WHY IS THERE FEMALE UNDER- REPRESENTATION AMONG PHILOSOPHY MAJORS?

EVIDENCE OF A PRE-UNIVERSITY EFFECT

SAM BARON

University of Western Australia

TOM DOUGHERTY

University of Cambridge

KRISTIE MILLER

*University of Sydney*¹

Why does female under-representation emerge during undergraduate education? At the University of Sydney, we surveyed students before and after their first philosophy course. We failed to find any evidence that this course disproportionately discouraged female students from continuing in philosophy relative to male students. Instead, we found evidence of an interaction effect between gender and existing attitudes about philosophy coming into tertiary education that appears at least partially responsible for this poor retention. At the first lecture, disproportionately few female students intended to major. Further, at the first lecture, female students were less interested in philosophy, were less self-confident about philosophy, and were less able to imagine themselves as philosophers. Similarly, female students predicted they would feel more uncomfortable in philosophy classes than male students did. Further study with a control is warranted to determine whether this interaction effect is peculiar to philosophy, or whether it is indicative of a more general gendered trend amongst first year undergraduate students.

1. Authors listed alphabetically by surname.

1. Introduction

1.1. Hypotheses Concerning Female Under-Representation in Philosophy

Female underrepresentation in the philosophy profession emerges during students' tertiary education in the United States (Paxton, Figdor, & Tiberius 2012), in Australia (Goddard, Dodds, & Burns 2008), and to a lesser degree in the United Kingdom (Beebee & Saul 2011).² Without positive action, under-representation among students will translate into under-representation in philosophy employment, which is widely recognised as problematic (Jenkins & Hutchison 2013; Friedman 2013). Solving the problem of under-representation requires understanding why the problem emerges. Several explanatory hypotheses have been proposed, which could point to either a single cause, or several causes that combine to form a “perfect storm” (Antony 2012). We survey these hypotheses in detail elsewhere (Dougherty, Baron, & Miller in press) and so we will discuss them only briefly here.

The first group of hypotheses, the *course content hypotheses*, hold that the content of philosophy courses fails to be sufficiently inclusive of women and their interests and thereby causes under-representation (Walker 2005; Superson 2011; Schouten 2015). One of these, the *role model hypothesis*, is that female students feel that they do not belong in philosophy as the result of lacking female role models in philosophy, either because of an absence of female instructors or female authors on syllabi (Hall 1993; Paxton et al. 2012). Similarly, the *subject matter hypothesis* maintains that men and women have different interests, and philosophy courses cater particularly to men's interests.

The second group of hypotheses, the *teaching methods hypotheses* focus on how teaching styles, classroom atmospheres and teacher behaviour discourage female students from studying philosophy. The *gendered intuitions hypothesis* is that male and female students have different philosophical intuitions and “male” intuitions are validated as “correct” in the classroom (Buckwalter & Stich 2014). The *learning styles hypothesis* is that philosophy is taught in a way that is ill-suited to learning styles that are disproportionately favored by women, such as styles that focus on everyday examples rather than abstract and artificial ones (Dodds & Goddard 2013) or thought experiments (Turri & Buckwalter 2015). The *aggressive argumentation hypothesis* is that philosophy frequently has an aggressive argumentative style and that this style disproportionately discourages female students (Hall 1993; Moulton 1989; Dotson 2011; Wylie 2011).

The third group, the *hostile atmosphere hypotheses* focus on the social atmo-

2. In the UK, female under-representation in philosophy becomes pronounced by graduate studies (Beebee & Saul 2011: 4).

sphere in philosophy education. The *coping methods hypothesis* is that the atmosphere is problematic insofar as there is a lack of social support networks that help students to implement social coping methods, which are particularly favored by women (Morganson, Jones, & Major 2010). The *sexist mistreatment hypothesis* posits that within philosophy female students are the victims of disrespectful, discriminatory, sexist or sexually harassing behavior by teachers or other students (Steele, James, & Barnett 2002; Haslanger 2008; Beebe & Saul 2011).

The fourth group, the *internalized stereotype / gender schema hypotheses* hold that students have internalised stereotypes or gender schemas (Valian 1998), which directly code philosophy as a male discipline (Haslanger 2008; Calhoun 2009),³ or indirectly code it as male, e.g., through the combination of a “field-specific ability belief” that philosophy requires natural brilliance and a societal stereotype of women as lacking this brilliance (Leslie, Cimpian, Meyer, & Freeland 2015). If philosophy is directly or indirectly coded as male through stereotypes or gender schemas, then this may:

- (i) make it harder for female students to imagine themselves as members of this discipline (Calhoun 2009);
- (ii) reduce female students’ interest in the subject matter of the discipline (Lupart, Cannon, & Telfer 2004);
- (iii) result in female students who are successful at the discipline being considered less likeable (Hill, Corbett, & Rose 2010: xvi);
- (iv) lead to female students holding themselves to disproportionately high standards in the discipline (Correll 2004);
- (v) leave female students vulnerable to stereotype threat in the discipline (Steele & Aronson 1995);⁴
- (vi) lead female students to have disproportionately low self-confidence in their ability to succeed in the discipline—a problem that would be exacerbated by a “fixed mindset” that sees this ability as innate, which contrasts with a “growth mindset” that sees ability as dependent on effort (Good, Aronson, & Inzlicht 2003, Dweck 2006, Dweck 2008); or
- (vii) lead to students experiencing anxiety when studying the subject and hence withdrawing from the discipline (McKinnon 2014; Schouten 2015).

The fifth group is a singleton — the *impractical subject hypothesis* is that female students disproportionately choose not to study philosophy on the grounds that it is not helpful for their life goals. This could be because male and female

3. A gender schema is a set of implicit hypotheses and stereotypes about the behaviors, traits and preferences of men and women (Valian 1998).

4. We include stereotype threat among schema hypotheses since Valian’s conception of a schema aims to be a refinement on the notion of a stereotype.

students have differing beliefs concerning how useful philosophy is, because they place a differential importance on the utility of disciplines when choosing majors, or because they have different types of goals (Calhoun 2015).

1.2. *A Chronological Taxonomy of the Hypotheses*

For investigating the causes of female under-representation, we think that it is helpful to distinguish hypotheses according to the stage in undergraduates' education at which the hypothesis predicts that under-representation will increase. Specifically, we propose distinguishing *pre-university effect hypotheses* from *classroom effect hypotheses*. The former postulate causes that increase female under-representation among prospective students who intend to major in philosophy, even before they have begun university. The latter postulate causes that will increase female under-representation among intending majors only during students' university experience.

The reason why it is helpful to draw this distinction is that it facilitates the following investigatory strategy. Pre-university effect hypotheses predict there will be under-representation among students who intend to major at the beginning of their university careers. Consequently, we can test these hypotheses by inquiring into whether there is this under-representation at this stage. By contrast, classroom effect hypotheses predict that female under-representation increases during students' university careers. We can test this by comparing female representation among intending or actual philosophy majors at various points in their university careers. Of course, given the possibility that multiple causal factors combine in a "perfect storm" (Antony 2012), evidence in favour of pre-university effect hypotheses is not *ipso facto* evidence against classroom effect hypotheses, and vice versa.

How do the aforementioned hypotheses fit into these two categories? Some hypotheses can only be classroom effect hypotheses: the course content hypotheses, the teaching methods hypotheses and the hostile atmosphere hypotheses. All of these hypotheses require students to experience philosophy education in order for the proposed increase in female under-representation to occur. However, the internalized stereotype / gender schema hypotheses could be formulated either as a pre-university effect hypothesis, or a classroom effect hypothesis. If such a hypothesis maintained that a stereotype or schema was internalised before university and this affected female students' intentions to major before university, then it would be a pre-university effect hypothesis. By contrast, if such a hypothesis maintained either that the stereotype or gender schema was internalised during classroom experience, or that a previously internalised stereotype or gender schema was activated during classroom experience (e.g., stereotype threat), then it would be a classroom effect hypothesis. Similarly, the impractical subject hypothesis could be formulated as a pre-university effect hypothesis, e.g., because women arrive at university al-

ready considering philosophy unhelpful for their goals, or as a classroom effect hypothesis, e.g., because students' classroom experience indicates that philosophy is more helpful for achieving certain goals that are disproportionately held by men.

1.3. Our Investigation Based on This Chronological Taxonomy

Currently, there are few studies into why female under-representation emerges among philosophy undergraduates.⁵ At the University of Sydney, we have taken a step along this road by surveying undergraduates before and after their first philosophy course. We had two key aims. Our first aim was to investigate whether a pre-university effect had already occurred. Accordingly, it was hypothesised that (1) at the beginning of a first-year philosophy course, there would be a difference between female students' and male students' attitudes toward philosophy. Consequently, we aimed to investigate whether there is already a gender imbalance in students' intentions to major before they take their first philosophy class, and in students' attitudes towards philosophy. Our second aim was to investigate whether the introductory course affected students' attitudes towards philosophy by comparing their attitudes before and after this course. Accordingly, it was hypothesised that (2) a first-year undergraduate philosophy course would have a more negative effect on female students' attitudes toward philosophy than on male students'. We aimed to investigate whether we could find evidence in support of at least one member of the set of classroom effect hypotheses, by investigating whether the gender ratio of students intending to major differed at the beginning and the end of the course; if female under-representation among intending majors increased over the course, then this would be evidence in support of at least one member of the set of classroom effect hypotheses. In addition, we aimed to test the following predictions of some of the aforementioned hypotheses. Versions of the internalized stereotype / gender schema hypothesis predict that in the first lecture there will be gender differences in students' self-confidence, interest in philosophy, ability to imagine themselves as philosophers and predictions of their comfort in class. A version of the impractical subject hypothesis predicts that female students will see philosophy as less useful for achieving their life goals in the first lecture. The learning styles hypotheses predict that in the last lecture female students would feel that the course suited their style of learning worse than male students felt this. The sexist mistreatment hypothesis (which is one of the hostile climate hypotheses) predicts that female students would feel that they are treated less fairly or with less

5. For an overview, see (Dougherty et al. in press) which surveys the evidence provided by (Calhoun 2009; Buckwalter & Stich 2014; Beebe & Saul 2011; Paxton et al. 2012; Adleberg, Thompson, & Nahmias 2014; Thompson, Adleberg, Sims, & Nahmias 2015). For a survey of research into female under-representation in science, technology, engineering and mathematics, see (Hill et al. 2010).

respect than male students in the last lecture. Both the hostile climate hypotheses, and the aggressive argumentation hypotheses predict that the course would have a disproportionately negative effect on female students' comfort in class.

2. Method

2.1. *Participants*

609 first-year undergraduate students from the University of Sydney, Australia were recruited for the study. Students were selected based on their attendance at the first and last lecture of an introductory philosophy course, PHIL1011 Reality, Ethics and Beauty, in 2013.

At the University of Sydney, PHIL1011 is offered in the first semester of the academic year, and both PHIL1012 Introductory Logic and PHIL1013 Society, Self and Knowledge are offered in the second semester. Students have to take two of these three courses in order to major in philosophy, and the majority of students do so by taking PHIL1011 and one of the other two courses. Students take these first year humanities courses in order to decide which humanities subject to major in or to form part of a double major, e.g., "Arts & Law." PHIL1011 had sequential components in ethics, metaphysics and aesthetics. For each of the three components, a different male instructor gave two large lectures each week. In addition, each student attended a weekly discussion section instructed by a single tutor in a class of 20 to 25 students. In the course as a whole, there were multiple tutors of both genders.

Ethics approval for the study was obtained from the Human Research Ethics Committee at the University of Sydney on 1st March 2013 (Project No.: 2013/095). Participation in the study was voluntary. All participants were over the age of 18.

2.1. *Procedure*

Data were collected at two times: during the first lecture of the semester and then during the last lecture of the semester. Each student survey was identified via a unique anonymizing code derived from the student's birthdate and mother's maiden name. This enabled us to match the response of a student who completed the first lecture survey, with his or her response in the last lecture survey. That, in turn, enabled us to gauge the extent to which a student's response to a particular question changed across the length of the course. (More on this in the results section below, under §3.1). At the first lecture, each participant was given a 15 question survey

to fill. Each participant was given 5 minutes to complete the survey.⁶ Surveys were administered by volunteer staff from the philosophy department. Following first lecture data collection, participants completed a 13 week introductory philosophy course covering three core components: ethics, metaphysics and aesthetics. At the last lecture, students completed a 20 question survey. Each participant was given 5 minutes to complete the survey. As at the first lecture, surveys were administered by volunteer staff from the philosophy department.

2.2. Materials

First lecture and last lecture surveys included a range of questions regarding the participants' views on philosophy. Both surveys included demographic questions concerning age and gender.

The first lecture survey included 11 further questions on attitudes toward philosophy, including intention to major, intention to take more philosophy classes, usefulness of philosophy for life goals, perceived ability in philosophy, ability to imagine becoming a philosopher, relationship between ability and natural talent in philosophy, interest in philosophy, personal meaningfulness of philosophy, class participation, and ability to overcome obstacles. Last lecture surveys included the same 11 questions. All 11 questions were scored on a 5-point Likert scale, where '1' was the highest (strongly agree), and '5' the lowest (strongly disagree).

Two further questions were included in the first lecture survey and last lecture survey concerning (i) perceived factors important for choosing a major and (ii) perceived reasons for not contributing in class discussion. Participants were given a choice of 5 options for each question, and asked to rank each in order of importance.

Last lecture surveys included 5 extra questions on: (i) respectfulness of teaching staff in philosophy; (ii) relationship between interactions with teaching staff and desire to do philosophy; (iii) relationship between interactions with other students and desire to do philosophy; (iv) relationship between philosophy and learning style; (v) performance within the course. Complete first lecture and last lecture surveys are provided in Appendix A.

3. Results

The results are divided into four sections. Section 3.1 contains descriptive statistics, which provide general demographic information concerning the study's par-

6. Based on practice filling out the survey by volunteer staff prior to administration, five minutes provided sufficient time to complete the survey successfully.

ticipants. The next two sections are then oriented toward each of the two main classes of hypotheses outlined in the introduction. In Section 3.2, we outline results relevant to pre-university effect hypotheses, focusing on gender differences in attitudes toward philosophy in the first lecture. In Section 3.3, we outline results relevant to classroom effect hypotheses focusing on differences between male and female students with respect to how their attitudes toward philosophy changed across the course. Section 3.4 reports effect sizes.

3.1. *Descriptive Statistics*

596 participants successfully completed the survey at the first lecture.⁷ 8 participants were not included in the analysis based on being neither male nor female, leaving a final sample of 588. Of these, 230 were male and 357 were female. 252 participants completed the survey at the last lecture. 8 participants were not included in the analysis based on being neither male nor female, leaving a final sample of 244. Of these 96 were male and 148 were female. The gender ratio at the first lecture survey was the same as the gender ratio at the last lecture survey (1 man to 1.5 women, see Table 1). So while there was attrition between the first and last lectures, men and women left the course in equal numbers.

{Finally, by using the anonymised code provided by students who completed the first lecture survey and the last lecture survey, we were able to match participants' responses to the first lecture survey with their responses to the last lecture survey. 125 students in total were matched in this way, and so 125 completed both the first lecture and last lecture surveys. Of these participants, 47 were male and 78 were female. The gender ratio of those who completed the first and last lecture surveys was not substantially different from the gender ratio for those who completed the first survey and those who completed the last survey (see Table 1). Note that not every member of the 125 completed every question on both surveys. Note also, that the data from questions 3 and 6 on both surveys were not used due to a systematic error in the way students responded to both questions.⁸

7. For present purposes, 'successful completion' means answering the majority of questions on the survey. The degrees of freedom column in Tables 8, 9 and 10 in Appendix B show variance in the number of students who answered each question. To find n (the number of students who answered each question) for a t-test simply add 2 to the degrees of freedom.

8. The systematic error at issue involved a fundamental misunderstanding concerning how to rank the items associated with this question. For example, some students ranked everything as '1', or ranked 'other' above all other options, making statistical comparison difficult to say the least. This was largely due to a lack of detailed instruction on how to perform the ranking task.

TABLE 1. Gender proportions between first and last lecture: the ratio of men to women remained constant throughout the course, despite a student attrition rate of 42%

Lecture	Number of Men	Number of Women	Ratio M:F
Attended First lecture Only	230	357	1:1.54
Attended Last lecture Only	96	148	1:1.54
Attended Both First and Last Lecture	47	78	1:1.66

3.2. Pre-University Effect Hypotheses

According to pre-university effect hypotheses, there are gender differences in attitudes toward philosophy that are exogenous to tertiary level study in philosophy. Testing of this claim proceeded in two phases. In phase one, independent samples t-tests were carried out for all questions at the first lecture to compare the average response for male students with the average response for female students. T-tests were used because they provide information about the degree to which the average for men differs from the average for women, which allows us to test for gender differences with respect to the responses for each question. Very roughly, the further apart the means for each group are, the more likely the test is to yield significance. Because multiple comparisons were used on this data set (i.e., 10 tests were carried out), a Bonferroni correction was performed, reducing the alpha level from the standard .05 and .01 levels to .005 and .001 respectively.⁹ Note that it is commonly recognised that the Bonferroni correction is an extremely conservative way to handle the compounded chance of error due to multiple comparisons. While it rules out false positives, it also generates false negatives. Accordingly, we have included the uncorrected significance results as well, as we believe these still to be of interest.

The results of the t-tests were as follows. At the first lecture, the mean difference between male and female participants with regard to perceived ability to do well in philosophy was significant ($t_{583} = -4.072$, $p < .01$). Because the average for female students was higher than the average for male students ($\mu_F = 2.32 > \mu_M = 2.08$), female students were less likely than male students to believe that they have the ability to do well in philosophy. In addition, male students could imagine themselves becoming a philosopher significantly more than female students could ($t_{579} = -5.183$, $p < .005$), with the male average at 3.05, and the female average at 3.48. Moreover, male students found philosophy more interesting than female

9. Basically, an alpha value is the threshold probability that a test yielded results that are compatible with the null hypothesis. It is standard to use an alpha value of .05 (5 % chance that the results are due to the null) as the threshold for significance. When multiple comparisons are carried out, the chance of false positives is increased and so the alpha value is often reduced; the test for significance is made more stringent.

students did ($t_{586} = -2.903$, $p < .005$), with female students on average less willing to agree that philosophy is interesting than male students ($\mu_F = 1.69 > \mu_M = 1.54$). Female students were also, on average, significantly less inclined than male students to predict they would feel comfortable about participating in class discussion in philosophy ($t_{584} = -6.194$, $p < .005$, $\mu_F = 2.49 > \mu_M = 2.07$). Though statistical significance was not detected for the difference between male students and female students with respect to the personal meaningfulness of philosophy, a trend was detected prior to Bonferroni correction ($t_{584} = -1.854$, $p = 0.064$, $\mu_F = 1.91 > \mu_M = 1.80$) (see Table 2 for a brief account of the results, see Table 8 in Appendix B for results in full).

TABLE 2. Gender Differences at First Lecture (summary).

Question	P-value	Female Mean	Male Mean	Significant? (Y/N)
Q1. Intend to Major	.486	3.29	3.23	N
Q2. Intend to Take More Courses	.546	2.41	2.36	N
Q4. Useful for Life Goals	.788	2.41	2.39	N
Q5. Self-Confidence	.000	2.32	2.08	Y
Q7. Can Imagine Self as Philosopher	.000	3.48	3.05	Y
Q8. Talent Rather than Effort	.732	3.42	3.44	N
Q9. Philosophy is Interesting	.004	1.69	1.54	Y
Q10. Philosophy is Personally Meaningful	.064	1.91	1.80	N
Q11. Comfort in Class	.000	2.49	2.07	Y
Q12. Ability to Cope	.138	2.31	2.23	N

A simplified version of Table 8 in Appendix B with the averages for the two genders, and a determination of whether the differences between the averages are significant.

According to the first row of Table 2, no difference between male students and female students for intention to major was detected via an independent samples t-test.

While there was no significant difference in the average response for men and women, it was determined that the data for this question was skewed in a way that could not be detected via a t-test. The parity in the means for intention to major was, in effect, masking a difference in the *proportions* of men versus women who intend to major. The number of women who intended to major was proportionally smaller than what was expected, given the number of women who attended the course, and assuming that women and men are equally likely to major. This could only be seen, however, once ‘intention to major’ was converted into a categorical variable. By narrowing down onto those male students who intend to major in philosophy and treating this as a single category (made up of those who strongly agree and agree) and those female students who intend to major in philosophy and

Figure 1. Gender Difference for Intention to Major. Observed gender proportions for question one versus expected gender proportions. Expected gender proportions for female students are higher than for male students due to the proportion of female students to male students enrolled in the course (roughly 2:1). Error bars represent 95% confidence interval.

treating this as a single category (made up of those who strongly agree and agree), a significant difference in the gender proportions for intention to major was detected using a chi-square goodness of fit test prior to Bonferroni correction (χ^2 (1, $N = 118$) = 7.618, $p = .006$).¹⁰ We used these tests to compare observed frequencies with expected frequencies. Expected frequencies in this case were determined via gender proportions of the overall sample. The further that expected frequencies are from observed frequencies, the more likely the test is to yield significance. The results of this test tell us, roughly, that the gender proportions for intention to major were not what we would expect, given the overall gender break-down of the course. Because there were more female students in the course than male students, we would expect there to be more female students intending to major than male students. (There should be 72 female students and 46 male students intending to major, see Figure 1). What we observed, however, was a near equality in number

10. It may seem ad hoc to deploy a chi square test when the t-test has failed. It is important to recognise, however, that the two kinds of tests are looking at different statistical phenomena: t-tests consider only differences in means; chi-square tests look at differences in proportions. A significant difference in proportion may well not have a difference on the mean nor vice versa. Since proportions are, in this case at least, important, it is reasonable to deploy both kinds of test in order to see if there are any significant differences between the genders.

of male students and female students intending to major (61 male students and 57 female students, see Figure 1). Note that while the difference between the observed gender break-down and the expected gender break-down was significant prior to Bonferroni correction, post correction the significance disappears. Nonetheless, we believe there remains some defeasible evidence that female students were less inclined to major than male students. Again, Bonferroni corrections are very conservative, and so a less conservative correction is likely to preserve the significance result.

In phase two of testing, we carried out a general correlational test looking for broad statistical relationships of potential interest in order to gain an idea of the interplay between the various factors probed by the survey. Correlational tests of this kind look for patterns in the data by considering the extent to which a response to one question can be used to predict responses on other questions. We looked for correlations between all questions, this time factoring out gender. By doing so, we were able to identify broader patterns of response across multiple questions, determining the extent to which a respondent's attitudes toward philosophy came as a 'package deal', with his or her responses to multiple questions moving together. A package deal of responses is what we would expect were a schema of some kind responsible for student's responses, and so the correlational tests are important for probing this aspect of the pre-university effect hypotheses.

The results of the first-pass correlational test are depicted in Table 3. Most of the questions are correlated with one another; the data displays a strong pattern of response.

3.3. *Classroom Effect hypotheses*

According to classroom effect hypotheses, there are gender differences in attitudes toward philosophy that are endogenous to tertiary level study in philosophy. To test this claim, we deployed a repeated measures design. We first narrowed the data down to only those students who responded to both surveys. With the data narrowed, we then compared the change in attitudes toward philosophy displayed by male students with the change in attitudes toward philosophy displayed by female students across the course. We did this by (i) identifying, for each question, the mean response for male students and female students in the first lecture; (ii) identifying, for each question, the mean response for male students and female students in the last lecture; (iii) subtracting, for each question, the mean response for each gender in the first lecture from the mean response for each gender in the last lecture to obtain the mean *difference* for each gender: the mean difference represents the extent to which male students' and female students' responses to these questions changed across the course—i.e., the degree to which their averages moved for each question; and (iv) comparing, for each question, the mean differ-

TABLE 3. Correlations at First Lecture

Question	Q1	Q2	Q4	Q5	Q7	Q8	Q9	Q10	Q11	Q12
Q1. Intend to Major	–									
Q2. Intend to Take More Courses	.479**	–								
Q4. Useful for Life Goals	.251**	.287**	–							
Q5. Self-Confidence	.075	.203**	.232**	–						
Q7. Can Imagine Self as Philosopher	.403**	.282**	.354**	.216**	–					
Q8. Talent Rather than Effort	.032	–.008	.021	–.044	.063	–				
Q9. Philosophy is Interesting	.253**	.369**	.386**	.266**	.300**	–.016	–			
Q10. Philosophy is Personally Meaningful	.260**	.365**	.454**	.232**	.301**	–.020	.693**	–		
Q11. Comfort in Class	.136*	.222**	.173**	.341**	.192**	–.015	.280**	.318**	–	
Q12. Ability to Cope	.034	.115^^	.065	.265**	.087^	–.050	.146**	.118^^	.297**	–

Each row and each column represents a question. Each cell, which is the intersection between two questions, represents the correlation between those two questions, the extent to which the response to one question predicts the response to the other. Positive values represent positive correlations, which tell us that a higher value on one question in a two-question pair predicted a higher value on the other question in that pair. A negative value represents a negative correlation, which tells us that a higher value on one question in a two-question pair predicted a lower value on the other question in that pair.

** . Correlation is significant at the 0.0002 level (2-tailed).

* . Correlation is significant at the 0.001 level (2-tailed).¹¹

^ Correlation is significant at the 0.01 level prior to Bonferroni correction.

^^ . Correlation is significant at the 0.05 level prior to Bonferroni correction.

ence for male students with the mean difference for female students using a t-test in order to determine whether one gender's responses had changed more than the other gender's responses.¹²

11. Because multiple comparisons were used on this data set (i.e., 45 tests were carried out), a Bonferroni correction was carried out, reducing the alpha level from the standard .05 and .01 levels to .001 and .0002 respectively

12. Note that for this last step, step (iv), we deployed an independent samples t-test. A paired samples t-test has the power to tell us *for women* whether their views changed across the course, or *for men* whether their views changed across the course. What it won't tell us, however, is whether the change experienced by men was significantly different to the change experienced by women. To determine the interaction effect here, we had to first identify the difference between men's views at the start and at the end of the course, then identify the difference between women's views at the start

TABLE 4. Gender Differences Between First and Last Lectures (summary)

Question	P-value	Female Mean (First)	Female Mean (Last)	Male Mean (First)	Male Mean (Last)	Significant? (Y/N)
Q1. Intend to Major	.803	3.29	3.655	3.23	3.323	N
Q2. Intend to Take More Course	.969	2.41	2.642	2.36	2.469	N
Q4. Useful for Life Goals	.877	2.41	2.72	2.39	2.41	N
Q5. Self-Confidence	.750	2.32	2.5	2.08	2.167	N
Q7. Can Imagine Self as Philosopher	.890	3.48	3.8	3.05	3.31	N
Q8. Talent Rather than Effort	.255	3.42	3.21	3.44	3.40	N
Q9. Philosophy is Interesting	.053	1.69	1.75	1.54	1.71	N
Q10. Philosophy is Personally Meaningful	.740	1.91	2.18	1.80	1.97	N
Q11. Comfort in Class	.688	2.49	2.58	2.07	2.14	N
Q12. Ability to Cope	.588	2.31	2.6	2.23	2.26	N

A simplified version of Table 9 in Appendix B with the averages for the two genders, and a determination of whether the differences between the averages are significant.

Stage (iv) of the process allowed us to determine whether the mean difference for female students was greater than the mean difference for male students and thus whether female students' attitudes toward philosophy changed more than male students'. No significant differences in gender were detected between the first lecture and the last lecture (see Table 5). However, a trend was detected for Question 9 (philosophy is interesting) ($t_{123} = 1.956$, $p = 0.053$) prior to Bonferroni correction, with the mean difference for female students coming out greater than for male students—i.e., a trend towards female students becoming less interested in philosophy than male students. This trend in our results was not significant, though this may have been due to the relatively small sample size of our study.

The last lecture survey also introduced five new questions on: (i) the extent to which students felt that interactions between staff and students were fair (Q13); (ii) the extent to which interactions with staff motivated students to take more

and at the end of the course, and then compare these differences. Since these differences did not form a single comparison group, a paired samples t-test would have been improper. An alternative option would have been to use a 2x2 ANOVA, with classroom exposure as the repeated measure and gender as the between-group factor. However, given that no interaction effect showed up in the independent t-test performed, it is doubtful that the ANOVA would have revealed anything either.

TABLE 5. Gender Differences at Last Lecture (summary)

Question	P-value	Female Mean	Male Mean	Significant? (Y/N)
Q1. Intend to Major	0.34	3.655	3.323	N
Q2. Intend to Take More Courses	.295	2.642	2.469	N
Q4. Useful for Life Goals	.031	2.72	2.41	N
Q5. Self-Confidence	.003	2.5	2.167	Y
Q7. Can Imagine Self as Philosopher	.001	3.8	3.31	Y
Q8. Talent Rather than Effort	.159	3.21	3.40	N
Q9. Philosophy is Interesting	.706	1.75	1.71	N
Q10. Philosophy is Personally Meaningful	.057	2.18	1.97	N
Q11. Comfort in Class	.001	2.58	2.14	Y
Q12. Ability to Cope	.002	2.6	2.26	Y
Q13. Interactions were Fair	.352	1.88	1.79	N
Q14. Interactions with Staff	.106	2.49	2.28	N
Q15. Interactions with Students	.352	2.76	2.65	N
Q16. Learning Style	.131	2.8	2.60	N
Q17. Performance of Self vs. Others	.090	2.88	2.65	N

A simplified version of Table 10 in Appendix B with the averages for the two genders, and a determination of whether the differences between the averages are significant.

philosophy (Q14); (iii) the extent to which interactions with other students were a motivational factor in taking more philosophy (Q15) (iv) student learning styles (Q16) and (v) comparisons of one's own performance with other students' performance (Q17). No significant gender differences were discovered for these five questions (see Table 4). It is notable that, in the first survey, questions 5, 7, 9 and 11 yielded significant gender differences whilst on the final survey questions 5, 7, 11 and 12 yielded significant gender differences. This might seem odd, given that there was no significant difference in the change in attitudes toward men and women toward philosophy across the course (see Table 5). However, the difference here is likely due to some further factor, such as the drop off in number of participants between the first and last surveys.

Once gender was factored out, a similar pattern of correlations as in the first lecture was observed, with student responses to, for instance, intention to major being highly correlated with all other questions (see Table 6).

TABLE 6. Correlations at Last Lecture

Question	Q1	Q2	Q4	Q5	Q7	Q8	Q9	Q10	Q11	Q12	Q13	Q14	Q15	Q16	Q17
Q1. Intend to Major	–														
Q2. Intend to Take More Courses	.655**	–													
Q4. Useful for Life Goals	.482**	.532**	–												
Q5. Self-Confidence	.378**	.418**	.434**	–											
Q7. Can Imagine Self as Philosopher	.635**	.572**	.501**	.477**	–										
Q8. Talent Rather than Effort	.042	.095	–.040	–.118	.156^	–									
Q9. Philosophy is Interesting	.312**	.436**	.432**	.332**	.310**	–.065	–								
Q10. Philosophy is Personally Meaningful	.299**	.413**	.434**	.261**	.283**	–.011	.730**	–							
Q11. Comfort in Class	.305**	.365**	.269**	.546**	.403**	–.031	.259**	.258**	–						
Q12. Ability to Cope	.381**	.404**	.338**	.566**	.435**	–.073	.399**	.376**	.514**	–					
Q13. Interactions were fair	.256**	.279**	.315**	.379**	.300**	.025	.399**	.299**	.359**	.396**	–				
Q14. Interactions with staff	.464**	.503**	.461**	.417**	.432**	–.022	.502**	.410**	.444**	.492**	.553**	–			
Q15. Interactions with students	.341**	.401**	.345**	.363**	.376**	.010	.352**	.263**	.449**	.422**	.339**	.535**	–		
Q16. Learning Style	.390**	.372**	.308**	.413**	.307**	–.191^^	.383**	.370**	.437**	.555**	.395**	.552**	.467**	–	
Q17. Performance of self vs. others	.403**	.392**	.321**	.609**	.393**	–.072	.345**	.305**	.450**	.556**	.315**	.448**	.364**	.500**	–

** . Correlation is significant at the 0.000095 level (2-tailed).

* . Correlation is significant at the 0.000476 level (2-tailed).

^ . Correlation is significant at the 0.01 level prior to Bonferroni correction.

^^ . Correlation is significant at the 0.05 level prior to Bonferroni correction.

Each row and each column represents a question. Each cell, which is the intersection between two questions, represents the correlation between those two questions, the extent to which the response to one question predicts the response to the other. Positive values represent positive correlations, which tell us that a higher value on one question in a two-question pair predicted a higher value on the other question in that pair. A negative value represents a negative correlation, which tells us that a higher value on one question in a two-question pair predicted a lower value on the other question in that pair.

TABLE 7. Effect Size

Question	Result	Effect Size (d)
Q1. Intend to Major (First Lecture)	$\chi^2 (1, N = 118) = 7.618, p = .006$	0.3521038
Q5. Self-confidence (First Lecture)	$t_{583} = -4.072, p < .01$	0.4364008
Q7. Can Imagine Self as Philosopher (First Lecture)	$t_{579} = -5.183, p < .005$	0.2332713
Q9. Philosophy is Interesting (First Lecture)	$t_{586} = -2.903, p < .005$	0.5203307
Q11. Comfort in Class (First Lecture)	$t_{584} = -6.194, p < .005$	0.5254
Q5. Self-Confidence (Last Lecture)	$t_{242} = -3.032, p < .005$	0.3987777
Q7. Can Imagine Self as Philosopher (Last Lecture)	$t_{241} = -3.377, p < .005$	0.4452028
Q11. Comfort in Class (Last Lecture)	$t_{242} = -3.464, p < .005$	0.4581778
Q12. Ability to Cope (Last Lecture)	$t_{240} = -3.073, p < .005$	0.4196813

The final column represents the size of the effect detected. Each row represents one of the statistical tests that yielded a significant result in the study. Effect sizes are generally strong, between medium and large.

3.4. Effect Sizes

When developing empirical research of the kind that we are engaged with here—that is, research into real populations—it is important to report the size of the effect(s) detected by statistical analysis. This trend is prevalent in empirical psychology, medicine and biology *inter alia*. Any empirical research that yields statistical results must also report effect sizes, since even if a result is statistically significant, if the effect size is small, then the result is considered to be trivial. Very roughly, an effect size tells us how large and thus how important the phenomenon is that has been detected by the relevant statistical tests carried out. It is standard to consider an effect size (d) of 0.2 small; an effect size of 0.5 medium and an effect size of 0.8 or more to be large.¹³ Effect sizes for our study are outlined below (Table 7). It is notable that the effects detected are medium in size, suggesting that the gender difference detected between men and women at the first lecture and the gender difference detected between men and women at the last lecture were both non-trivial (anything below 0.2 is considered to be trivial).

¹³ The scale of the effect size, (d), is a proportion of the number of standard deviations by which the means of the two groups differ. So if (d) = 1, then the two means differ by one standard deviation. Because there is no theoretical limit to the number of standard deviations that two means can differ by, there is no highest value for (d). Accordingly, (d) should not be confused with a probability measure, which has a minimum value of 0 and a maximum value of 1. Note that while values for (d) have no theoretical upper limit, it is unusual for (d) to take high values (i.e. values > 1).

4. Discussion

This study aimed to determine the extent to which pre-university effect hypotheses or classroom effect hypotheses explain female under-representation in undergraduate philosophy. Specifically, it was hypothesised that at the beginning of a first-year philosophy course there would be a difference between female students' and male students' attitudes toward philosophy and second, it was hypothesised that a first-year undergraduate philosophy course would have a more negative effect on female students' attitudes toward philosophy than for male students'.

4.1. *Evidence of Pre-University Effect Hypotheses*

We found significant differences at the first lecture between the attitudes of male students and female students to the study of philosophy. In particular, fewer female students than expected intended to major in philosophy, given the overall gender ratio of the course. Since these differences were present in the first lecture, our results provide some support for the claim that one of the pre-university effect hypotheses identifies a significant cause of female under-representation in our sample. Our results also brought some nuance to the issue, given the unexpected finding that among those taking the course, female students were over-represented. We are unsure about the explanation of this finding, or whether it would be found at other universities. All the same, the fact that disproportionately few female students taking the course intended to major indicates that some pre-university effect had occurred.¹⁴

This raises the question of *which* pre-university effect had occurred. Our results failed to provide support for a version of the impractical subject hypothesis, which predicts that in the first lecture there would be a significant gender difference with respect to how useful students saw philosophy for achieving their goals in life.¹⁵ We failed to find any such difference to support this prediction.

On the other hand, our results provided support for versions of the internal-

14. There is an alternative explanatory hypothesis, which is that female and male students were equally likely to intend to major in philosophy, but out of all the students who did not intend to major in philosophy, disproportionately many female students chose to take the first-year course, e.g. out of pure interest in exploring a subject outside of their major. This hypothesis could explain the even gender balance among students intending to major, and the gender imbalance of students taking the course, which together would give rise to the fact that, out of the students taking the course, unexpectedly few female students intended to major. We consider this hypothesis unlikely, however, given our other findings that female students' had more negative attitudes towards philosophy than male students in the first lecture.

15. This is not the only form that the hypothesis may take. In addition, it may be that male and female students have similar beliefs about how useful philosophy is for achieving their life goals, but female students place greater weight on these considerations when choosing their majors. Thanks to an anonymous reviewer for making this point.

ized stereotype / gender schema hypotheses. As we discussed earlier, some versions of this hypothesis predict gender differences in students' interest in philosophy, self-confidence, ability to imagine themselves as philosophers, and comfort in class. This study found gender differences in all four areas. First, male students found philosophy significantly more interesting than female students.¹⁶ Second, female students saw themselves as less able to do well in philosophy than did male students. Third, female students were less able than male students to imagine themselves as philosophers. Fourth, female students were less likely than male students to predict that they would be comfortable in classroom discussion.

Another feature of our results that supports versions of the internalized stereotype / gender schema hypotheses is that answers to the questions in the first survey that were significantly correlated with gender were also significantly correlated with one another. Gender predicts one's answers to questions concerning intention to major, perception of ability in philosophy, ability to imagine becoming a philosopher, interest in philosophy and comfort in philosophy class. Similarly, responses to each of these questions predict responses to each of the others. This suggests that students' attitudes toward these issues come as a 'package deal.' This is what we would expect if those attitudes are ultimately influenced by a structured schema of some kind, which encodes attitudinal responses to a set of interrelated questions.

Along these lines, we would like to indicate a potential affinity between our results and some other important recent results of investigation into female under-representation in philosophy and other disciplines, carried out principally by Sarah-Jane Leslie and Andrei Cimpian (Leslie et al. 2015). This investigation involved a nation-wide survey of faculty, postdoctoral fellows and graduate students at public and private research universities across the United States. The survey presented respondents with claims concerning what is required for success in their field, e.g., "Being a top scholar of [discipline] requires a special aptitude that just can't be taught" (Leslie et al 2015: 262), and asked them to indicate the extent to which they agreed with the claim, and the extent to which they believed people in their field would agree with the claim. These answers were averaged to produce a measure of how much the field emphasized raw talent. Leslie et al. found that the more a field valued giftedness, the fewer female PhDs there were in the field. As such, Leslie et al. found powerful evidence in support of what they call the *field-specific ability belief hypothesis*.

It is important to note that Leslie et al.'s focus was on representation at the

16. Moreover, interest in philosophy is not only significantly correlated with gender, it is also significantly correlated with every other question, except for judgements about the extent to which success in philosophy depends on natural talent. It is notable, however, that the significant correlation between gender and interest in philosophy only appears at the first lecture. In the last lecture, the significant correlation is not present.

PhD level, rather than the undergraduate level, and that the field-specific ability belief hypothesis is compatible with multiple mechanisms linking the field-specific ability belief with under-representation.¹⁷ We would like to draw attention to one such mechanism—before arriving at university, students adopt a field-specific ability belief and internalize a stereotype of women as lacking “raw brilliance” in philosophy. This combination could lead women to fail to identify with philosophy, and could reduce their self-confidence and their interest in philosophy. Moreover, this mechanism would postulate a pre-university effect of increasing female under-representation. As such, if this mechanism is operative, then our results may lend further support to the field-specific ability belief hypothesis.

4.2. Absence of Evidence of Classroom Effect Hypotheses

Our second hypothesis was that a first-year undergraduate philosophy course would have a more negative effect on female students’ attitudes toward philosophy than on male students’. If any of the classroom effect hypotheses are right, then we would expect to find that female students’ attitudes towards philosophy are disproportionately affected by their classroom experience.

This study did not find evidence to support any of these classroom effect hypotheses. No significant differences were detected between first lecture and last lecture attitudes between male students and female students. Overall, the mean response for female students with regard to intention to major, ability to do philosophy, comfort in class, and interest in philosophy did not change significantly as compared to the mean response for male students with regard to these issues.

One might wonder if the absence of a classroom effect is the result of tracking students’ responses to the question of whether they are intending to major in philosophy. The worry here is that, plausibly, a small proportion of students have made a decision about their major, and fewer still have decided to major in philosophy. Thus there is a large pool of students who are unsure as to their major. Perhaps, then, women’s experiences through the course will drastically influence how likely it is that they are to major as compared to men’s experiences through the course in one of two senses: (1) women who are unsure at the beginning of the course are more likely to decide not to major by the end of the course as compared to men; (2) men who are unsure at the beginning of the course are more likely to decide to major by the end of the course as compared to women. The same worry could be raised in the context of male and female students reporting whether or not they intend to take further philosophy subjects if they are not intending to ma-

17. As well as the mechanism that we go on to discuss, Leslie et al. (2015) note that another mechanism is the activation of stereotype threat in the classroom, and yet another mechanism is that teachers are influenced by the combination of the field-specific belief and a stereotype of women as lacking “raw brilliance.” Both of these would lead to “classroom effects” in our terminology.

jor. Perhaps many students are initially unsure, and female students are more likely to become sure that they will not take further philosophy units than male students as a result of their classroom experiences.

The present study does not vindicate either of the two hypotheses outlined in the previous paragraph. We tracked the mean responses to each question on a Likert scale for both men and women across time. What we saw was that the average response for women across the course did not change significantly more than the average response for men. If more women who were unsure at the beginning of the course were deciding not to major in philosophy than men, then we would expect to see more movement around the mean. In particular, the mean for women should have dropped by a higher degree than the mean for men. This did not happen. Similarly, if more men who were unsure at the beginning of the course were deciding to major in philosophy than women, then we would expect to see the mean for men rise higher than the mean for women. Again, that was not supported by the results: the mean response for men and the mean response for women changed to the same degree across the course, suggesting that women were not disproportionately put off philosophy as compared to men due to their experiences in the course. More generally, we found no evidence that women's experiences through a philosophy course were more likely to determine their intention to major as compared to men's experiences. The same is true for those students who were unsure whether they would take additional philosophy units: we found no significant difference in the movement around the mean for women as compared to men with respect to whether they will take additional philosophy subjects. Again, if women who were unsure at the beginning of the course were more likely to be put off by the end of the course than men, then this should show in the means.

This study included questions that focused on the learning styles hypotheses, the hostile climate hypotheses and the aggressive argumentation hypothesis, but failed to find specific support for any of these hypotheses. The learning styles hypotheses predicts that in the last lecture, female students would disproportionately report that the class poorly suited their learning styles, but we found no significant gender difference here. The sexist mistreatment hypothesis (one of the hostile climate hypotheses) predicts that in the last lecture female students would feel that they were treated less fairly or with less respect than male students, but again we found no evidence of this. Both the hostile climate hypotheses, and the aggressive argumentation hypothesis predict that the course would have a disproportionately negative effect on female students' comfort in class. But while this study found that there was a difference in comfort levels between male and female students in the first lecture, female students did not become significantly more uncomfortable in class over time.

4.3. *Limitations*

A shortcoming of the present study is that it lacked a control, in this sense: attitudes toward philosophy reported by first year undergraduate students were not compared with attitudes amongst first year undergraduates toward other subjects at university. Such a comparison is needed to rule out that the possibility that gendered differences in attitudes toward philosophy are simply indicative of gendered attitudes more widely. For instance, consider the relationship between self-confidence and philosophy, which was probed by Question 5 on the first lecture survey. Statistical analysis revealed a significant difference between men and women with respect to how confident they feel toward philosophy. What we may be observing, however, is simply a difference in confidence between males and females at the relevant ages, and not one that is peculiar to philosophy *per se*. Or, at least, that cannot be ruled out without comparing our cohort with a range of similar cohorts across the university, each involved in a different first-year undergraduate unit.

We recognise, then, that there is a real need to perform a controlled study.¹⁸ Given this limitation, then it is appropriate to be circumspect about our results, and so everything we say should be read with that caveat in mind. Still, the results are striking and strongly suggest that further testing with control is warranted.

There is also a variety of reasons for caution in inferring from our results that philosophy courses have no gender-specific effects on students' attitudes. It may be that courses have these effects, but only across a longer time-span than a one semester course. For example, it may be that female students are discouraged from studying philosophy after the accumulation of many "micro-inequities"—small-scale injustices that in themselves have insignificant effects, but combine to produce significant effects in the aggregate (Brennan 2013). It may be that there were opposing classroom influences that counter-balanced one another. For example, it may be that exposure to philosophy, and successful performance on assessments, increased female students' interest in philosophy and self-confidence,¹⁹ but that this was cancelled out by, for example, their dislike of an aggressive classroom atmosphere. Although we do not know of any evidence to support this hypothesis, it may be that in general female students are more likely to hedge by saying "unsure" rather than "agree" or "disagree" when expressing their intentions about their fu-

18. We note, however, that such a study would be difficult to say the least. We would need to find out, for example, the background declaration-of-major rates for males and females in the relevant age groups across the board at the University of Sydney. This would be a significant undertaking, and would require surveying more students than first year philosophy students alone.

19. The thought would be that a misleading gender schema leads female students to mistakenly assume that they would perform badly in philosophy, and would not find it interesting. But they later revise this assumption on the basis of their actual experience of finding philosophy interesting, and performing well. Thanks to Julia Markovits for suggesting this hypothesis.

ture majors;²⁰ this would be significant in light of the fact that the gender imbalance we detected was constituted by the fact that a disproportionate number of students who positively expressed an intention to major in philosophy were male. Further, there may be other changes in students' attitudes that the current study did not test, either because it did not investigate these attitudes or because the survey methodology was unable to investigate them properly. For example, it could be that staff and fellow students were acting in ways that discriminated against female students, but these students did not consciously identify this as discrimination. Alternatively, it may be that students did not interpret our questions in the ways that we assume they interpreted them: a weakness of our study is that we did not pilot the study with focus groups to investigate how students themselves interpret the questions.

Another significant limitation of the study concerns the manner in which students were sampled. Students were administered surveys during lectures. Therefore, only those students who were still attending lectures in the last week of semester received the second survey. There was, however, a significant attrition ratio between the two lectures: 596 students attended the first lecture, whereas only 252 students attended the last lecture. It may be that those female students whose attitudes changed significantly more than their male counterparts were those who were absent in the last week of semester. This may potentially explain the null results obtained with respect to the classroom effect hypotheses. That said, there are two aspects of our study methodology that go some way toward addressing this issue. First, through the use of anonymous coding of student surveys, we were able to match student responses to the first survey with their responses to the last survey. This enabled us to test the change in attitudes for those students. This gives us at least some sense of how men's and women's attitudes toward philosophy changed, even despite the attrition rate. Second, it is notable that the gender proportions in the first lecture survey and the last lecture survey were the same (see Table 1 in Section 3). This provides some (albeit weak) evidence against the idea that women were disproportionately discouraged with respect to philosophy by the end of the course. Of course, it may be that while men and women failed to attend the last lecture at the same rate, they did so for different reasons, a difference that may ultimately support the classroom effect hypotheses. Further testing is therefore required in order to determine whether this is the case. One difficulty with any such future testing, however, is that attrition rates between first and last lectures in large undergraduate courses are typically high. Accordingly, it may be difficult to control for the attrition rate if a lecture-based survey methodology is

20. Also possible, though to our minds less plausible, is the claim that in general female students are more likely to hedge their answers to all questions. This claim is unsupported by our findings insofar as we found that male students and female students answered the majority of questions similarly.

used. On the other hand, practically speaking, it is difficult to know how else to apply a survey of the right kind to a large group of undergraduates except by way of the lectures, given the low response rates of internet surveys.

In addition to the above limitations, it is worth pointing out that while gender issues were not explicitly discussed during the first year philosophy course that is the subject of this study, one of the three lecturers reported that, “In my lectures, I made sure most of the images were of women, and that the examples were always ‘women-friendly’. I also tried to emphasise that success in philosophy involves practising skills, not innate genius.” This may undermine the extent to which our study tested, for example, the course content hypothesis. Although we consider this possibility unlikely, our results do not rule out the possibility that the teaching approach used within the course suppressed a classroom effect that otherwise would have been operative.

Finally, because our study was carried out in Sydney, Australia, the results possess a limited ecological validity due to cultural idiosyncrasies that are no doubt present but difficult to control for. Care must therefore be taken in extrapolating our conclusions outward to, e.g., North America or the United Kingdom. In this respect, we think that the value of these studies lies in the gradual accretion of data. We do not suppose that the results of this survey can simply be extrapolated to other countries. Rather, our hope is that similar work will take place in other locations and that the accumulation of these investigations will shed light on underrepresentation throughout the philosophy profession. The present study is only one piece of that overall picture, but, we think, an important one.

However, in order to get a better sense of the extent to which the Australian university and pre-university educational system is similar to, and different from North America and the United Kingdom, it is worth making some general comments. First, it is notable that although students typically had opinions about philosophy as a discipline when surveyed in the first lecture, very few of those students would have encountered philosophy prior to university. Very few high schools in Australia offer philosophy as a subject. So an overwhelming majority of students must have formed opinions about philosophy without having formally studied it. (In this we expect Australia is similar to North America, but perhaps dissimilar to the United Kingdom, where some high school students take a “Philosophy and Religious Studies” A-Level). Second, it is worth saying something brief about the structure of Australian degrees (most relevantly Arts degrees) since appreciating whether or not students are inclined to want to take philosophy units outside of their non-philosophy major is sensitive to how flexible the degree structure is, and to how many non-major units students are able to take. At the University of Sydney, students can take a double major, a major, or a minor in subjects such as philosophy. Within the Faculty of Arts, students can take as many as 6 first year units outside their major subject, and as many as 10 second and third year (combined)

units outside their major subject. Thus, in total, students can take 16 units that do not fall within their major subject. To put that in perspective, students must complete 6 second and third year (combined) units within their major subject leaving the remaining 10 to be taken from other disciplines. There is, therefore, substantial flexibility for students to enroll in philosophy courses even if they are not intending to major in philosophy.

With those caveats in place, the robustness of attitudes toward philosophy at the first lecture and the last lecture seem to us striking and important. On the face of it, this suggests that insofar as gender representation in philosophy disproportionately favours male students after first year philosophy at the University of Sydney, this appears at least partially due to an external, pre-university effect. (Though, again, further study with control would be needed to establish this with more conviction). By contrast, we failed to find evidence of a classroom effect on female-under-representation that occurs during introductory courses. Not only did the ratio of attitudes of male students and female students towards philosophy not significantly change, neither did their intention to major in philosophy. Insofar as students' intentions to major are good evidence for their future choices, this study failed to find evidence that our first year philosophy course was a significant cause of female students' under-representation among philosophy majors, but did find evidence that a pre-university effect was a significant cause.

5. Tackling Female Under-Representation

If a pre-university effect is at least partially responsible for the gender imbalance in philosophy, as our study suggests, then it is not enough for professional philosophers to avoid making the problem of under-representation worse. Instead, active interventions are needed to remedy the problem. So what, if anything, can be done? We cannot pretend to have the answer to that question. But drawing on previous discussions, we have some tentative suggestions both for follow up studies and for pedagogical practice.

In follow up studies, it would be useful to investigate the extent to which pre-existing attitudinal differences between male students and female students are due to conflict between gender and disciplinary schemas, and to clarify what content these schemas have. An excellent example of such research are the philosophy-specific implicit association tests that Jennifer Saul reports that she is currently designing and implementing with psychologists at the University of Sheffield (2013). Additionally, it would also be helpful to investigate the mechanisms by which gender schemas might influence major choice. This also seems a place where it might be helpful to gather qualitative data that shines light on students' explicit or implicit attitudes concerning the relationship of gender and philosophy. Perhaps most

importantly, given the impressive evidence found by Leslie et al. in support of the field-specific ability hypothesis (2015), it would be helpful to investigate further the specific mechanism whereby field-specific ability beliefs influence female under-representation.

If gender schemas are a key causal factor behind female under-representation in philosophy, then there are two ways of redressing this. The first is to change philosophy's image in society, and hence directly confront problematic gender schemas before women enter tertiary education. The second is to challenge these schemas in female students' minds after they enter university. Interestingly, both approaches may require the same interventions. This is because it may turn out that the most effective way to change philosophy's image in society is to begin by changing its image in the classroom, in the hope that students go on to spread this image after they leave university (Mackenzie & Townley 2013). Moreover, interventions within the classroom are most likely to remedy female under-representation in the short-term. We have some reason to believe interventions may be effective because it may well be that many students do not decide what to major in until the end of their first year. Accordingly, throughout this uncertain period, there is a great deal of scope for challenging gender schemas surrounding philosophy with the aim of increasing female representation in undergraduate philosophy.

This, of course, raises the question: how might we disrupt gender schemas that code philosophy as male? As we mentioned above, at least one of the lecturers made a conscious effort to structure the course in a 'female friendly' fashion, for example by using images of women in his ethics lectures, which were compulsory for all students taking the course. Since this minimal intervention did not appear to make a difference to gender related attitudes toward philosophy, a more radical restructuring of entry-level courses would appear to be required. Drawing on work by Blair (2002) and Kang and Banaji (2006), Jennifer Saul proposes that a key way to challenge the gender schema is simply to expose people to female philosophers:

There are many ways in which this can be done: making sure to invite women speakers to departmental seminars and to conferences; including women in invited volumes; putting women philosophers' pictures up in departments and on websites; ensuring that reading lists include women philosophers. (Saul 2013: 51)

The last recommendation may be particularly relevant to our introductory course, because the course had a significantly higher percentage of readings by men.²¹ Second, because the course was taught entirely by male lecturers, the gen-

21. For further endorsement of this recommendation, see among others (Dodds & Goddard 2013; Thompson et al. 2015; Schouten 2015).

der balance of the teaching team could be made more even (Mackenzie & Townley 2013; Schouten 2015). Both approaches would make female philosophers more visible. We suspect, however, that more radical changes will be required. As Cheshire Calhoun notes,

including women in one's program, however is not the same as *degendering* the cultural schema for "philosopher." Most obviously, there are ways of including women that only reinforce the gendered cultural schema for philosophy; for example, tokenistic inclusion of women faculty, or visiting lecturers, or texts by women. (Calhoun 2009: 221)

Instead, Calhoun proposes

actively courting cognitive dissonance . . . : use images of women to represent philosophy on one's website and announcement boards; teach an intro, ethical theory, or epistemology course using only texts by women (and without structuring it as a feminist ethics or feminist epistemology course); call philosophy club meetings philosophy "teas." Construct a visiting lecture series where there is only one man. (2009: 221-2)

These strike us as promising interventions, but we can report from personal experience that the strategy of using an exclusively female syllabus for an ethics course does not by itself necessarily produce cognitive dissonance. When one of us tried this, students used the pronoun "he" roughly as often as they used the pronoun "she" to refer to female authors.

If we assume that female students' first lecture intentions to major in philosophy are at least partially determined by their other general attitudes towards philosophy, then this suggests four possible targets for pedagogical interventions. First, female students report being less interested in philosophy than male students; second, they report feeling less capable of doing well in philosophy than male students; third, female students report being less comfortable in class discussion than male students, and fourth, female students are less able to imagine themselves as philosophers than are male students. For each of these results, there are possible interventions that could be made to a first year course with a view to changing these attitudes. For example, Saul notes that simple encouragement is a way of building students' self-confidence. This encouragement may not need to be directed only at female students. From personal communication with Rae Langton, Saul reports that Monash University's policy of writing to all high-achieving third year students to encourage them to stay on for an optional fourth year raised the percentage of female students in the fourth year from 20% to 50% (Saul 2013: 52). Moreover, increasing someone's self-confidence in an activity can increase his or

her interest in this activity (Valian 1998: 152). Along similar lines, Gina Schouten (2015) has suggested that both assessments and classroom teaching might be designed in ways that encourage students to reflect on personally important values, with the goal that doing so insulates female students from gendered stereotypes.²² There may be particular reason to focus on interest, since interest is so strongly correlated with learning outcomes and vocational choices (Hidi 1990; Schiefele 1991; Tobias 1994; Krapp 1999). Consequently, focused programs to build interest in philosophy amongst female undergraduates may have a significant effect on female representation. Since these are the kinds of interventions to which those of us in the profession have access, a further task is to develop courses that explicitly address these issues and then test to see whether (a) female students' attitudes disproportionately change relative to male students' and (b) whether there are better retention rates for female students given this change of attitudes.

6. Conclusion

This study has failed to find any evidence that introductory philosophy courses at the University of Sydney disproportionately discourages female students from continuing in philosophy relative to male students. Instead, we have found defeasible evidence that existing attitudes coming into tertiary education are responsible for this poor retention. Because the present study lacks a control, it can be thought of as a pilot study indicating where further research is needed. This study suggests that, in particular, further investigation of pre-university effect hypotheses with control is warranted. With respect to the question of how to change female students' attitudes towards philosophy, this study provides some clues as to which attitudes to target and as such it provides some basis upon which to choose between different possible pedagogical interventions.

22. For evidence that this intervention has been successful at combating stereotype threat in STEM subjects, Schouten cites (Miyake et al. 2010; Cohen & Garcia 2008; Schiebinger 2001; Cohen, Garcia, Apfel, & Master 2006). Schouten also makes proposals for removing the effects of stereotype threat on assessments, based on interventions that have proved effective in STEM subjects, such as telling students that exams are gender-fair or that men and women perform equally well on exams.

Appendix A

Unless otherwise specified, all questions asked students to state their level of agreement/disagreement on a 5 point Likert Scale:

Strongly Agree / Agree / Unsure / Disagree / Strongly Disagree

First Lecture Survey Questions

Q1. I am likely to major in philosophy.

Q2. If I major in another subject, I am likely to take philosophy courses past the first year.

Q3. What factors are important to you in choosing your major? (Please number the following options in order of importance, using “1” for most important.)

1. My personal interest
2. My career prospects
3. My talents are well suited to this subject
4. I want to make a difference to the world
5. Other (please specify)

Q4. Studying philosophy is useful for achieving my goals in life.

Q5. I have the ability to do well in philosophy.

Q6. If I don’t contribute to class discussions, it is because . . . (Please number the following options in order of importance, using “1” for most important.)

1. I was unsure what to think
2. I got nervous speaking in front of the class
3. I wanted to learn from what other students say
4. I did not feel included in discussions
5. Other (please specify)

Q7. I can imagine myself becoming a philosopher.

Q8. Doing well in philosophy depends more on natural talent than hard work.

Q9. I find philosophy interesting.

Q10. Topics discussed in philosophy are meaningful to me.

Q11. I will feel comfortable participating in class discussions about philosophy.

Q12. If I encounter difficulties when studying philosophy, I will be able to cope with them.

Q13. How do you learn best? (Please rank in order of importance: 1, 2 . . .)

1. Debating the reasons for holding different views
2. Cooperating with other students to answer questions together
3. Answering questions on my own
4. Reading
5. Attending Lectures
6. Assessments
7. Other

Q14. What is your gender?

Q15. How old are you?

Last Lecture Survey Questions

Q1. I am likely to major in philosophy.

Q2. If I major in another subject, I am likely to take philosophy courses past the first year.

Q3. What factors are important to you in choosing your major? (Please number the following options in order of importance, using “1” for most important.)

1. My personal interest
2. My career prospects
3. My talents are well suited to this subject
4. I want to make a difference to the world
5. Other (please specify)

Q4. Studying philosophy is useful for achieving my goals in life.

Q5. I have the ability to do well in philosophy.

Q6. If I didn’t contribute to class discussions, it was because . . . (Please number the following options in order of importance, using “1” for most important.)

1. I was unsure what to think
2. I got nervous speaking in front of the class
3. I wanted to learn from what other students say
4. I did not feel included in discussions
5. Other (please specify)

Q7. I can imagine myself becoming a philosopher.

Q8. Doing well in philosophy depends more on natural talent than hard work.

Q9. I found philosophy interesting.

Q10. Topics discussed in philosophy were meaningful to me.

Q11. I felt comfortable participating in class discussions about philosophy.

Q12. If I encountered difficulties when studying philosophy, I was able to cope with them.

Q13. Philosophy lecturers, tutors and students treated me fairly and respectfully.

Q14. My interactions with philosophy lecturers and tutors have made me want to take more philosophy courses.

Q15. My interactions with other students have made me want to take more philosophy courses.

Q16. My philosophy classes fitted the ways in which I learn best.

Q17. I feel happy about how my performance in my philosophy classes compares to other students' performances.

Q18. How do you learn best? (Please number the following options in order of importance, using "1" for most important.)

Debating the reasons for holding different views

Cooperating with other students to answer questions together

Answering questions on my own

Reading

Attending Lectures

Assessments

Other (please specify)

Q19. What is your gender?

Q20. How old are you?

Appendix B

Appendix B. Table 8 Gender Differences at First Lecture (detail)

Question	t	Degrees of freedom	Significance	Mean Difference	Standard Error Difference	99.5 % Confidence In- terval of the Difference	
						Lower	Upper
Intend to Major (Q1)	-.697	585	.486	-.064	.091	-0.321	0.194
Intend to Take More Courses (Q2)	-.604	582	.546	-.049	.081	-0.276	0.179
Useful for Life Goals (Q4)	-.269	585	.788	-.019	.071	-0.218	0.18
Self-Confidence (Q5)	-4.072	583	.000**	-.233	.057	-0.394	-0.072
Can Imagine Self as Philosopher (Q7)	-5.183	579	.000**	-.430	.083	-0.664	-0.196
Talent Rather than Effort (Q8)	.343	584	.732	.025	.072	-0.178	0.227
Philosophy is Interesting (Q9)	-2.903	586	.004**	-.158	.054	-0.311	-0.005
Philosophy is Personally Meaningful (Q10)	-1.854	584	.064	-.112	.060	-0.282	0.058
Comfort in Class (Q11)	-6.194	584	.000**	-.425	.069	-0.618	-0.232
Ability to Cope (Q12)	-1.485	577	.138	-.080	.054	-0.233	0.072

** . Result is significant at the 0.001 level (2-tailed).

* . Result is significant at the 0.005 level (2-tailed).

Each row represents an independent samples t-test carried out on the average response for male students versus the average response for female students for that question. Column Three, the significance column, tells us whether, for each question, there was a significant difference between the genders for answers to that question. Significant differences between the genders were detected for Q5, Q7, Q9 and Q11.

TABLE 9. Gender Differences Between First Lecture and Last Lecture (detail)

Question	t	Degrees of freedom	Significance	Mean Difference	Standard Error Difference	99.5 % Confidence Interval of the Difference	
						Lower	Upper
Intend to Major (Q1)	.250	123	.803	.05456	.21830	−0.56948	0.67859
Intend to Take More Courses (Q2)	−.039	122	.969	−.00948	.24271	−0.70339	0.68443
Useful for Life Goals (Q4)	.155	119	.877	.02895	.18729	−0.50675	0.56464
Self-Confidence (Q5)	.320	123	.750	.05101	.15952	−0.40499	0.50701
Can Imagine Self as Philosopher (Q7)	−.139	121	.890	−.02604	.18743	−0.56198	0.5099
Talent Rather than Effort (Q8)	1.145	118	.255	.23502	.20531	−0.35234	0.82237
Philosophy is Interesting (Q9)	1.956	123	.053	.29760	.15217	−0.13739	0.73259
Philosophy is Personally Meaningful (Q10)	.333	120	.740	.05504	.16535	−0.41785	0.52792
Comfort in Class (Q11)	.402	122	.688	.07525	.18702	−0.45944	0.60994
Ability to Cope (Q12)	−.543	120	.588	−.09106	.16871	−0.57098	0.38885

**. Result is significant at the 0.001 level (2-tailed).

*. Result is significant at the 0.005 level (2-tailed).

Each row represents an independent samples t-test carried out on the mean difference between the first and last lectures for male students and the mean difference between the first and last lectures for female students for each question. Column Three, the significance column, tells us whether, for each question, there was a significance difference between the change in male students' responses to a question across the course and the change in female students' responses to that question across the course. No significant differences between the two genders were detected.

TABLE 10. Gender Differences at Last Lecture (detail)

Question	t	Degrees of freedom	Significance	Mean Difference	Standard Error Difference	99.77 % Confidence Interval of the Difference	
						Lower	Upper
Intend to Major (Q1)	-2.133	242	0.34	-.3325	.1559	-0.8128	0.1478
Intend to Take More Courses (Q2)	-1.049	242	.295	-.1731	.1651	-0.6817	0.3354
Useful for Life Goals (Q4)	-2.172	239	.031	-.306	.141	-0.741	0.128
Self-Confidence (Q5)	-3.032	242	.003*	-.3333	.1099	-0.6721	0.0054
Can Imagine Self as Philosopher (Q7)	-3.377	241	.001*	-.492	.146	-0.941	-0.043
Talent Rather than Effort (Q8)	1.413	238	.159	.186	.132	-0.22	0.592
Philosophy is Interesting (Q9)	-.378	242	.706	-.042	.110	-0.382	0.298
Philosophy is Personally Meaningful (Q10)	-1.911	241	.057	-.215	.112	-0.561	0.132
Comfort in Class (Q11)	-3.464	242	.001*	-.446	.129	-0.842	-0.049
Ability to Cope (Q12)	-3.073	240	.002*	-.335	.109	-0.672	0.001
Interactions were Fair (Q13)	-.932	240	.352	-.088	.094	-0.379	0.203
Interactions with Staff (Q14)	-1.623	242	.106	-.212	.131	-0.614	0.19
Interactions with Students (Q15)	-.932	242	.352	-.118	.126	-0.507	0.271
Learning Style (Q16)	-1.514	241	.131	-.197	.130	-0.599	0.204
Performance of Self vs. Others (Q17)	-1.703	239	.090	-.230	.135	-0.646	0.186

**. Result is significant at the 0.0033 level (2-tailed).

*. Result is significant at the 0.00066 level (2-tailed).

Each row represents an independent samples t-test carried out on the average response for male students versus the average response for female students for that question. Column Three, the significance column, tells us whether, for each question, there was a significant difference between the genders for answers to that question. Significant differences between the genders were detected for Q5, Q7, Q11 and Q12.

Acknowledgements

For research assistantship, the authors would like to thank Lesley Wright. For data entry, the authors would like to thank Elena Walsh and Pierrick Bourrat. For helpful comments and discussion, the authors would like to thank Toni Adleberg, Louise Antony, David Braddon-Mitchell, Rachael Briggs, Cheshire Calhoun, Mark Colyvan, Helena De Bres, Nina Emery, Paul Griffiths, Sally Haslanger, Sophie Horowitz, Katrina Hutchison, Fiona Jenkins, Karen Jones, Colin Klein, Rae Langton, Sarah-Jane Leslie, Maureen O'Malley, Ned Markosian, Julia Markovits, Carla Merino, Sara Mrsny, Eddie Nahmias, Jenny Saul, Miriam Schoenfield, Amia Srinivasan, Morgan Thompson, Christina Van Dyke and an anonymous referee for *Ergo*. The authors would like to acknowledge the funding support of Australian Research Council Discovery Grant DP0987186.

References

- Adleberg, Toni, Morgan Thompson, and Eddy Nahmias (2014). Do Men and Women Have Different Philosophical Intuitions? Further Data. *Philosophical Psychology*. Advance online publication. doi:10.1080/09515089.2013.878834
- Antony, Louise (2012). Different Voices or Perfect Storm: Why Are There So Few Women in Philosophy? *Journal of Social Philosophy*, 43(3), 227–255.
- Beebe Helen, and Jennifer Saul (2011). *Women in Philosophy in the UK. A Report by the British Philosophical Association and the Society for Women in Philosophy UK*.
- Blair, Irene V. (2002). The Malleability of Automatic Stereotypes and Prejudice. *Personality and Social Psychology Review* 6(3), 242–261.
- Brennan, Samantha (2013). Rethinking the Moral Significance of Micro-Inequities: The Case of Women in Philosophy. In Katrina Hutchison and Fiona Jenkins (Eds.), *Women in Philosophy: What Needs to Change?* (180–197). Oxford University Press.
- Buckwalter, Wesley, and Stephen Stich (2014). Gender and Philosophical Intuition. In Joshua Knobe and Shaun Nichols (Eds.), *Experimental Philosophy* (Vol. 2, 307–346). Oxford University Press.
- Calhoun, Cheshire (2009). The Undergraduate Pipeline Problem. *Hypatia*, 24(2), 216–22.
- Calhoun, Cheshire (2015). Precluded Interests. *Hypatia*. Advance online publication. doi: 10.1111/hypa.12150
- Cohen, Geoffrey, and Julio Garcia (2008). Identity, Belonging, and Achievement: A Model, Interventions, Implications. *Current Directions in Psychological Science*, 17(6), 365–369.
- Cohen, Geoffrey, Julio Garcia, Nancy Apfel, and Allison Master (2006). Reducing the Racial Achievement Gap: A Social-Psychological Intervention. *Science*, 313(5791), 1307–1310.
- Correll, Shelley J. (2004). Constraints into Preferences: Gender, Status, and Emerging Career Aspirations. *American Sociological Review*, 69(1), 93–113.
- Dodds, Susan, and Eliza Goddard (2013). Not Just a Pipeline Problem: Improving Women's Participation in Philosophy in Australia. In Katrina Hutchison and Fiona Jenkins

- (Eds.), *Women in Philosophy: What Needs to Change?* (143–163). Oxford University Press.
- Dotson, Kristie (2011). Concrete Flowers: Contemplating the Profession of Philosophy. *Hypatia*, 26(2), 403–409.
- Dougherty, Tom, Samuel Baron, and Kristie Miller (in press). Female Under-Representation among Philosophy Majors: A Map of the Hypotheses and a Survey of the Evidence. *Feminist Philosophy Quarterly*.
- Dweck, Carol (2006). Is Math a Gift? Beliefs that Put Females at Risk. In Stephen J. Ceci and Wendy M. Williams (Eds.), *Why Aren't More Women in Science? Top Researchers Debate the Evidence* (47–55). American Psychological Association.
- Dweck, Carol (2008). *Mindsets and Math/Science Achievement*. Carnegie Corporation of New York, Institute for Advanced Study, Commission on Mathematics and Science Education.
- Friedman, Marilyn (2013). Women in Philosophy: Why Should We Care? In Katrina Hutchison and Fiona Jenkins (Eds.), *Women in Philosophy: What Needs to Change?* (21–38). Oxford University Press.
- Goddard, Eliza, Susan Dodds, Lynda Burns, Mark Colyvan, Frank Jackson, Karen Jones, and Catriona Mackenzie (2008). Improving the Participation of Women in the Philosophy Profession; Report C: Students by Gender in Philosophy Programs in Australian Universities. *Australasian Association of Philosophy*.
- Good, Catherine, Joshua Aronson, and Michael Inzlicht (2003). Improving Adolescents' Standardized Test Performance: An Intervention to Reduce the Effects of Stereotype Threat. *Applied Developmental Psychology*, 24, 645–62.
- Hall, Pamela C. (1993). From Justified Discrimination to Responsive Hiring: The Role Model Argument and Female Equity Hiring in Philosophy. *Journal of Social Philosophy*, 24(1), 23–45.
- Haslanger Sally (2008). Changing the Ideology and Culture of Philosophy: Not by Reason (Alone). *Hypatia*, 23(2), 210–223.
- Hidi, Suzanne (1990). Interest and its Contribution as a Mental Resource for Learning. *Review of Educational Research*, 60(4), 549–571.
- Hill, Catherine, Christianne Corbett, and Andresse St. Rose (2010). *Why So Few? Women in Science, Technology, Engineering, and Mathematics*. AAUW.
- Jenkins, Fiona, and Katrina Hutchison (2013). Introduction: Searching for Sofia: Gender and Philosophy in the 21st Century. In Katrina Hutchison and Fiona Jenkins (Eds.), *Women in Philosophy: What Needs to Change?* (1–20). Oxford University Press.
- Kang, Jerry, and Mahzarin R. Banaji (2006). Fair Measures: A Behavioral Realist Revision of 'Affirmative Action'. *California Law Review*, 94(4), 1063–1118.
- Krapp, Andreas (1999). Interest, Motivation and Learning: An Educational-Psychological Perspective. *European Journal of Psychology of Education*, 14(1), 23–40.
- Leslie, Sarah-Jane, Andrei Cimpian, Michelle Meyer, and Edward Freeland (2015). Expectations of Brilliance Underlie Gender Distributions across Academic Disciplines. *Science*, 347 (6219), 262–265.
- Lupart, Judy, Elizabeth Cannon, & Jo Ann Telfer (2004). Gender Differences in Adolescent Academic Achievement, Interests, Values and Life-Role Expectations. *High Ability Studies*, 15(1), 25–42.
- Mackenzie, Catriona, and Cynthia Townley (2013). Women in and out of Philosophy. In Katrina Hutchison and Fiona Jenkins (Eds.), *Women in Philosophy: What Needs to Change?* (164–179). Oxford University Press.

- McKinnon, Rachel (2014). Stereotype Threat and Attributional Ambiguity for Trans Women. *Hypatia*, 10(10), 1–16.
- Miyake, Akira, Lauren Kost-Smith, Noah Finkelstein, Steven Pollock, Geoffrey Cohen, and Tiffany A. Ito (2010). Reducing the Gender Achievement Gap in College Science: A Classroom Study of Values Affirmation. *Science*, 330, 1234–1237.
- Morganson, Valerie J., Meghan P. Jones, and Debra A. Major (2010). Understanding Women's Underrepresentation in Science, Technology, Engineering, and Mathematics: The Role of Social Coping. *Career Development Quarterly*, 59(2), 169–79.
- Moulton, Janice (1989). A Paradigm of Philosophy: The Adversary Method. In Ann Garry & Marilyn Pearsall (Eds.), *Women, Knowledge Reality: Explorations in Feminist Philosophy* (5–20). Routledge Press.
- Paxton, Molly, Carrie Figdor, and Valerie Tiberius (2012). Quantifying the Gender Gap: An Empirical Study of the Underrepresentation of Women in Philosophy. *Hypatia*, 27(4), 949–957.
- Saul, Jennifer (2013). Implicit Bias, Stereotype Threat and Women in Philosophy. In Katrina Hutchison, & Fiona Jenkins (Eds.), *Women in Philosophy: What Needs to Change?* (39–60). Oxford University Press.
- Schiefele, Ulrich (1991). Interest, Learning, and Motivation. *Educational Psychologist*, 26(3-4), 299–323.
- Schiebinger, Londa (2001). *Has Feminism Changed Science?* Harvard University Press.
- Schouten, Gina (2015). The Stereotype Threat Hypothesis: An Assessment from the Philosopher's Armchair, for the Philosopher's Classroom. *Hypatia*, Advance online publication. doi: 10.1111/hypa.12148
- Steele, Jennifer, Jacquelyn B. James, and Rosalind C. Barnett (2002). Learning in a Man's World: Examining the Perceptions of Undergraduate Women in Male-Dominated Academic Areas. *Psychology of Women Quarterly*, 26(1), 46–50.
- Steele, Claude M., and Joshua Aronson (1995). Stereotype Threat and the Intellectual Test Performance of African Americans. *Journal of Personality and Social Psychology*, 69(5), 797–811.
- Superson, Anita (2011). Strategies for Making Feminist Philosophy Mainstream Philosophy. *Hypatia*, 26(2), 410–18.
- Thompson, Morgan, Toni Adleberg, Sam Sims, and Eddy Nahmias (2015). Why Do Women Leave Philosophy? Manuscript in preparation.
- Tobias, Sigmund (1994). Interest, Prior Knowledge, and Learning. *Review of Educational Research*, 64(1), 34–54.
- Turri, John, and Wesley Buckwalter (2015). Perceived Weaknesses of Philosophical Inquiry: A Comparison to Psychology. Manuscript in preparation.
- Valian, Valerie (1998). *Why So Slow? The Advancement of Women*. MIT Press.
- Walker, Margaret U. (2005). Diotima's Ghost: The Uncertain Place of Feminist Philosophy in Professional Philosophy. *Hypatia*, 20(3), 153–164.
- Wylie, Alison (2011). Women in Philosophy: The Costs of Exclusion—Editor's Introduction. *Hypatia*, 26(2), 374–382.