

This article was downloaded by: [Toni Adleberg]

On: 19 February 2014, At: 03:16

Publisher: Routledge

Informa Ltd Registered in England and Wales Registered Number: 1072954 Registered office: Mortimer House, 37-41 Mortimer Street, London W1T 3JH, UK



## Philosophical Psychology

Publication details, including instructions for authors and subscription information:

<http://www.tandfonline.com/loi/cphp20>

### Do men and women have different philosophical intuitions? Further data

Toni Adleberg, Morgan Thompson & Eddy Nahmias

Published online: 14 Feb 2014.

To cite this article: Toni Adleberg, Morgan Thompson & Eddy Nahmias, Philosophical Psychology (2014): Do men and women have different philosophical intuitions? Further data, Philosophical Psychology, DOI: [10.1080/09515089.2013.878834](https://doi.org/10.1080/09515089.2013.878834)

To link to this article: <http://dx.doi.org/10.1080/09515089.2013.878834>

PLEASE SCROLL DOWN FOR ARTICLE

Taylor & Francis makes every effort to ensure the accuracy of all the information (the "Content") contained in the publications on our platform. However, Taylor & Francis, our agents, and our licensors make no representations or warranties whatsoever as to the accuracy, completeness, or suitability for any purpose of the Content. Any opinions and views expressed in this publication are the opinions and views of the authors, and are not the views of or endorsed by Taylor & Francis. The accuracy of the Content should not be relied upon and should be independently verified with primary sources of information. Taylor and Francis shall not be liable for any losses, actions, claims, proceedings, demands, costs, expenses, damages, and other liabilities whatsoever or howsoever caused arising directly or indirectly in connection with, in relation to or arising out of the use of the Content.

This article may be used for research, teaching, and private study purposes. Any substantial or systematic reproduction, redistribution, reselling, loan, sub-licensing, systematic supply, or distribution in any form to anyone is expressly forbidden. Terms & Conditions of access and use can be found at <http://www.tandfonline.com/page/terms-and-conditions>

# Do men and women have different philosophical intuitions? Further data

Toni Adleberg, Morgan Thompson and Eddy Nahmias

*To address the underrepresentation of women in philosophy effectively, we must understand the causes of the early loss of women. In this paper we challenge one of the few explanations that has focused on why women might leave philosophy at early stages. Wesley Buckwalter and Stephen Stich (2014, *Experimental philosophy*. Oxford: Oxford University Press) offer some evidence that women have different intuitions than men about philosophical thought experiments. We present some concerns about their evidence and we discuss our own study, in which we attempted to replicate their results for 23 different responses (intuitions or judgments) to 14 scenarios (thought experiments). We also conducted a literature search to see if other philosophers or psychologists have tested for gender differences in philosophical intuitions. Based on our findings, we argue that that it is unlikely that gender differences in intuitions play a significant role in driving women from philosophy.*

*Keywords:* Gender; Intuitions; Thought Experiments; Underrepresentation; Women

## 1. Introduction

The underrepresentation of women in philosophy is worse than in any field in the humanities or social sciences, and it is as bad or worse than in most STEM fields (Haslanger, 2013). Of all academic disciplines, philosophy has a better gender ratio of PhD recipients than only three STEM fields: computer science, engineering, and physics (Healy, 2011). However, the problem has been discussed and addressed more fully in most of these other fields. In philosophy the discussion has typically focused on the problem of women leaving philosophy at relatively late career stages, such as graduate school. Yet, in the U.S., the most substantial drop in women's enrollment in philosophy courses occurs early, between introductory courses and choosing a major

---

Toni Adleberg is a graduate student at the University of California, San Diego.

Morgan Thompson is a graduate student at the University of Pittsburgh.

Eddy Nahmias is an Associate Professor of Philosophy at Georgia State University.

Correspondence to: Morgan Thompson, Department of History and Philosophy of Science, University of Pittsburgh, 4200 Fifth Ave, Pittsburgh, PA 15260, U.S.A. Email: [mot14@pitt.edu](mailto:mot14@pitt.edu)

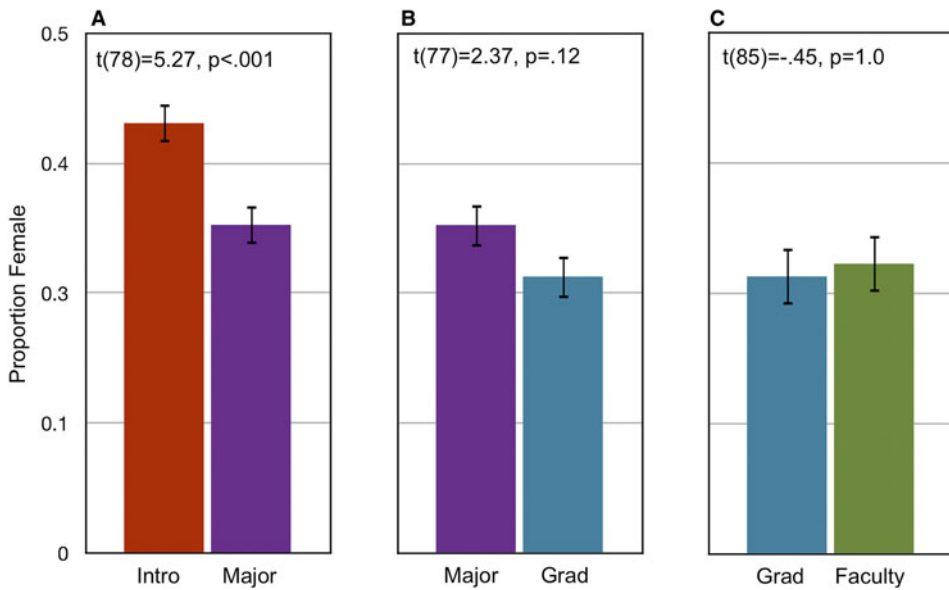
(Paxton, Figdor, & Tiberius, 2012). To address the underrepresentation of women in philosophy most effectively, efforts must target this significant early loss of women, not just later ones. We emphasize that such efforts obviously do not preclude the need to address other problems women in philosophy face at later stages, nor the problems faced by any other groups who are underrepresented in philosophy.

In section 2, we will briefly explain why we think the underrepresentation of women in philosophy is a problem that should be addressed. We must understand the causes of the problem in order to address it successfully. In this paper we challenge one of the few explanations that has focused on why women might leave philosophy at early stages. In a recent, but already widely discussed article, Wesley Buckwalter and Stephen Stich (2014) offer some evidence that women have different intuitions than men about philosophical thought experiments. We explain their hypothesis in section 3 and present some concerns about their evidence. Because of these concerns, we ran our own study to try to replicate their results for 23 different responses (intuitions or judgments) to 14 scenarios (thought experiments). In section 4, we present our methods and results. We found no statistically significant gender differences for the thought experiments we tested. We also conducted an extensive literature search to see if other philosophers or psychologists have tested for gender differences in philosophical intuitions, and if so, what they found. We present these results in section 5. In section 6, we discuss other problems with Buckwalter and Stich's hypothesis. And in section 7, we present our conclusions and some suggestions for future research.

## 2. The Underrepresentation Problem

Only 16.6% of full-time philosophy faculty members in the U.S. are women, and only 21% of all professional philosophers are women (Norlock, 2011). Recent evidence that Paxton et al. (2012) collected from over 50 U.S. colleges indicates that the greatest drop in women's enrollment in philosophy in those institutions occurs after initial philosophy courses (see Figure 1). Women and men sign up for introductory courses in roughly the same numbers. But women are significantly less likely to major in philosophy, while the proportion of women in philosophy between undergraduate and graduate levels and between graduate and faculty levels does not decrease significantly (but see Cherry, 2013 and Spencer, 2013 on the generalizability of these results to black women). Paxton et al. conclude that "after the initial drop from the introductory level to the major level, women tend to stay in philosophy at the same rate as their male counterparts" (2012, p. 954).<sup>2</sup>

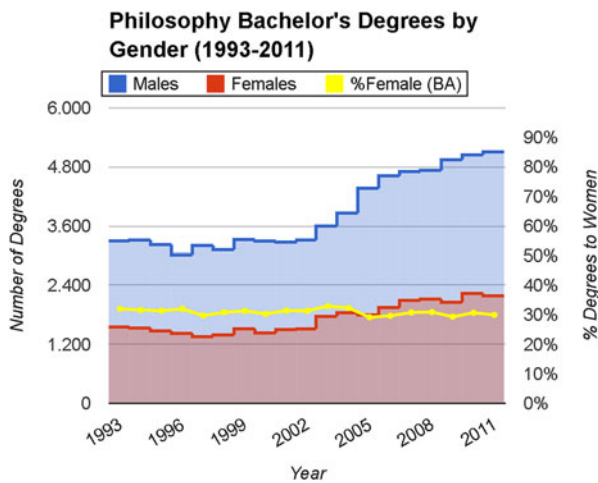
Trends at the institution where we conducted our surveys, Georgia State University, are consistent with this data. About 55% of students enrolled in Introduction to Philosophy are women (roughly in proportion with the 60% female undergraduate population). However, only 33% of philosophy majors are women.<sup>3</sup> Buckwalter and Stich (2014, p. 333) report similar trends at Rutgers University, where from 1999–2010, the percentage of women in philosophy courses drops from 46.2% in introductory courses to 40.38% at the 200 level, 36.5% at the 300 level, 29.31% at the



**Figure 1.** Proportion of various groups in philosophy that are women. Reproduced from Paxton et al. (2012).

400 level, and 26.2% at higher levels. Furthermore, the gender ratio of students receiving Bachelor’s degrees in philosophy has remained stable from 1993–2011 with women accounting for only about 30% (see Figure 2).

We find the underrepresentation of women at all stages to be problematic for several reasons. First, it may be a reflection of unjust practices. When we conducted a climate survey of undergraduates at Georgia State University, we were happy to find that very



**Figure 2.** Philosophy bachelor’s degrees by gender (1993–2011). Graph by Elena Spitzer (2013). Data from: National Center for Education Statistics.

few students reported incidents of explicit bias or discriminatory practices in the classroom (Thompson, Adleberg, Nahmias, & Sims, [unpublished manuscript](#)). Still, anecdotal evidence and online forums like the blog “What is it Like to be a Woman in Philosophy?” suggest that explicit bias is a problem that many female philosophers face.

Second, even if students are not aware of explicit bias, research on implicit bias suggests that women may be discouraged from pursuing academic careers because their work is often undervalued (Saul, 2013). For example, when one ecology journal recently decided to implement anonymous review (which is not standard practice in ecology and evolution journals), they saw a 33% rise in female authors (Budden et al., 2008). Research also indicates that both men and women tend to evaluate the same academic credentials more positively if they are associated with a man’s name rather than a woman’s (Moss-Racusin, Dovidio, Brescoll, Graham, & Handelsman, 2012).

Third, due to stereotype threat, the underrepresentation of women in philosophy may lead women to underperform (Good, Aronson, & Inzlicht, 2003) and even to avoid situations where the stereotype may be active (McKinnon, [forthcoming](#)). There is some evidence that counter-stereotype exemplars can counteract both implicit biases and stereotype threat (Blair, 2002). So, by attracting more women to philosophy, the effects of implicit bias and stereotype threat would be diminished. Finally, if women are leaving philosophy for preventable reasons, then philosophy is likely being deprived of high-quality work that they would have done (Beebe & Saul, 2011).

We are not suggesting that there is necessarily a problem with philosophy unless and until the proportion of women at each stage equals the proportion of women in the population. Also, we are not suggesting that women are necessarily worse-off for leaving philosophy. Leaving may be wise if it is for lack of interest in philosophy or the pursuit of a better career option. We are concerned, however, that many women may be leaving the field because their work is being undervalued or because introductory courses present philosophy in a way that disproportionately fails to interest women (and perhaps minorities). While the disproportionate loss of women from philosophy occurs early on, evidence and anecdote strongly suggest that women continue to face barriers in the field at every stage. And now we are discovering that women find philosophy less interesting than men right from their initial exposure to it (Thompson et al., [unpublished manuscript](#)). We find it implausible that there is something essential to the discipline of philosophy that makes it twice as likely that men want to major in it than women, yet that cannot be changed in such a way that makes philosophy better for everyone.

Until contrary arguments or evidence are produced, we begin with the assumption that there are ways to attract more women to philosophy that are also likely to make philosophy more engaging and attractive to all students. For instance, female (and minority) students might be more concerned than male students about the practicality of their major, or its relevance to their lives. If so, making introductory philosophy courses more relevant and interesting, and providing more information about the value of philosophy, including for practical issues like getting a job, may stem some of the exodus by women and minorities. From our perspective, such attempts to make

philosophy engaging, relevant, and useful would likely make introductory courses better independently of their value for retaining women and minority students.

Many of the solutions that have been proposed to attract and retain women in philosophy focus on issues at the upper end of the academic ladder. For example, the Gendered Conference Campaign aims to increase the number of women invited to be speakers at conferences. Some programs have focused on revising hiring practices or attracting more female graduate students. Graduate programs have conducted climate surveys to reveal and address problems. While these initiatives should improve the quality of life for women already in philosophy and the proportion of women who remain in philosophy in graduate school and beyond, they will likely impact the number of women who choose to major in philosophy only indirectly and over the long term. Our profession needs to understand why so many women leave philosophy after their initial exposure to it and to find solutions to *this* problem. Indeed, increasing the number of female philosophy majors, and hence the pool of students available to go into graduate school, will be a necessary step for any substantive increases in the proportion of female graduate students and professors.

### 3. The Different Intuitions Hypothesis

One possible explanation for the early loss of women in philosophy is that women and men have different philosophical intuitions. As far as we know, Buckwalter and Stich (2014) were the first to consider this possibility and certainly the first to explore it empirically. We commend them for bringing attention to the initial drop-off of women in philosophy and potential explanations for it. They propose that a disproportionate number of women leave philosophy early because their philosophical intuitions differ from those accepted as standard by their instructors and by the authors they read in their classes. As Buckwalter and Stich describe it,

The more courses a woman takes, the more likely it is that she will be exposed to thought experiments on which her intuitions and those of her instructor diverge—and the more likely it is that she will decide not to take another course. (2014, p. 333)

We take them to be making the following general argument:

1. If women have different intuitions about philosophical thought experiments than men, then women will be less likely than men to take more philosophy classes.
2. Women do have different intuitions about philosophical thought experiments than men.
3. So, women are less likely than men to take more philosophy classes, which is one cause of the underrepresentation of women in philosophy.

In this section and the next, we will be addressing premise 2, though we also suggest reasons for doubting premise 1 in section 6. We also do not address another problem Buckwalter and Stich consider (2014, pp. 336–337), namely, that women may be more likely than men to have a “fixed mindset” about the abilities required for philosophy and

more likely to perceive philosophy as requiring such fixed abilities (Dweck, 2006). We find this possibility plausible and are exploring it in other work, but we doubt that it is related to potential gender differences in philosophical intuitions.

In support of premise 2, Buckwalter and Stich present evidence of gender differences in intuitions about the following thought experiments: the Gettier Problem; Compatibilism; Physicalism (Mary); Dualism (robot experience); Thomson's Violinist; the Mob and the Magistrate; the Trolley problem; Causal Deviance; Epistemic Side-Effect Effect; Brain-in-a-Vat; Twin Earth; Chinese Room; and the Plank of Carneades. They present data from their own research as well as data that they solicited from other researchers (Buckwalter & Stich, 2014, p. 312).

We have methodological concerns about the collection of both sets of data and do not think they adequately support premise 2. Since it appears that they solicited data from other researchers only when it indicated significant gender differences, Buckwalter and Stich do not report, and likely do not know, the total number of experiments that were checked by these researchers, including the ones that did *not* indicate significant gender differences. Without knowing the total number of measures examined for gender differences, Buckwalter and Stich cannot adjust their p-values for multiple comparisons to avoid the risk of concluding that a difference in women's and men's mean responses is statistically significant when it may instead be due to chance. If, for example, 100 sets of survey responses were analyzed for gender differences with the standard significance level of 0.05, we should expect that five of those responses would suggest significant gender differences by chance alone. When performing many statistical tests, it is common practice for statisticians to perform a Bonferroni or Sidak correction to avoid increasing the Type-I error rate. It seems that Buckwalter and Stich did not perform either correction except, perhaps, on one set of cases.<sup>4</sup>

Buckwalter and Stich also ran their own series of thought experiments on Amazon's mTurk with a total 1,836 participants (and they excluded those who had taken a prior philosophy course, but do not report how many were excluded). They report four cases with gender differences, which used 384 participants. They may have used 15 or more scenarios, of which only four were reported as showing evidence of gender differences. It is unclear how many of the putative gender differences would be statistically significant had Buckwalter and Stich accounted for multiple comparisons. Given these methodological concerns, we decided to replicate Buckwalter and Stich's experiments to obtain more data regarding the different intuitions hypothesis.<sup>5</sup>

## 4. Replication Studies

### 4.1. Method

We replicated nearly all of Buckwalter and Stich's scenarios (with the exception of Mary and the Epistemic Side-Effect Effect).<sup>6</sup> Participants were undergraduate students enrolled in a critical thinking course at Georgia State University over the summer term in 2012, who received extra credit for completing a survey online through QuestionPro. One group (n = 136; female = 84 and male = 52) was presented with the following

seven scenarios in this order: Compatibilism; Twin Earth; Violinist; Gettier case; Positive Normal; Negative Deviant; and Trolley (switch). The second group ( $n = 158$ ; female = 87 and male = 71) was presented with the following seven scenarios in this order: Chinese Room; Plank of Carneades; Magistrate and the Mob; Brain in a Vat; Positive Deviant; Negative Normal; and Dualism.<sup>7</sup> Our results are presented below in the order that Buckwalter and Stich (2014) present the scenarios, which is not the order of the scenarios that our participants saw. After providing consent, participants read that they would be answering questions about a series of scenarios that are not related in any way, and they were reminded between scenarios that they are not related. Participants were asked a basic attention check question; data from those who missed the attention check question were not analyzed.<sup>8</sup> Demographic information was collected at the end of the survey. We conducted a post hoc power analysis to test whether our study had enough statistical power to detect significant gender differences.<sup>9</sup>

#### 4.2. Results

Our results do not indicate that women have different intuitions than men about this set of philosophical thought experiments. We will discuss the thought experiments in the order they appear in Buckwalter and Stich (2014). Except where noted, the wording of the scenarios and questions remains the same as theirs.<sup>10</sup> Note that the college student participants in our sample are more representative of the population of undergraduate students taking their first philosophy course than the participants in Buckwalter and Stich's online studies using Amazon's mTurk, and hence more useful for exploring the hypothesis that women stop taking philosophy courses because they have different philosophical intuitions than men. We had no reason to suspect that many of these students had been previously exposed to these thought experiments. However, we asked the students whether they had taken (or were taking) any philosophy courses. We found no significant differences (using independent-samples *t*-tests) between the minority of students who had philosophy course(s) and those who hadn't. For lack of space, we focus here on the relevance of the results to Buckwalter and Stich's hypothesis. However, the results also raise interesting questions about pre-philosophical intuitions about important thought experiments and potential challenges or support for the arguments that use these thought experiments. We leave it as an exercise for the reader to consider these questions (see also the findings from Seyedsayamdost, this issue).

##### 4.2.1. Gettier case (watch)

Participants read a scenario in which a burglar steals Peter's watch from a table and replaces it with a cheap replica without Peter's knowledge. They were asked: "Does Peter really know that there is a watch on the table or does he only believe it?" In response, participants were asked to choose either: "Peter really knows it" or "Peter only believes it."

Buckwalter and Stich report that according to Christina Starman and Ori Friedman's results, women were more likely than men to attribute knowledge to Peter.



However, Buckwalter and Stich also report that Starmans and Friedman have failed to replicate this finding (see the literature review in section 5 below).

We also failed to replicate this finding. In our study, 57.1% of women and 44.2% of men responded that Peter “really knows” there is a watch on the table (see [Figure 3](#)). Our results are not statistically significant,  $p = .16$ , Fisher’s exact test, two-tailed (note that in psychology,  $p$ -values  $< .05$  are taken to be statistically significant).

#### 4.2.2. *Compatibilism*

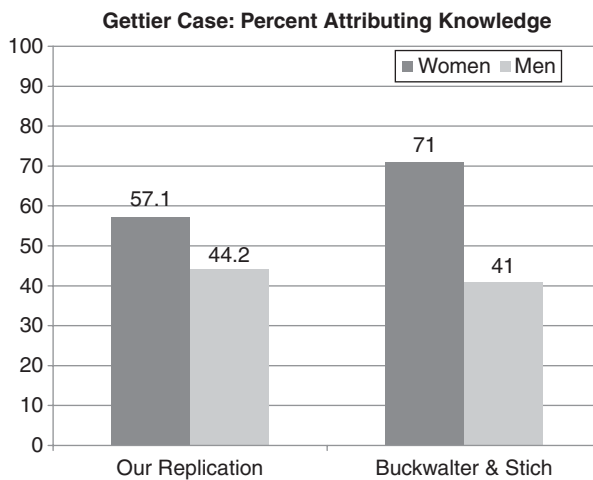
Participants read a scenario that describes a deterministic universe in which “everything that happens, has to happen exactly that way because of the laws of physics and everything that’s come before,” and responded to the following question: “In this case, is a person free to choose whether or not to murder someone?” Response choices were ‘yes’ and ‘no’.

Buckwalter and Stich report Holtzman’s (2013) finding that women were more likely than men to agree that a person in the scenario is free to choose. Holtzman also found gender differences on Dualism (reported below) and the Mary scenario (but see notes 4 and 14).

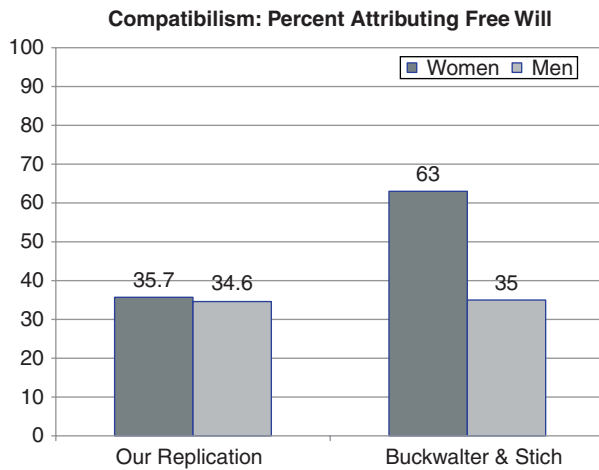
When we retested the Compatibilism scenario, we found that 35.7% of women and 34.6% of men answered ‘yes’, that a person in a deterministic universe described in this way is free to choose (see [Figure 4](#)). This is not a significant difference,  $p = 1.00$ , Fisher’s exact test, two-tailed.

#### 4.2.3. *Dualism (robot experience)*

After reading a brief description of a robot with a complete electronic replica of a human brain, participants were asked to respond to the question: “Could this robot experience love?” They could respond ‘yes’ or ‘no’.



**Figure 3.** Gettier Case. This figure represents the percent of respondents by gender from both Buckwalter and Stich’s reported study and our replication who agreed that Peter ‘really knows’ the watch is on the table.

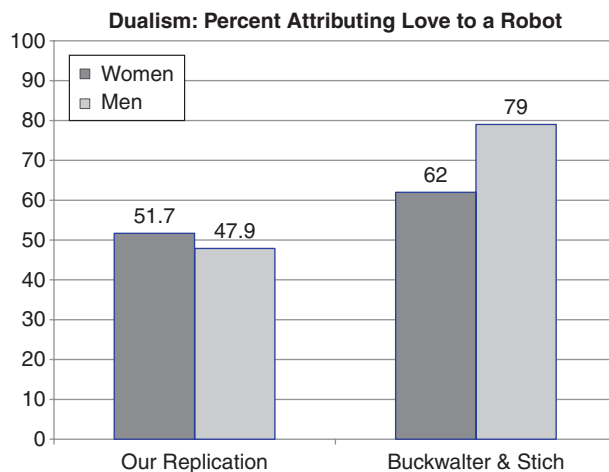


**Figure 4.** Compatibilism. This figure shows the percent of respondents by gender for Buckwalter & Stich’s report and our replication answering “yes” that the described person is free to choose whether or not to murder someone.

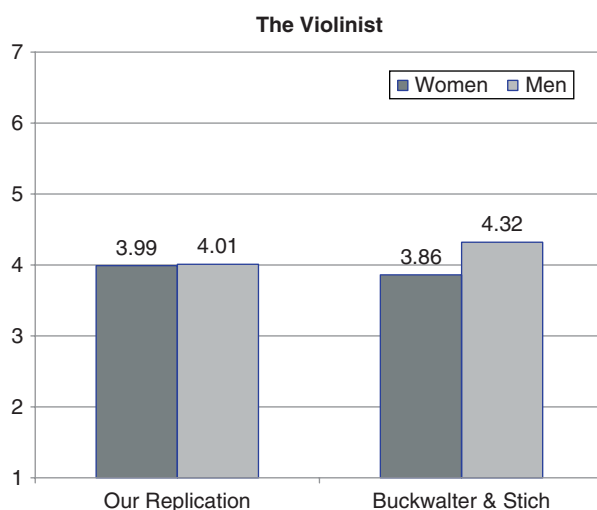
In Holtzman’s study, Buckwalter and Stich report that women were less likely than men to agree that a robot can experience love. In our study, however, 51.7% of women and 47.9% of men answered ‘yes’, which does not suggest a significant gender difference,  $p = .75$ , Fisher’s exact test, two-tailed (see Figure 5).

4.2.4. *The violinist (abortion)*

Participants read a short description of J. J. Thomson’s violinist case and were asked to finish the following sentence: “Jill’s pulling the plug was:” and they were offered a



**Figure 5.** Dualism. This figure represents the percent of respondents by gender for Buckwalter and Stich’s report and our replication answering “yes” that the robot could experience love.



**Figure 6.** The Violinist. This figure represents participants' mean responses by gender to whether Jill's pulling the plug was forbidden (1), permissible (4), or obligatory (7), for both the study reported by Buckwalter and Stich and our replication.

seven-point scale from 1 = 'forbidden' to 7 = 'obligatory'. The midpoint was labeled 'permissible'.

Buckwalter and Stich report Fiery Cushman's finding that in the Violinist case, women were more likely to find Jill's pulling the plug to be forbidden. Our results, however, do not suggest that there is a significant gender difference. For women,  $M = 3.99$ ,  $SD = 1.27$ , and for men,  $M = 4.01$ ,  $SD = 1.37$ ,  $t(134) = -0.39$ ,  $p = 0.70$  (see Figure 6).

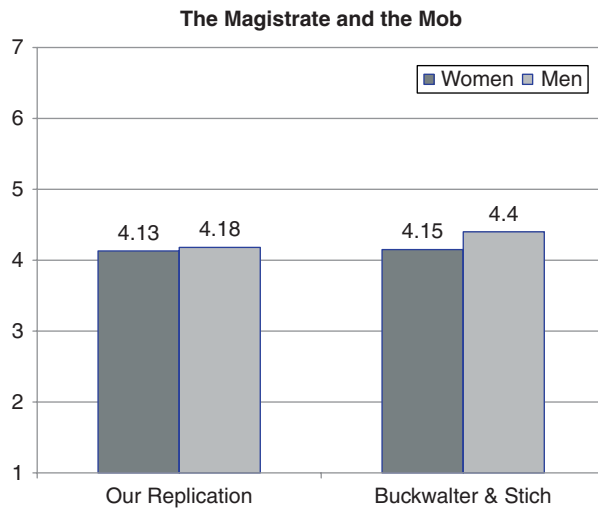
#### 4.2.5. *The Magistrate and the mob (utilitarianism)*

Participants read a description of a police chief, Steve, who chooses not to frame his little brother for a crime to prevent mob violence. Buckwalter and Stich report Cushman's Magistrate and the Mob case, in which participants were asked to fill in the sentence: "The choice Steve made was ..." on a slider scale from 'good' (-225) to 'bad' (225). While both women and men agreed that Steve's choice was good, women agreed less strongly, on average.

We did not use the slider scale described above; rather, we asked participants to complete the following sentence: "The choice Steve made was:" They were given a five-point scale of responses from 1 = 'bad' to 5 = 'good', where the midpoint was labeled 'in between'. For women,  $M = 4.13$ ,  $SD = 1.26$ , and for men,  $M = 4.18$ ,  $SD = 1.23$ ,  $t(156) = -0.28$ ,  $p = 0.78$  (see Figure 7). Hence, our results do not suggest that there is a significant gender difference in responses to this case.

#### 4.2.6. *Trolley case (switch)*

We tested only one version of the trolley scenario, in which the one person on the side track is a stranger. We did not try to test the other three versions of the trolley scenario



**Figure 7.** The Magistrate and the Mob. This figure shows participants' mean responses by gender to whether what Stephen did was "bad" (1), "in between" (3), or "good" (5). Data from Buckwalter and Stich (2014) has been converted to match our scale of 1 to 5.

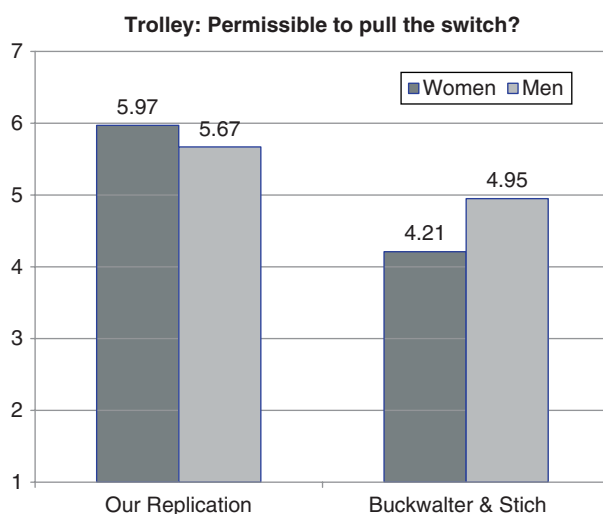
reported by Buckwalter and Stich (2014) (e.g., in which the person on the track is a 12-year-old boy), because we agree with Antony (2012) that gender differences in intuitions based on the gender of the person on the track are unlikely to be indicative of substantive philosophical disagreements between women and men. After reading the scenario in which the trolley is heading towards five people and can be switched to a side-track with one person, participants were asked first to rate their agreement on a seven-point scale from 'strongly disagree' (1) to 'strongly agree' (7) with the following sentence: "It is morally acceptable for me to pull the switch."

Buckwalter and Stich present results from Jennifer Zamzow and Shaun Nichols suggesting that women were more likely to agree that it would be acceptable pull the switch. Our results, however, do not suggest a significant difference: for women,  $M = 5.97$ ,  $SD = 1.60$ , and for men,  $M = 5.67$ ,  $SD = 1.93$ ,  $t(134) = 0.26$ ,  $p = 0.34$  (see Figure 8).

Participants were then asked to answer 'yes' or 'no' to the following question: "Would you pull the switch?" 73.8% of women and 71.2% of men answered 'yes'. This does not suggest a statistically significant gender difference,  $p = .84$ , Fisher's exact test, two-tailed. Buckwalter and Stich report (2014, p. 318) that Zamzow and Nichols also found no gender difference on this measure.

#### 4.2.7. Moral responsibility and causal deviance

There are four causal deviance scenarios: Positive Normal; Positive Deviant; Negative Normal; and Negative Deviant. In our Positive Normal scenario, Tom saves a choking man by performing the Heimlich maneuver. In our Positive Deviant scenario, Samantha saves a choking man because she had a seizure when she intended to help him. In our Negative Normal Scenario, Tom intentionally shoots and kills his enemy. In our



**Figure 8.** Trolley: Permissible to pull the switch? This figure shows participants' mean responses by gender on a seven-point scale, from "strongly disagree" (1) to "strongly agree" (7), that it would be morally acceptable to pull the switch.

Negative Deviant Scenario, Samantha's anxiety about her intent to kill causes a seizure that leads her to pull the trigger. The gender of the agent in two of the scenarios was changed from the cases reported by Buckwalter and Stich (from Tom to Samantha) to clearly distinguish the two agents in the two scenarios each participant read.

One group of our participants read the Positive Deviant and Negative Normal scenarios and the other group read the Positive Normal and Negative Deviant scenarios. For each scenario, participants were asked three questions. First, they were asked to respond to the question, "How [moral/immoral] was [Tom/Samantha]?" on a nine-point scale from 'completely immoral' (coded as '1') to 'completely moral' (coded as '9'). Second, they were asked to respond to the question, "How much [praise/blame] should [Tom/Samantha] receive for [his/her] actions?" on a nine-point scale from 'extreme blame' (coded as '1') to 'extreme praise' (coded as '9'). Third, they were asked to respond to the question, "How [positively/negatively] should [Tom/Samantha] be judged?" on a nine-point scale from 'extremely negatively' (coded as '1') to 'extremely positively' (coded as '9').

Buckwalter and Stich report on causal deviance studies from Pizarro, Uhlmann, and Bloom (2003). Following Pizarro, Uhlmann, and Bloom, we averaged each participant's responses to the three questions to create a "moral sanction index" score for each scenario. We subtracted 5 from our moral sanction index scores so that our scale of 1 to 9 is more easily comparable to Pizarro, Uhlmann, and Bloom's scale of  $-4$  to  $4$ . For the negative scenarios, responses were multiplied by  $-1$ , so higher scores in a negative scenario indicate that the agent was seen as more immoral, blameworthy, and negative, whereas higher scores in a positive scenario indicate that the agent was seen as more moral, praiseworthy, and positive.

Buckwalter and Stich report that according to the data from Pizarro, Uhlmann, and Bloom, women attributed significantly less moral responsibility to Negative Deviant cases than to Negative Normal cases and men attributed significantly less moral responsibility to Positive Deviant cases than to Positive Normal cases. In other words, causal deviance seems to drive women to attribute less blame in the negative scenarios, whereas it seems to drive men to attribute less praise in the positive scenarios (Buckwalter & Stich, 2014, p. 340, note 19).

We ran two-tailed independent samples t-tests to see whether our data supported the pattern discovered by Pizarro, Uhlmann, and Bloom and reported by Buckwalter and Stich.<sup>11</sup> For both women and men, we failed to find evidence that causal deviance plays a role in attributions of moral responsibility for positive or negative scenarios. For women: Negative Deviant ( $M = 1.77$ ), Negative Normal ( $M = 2.02$ ),  $t(166) = -1.06$ ,  $p = .29$ . Positive Normal ( $M = 2.77$ ), Positive Deviant ( $M = 3.09$ ),  $t(168) = -1.71$ ,  $p = .09$ . Our data actually indicates a trend that men seem to attribute *less* moral responsibility to Positive Normal cases than to Positive Deviant cases, but the trend is not significant using our Sidak-adjusted significance level of .002. For men: Positive Normal ( $M = 2.56$ ), Positive Deviant ( $M = 3.03$ ),  $t(117) = -2.05$ ,  $p = .04$ . Negative Deviant ( $M = 1.65$ ), Negative Normal ( $M = 1.60$ ),  $t(120) = .183$ ,  $p = .86$ . Thus, our data does not support the pattern of results reported by Buckwalter and Stich.

#### 4.2.8. *Brain in a vat (skepticism)*

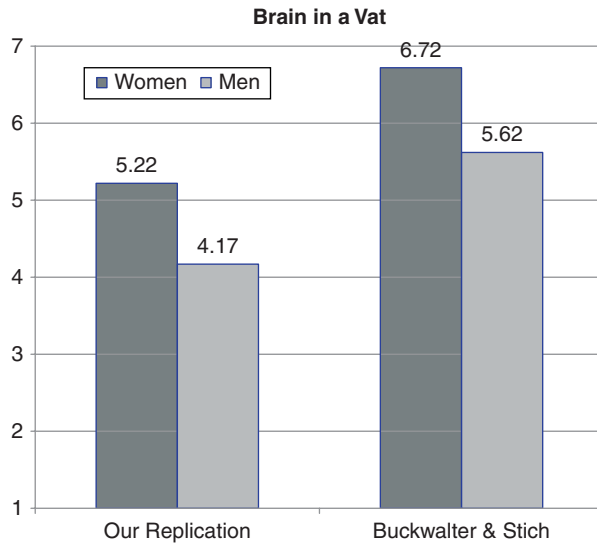
After reading about a conversation between George and Omar in which they describe the possibility of being a brain in a vat, participants were asked to rate their agreement with the following statement: “George knows that he is not a virtual-reality brain” from 1 = ‘completely disagree’ to 7 = ‘completely agree’.

Buckwalter and Stich report that women were more likely to agree that “George knows.” According to our study, for women,  $M = 5.22$ ,  $SD = 1.83$ , and for men,  $M = 4.17$ ,  $SD = 2.37$ ,  $t(129.74) = 3.06$ ,  $p = 0.003$  (see Figure 9). This is a strong trend in our data that is in the same direction as difference reported by Buckwalter and Stich (2014). However, it is not statistically significant using a Sidak-corrected significance level of .002.

#### 4.2.9. *Twin Earth (reference)*

Participants read a long description of twin earth which differs from earth only in that its watery substance is XYZ, and responded to the following question: “When Oscar and Twin-Oscar say ‘water’ do they mean the same thing, or different things?” Responses were on a seven-point scale from 1 = ‘they mean different things’ to 7 = ‘they mean the same thing’.

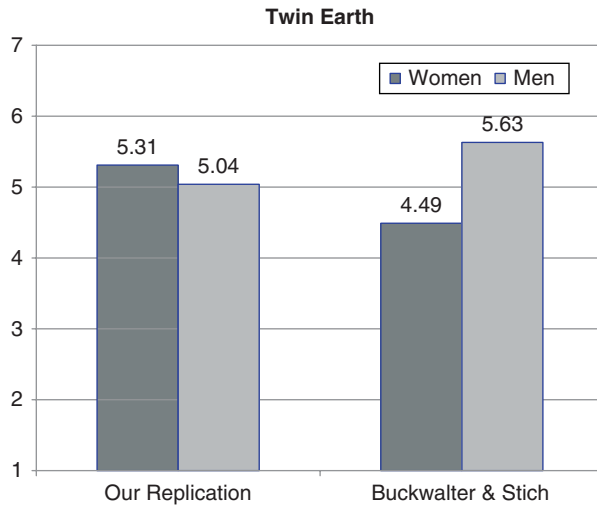
Buckwalter and Stich report a lower mean response for women on the twin earth scenario. We did not find evidence of this gender difference. According to our data, for women,  $M = 5.31$ ,  $SD = 2.09$ , and for men,  $M = 5.04$ ,  $SD = 2.38$ ,  $t(134) = .69$ ,  $p = 0.49$  (see Figure 10).



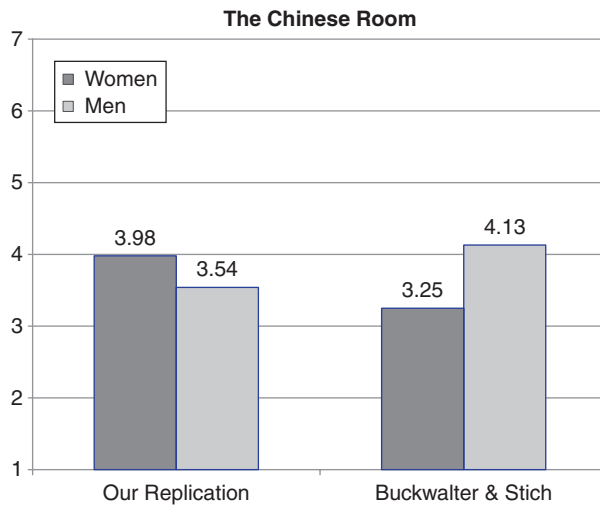
**Figure 9.** Brain in a Vat. This figure shows participants’ mean agreement by gender that George knows he is not a virtual reality brain.

4.2.10. *The Chinese room (artificial intelligence)*

Participants read a version of Searle’s scenario in which Jenny is the person in the room who uses a manual to translate Chinese symbols she does not understand to answer questions, and then they rated their agreement with the statement, “The computational system consisting of Jenny and her instruction manual understands the Chinese written on the notes” on a seven-point scale from 1 = ‘completely disagree’ to 7 = ‘completely agree’.



**Figure 10.** Twin Earth. This figure shows participants’ mean responses by gender on a seven-point scale from “they mean different things” (1) to “they mean the same thing” (7).



**Figure 11.** The Chinese Room. This figure shows participants' mean agreement by gender on a seven-point scale with the statement, "The computational system consisting of Jenny and her instruction manual understands the Chinese written on the notes."

Buckwalter and Stich report that women were less likely to agree that the Chinese room understands Chinese. Our study does not suggest a significant gender difference for the Chinese room. For women,  $M = 3.98$ ,  $SD = 1.95$ , and for men,  $M = 3.54$ ,  $SD = 1.85$ ,  $t(156) = 1.45$ ,  $p = 0.15$  (see [Figure 11](#)).

#### 4.2.11. *The plank of Carneades*

After reading a scenario in which a shipwrecked sailor, Ricki, pushes another, Jamie, off a plank so that he is saved while Jamie drowns, participants were asked: "How morally blameworthy is Ricki for what he did?" Responses were on a seven-point scale from 1 = 'not blameworthy at all' to 7 = 'extremely blameworthy'.

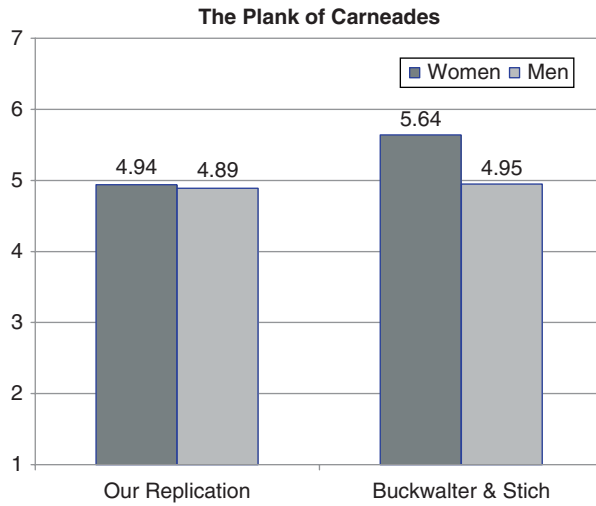
Buckwalter and Stich report that women attributed more blame to Ricki. We did not find evidence of this difference. In our study, for women,  $M = 4.94$ ,  $SD = 2.04$ , and for men,  $M = 4.89$ ,  $SD = 1.72$ ,  $t(156) = 0.19$ ,  $p = 0.87$  (see [Figure 12](#)).

### 4.3. *Summary of Results*

We did not replicate any of the gender differences reported by Buckwalter and Stich. One potential concern with our study is that there may be gender differences we did not detect because we had insufficient statistical power, so our colleague Sam Sims conducted a post hoc power analysis of our study using GPower.

Social scientists typically aim to have at least 80% power to detect an effect of a given size (Cohen, 1992). Although some argue that psychologists should require 95% power, analogous to using  $\alpha = 0.05$  (Machery, 2012), we will follow the current norms for reporting statistical power in psychology. Effect sizes can be small ( $|r| = 0.1$ ),





**Figure 12.** The Plank of Carneades. This figure shows participants' mean responses by gender to how morally blameworthy they find Ricki.

medium ( $|r| = 0.3$ ), or large ( $|r| = 0.5$ ) (Cohen, 1992). In order to conclude that we found no evidence of any gender differences for the scenarios we tested, we would need at least 80% power for small effect sizes. Given our sample sizes, the t-tests that we conducted have only 80% power for effect sizes that are slightly larger than medium. That is, for any gender difference that explains at least 9% of the variance in responses to a given thought experiment, our study has at least an 80% chance of detecting that gender difference.

Though we may not have enough power to detect gender differences of medium or smaller size, we suggest that such differences would likely be inadequate to explain the underrepresentation of women in philosophy. In other words, a gender difference that explains less than 9% of the variance in responses to a given thought experiment is unlikely to be an important factor leading women to stop taking more philosophy courses (section 6). Nonetheless, it would be helpful to have more data on whether women and men have different philosophical intuitions, and like Buckwalter and Stich, we encourage experimental philosophers and psychologists to test for such differences, especially since, as we will now explain, we looked at a number of studies, and very few report testing for gender differences.

### 5. Is There Evidence of Gender Differences in Philosophical Intuitions from Other Experiments?

Our results, as well as our methodological concerns about Buckwalter and Stich's results, suggest that their hypothesis that women leave philosophy because they have different intuitions about philosophical thought experiments is unlikely to be correct.

But their results (and hence ours) sampled only a subset of the sorts of thought experiments students are likely to be exposed to in philosophy classes (in fact, some of them, such as causal deviance, likely come up in very few introductory classes), and the statistical power of our analyses does not allow us to rule out the existence of some (relatively small) gender differences. So, we decided to do a literature search to see if experimental philosophers or psychologists have tested for gender differences in philosophical intuitions, and if so, what they found. As it turns out, very few report testing for such effects.<sup>12</sup> Of the few who have, the majority find no gender differences (questions of statistical power might arise for some of these studies as well), and such differences typically show up in studies testing for moral judgments. We summarize our findings in [Table 1](#).

### 5.1. Literature Review

Our literature review was methodical and extensive. We checked whether researchers tested for gender differences by examining the methods and results sections of every paper on the “Experimental Philosophy Page” website<sup>13</sup> to which we could get access, which included over 200 papers. We also did searches, using the search terms ‘philosophy’ AND ‘gender difference’ OR ‘sex difference’, at [philpapers.org](http://philpapers.org) and on most of the major journals in which relevant papers have been published, including *Cognition*, *Consciousness & Cognition*, *Mind & Language*, *Review of Philosophy and Psychology*, and *Philosophical Psychology*. Of the papers we examined, the vast majority of the studies did not include any tests for gender or sex differences. [Table 1](#) indicates all the papers we found that reported such analyses.

The results of our literature review do not suggest that there are systematic or substantial differences in the philosophical intuitions of male and female students, consistent with our failed replication of Buckwalter and Stich’s findings. Most of the studies that report any gender differences are studies of moral judgments. While such judgments clearly count as relevant to philosophical intuitions, and what gender differences exist in this subset of philosophical intuitions could potentially influence whether some women are more or less likely to want to study more philosophy, we suggest that the findings are not stable or pervasive enough to support Buckwalter and Stich’s hypothesis.

The results from John Turri’s meta-analysis of epistemic judgments suggest that there are gender differences in that subset of philosophical intuitions (some of his cases might include the brain-in-a-vat-type skepticism cases where our results came closest to indicating a gender difference). Nonetheless, these differences are still relatively small in absolute terms (4.2%), suggesting that there will be very few women, relative to the number of men, who would be introduced to an epistemic thought experiment and make a judgment that strongly conflicted with the judgment presented by the instructor (or text) as correct. Indeed, once we look more closely at the body of results Buckwalter and Stich discuss and the details of their proposed hypothesis, as we will now do, we find further reasons to question whether it is a plausible explanation for the underrepresentation of women in philosophy.

**Table 1.** Studies reporting tests for gender differences in philosophical intuitions.

Paper reference	Topic of thought experiment	Number of participants	Results reported
Bartels (2008)	Moral judgments in 14 dilemmas	75 total (26 men, 45 women), plus two other studies reported	No gender differences in moral judgments (p. 393), but see note 1 for other differences.
Bartels and Pizarro (2011)	Moral judgments in 14 scenarios	208 (107 men, 101 women)	They controlled for gender to test other interactions; gender did not predict utilitarian judgments (p. 117).
Beebe (2013)	Weakness of will judgments in six scenarios	180 (36.6% women); 60 (53% women); 240 (63% women); 360 (63% women)	No gender differences.
Bourget and Chalmers (forthcoming)	Self-reported philosophical views	3,226 participants, mostly philosophers (17.4% women, 77.2% men, 5.3% unspecified)	A number of differences, including women being less likely to agree one should pull the switch in a trolley case (p. 17).
Charman, Ruffman, and Clements (2002)	Extensive research on theory of mind development	Dataset 1: 375 children (183 girls, 192 boys) Dataset 2: 1093 children (558 girls, 535 boys)	On average, in children under 4 <i>only</i> ; girls had a weak advantage in theory of mind tasks (p. 7).
Cokely and Feltz (2009)	Intentionality judgments (Knobe effect)	95 total (genders not reported)	No gender differences in judgments (p. 20), but replicated gender difference in order effects reported in Feltz and Cokely (2007).
Feltz and Cokely (2007)	Intentionality judgments (Knobe effect)	161 total (genders not reported)	Order effect found in women, not men (but sample size for men might have been too small to detect; p. 1748).
Feltz and Cokely (2009)	Free will judgments (compatibilism)	56 total (genders not reported)	No gender differences (p. 346).
Feltz and Cokely (2011)	Intentionality judgments (Knobe effect)	110 total (48 men, 62 women)	Gender did not interact with the order effect (pp. 346–347).
Feltz and Cokely (2012)	Three studies of judgments of ignorance and virtue	674 total (genders not reported)	No gender differences (p. 345).
Cushman, Knobe, and Sinnott-Armstrong (2008)	Doing/allowing judgments and moral judgments	300 total (genders not reported)	Gender difference in allowing judgment about termination of pregnancy but not in moral judgments or attitudes about abortion (p. 287).
Hauser, Cushman, Young, Kang-Xing Jin, and Mikhail (2007)	Judgments of moral permissibility	2646 total (1490 men, 1156 women)	No gender differences when comparing responses to switch vs. push and loop versus loop weight (pp. 12–13).

Inbar, Pizarro, Knobe, and Bloom (2009)	Disgust judgments and moral judgments	44 total (14 men, 30 women)	Men less disgust sensitive than women but no interactions between gender, disgust sensitivity, and moral judgments (p. 436).
Knobe (2010) reporting Keys and Pizarro (unpublished data)	Intention and morality judgments (Knobe Effect)	Not reported	“[A] significant gender x character interaction, whereby women tended to regard the act as more intentional when the agent had a bad character, while men tended to regard the act as more intentional when the agent had a good character” (p. 328 note 1).
Lombrozo (2009)	Moral judgments	336 (112 men, 224 women)	“[S]mall but significant effect of sex ... with men generating more consequentialist judgments” on 3 of 5 dilemmas (p. 278), no sex differences in trolley cases (p. 270)
Nagel, San Juan, and Mar (2013)	Knowledge ascriptions, belief ascriptions, and evaluations of justification	208 total (37 men, 171 women)	“[M]en and women did not respond differently for most cases, although there was a trend for women to attribute more knowledge in Skeptical Pressure cases” (p. 7).
Royzmann, Leeman, and Baron (2009)	Judgments of etiquette, ethics, and disgust (based on Nichols, 2002)	1) 35 (14 men, 21 women) 2) 165 (44 men, 117 women, 4 not identified)	1) No gender differences (p. 166) 2) Women’s responses were more punitive and indicated higher levels of disgust (p. 170–171)
Schwitzgebel and Rust (forthcoming)	Moral judgments about a number of cases (e.g., theft, vegetarianism, organ donation)	573 total (gender not reported)	No gender differences on a number of questions, but women were more likely than men to rate eating meat as bad (p. 58).
Sinnott-Armstrong, Mallon, McCoy, and Hull (2008)	Moral judgments in three trolley scenarios	190 total (52% women)	No gender differences (p. 96)
Strandberg and Björklund (2013)	Moral internalism judgments in 6 scenarios	176 total (58% women)	No gender differences (p. 324)
Swain, Alexander, and Weinberg (2008)	Knowledge attributions about four different cases	220 (136 men, 83 women)	In one of the four cases (the coin flip case), men were slightly less likely than women to attribute knowledge, but almost no participants attributed knowledge, male or female (p. 147).
Turri (unpublished) 7/13/2013 post at Experimental Philosophy blog	Meta-analysis of epistemic judgments in over 30 experiments	4967 (3014 men, 1953 women)	Percent attributing knowledge: male = 62.8%, female = 67%, Fisher’s exact test, $p = .003$ , two-tailed, Cramer’s $V = .043$ .
Wright (2010)	1) Epistemic judgments (and order effects) 2) Ethical judgments (and order effects)	1) 188 (87 men, 101 women) 2) 181 (33 men, 148 women)	1) No gender differences (p. 492) 2) No gender differences (p. 496)

## 6. Further Problems for the Different Intuitions Hypothesis

Our replication of the studies presented by Buckwalter and Stich (2014) did not provide evidence of the gender differences they report. Our extensive literature review further suggests that there are not systematic or substantial differences in the philosophical intuitions of female and male students.

Nonetheless, suppose we focus on the gender differences in philosophical intuitions that have turned up, and even allow that there may be others, perhaps ones that are too small for our replication to be likely to detect. We still take issue with the first premise in Buckwalter and Stich's argument:

1. If women have different intuitions about philosophical thought experiments than men, then this would likely lead more women than men to stop taking more philosophy classes.

Here, we present three reasons to doubt that women would be likely to drop out of philosophy even if some such differences existed.

First, in cases where Buckwalter and Stich report a small gender difference in responses, it is unclear that the difference implies that women and men actually make different judgments. For example, they report a gender difference on the Plank of Carneades thought experiment in which one shipwrecked sailor, Ricki, pushes another sailor, Jamie, off a plank that could not support them both. On a seven-item scale, women attributed a greater degree of blameworthiness to Ricki than men did (though we did not replicate this difference). Yet, both women and men *agreed* that Ricki is morally blameworthy. Similar patterns occur in Physicalism, Dualism, Magistrate and the Mob, all causal deviance and moral responsibility scenarios, Epistemic Side-Effect Effect, and Brain in a Vat. That is, in each case, Buckwalter and Stich report differences in the *degree* to which women and men make certain judgments about the cases, but not that women and men make categorically different judgments. Meanwhile, responses they report to the Violinist case are clustered around the midpoint labeled 'permissible'. Responses they report to the Trolley switch case are also clustered around the midpoint labeled 'in between'. So, their studies may indicate slight gender differences in the *degrees* of (dis)agreement without indicating any differences in whether women and men make different judgments, or have different intuitions, about the cases.

Second, in many cases, there is no accepted philosophical intuition in response to the thought experiment. Buckwalter and Stich present a gender difference in intuitions about compatibilism regarding free will and determinism. However, there is heated debate between compatibilists and incompatibilists, and it is unlikely that most instructors present one set of intuitions about the issue as clearly mistaken. Similarly, it is unlikely that the diverging intuitions regarding the Chinese Room scenario are presented as correct or incorrect.

Third, when there *is* an accepted intuition by the philosophy profession, the gender difference reported sometimes suggests that women, rather than men, have the accepted intuition. Take Putnam's Twin Earth thought experiment; women are reported to be *less likely* to agree that Oscar and Twin-Oscar mean the same thing when they say 'water' when compared to men, which is the view that Putnam defends and is presumably

presented by most instructors as correct. Furthermore, unpublished data from the study by Holtzman (2013) call into question whether women have intuitions that differ from the philosophical mainstream.<sup>14</sup> At a minimum, these results suggest that if and when women and men have different philosophical intuitions, it is not the case that women have intuitions that would lead them to feel out of place, mistaken, or confused, and hence more inclined to lose interest in philosophy.

Because we greatly appreciate Buckwalter and Stich's attempt to explain the early drop in women's enrollment in philosophy courses, but found their hypothesis problematic, we decided to look for other explanations. We developed a climate survey for undergraduates in Introduction to Philosophy at Georgia State University. We found many differences between genders (and also between black students and white students) in their perceptions of their introductory class, some of which provide clues for where to look further (Thompson et al., unpublished manuscript).

One interesting finding from our results is that the female students we surveyed were actually *less* likely than the male students to agree that their opinions differed from their peers. Furthermore, students who perceived themselves as having different opinions from their classmates reported being more likely to take more philosophy classes. This finding also provides evidence against Buckwalter and Stich's (2014, p. 338) claim that "*differences in intuition tout court*" makes one less likely to continue in philosophy.

Results from our climate survey suggest other potential causes among the many that likely contribute to the loss of female philosophy students. In concluding, we briefly suggest some of these.

## 7. Conclusions

Women, more than men, appear to leave philosophy soon after first being introduced to it, which obviously plays a significant role in the underrepresentation of women among philosophy graduate students and professors. We find this trend troubling for a variety of reasons, and seek to discover some of the surely complex web of causes that contribute to it (Antony, 2012). Buckwalter and Stich offered an interesting hypothesis—that women have different intuitions than men regarding the sorts of philosophical thought experiments to which they are introduced when first studying philosophy, and these differences lead them to feel "puzzled or confused or uncomfortable or angry or just plain bored . . . [or] convinced that they aren't any good at philosophy" (2014, p. 333). From the armchair, we were dubious that such differences were likely to be one of the more significant contributing causes to women leaving philosophy at a higher rate than men. So, we left the armchair and tested their results to see if we could replicate them.

Due to our failure to replicate Buckwalter and Stich's data, the few gender differences turned up in our literature search, and our doubts that any gender differences in intuitions that might exist play a significant role in driving women to leave philosophy, we started to look for other possible explanations for why so many women leave philosophy after introductory courses. In future research, we aim to explore other potential explanations for the early drop-off of women in philosophy,

and we hope others will join this effort. We also aim to explore whether and why there may be similar drop-offs in enrollment for other groups of students, such as black students, and whether the problems for black women are qualitatively different than the problems for white women or black men (Cherry, 2013; Spencer, 2013). In addition to testing Buckwalter and Stich's suggestion that women have fixed mindsets about intelligence in philosophy, we are testing whether *instructors* are likely to believe that philosophy requires innate talents or that female students perceive themselves as having a fixed amount of talent in philosophy.

We are also testing a number of other potential explanations for the early drop-off. Perhaps women do not encounter enough female role models in philosophy, due to the disproportionate gender ratio among faculty. Paxton et al. (2012) found a correlation between the number of female faculty members in a department and the number of female philosophy majors in that same department. We are testing this hypothesis by comparing students' responses to our climate survey based on both gender of their instructor and the gender ratio of authors on their course's syllabus (Thompson et al., unpublished manuscript). We are also testing whether explaining the relevance of philosophy to students' daily lives or the practical advantages of the philosophy major will encourage more women to continue on in philosophy. Finally, the hypothesis that women are uncomfortable with the combative or confrontational nature of some philosophical discussions has come up in many debates in the profession about why women leave philosophy. We're interested in testing whether female students actually find the philosophical discussions in their introductory courses to be excessively combative. When appropriate, we also hope to propose solutions to these problems.

One interesting line of inquiry looks to the similarities and differences between the underrepresentation of women in philosophy and in STEM fields. If the two problems result from a convergence of some similar factors, then philosophers may be able to adopt the strategies and solutions already pursued in these STEM fields. In our opinion, a particularly plausible hypothesis is that women may be less impressed by certain philosophical methodologies than men. Buckwalter and Turri (unpublished manuscript) have found that women's preference for observational methods over methods based on intuitions is stronger than men's preference for observational methods. So, when philosophy courses focus on intuition-based methods, male students may become more interested than female students in taking another philosophy course.

In any case, we hope that more—and more systematic—efforts will be made to understand the complex set of causes that lead women to leave philosophy at a higher rate than men, and where appropriate, to counteract them. We suspect that most of these efforts will make undergraduate philosophy courses more relevant, useful, and enjoyable for *all* students.

### Acknowledgements

For valuable help on this project, we would like to thank Sam Sims. We would also like to thank Jason Shepard and Shane Reuter. We are grateful for the helpful comments we received from Liam Bright, Wesley Buckwalter, David Colaço, Fiery Cushman, Adam Feltz, Carrie Figdor, Geoff Holtzman, Joshua Knobe, Edouard Machery, Dan

Malinsky, Christian Mott, Shaun Nichols, David Pizarro, Eric Schwitzgebel, Chandra Sripada, John Turri, Virginia Valian, and Jennifer Wright. We would also like to thank the audiences at: the Implicit Bias, Philosophy, and Psychology conference at the University of Sheffield and the Leverhulme Trust; the Diversity in Philosophy conference at the University of Dayton; the 2013 Society for Philosophy and Psychology conference at Brown University; and Georgia State University. We would also like to thank the Pittsburgh feminism reading group for helpful comments. Finally, we would like to thank George Rainbolt and Tim O’Keefe for their support and encouragement throughout this project.

## Notes

- [1] Toni Adleberg and Morgan Thompson are the primary authors, and made equal contributions to the paper.
- [2] Independent samples t-test,  $t(78) = 5.27$ ,  $p < .001$  (Paxton et al., 2012). Their study was initiated as part of a diversity initiative of the Society for Philosophy and Psychology. Their data also suggest that at some universities, women may enroll in fewer introductory courses than men, relative to the proportion of women in the undergraduate population. If so, it might be that something about the way incoming students perceive philosophy already makes women less interested than men in studying it.
- [3] We should also note that at our university black students make up 35% of Introduction to Philosophy students, roughly in proportion to the number of black students enrolled in undergraduate studies (38%). However, like female students, black students are significantly less likely than white students to major in philosophy; just 20% of philosophy majors are black, and we suspect that this trend is not unusual among American colleges and universities. We are examining possible causes for this drop-off among black students, including whether they might overlap with some of the causes for the drop-off among female students (Thompson et al., unpublished manuscript). Interestingly, Asian students make up 12% of the undergraduate population and are approximately equally represented in Introduction to Philosophy (13%) and in the philosophy major (10%).
- [4] It appears that either Holtzman or Buckwalter and Stich have corrected for multiple comparisons on Holtzman’s cases and that the difference on the Dualism case is not actually significant. In a footnote, Buckwalter and Stich state that with the exception of the Dualism case, “significance values in the experiments we have recounted remain at the  $p < .05$  level after correcting by a factor of 9” (2014, p. 340, note 13). While we think it is a good idea to correct for multiple comparisons as they did, we suggest that it would have been appropriate to correct for the *total* number of measures performed by Buckwalter, Stich, and their colleagues, rather than just for Holtzman’s 9 measures. Given the large total number of cases that were analyzed for gender differences, it may be that even the differences on Compatibilism and Physicalism are not statistically significant.
- [5] For further methodological concerns, see Antony (2012). For other evidence relevant to Buckwalter and Stich’s argument, see Chernkhovskaya (unpublished manuscript), who replicated five of Buckwalter and Stich’s scenarios with slightly different wording and failed to find any gender differences; Seyedsayamdost (this issue), who failed to replicate most of Buckwalter and Stich’s gender differences; and Turri (forthcoming), who failed to replicate the gender difference on the Gettier case (but see section 5).
- [6] The Absence Causation case described in Buckwalter and Stich (2014) was not reported at the time we ran our study. The Mary and the Epistemic Side-Effect Effect cases were tested for gender differences from data previously collected by Shane Reuter. We did not find evidence of significant



gender differences for intuitions about either case. Participants were 79 women and 56 men for both scenarios. Data from participants who took less than 500 seconds to complete the survey were excluded. For the Mary case: for women,  $M = 2.53$ ,  $SD = 1.53$  and for men,  $M = 2.75$ ,  $SD = 1.84$ ,  $t(105) = -0.73$ ,  $p = 0.47$ . For the Epistemic Side-Effect Effect: for women,  $M = 5.14$ ,  $SD = 1.88$  and for men,  $M = 5.57$ ,  $SD = 1.70$ ,  $t(125) = -1.39$ ,  $p = 0.17$ . Note that the response question tested by Reuter for the Epistemic Side-Effect Effect was different than the response tested by Beebe and Buckwalter and reported by Buckwalter and Stich (2014). Reuter asked participants to rate their agreement with the statement: “The chairman intentionally [helped/harmed] the environment.” Buckwalter and Stich report participants’ responses to the question: “Did the chairman know that the new program would [help/harm] the environment?” (Buckwalter & Stich, 2014, p. 321).

- [7] Because the order of scenarios was not randomized, we did not check for order effects in participants’ responses.
- [8] Twenty participants in the first group and eight participants in the second group were excluded for missing the attention check question. We were also unable to use data from three participants in the first group and one participant in the second group because they did not report their gender.
- [9] For the scenarios we tested in group 1 ( $n = 136$ ), the smallest detectable effect size  $|r|$  for 80% power and a Sidak-corrected significance level of .002 was .31. For the scenarios we tested in group 2 ( $n = 158$ ), the smallest detectable effect size  $|r|$  for 80% power and a Sidak-corrected significance level of .002 was .32.
- [10] For complete descriptions of the scenarios, see Buckwalter and Stich (2014).
- [11] We ran independent-samples t-tests rather than paired-samples t-tests because, in our study, the presentation of the four scenarios was split between two groups.
- [12] There may be many reasons why researchers do not report testing for gender differences, including that they do not want to find any differences that might exist (e.g., to avoid having to analyze groups separately and lose power) or that they do not find potential gender differences interesting.
- [13] <http://pantheon.yale.edu/~jk762/ExperimentalPhilosophy.html>
- [14] We are grateful to Geoff Holtzman for allowing us to cite his data. Note that he has not expressed any stance for or against our theoretical views. Three of Holtzman’s findings (from t-tests analyzing male and female intuitions about compatibilism, physicalism, and dualism) were reported by Buckwalter and Stich (2014). Holtzman analyzed these and six other thought experiments for the effects of both gender and philosophical training (having a PhD versus not having a PhD). The philosophy PhDs ( $n = 234$ ) participating in Holtzman’s study (2013) were primarily male (82%), white (94%) and from Western countries (100%). His results suggest that, when male and female intuitions do diverge, it is more often the *male* students, not the female students, who have different intuitions from philosophers with PhDs. His findings suggest that female students are more likely than male students to agree with their professors in the scenarios described above for compatibilism and dualism (robot experience), as well as scenarios asking whether science can reductively explain taste experiences, whether thinking requires a body, and whether it is fair to “take one for the team,” while women were less likely than men to agree with their professors in a trolley case and a Gettier case. Given that we did not replicate some of the gender differences described in Holtzman’s responses and that our literature search provides conflicting evidence about others (and see Seyedsayamdost, [this issue](#)), we do not know whether they indicate stable or systematic gender differences in philosophical intuitions.

## References

- Antony, L. (2012). Different voices or perfect storm: Why are there so few women in philosophy? *Journal of Social Philosophy*, 43(3), 227–255.

- Bartels, D. (2008). Principled moral sentiment and the flexibility of moral judgment and decision making. *Cognition*, 108, 381–417.
- Bartels, D., & Pizarro, D. (2011). The mismeasure of morals: Antisocial personality traits predict utilitarian responses to moral dilemmas. *Cognition*, 121, 154–161.
- Beebe, J. R. (2013). Weakness of will, reasonability, and compulsion. *Synthese*, 190, 4077–4093.
- Beebe, H., & Saul, J. (2011). *Women in philosophy in the UK* (Research Report). Retrieved from the Society for Women in Philosophy UK website: <http://www.swipuk.org/notices/2011-09-08/>
- Blair, I. (2002). The malleability of automatic stereotypes and prejudice. *Personality and Social Psychology Review*, 6, 242–261.
- Bourget, D., & Chalmers, D. (forthcoming). What do philosophers believe? *Philosophical Studies*.
- Buckwalter, W., & Stich, S. (2014). Gender and philosophical intuition. In J. Knobe & S. Nichols (Eds.), *Experimental philosophy* (Vol. 2, pp. 307–346). Oxford: Oxford University Press.
- Buckwalter, W., & Turri, J. (unpublished manuscript). Some perceived weaknesses of philosophical inquiry and method.
- Budden, A., Tregenza, T., Aarssen, L., Koricheva, J., Leimu, R., & Lortie, C. (2008). Double-Blind review favours increased representation of female authors. *Trends in Ecology & Evolution*, 23(1), 4–6.
- Charman, T., Ruffman, T., & Clements, W. (2002). Is there a gender difference in false belief development? *Social Development*, 11, 1–10.
- Chernykhovskaya, Y. (unpublished manuscript). Why so few women philosophers? Do gender differences in philosophical intuition account for women's underrepresentation in philosophy departments?
- Cherry, M. (2013, May). *The state of black women in philosophy*. Paper presented at the Diversity in Philosophy Conference, University of Dayton, Dayton, OH.
- Cohen, J. (1992). A power primer. Quantitative methods in psychology. *Psychological Bulletin*, 112(1), 155–159.
- Cokely, E. T., & Feltz, A. (2009). Individual differences, judgment biases, and theory-of-mind: Deconstructing the intentional action side effect asymmetry. *Journal of Research in Personality*, 43, 18–24.
- Cushman, F., Knobe, J., & Sinnott-Armstrong, W. (2008). Moral appraisals affect doing/allowing judgments. *Cognition*, 108(2), 353–380.
- Dweck, C. S. (2006). *Mindset: The new psychology of success*. New York: Random House.
- Feltz, A., & Cokely, E. T. (2007). An anomaly in intentional action ascriptions: More evidence of folk diversity. In D. S. McNamara & J. G. Trafton (Eds.), *Proceedings of the 29th Annual Cognitive Science Society* (p. 1748). Austin, TX: Cognitive Science Society.
- Feltz, A., & Cokely, E. T. (2009). Do judgments about freedom and responsibility depend on who you are? Personality differences in intuitions about compatibilism and incompatibilism. *Consciousness & Cognition*, 18, 342–350.
- Feltz, A., & Cokely, E. T. (2011). Individual differences in theory-of-mind judgments: Order effects and side effects. *Philosophical Psychology*, 24, 343–355.
- Feltz, A., & Cokely, E. T. (2012). The virtues of ignorance. *The Review of Philosophy and Psychology*, 3, 335–350.
- Good, C., Aronson, I., & Inzlicht, M. (2003). Improving adolescents' standardized test performance: An intervention to reduce the effects of stereotype threat. *Journal of Applied Developmental Psychology*, 24(6), 645–662.
- Haslanger, S. (2013, September 2). Women in philosophy? Do the math. *The New York Times*. Retrieved from [http://opinionator.blogs.nytimes.com/2013/09/02/women-in-philosophy-do-the-math/?\\_r=0](http://opinionator.blogs.nytimes.com/2013/09/02/women-in-philosophy-do-the-math/?_r=0)
- Hauser, M., Cushman, F., Young, L., Kang-Xing Jin, R., & Mikhail, J. (2007). A dissociation between moral judgments and justifications. *Mind & Language*, 22(1), 1–21.

- Healy, K. (2011, February 4). Gender divides in philosophy and other disciplines [Web log message]. Retrieved from <http://kieranhealy.org/blog/archives/2011/02/04/gender-divides-in-philosophy-and-other-disciplines/>
- Holtzman, G. (2013). Do personality effects mean philosophy is intrinsically subjective? *Journal of Consciousness Studies*, 20, 27–42.
- Inbar, Y., Pizzaro, D., Knobe, J., & Bloom, P. (2009). Disgust sensitivity predicts intuitive disapproval of gays. *Emotion*, 9(3), 435–439.
- Knobe, J. (2010). Person as scientist, person as moralist. *Behavioral and Brain Sciences*, 33, 315–329.
- Lombrozo, T. (2009). The role of moral commitments in moral judgment. *Cognitive Science*, 33, 273–286.
- Machery, E. (2012). Power and negative results. *Philosophy of Science*, 79(5), 808–820.
- McKinnon, R. (forthcoming). Stereotype threat and attributional ambiguity for trans women. *Hypatia*.
- Moss-Racusin, C., Dovidio, J., Brescoll, V., Graham, M., & Handelsman, H. (2012). Science faculty's subtle gender biases favor male students. *Proceedings of the National Academy of Sciences*, 109(41), 16474–16479.
- Nagel, J., San Juan, V., & Mar, R. A. (2013). Lay denial of knowledge for justified true beliefs. *Cognition*, 129, 652–661.
- Nichols, S. (2002). Norms with feeling: Towards a psychological account of moral judgment. *Cognition*, 84, 221–236.
- Norlock, K. (2011). Women in the profession: A report to the CSW: 2011 Update (Research Report). Retrieved from the American Philosophical Association Committee on the Status of Women website: <http://www.apaonlinecsw.org/workshops-and-summer-institutes>
- Paxton, M., Figdor, C., & Tiberius, V. (2012). Quantifying the gender gap: An empirical study of the underrepresentation of women in philosophy. *Hypatia*, 27(4), 949–957.
- Pizarro, D., Uhlmann, E., & Bloom, P. (2003). Causal deviance and the attribution of moral responsibility. *Journal of Experimental Social Psychology*, 39, 653–660.
- Royzman, E., Leeman, R., & Baron, J. (2009). Unsentimental ethics: Towards a content-specific account of the moral-conventional distinction. *Cognition*, 112, 159–174.
- Saul, J. (2013). Implicit bias, stereotype threat, and women in philosophy. In F. Jenkins & K. Hutchison (Eds.), *Women in philosophy: What needs to change?* (pp. 39–60). Oxford: Oxford University Press.
- Schwitzgebel, E., & Rust, J. (forthcoming). The moral behavior of ethics professors: Relationships among self-reported behavior, expressed normative attitude, and directly observed behavior. *Philosophical Psychology*.
- Seyedsayamdost, H. (this issue). On gender and philosophical intuition: Failure of replication and other negative results. *Philosophical Psychology*.
- Sinnott-Armstrong, W., Mallon, R., McCoy, T., & Hull, J. (2008). Intention, temporal order, and moral judgments. *Mind & Language*, 32(1), 90–106.
- Spencer, Q. (2013, May). *How's it going for blacks in philosophy?* Paper presented at the Diversity in Philosophy Conference, University of Dayton, Dayton, OH.
- Spitzer, E. (2013). *Women & philosophy*. Retrieved from <http://www.elenaspitzer.info/women-philosophy/>
- Strandberg, C., & Björklund, F. (2013). Is moral internalism supported by folk intuitions? *Philosophical Psychology*, 26(3), 319–335.
- Swain, S., Alexander, J., & Weinberg, J. (2008). The instability of philosophical intuitions: Running hot and cold on Truetemp. *Philosophy and Phenomenological Research*, 76(1), 138–155.
- Thompson, M., Adleberg, T., Nahmias, E., & Sims, S. (unpublished manuscript). Why do women leave philosophy?
- Turri, J. (forthcoming). A conspicuous art: Putting Gettier to the test. *Philosopher's Imprint*.

- Turri, J. (unpublished manuscript). The modest but real gender-effect on knowledge attributions. Retrieved July 13, 2013 from <http://philosophycommons.typepad.com/xphi/2013/07/the-modest-but-real-gender-effect-on-knowledge-attributions.html>
- Wright, J. (2010). On intuitional stability: The clear, the strong, and the paradigmatic. *Cognition*, 115, 491–503.