degrees in philosophy tend to be more cognitively reflective at the start of their undergraduate training, then we would have reason to believe that philosophical training selects for cognitive reflectivity.

We think that some simple short- and long-term observations could help better determine the relation between philosophical training and cognitive reflectivity. Take four groups of college freshmen in the two-by-two Solomon design (1949) described in Table 3.

	Philosophical Training	No Philosophical Training
Pre-Test	Group I	Group III
No Pre-Test	Group II	Group IV

Table 3: Two-by-two Solomon design for a longitudinal study on philosophical training and reflectivity.

In the short-term, all four groups would receive the CRT at the end of one semester (the CRT post-test), but only Groups I and III would also receive a CRT pre-test.²⁰ The Solomon design controls for the confounding possibility that the act of measuring CRT score in a pre-test changes participants' scores on subsequent measurements (Hultsch & Deutsch, 1981; Rosenthal & Rosnow, 1984). The Solomon design allows us to (1) estimate the effect pre-testing would have had on Groups II and IV by looking at data from Groups I and II, (2) ensure that Group II's and Group IV's post-tests are not influenced by learning from being given the pre-test, and (3) determine if pre-testing had an effect on subsequent judgments by comparing the post-test scores of Groups I and III with those of Groups II and IV (Rosenthal & Rosnow, 1984, 93-94). If, after correcting for such effects, Group I were to show greater improvement on the CRT than Group

20. Both the post-test and the pre-test might just consist in the three-item CRT; alternatively, additional items could be developed or the CRT problems divided between multiple tests.

III, then we would have reason to believe that the educationists are right and philosophical training causes reflectivity.

In the long-term, the philosophical training of participants should be tracked to see whether those with high initial CRT scores are more likely to take philosophy classes or obtain a philosophy degree than participants with low initial CRT scores. If participants with high initial CRT scores were more likely to obtain philosophical training than those with low initial CRT scores, then we would have reason to believe that the selectionists are right and philosophical training selects for higher cognitive reflectivity. Were we to see positive results in *both* the short- and long-term observations, then we would have evidence of feedback between philosophical training and cognitive reflection.²¹

6. Conclusion

Philosophers have worried about just what philosophy is, often looking for an answer by considering what philosophers do. We have taken a different approach in this article, asking what philosophers are like. Based on an extensive internet study of both philosophers and non-philosophers, we have argued that in comparison to their peers, philosophers are especially reflective. We then suggested that this is an important commonality amongst philosophers, forming one aspect of the philosophical temperament. Accepting that reflectivity is an important aspect of the philosophical temperament, a follow-up question naturally arises: How did we get to be this way? We considered what our data can tell us about this question, looking at a range

21. It is worth noting that cohort effects can be a problem for longitudinal studies and that our proposed study is no exception. In a simple longitudinal study, only one cohort is studied over time (e.g., students as they enter into college) and that group of individuals may have unique, shared experiences that make them non-representative of other cohorts. One way to control for this is to use a “sequential” experimental design (Hultsch and Deutsch, 1981; Rosenthal and Rosnow, 1984). In sequential designs, one conducts several longitudinal experiments using different cohorts. Because the participants belong to different cohorts, it is possible to determine differences that are likely due to differences in shared experiences unrelated to the question of interest. Moreover, because different cohorts are used, it is more likely that effects found in both studies are generalizable to other cohorts.

of causal models, but unfortunately were forced to conclude that our data are unable to offer a compelling answer to the question. The task is not hopeless, however, and we suggested how the question can be answered by conducting a longitudinal study.

References

- Appiah, K. (2008). *Experiments in Ethics*. Cambridge: Harvard University Press.
- Cokely, E. & Feltz, A. (2009). Adaptive variation in folk judgment and philosophical intuition. *Consciousness and Cognition*, 18, 355-357.
- Cokely, E. & Kelley, C. (2009). Cognitive abilities and superior decision making under risk: A protocol analysis and process model evaluation. *Judgment and Decision Making*, 4, 20-33.
- Cook, T. & Campbell, D. (1979). *Quasi-Experimentation: Design and Analysis Issues for Field Settings*. Chicago: Rand McNally.
- Dennett, D. (2005). *Sweet Dreams*. Cambridge: MIT Press.
- Doris, J. & Stich, S. (2006). Moral psychology: empirical approaches. *Stanford Encyclopedia of Philosophy*, <http://plato.stanford.edu/entries/moral-psych-emp/>
- Franklin, A. (2002). *Selectivity and Discord: Two Problems of Experiment*. Pittsburgh: University of Pittsburgh Press.
- Franklin, A. (2009). Experiments in Physics. *Stanford Encyclopedia of Philosophy*, <http://plato.stanford.edu/entries/physics-experiment/>
- Frederick, S. (2005). Cognitive Reflection and Decision Making. *Journal of Economic Perspectives*, 19(4), 25–42.
- Gigerenzer, G. & Regier, T. (1996). How do we tell an association from a rule? *Psychological Bulletin*, 119, 23–26.
- Hintikka, J. (1999). The emperor's new intuitions. *Journal of Philosophy*, 96 , 127–147.
- Hofstadter, D. & Dennett, D. (1981). *The Mind's Eye*. New York: Basic Books.
- Hultsch, D. & Deustch, D. (1981). *Adult Development and Aging: A Life-span Perspective*. New York: McGraw-Hill.
- Kripke, S. (1972/1980). *Naming and Necessity*. Cambridge: Harvard University.

- Lewis, D. (1983). *Philosophical Papers, Volume I*. New York: Oxford University Press.
- Machery, E. (2009). *Doing without Concepts*. New York: Oxford University Press.
- Machery, E. (forthcoming). The bleak implications of moral psychology. *Neuroethics*.
- Nichols, S. & Knobe, J. (2007). Moral responsibility and determinism: The cognitive science of folk intuitions. *Nous*, 41(4), 663–685.
- Popper, K. (1952). The nature of philosophical problems and their roots in science. *British Journal for the Philosophy of Science*, 3(10), 124–156.
- Rosenthal, R. & Rosnow, R. (1984). *Essentials of Behavioral Research: Methods and Data Analysis*. New York: McGraw-Hill.
- Spirtes, P., Glymour, C., & Scheines, R. (2000). *Causation, Prediction, and Search*, 2nd Edition, MIT Press.
- Solomon, R. (1949). An extension of control group design. *Psychological Bulletin*, 46, 137–150.
- Stanovich, K. & West, R. (2000). Individual differences in reasoning: Implications for the rationality debate? *Behavioral and Brain Sciences*, 22(5), 645–726.
- Weinberg, J. (2007). How to challenge intuitions empirically without risking skepticism. *Midwest Studies in Philosophy*, 31, 318–343.
- Williamson, T. (2007). *The Philosophy of Philosophy*. Oxford: Blackwell.