Ranking Exercises in Philosophy and Implicit Bias[1]

In recent decades, ranking exercises have become increasingly important to philosophy. And in recent years, philosophers have become increasingly concerned about the situation of under-represented groups in philosophy. The Australasian Philosophical Association published a report on women in philosophy in 2008 (Goddard 2008), and the British Philosophical Association and Society for Women in Philosophy in the UK did so in 2011 (Beebee and Saul 2011). The American Philosophical Association has a long-standing Committee on the Status of Women (http://www.apaonlinecsw.org/), and the Canadian Philosophical Association has an Equity Committee, both of which have published many reports (see, for example the APA Newsletters on Feminism and Philosophy Spring and Fall 2009; and various reports here: http://www.acpcpa.ca/en/equitycommittee.php). There is the new Women in Philosophy Task Force (http://web.mit.edu/wphtf/Welcome.html), and there have been a variety of international conferences and workshops in recent years on under-represented groups in philosophy, and many widely read papers (the best-known of which is Haslanger 2008). There have been articles on the

topic from The Philosophers Magazine to The New York Times, and there are blogs, campaigns and even songs devoted to these issues.[2]

Although most of this work has been specifically on women in philosophy, not all of it has. Haslanger's paper, for example, also calls attention to the situation of racial minorities in philosophy; the Cardiff conference (http://www.cf.ac.uk/encap/newsandevents/events/conferences/groups.html) was on under-represented groups more generally in philosophy; Penn State's Philosophy in an Inclusive Key Summer Institute (PIKSI) is for promising undergraduate members of all under-represented groups in philosophy (http://www.psu.edu/dept/rockethics/education/piksi/index.shtml); The Rutgers Summer Institute for Diversity in Philosophy (http://www.philosophy.rutgers.edu/events/summer-institute) is for talented undergraduate students whose "experiences and background foster greater diversity in the field of philosophy"; and the recently created Young Black Philosophers Association and Collegium of Black Women Philosophers (http://web.me.com/ktgphd/CBWP/Welcome.html) are also vital new initiatives. [check details, add references]

It seems a good time, then, to reflect on ways that ranking exercises may intersect with issues concerning under-represented groups, and that is the goal of this paper. More specifically, I will explore the implications for these exercises of the implicit biases against stigmatized groups that psychologists over the last few decades have shown to be widely held. I

---

[2] Baggini 2011; Anonymous 2009; < http://beingawomaninphilosophy.wordpress.com/>; < http://feministphilosophers.wordpress.com/gendered-conference-campaign/>; < http://feministphilosophers.wordpress.com/2010/11/29/i-like-to-see-the-ladies/>.

will do this by examining methodologies of the Philosophical Gourmet Report and the United Kingdom's REF.[3]

Before I start examining these methodologies, however, it is important to note that the world before these exercises came into being was very far from a bias-free world. As is frequently noted in defenses of these exercises, it was a world in which hierarchies and prejudices were given relatively free reign, though unofficial channels like gossip, "old boys networks", and time-honoured assumptions. One key good that has come out of these exercises is the partial disruption of the power formerly residing in these channels. This is clearly illustrated by the way that outstanding philosophy departments in less famous universities have come to be recognized as such.

What I will argue here, however, is that there is much more work to be done. These exercises give us, as a profession, the opportunity to attempt to systematically correct for a wide range of pernicious biases that hinder the accurate evaluation of work and that perpetuate stereotypes and unjust inequalities. We have an obligation to attempt to do this. The present paper is a first step toward tackling that obligation, by clearly stating some of the problems these biases currently raise for ranking exercises. Although I am only able to gesture at some preliminary ideas for improvement, my hope is that spelling out the

---

[3] I choose these because they are the most widely known and influential. The much newer Pluralists' Guide is, like the Gourmet Report, a reputational survey, and so prone to the problems of reputational surveys. However, I have not examined its methodology in detail and there may be particular problems or virtues that would need consideration for a considered judgment to be made. I am focusing solely on rankings of philosophical quality/reputation, so will not discuss climate for women surveys at all.

problems will spur the profession to begin to work collectively toward fuller solutions.

Part 1: What we know about implicit bias

Over the last few decades, psychologists have established very clearly that human beings, even those who hold strongly egalitarian ideals, are prone to a range of unconscious biases against members of groups that are stigmatized in certain areas. [4]This is true even of members of these groups. Most people, for example, are faster to pair black faces with bad adjectives than good ones. Most people, including but not limited to police officers, are more likely to misidentify a harmless object as a gun if it is held by a black person. Psychologists have established that these biases are not (readily) amenable to direct conscious control, and that they are not (in general) introspectively accessible to those that have them, even as they are acting on them. These seem to arise, broadly speaking, from being immersed in a society where certain biases and associations are widespread.[5] I have argued elsewhere that those who are prey to these biases should not be blamed merely for possessing, or even acting on them, although they may well be blameworthy for failing

---

[4] For a good review of this literature, see Jost et. al.
[5] The full picture, however, must be more complicated than this since it is not the case that everyone in a given society is prone to the same biases or to the same extent. This variation is not yet fully understood, although studies are ongoing. Part of the explanation, though, is surely that not everyone in a single society experiences the same environment: some environments are more integrated, some contain more overt expressions of prejudice, etc.

to take corrective measures once they become sufficiently aware of the biases and how to correct for them (Saul forthcoming).[6]

Some of these results are particularly relevant to rankings exercises in philosophy, and I outline these below. Although departmental ranking exercises have not, to my knowledge, been studied by psychologists, there is much research that bears on them. As we will see, some of the departmental ranking exercises take the form of assessing particular pieces of work and some of them take the form of assessing particular individuals or groups of individuals. Both of these sorts of assessments have been studied, as we'll see below.

1.1 Assessments of pieces of work

The studies that are most relevant here are studies of journal refereeing and of marking. In both cases we are dealing with a person who attempts to assign a judgment of quality to a piece of written work after a close study of it. In the United Kingdom, anonymous marking has come to be widely practiced in recent decades. Studies have clearly shown anonymous marking leads to higher grades for women students (Bradley 1984, 1993) The only sensible interpretation of this result is that non-anonymous marking is biased against female students. (One might expect similar biases in cases where a name is indicative of a stigmatized racial, ethnic or religious background, though to my knowledge this has not been studied.)

Journal article refereeing, however, provides a closer analogy to the sort of thing that goes on in some ranking exercises. In journal article

[6] Though for an opposing view, see Holroyd (this issue).

refereeing, an expert reviewer closely reads a piece of professional level work and makes a judgment of its quality. These judgments seem to be strongly affected by unconscious biases. The clearest cases of these are those of prestige bias.[7] A classic study of prestige bias took papers that had already been published in top psychology journals (which did not practice anonymous review) and resubmitted them to the same journals with false names and false, unprestigious affiliations. 90% of the papers were rejected, citing serious methodological errors (Lee and Schunn 2010, Peters and Ceci 1982).[8]

Finally, there may be some biases related to topic. Things associated with women or with femininity are often judged to be of lesser significance, quality or difficulty (e.g. when computing was dominated by women it was much less prestigious than it is now). This seems likely to hold true for areas of philosophy that have more women, such as feminism. (See Haslanger 2008.) Similar biases are likely to hold with

---

[7] The case of gender bias is less clear, perhaps due to the frequency of multiple author submissions in the fields that have been studies, and to complications arising from names of uncertain gender. For a study suggesting gender bias see Budden et. al. For an overview of concerns about studies showing gender bias see Lee (draft). Again, one might expect similar biases in cases where a name is indicative of a stigmatized racial or ethnic background, but this has not to my knowledge been studied. It is important to note, though, that even prestige bias may have a gendered impact, given that women tend to be concentrated disproportionately in less prestigious institutions (Valian 1999).

[8] A judgment this negative is far from typical in referee reports from psychologists (Lee and Schunn 2011), so it's not the case that this is simply a standard sort of rejection.

respect to areas of philosophy associated with non–whites, for example philosophy of race or post–colonial philosophy.

1.2 Implicit Bias and Assessments of Individuals

A variety of studies have shown that implicit bias strongly affects assessments of individuals as well. Some of the clearest of these involve the assessment of pairs of CVs, which are identical except for one purportedly irrelevant factor: (likely) gender of name, (likely) race of name, parental status and so on. These studies have found that the same CV is judged to be of significantly higher quality with a male name rather than a female name, a "white" name rather than a "black" name, and an indication of paternity rather than an indication of maternity. Women are generally stereotyped as less likely to be excellent, and less likely to be original (Valian 1999, 2005). More specifically to philosophy, women are stereotyped as less likely to excel in areas where a substantial mathematical/formal component is present (Nosek et. al. 2002). There's no reason to think this stereotype wouldn't have pernicious effects in at least some areas of analytic philosophy, due to the emphasis placed on logic and formal methodology (Saul forthcoming). Finally, and importantly, psychologists have also shown that a dimly remembered male name is judged to be more famous than a dimly remembered female name (Banaji and Greenwald 1995.)

1.3 How to avoid/minimize implicit bias

Research on how to reduce or minimize implicit bias is ongoing, and much at this stage is a bit speculative. But a few things are very clear. One is that a conscious, direct effort to simply not be biased is unlikely to succeed and may even make things worse (Blair 2002). We don't have

this sort of control over our biases and their functioning. Another is that anonymising greatly reduces the potential for implicit bias: if one does not know the name of the person who is the subject of a CV or the author of a paper one (usually) cannot be unconsciously biased against them due to their social category membership.[9]

There are also some interventions that can reduce implicit bias or its manifestations. One of the best is exposure to counterstereotypical exemplars—people who are members of a stigmatized social group, but who don't fit the stereotypes of that group (Blair 2002). Even fairly minimal exposure can make a difference (e.g. gazing at a picture of Martin Luther King before taking a race IAT reduces implicit race bias). But greater exposure is more effective.

Implicit biases manifest themselves most strongly when one is trying to make decisions at a rapid speed, so taking more time over decisions reduces the potential for bias (Women in Science and Engineering Leadership Institute 2006). Being required to discuss and justify one's decisions, and to explicitly reflect on the potential for stereotypes to play a role, can also reduce their harmful effects (Ibid.). (This does not rely on the idea that one has conscious control over biases: rather, it relies on the idea that this sort of explicit discussion may help to flag up at least some cases in which there is no defensible reason for a judgment. And some of these may be cases of bias.)

Finally, there is recent work on the potential for implementation intentions to reduce implicit bias. Implementation intentions are very

---

[9] Even this is not infallible: one may sometimes be able to work out the author's social category membership through contextual cues, and one may know who the author is because work has been previously circulated. But this is no reason not to try.

specific intentions of the form "If I'm in situation S, I will do F" (as opposed to "I'll try to to F as much as I can"[10]). Forming an intention like "if I see a black person, I will think safe" or "if I see Ina, I will ignore her gender" has been shown to reduce implicit bias (Stewart and Payne 2008, Webb et. al. 2011). It is far from clear, though, what whether any such intentions could work for exercises like those discussed here.

Finally, it seems that reflecting on one's past failures to be unbiased can help one to be less biased in the future. (Reflecting on past successes increases bias.) This seems to be, like implementation intentions, a way of *activating* a goal that might otherwise not affect one's behavior (Moskowitz and Li 2011).

1.4 The psychology of reputation-building behaviours
Reputation-building is a complex phenomenon. But a part of it involves one's public behavior as a speaker and as an audience member at conferences and the like. These are especially important in a profession like philosophy where "being quick on one's feet" is a highly valued characteristic. Some of what I discuss in this section involves implicit bias, but some does not, and the interaction of these factors is important enough that it is worth devoting a separate section to this cluster of factors, despite its heterogeneity.[11]

There is a great deal of evidence indicating that women are likely to have a more difficult time building reputations in philosophy than men are. Our focus here will be on reputation-building via public behaviour at conferences at the like. What is most likely to impress on cases like these is argumentative quickness: a speaker who is able to respond forcefully

---

[10] This latter sort of intention is a goal intention, and it is extremely *ineffectual.*
[11] Many thanks to Ruth Chang for encouraging me to include this.

decisively to all objections is generally viewed as more impressive, as is a questioner who is quick to the devastating counterexample. Now, there's no reason at all to suppose that women would be less capable of defending their arguments or generating counterexamples. But there are plenty of reasons to suppose that women will have a harder time doing these things at conferences in a way that impresses.

A first way that women will have a more difficult time building a reputation via conferences and the like is that there is good reason to believe they are less likely to be on the programme as invited speakers. There are quite a lot of philosophy conferences with no invited female speakers at all (http://feministphilosophers.wordpress.com/gendered-conference-campaign/).   It is far from clear why this is the case. One possible explanation is that women are less likely to leap to mind as possible speakers than men are, which is a highly plausible manifestation of implicit bias (Saul forthcoming). Even when women's names do come to mind, it is likely to take more accomplishments for a woman to be judged as excellent than for a man to be so judged (Wenneras and Wold 1997). There have, however, been many anecdotal reports indicating that—in addition-- women are less likely than men to accept invitations. Attempts to discern reasons for this (see < http://feministphilosophers.wordpress.com/2011/03/18/results-what-if-anything-prevents-women-from-accepting-conference-invitations/>) have suggested that women may be asked especially late; that women who are asked may have more conflicting professional commitments than men (perhaps because the same small group of women are receiving most of the invitations) and that women may be more likely to have conflicting personal commitments. Someone also avoid conferences due to their experiences with sexual harassment. Male social/professional networks may also play a role. The number one reason cited by women

themselves, however, is lack of funding—which makes sense if women are, as some studies have suggested, more likely to be in less senior posts at less research-intensive institutions (Valian 1999). Whatever the cause, if women are less likely to *be* at conferences, they are less likely to impress at conferences.

Once at a conference, a woman will face other obstacles to building reputations. Studies have shown that both men and women are less likely to call on women (Sadker and Sadker 1994; Sadker, Sadker and Zittleman 2009). But women may also be more reluctant than men to contribute. A key reason for this is stereotype threat, another very well-confirmed phenomenon discovered by psychologists. In cases of stereotype threat, members of groups that are stigmatized at some activity (e.g. blacks on tests of intellectual ability, whites on tests of sporting ability, girls on tests of maths ability) underperform when their group membership is made salient and they are under pressure (Steele 2010). Importantly, just being very much in the minority is enough to trigger stereotype threat—which is generally the situation for women (and racial minorities) at conferences. As a result, women may feel less confident and less able to contribute; they may also genuinely find it more difficult to think of something to say. Helen Beebee has argued (Beebee forthcoming) that this will be exacerbated by certain common behaviors in philosophy that are stereotypically associated with men, such as aggressive behavior in the seminar room or the use of combative metaphors ('defeating' the opponent, 'going for the jugular', 'demolishing' an argument, etc). And when women do engage in these behaviours, their contributions may be viewed more negatively than men's. Psychologists have shown that behavior which is viewed as appropriately assertive in a man is often viewed as inappropriately aggressive in a woman (Valian 1999, 2005).

More generally, Eric Schwitzgebel (2010) has argued, quite compellingly, that philosophers have a tendency to over-rely on judgments of who "seems smart", and that these judgments will have a tendency to favour precisely those who are not from groups udner-represented in philosophy.

> I have been collecting anecdotal data on seeming smart. One thing I've noticed is what sort of person tends spontaneously to be described, in my presence, as "seeming smart". A very striking pattern emerges: In every case I have noted the smart-seeming person has been a young white male....I would guess that there is something real behind that pattern, to wit:

> Seeming smart is probably to a large extent about activating people's associations with intelligence...And what do people associate with intelligence? Some things that are good: Poise, confidence (but not defensiveness), giving a moderate amount of detail but not too much, providing some frame and jargon, etc. But also, unfortunately, I suspect: whiteness, maleness, a certain physical bearing, a certain dialect (one American type, one British type), certain patterns of prosody -- all of which favor, I suspect, upper- to upper-middle class white men.

All of these factors put together make it likely that men will be better at building and enhancing their reputations through public venues like conferences than women are. And the same is very likely to be true as well to other groups that are under-represented in philosophy.

Part 2: The Research Excellence Framework (REF)

The Research Excellence Framework (REF) 2014 is the successor to the Research Assessment Exercise (RAE), which has been running since the 1990s. It is a UK-wide, government-run assessment of research quality of departments in all fields, in all universities. The government uses it as a guide to the dispersal of Higher Education Funding Council for England (HEFCE) research funding. In the past, REF returns have had very substantial financial consequences for departments, although there is now so little research money to distribute that is expected to be less financially important (though still important to esteem). In each field, a panel of experts is appointed who read (normally) four samples of work from (approximately) a five-year period for each individual whose work a department submits as part of its REF return. For each department, a profile is created with ratings of the work read (percentage of submissions that fall into each quality band) and also of "impact", a new measure designed to assess the significance of research outside academia. There are a lot of concerns that can be raised about this process, and particularly (though by no means exclusively) about the impact measure. But here I am interested only in discussing considerations related to implicit bias.

It is worth noting, however, that the REF has certain virtues: it focuses on written work, and not on ineffable matters of reputation. It is not, then, subject to the sorts of problems described above concerning the building of reputations at public venues (except in so far as they may unconsciously influence judgments of papers). It also focuses on recent work, and so gives guidance that is not based only on history. This allows it to be highly responsive to improvements not just in the makeup of a department, but also in an individual's own work.

2.1 Implicit Bias and decisions regarding who to submit

A first opportunity for implicit bias comes at the stage when departments (or universities) decide which members to submit. They are meant to submit work from all "research active staff", so a decision not to submit a staff member means a decision to consider them not research active. This is very detrimental to both career and morale for most who fall into this category (though it may not be for someone who identifies primarily as a teacher rather than a researcher). And yet departments (or universities) sometimes make this decision, if a person does not have enough outputs, or if their outputs are judged not to be of high enough quality. If a department gets this decision right, they will end up with a better research profile than they otherwise would. Since any funding is determined by both profile and number of staff, it's a delicate balancing act. And yet many departments do decide to exclude some staff members.

These decisions can easily be influenced by implicit bias. They are made by people who are well aware of the social category(ies) into which the researcher falls, so there is room to be unconsciously influenced by age, race, gender, maternity, class—indeed, pretty much anything that is a source of bias. How these decisions are made will vary from university to university and department to department. The decisions may involve reading submissions, but they may involve just looking at CVs. Place of publication is often very influential. (Anecdotally, I have heard of more than one department where papers in Hypatia, the leading journal for feminist philosophy, are judged to be not REF returnable.) Some places may adopt checks against implicit bias, but most will probably not. As a result of this, departments may be making biased decisions that wrongly damage individuals' careers, reduce morale in their department, and are

ultimately damaging to philosophical quality. After all, discouraging people whose research is actually strong is bad for philosophy, as is underestimating the worth of whole fields of philosophy.

2.2 Implicit Bias and Quality assessments

REF assessors read a very large quantity of work over a relatively short time period, while at the same time generally continuing with most of their ordinary responsibilities. There is likely, then, to be time pressure. The procedure is not anonymous: assessors are aware of the names on the work they are assessing. They are also highly likely to be aware of institutional affiliation, both because this usually appears on publications and because philosophy in the UK is a small world. Given the smallness of the philosophy world, they may also be aware of class background, race[12], maternity status and disabilities. It is very likely, then, that they will be affected by implicit biases regarding gender and prestige, leading them to under-rate work by women and people at less prestigious institutions. It is reasonably likely that in some cases they will also be affected by biases about class, race, maternity status and disabilities.

2.3 Efforts to combat implicit bias

The REF as a whole has an Equality and Diversity Advisory Group. This group has been very concerned about diversity of panel membership, and in fact they have done pretty well on this, at least with respect to gender.

---

[12] They may also get a indication of this just from the name.

The Philosophy Panel is 40% female.[13] Unfortunately, the literature shows very clearly that it is not only men who are subject to implicit bias. So while it is good to have a high representation of women on the panel, it is not enough on its own to eliminate concerns. The Advisory Group has not issued any guidance about implicit bias.

However, Alexander Bird, the Chair of the Philosophy Panel, is very concerned about implicit bias, and has asked me to advise on what can be done to reduce it. Obviously, the single most important thing one could do would be to ensure that submissions to the REF are presented anonymously. This would be a significant undertaking, requiring the removal or obscuring of names from four outputs for (nearly) every research-active academic in the UK. It would also, of course, be an imperfect effort, since in some cases names would undoubtedly be known to panel members. It would not be possible for the philosophy panel to carry out this task on its own, and it is not feasible to do this for the REF as a whole for 2014. However, we can begin to lobby HEFCE to put procedures like this in place for future exercises.[14]

Anonymity does not seem to be an option. But there are still some measures that can be taken. I have recommended the following:

1. Try to minimize time pressure: Implicit biases have their strongest effects when subjects are working very quickly. Obviously, time pressure is inevitable. But, at the very least desire to avoid implicit bias may provide an additional

---

[13] To view the panel membership, go here
http://www.hefce.ac.uk/research/ref/panels/members.htm, and select Main Panel D.
[14] I thank Jon Webber for insisting that one should press for this.

motivation not to save everything until late. (Valian 1999: 207–8)

2. Put lots of excellent work by women in the initial calibration phase. This is a period before the assessment period, during which members of the panel read and assess work from the previous RAE to try to arrive at common views about e.g. what constitutes a 3* paper.

   a. Implicit biases have less effect where members of the stigmatized group are more represented in the sample. So, in studies like the CV one, CVs with female names fare better where they are present in a higher proportion. You obviously don't have control over the number of women in the evaluation phase, but you can do something about the calibration phase. (Valian 1999: 309–10)

   b. The effect of implicit biases is also reduced when subjects are exposed to counterstereotypical exemplars. So, for example, we'd expect that exposure to excellent women philosophers would reduce the effects of any biases against women philosophers. (Blair 2002: 249) The presence of so many excellent women on the panel will also help, of course. But the more the better.

3. Trying not to think stereotypical thoughts is unlikely to be effective, and it can increase bias. However, it can be useful to be aware of stereotypes and explicitly reflect on (and even discuss) where they may be influencing one, and whether they are playing a role in particular cases. (In case the contrast is

unclear: trying not to think stereotypical thoughts is an exercise in suppression. Reflecting on the role stereotypes may play in one's judgments is an exercise in awareness.) It may be useful for each panel member to do a 5-minute exercise in which they compare papers by a woman and a man to which they gave a 2, and reflect on whether bias may have played a role in their judgment. If they think it may have, then can then revisit other judgments.[15]

Bird is planning to supply the panel with papers to read on implicit bias and to attempt to implement these recommendations.[16]

2.4 Solutions?

Although genuine and conscientious efforts are being made to combat implicit bias, it will nearly inevitably have some impact on the REF, as long as anonymity of submissions is not possible. It is not clear from the literature just how big a mitigating effect the steps suggested above may have. Psychologists have not studied the use of these interventions in an exercise of this kind. It is to be hoped that there will be some reduction in implicit bias, but it seems over-optimistic to suppose it could be eliminated without anonymity.

Until we are able to achieve anonymity, it's far from clear how much progress can be made toward improving the exercise. At this point, it seems to me the best course is to be aware of the limitations of the

---

[15] I thank Ruth Chang for suggesting this exercise.
[16] Margaret Crouch suggested after reading this paper that individual panel members' ratings could be periodically reviewed for patterns that might indicate bias. If such patterns are found, these could be—carefully and privately—raised with panel members, as reflecting on one's past biased judgments is an extremely effective way of reducing bias.

exercise, and to keep on top of developments from psychologists who are working very hard to find new ways to combat implicit bias. A further step, not so relevant to philosophy but nonetheless important, would be to get the Equality and Diversity Advisory Board to begin addressing these issues. But the most important thing is to keep the pressure on for anonymity.

Part 3: The Philosophical Gourmet Report

3.1 Methodology

The Philosophical Gourmet Report presents itself as a measure of both faculty quality and faculty reputation, designed to be used by undergraduates considering where to go for graduate school. The Gourmet Report surveys several hundred experts on different areas of philosophy. These experts are presented with a list of names for each department that is included in the survey. The list does not include the names of departments. Experts are asked to make judgments about the quality of departments (on the basis of the listed names of department members) in each field in which they are designated as an expert, and to make a judgment about the quality of the department overall. Those surveyed are not required to make judgments of all categories requested or of all departments listed. The instructions specify: "please give your opinion of the attractiveness of the faculty for a prospective student, taking into account (and weighted as you deem appropriate) the quality of philosophical work and talent on the faculty, the range of areas the faculty covers, and the availability of the faculty over the next few years." (Leiter 2011) The experts who are surveyed are chosen by a large advisory board containing members from a wide variety of areas of philosophy.

3.1.1 Prestige Bias and Quality Assessments

By not listing department names, the Gourmet Report attempts to inoculate against (or at least reduce) prestige bias. However, its method is imperfect. Lists of names of members of top departments are easily recognized as such. And, with this knowledge, lists of less-known names are likely to be recognized as from less prestigious departments. Moreover, many (if not most) rankers will also Google names, thereby acquiring more certain knowledge of department affiliation. This is a perfectly legitimate, and indeed often necessary thing to do. A name will often feel familiar, without one remembering what it is familiar for. Googling can give one a list of publications, allowing one to realize this is (or isn't) the person who wrote that wonderful paper one read six months ago. One important result of all of this is that there is still a great deal of room for prestige bias to affect quality judgments, leading to people (and so areas) at less prestigious departments being judged as of lower quality than they in fact are. Whole department judgments are even more directly affected, since one is less likely to have familiarity with the actual work of department members outside one's own field.

3.1.2 Stigmatized groups and quality assessments

In the CV studies discussed above, women and members of certain minority groups are judged to be of lower quality than white men. The same is true for individual pieces of work by women and members of certain minority groups (as can be seen from studies of anonymous review and marking). Although the particular methodology employed by the Gourmet Report has not been studied by psychologists, there is no reason to suppose that it would be immune to these effects. Indeed,

there are some reasons to suppose that the effects would be stronger: there are a lot of lists to rank, and the surveys are being filled out by busy people at a busy time of year. This means that judgments will probably be rushed, and rushed judgments are known to be more prone to distortion by implicit biases.

Moreover, many of the methods that are most successful in combating implicit bias are unavailable. Anonymity is very obviously not an option. Another method that psychologists have shown to be very helpful is requiring those making quality judgments to provide detailed reasons for these judgments and to discuss them with others. While this is possible in theory for a survey participant, in practice it would be very difficult to pull off. Certainly, participants are not urged in any way to do this. And as we've already noted making the judgments without time pressure, very slowly, is also likely to be impossible.

3.1.3 Fame judgments and gender

As we have already seen, a dimly remembered male name is judged to be more famous than a dimly remembered female name. This has consequences for the Gourmet Report's methodology. It has especially stark consequences for the whole department rankings. Those making these rankings are unlikely to have detailed knowledge of the work of very many members of the department. They will then be relying on the general sense of the reputation of these people have. Since fame judgments are known to be influenced by gender, we can expect that male names may artificially inflate a department's ranking, and female names may artificially deflate it.

3.1.4 Reputation and under-represented groups

Because the Gourmet Guide asks rankers to judge the quality of many lists of philosophers' names, one's judgments are highly likely to be influenced by knowledge not just of actual work but of reputation. And indeed the Gourmet Guide is meant to be in part a guide to reputation, so arguably one should be influenced by this. In addition to the fame and gender bias noted above, we have seen good reason to think that women and members of other under-represented groups will find it more difficult to build up strong reputations via the common method of impressing at conferences. If this is right, then members of these groups (and the lists containing their names) are likely to be underrated relative to their actual quality.

3.2 Measuring and Creating Reputation

It is important to note that the Gourmet Report does not just *measure* reputation. It also has a significant effect on reputation. Although it is departments and not individuals that are ranked, all interested parties know who, say, the two philosophers of language are at institution X, and the philosophy of language ranking the institution X receives cannot help but have an effect on the reputation of these individuals. Suppose, then, that the quality judgments of departments are affected by biases in the manner that I have suggested above. As a result of the rankings that are published, both individuals and departments have their reputations affected by these results, which have been influenced by implicit bias.

A couple of years later, it's time for another report. Now prestige bias, which was never really absent, does its work again. Those departments and individuals that were judged to be of high quality before experience

an unfair advantage, and those departments and individuals that did not fare so well experience an unfair disadvantage. Since these earlier judgment were affected by various forms of bias, prestige bias helps to perpetuate these earlier biases. And since the same sorts of biases against members of stigmatized groups will still be in effect, these biases are affecting judgments and at least two different ways: by directly affecting current judgments of quality; and by indirectly affecting them through their effect on prior judgments of quality.[17]

The result is that, through all the mechanisms described above, members of stigmatized groups are likely to receive inappropriately low assessments of relative quality. And departments that contain these people are likely to receive inappropriately low assessments of relative quality. Importantly, I am not making the claim that departments with higher percentages of women will receive lower quality rankings than those with higher percentages of women. Implicit bias and stereotype threat, after all, will make it harder for women to do well as undergraduates, less likely to be encouraged to continue in philosophy, harder for them to excel in graduate school, harder for their quality to be recognized in graduate school, less likely to get strong letters of recommendation, and less likely to be hired. (For more on this, see Saul forthcoming.) The women who, despite this, get hired at strong research departments are likely to be especially exceptional philosophers. My claim is that these exceptional philosophers will nonetheless be likely to be underrated. This might result in departments with a high percentage of women having lower rankings than departments with a low percentage of women. But it might also result in these departments having the same ranking, where the one with a high percentage of women in fact deserves

---

[17] For more on these sorts of effects of prestige bias, see the literature on the Matthew Effect, e.g. Merton 1996.

a higher ranking. (I'm sure the reader will be able to work out other scenarios as well.)

3.3 Practical Effects

The practical effects of this are even more disturbing. And here it is important to note that the Gourmet Report is only *intended* to serve as a starting point for those considering graduate school. Despite this limited intention, the report has come to play some rather influential unintended roles: most significantly, perhaps, as a guide for university administrators, who are known to demand that new hires raise a department's ranking and to criticize departments whose ranking declines.

To begin to appreciate the damaging effects of all of this, imagine that you're on the hiring committee for a department that has been told to raise its ranking in the Gourmet Report. You now find yourself faced with two candidates who seem to you equally good as researchers, as teachers, as colleagues and so on. One candidate is a woman and one candidate is a man. (We will ignore the fact that implicit bias is probably affecting your judgment that the two candidates are of equal quality.) What should you do? Well, if you're familiar with the literature that I have just discussed, and if you understand the working of the Gourmet Report, the answer is clear: you should hire the man. If the research of these two candidates is truly of equal quality, it is likely that the man's work will be judged to be of higher quality by those filling out Gourmet Report surveys. And when the survey participants are ranking whole departments and *his* name is one amongst many dimly remembered names, it is likely to be judged more famous that *her* name would be. The instruction from your dean, then, amounts to an instruction to discriminate against members of stigmatized groups.

3.4 Partial Solutions

It is far from obvious to me how a Gourmet-style ranking exercise could be done in a way that immunizes it from the unfairnesses and inaccuracies that come with implicit bias. In particular it seems impossible to achieve anonymity and very difficult to eliminate the problems from time pressure—although a larger time period for responses might help.   Nonetheless, I offer two ideas below for measures that might improve the report. I the present a very important suggestion that is not so much about the report but about how it used.

3.4.1 Present as measuring reputation, not quality

It seems to me that one useful thing to do is to think very carefully about how the report is presented and used. I have argued that implicit biases make it highly likely that the report is inaccurate guide to quality, and that they render its use as a guide to hiring both problematic and arguably discriminatory. Nonetheless, I think the report may well be a fairly accurate guide to reputation, and this is one of the things that it presents itself as. Reputation genuinely is in part a product of implicit biases. Because reputation is a matter of how people and work are *perceived*, the influence of the implicit biases on judgments of reputation does not render these judgments any less accurate. My suggestion, then, would be to present the guide very explicitly as a guide to reputation rather than quality. I think the psychological data pose serious problems for its use as a guide to quality but no problems at all for its use as a guide to reputation. (However, there is something of a paradox lurking: if those making the judgments take their task to be one of judging reputations rather than quality, they would have good reason to

artificially inflate their judgments of white men since white men will find it easier to build strong reputations for all the reasons indicated above. So it is probably important that rankers continue to attempt to track quality[18].)

3.4.2 Drop (or re-work) whole department rankings

As I have argued above, there is plenty of room for implicit bias to detrimentally affect rankings of both areas and whole departments. However, it seems to me this worry is much more acute in the case of whole department rankings. With that in mind, I offer what is sure to be a controversial suggestion: abandon the portion of the Gourmet Report that asks rankers to evaluate whole departments. It is almost never the case that a ranker has detailed familiarity with the work of all the staff members in a department, across all areas of philosophy. It is even less commonly the case that they have both this familiarity and the competence to judge quality across all areas of philosophy. As a result, judgments of whole departments are especially unlikely to be accurate guides to quality, and especially likely to be strongly influenced by all the pernicious biases discussed above.   Even if the report came to be presented as a guide to reputation rather than quality, this worry would remain—albeit in a slightly different form. It would still be very rare for a ranker to be familiar with the reputations of all the members of a department, across all areas of philosophy, after all. So, once more, a vague sense of a name's fame is likely to be far too influential.

One way of dealing with these worries would be to simply abandon the ranking of a whole departments, and offer the Gourmet Report only as a

---

[18] I thank Rae Langton for raising this point.

guide to areas of philosophy. Another would be to let the rankings of whole departments be determined by rankings of areas. The algorithm for calculating this would be complex and undoubtedly controversial. However, there would be significant benefits to this approach: whole department rankings would better track the views of experts about areas of philosophy, and would thereby become less subject to biases about the fame of names. Moreover, the publicizing of the algorithm for calculating department rankings from area rankings would allow individuals to form an accurate view of the relative weight given to expertise in the various areas of philosophy (rather than speculating about this in the way that they now do). Those who disagree with the weighting would be free to disregard this aspect of the exercise.

3.4.3 Do not use as a guide to hiring

It is very important that university administrators be made aware of the possible discriminatory effects of using likely Gourmet Report rankings as a method for making hiring decisions. Given the way that reputations are created and perpetuated, the workings of implicit bias, and methodology of the Gourmet Guide, it is highly likely that members of stigmatized groups will have less impressive reputations than they actually deserve. Using a reputational guide like the Gourmet Report as a guide to hiring, then, is likely to have detrimental effects on members of stigmatized groups. Worse yet, one of the best ways to follow the advice "hire someone who will improve our Gourmet ranking" would be to deliberately discriminate. Because of this, it cannot in good conscience be used as a guide to hiring. It would be worth investigating the implications of this legally as well.

4. Longer-Term Solution

The longer-term solution to problems for ranking exercises that stem from implicit bias is to work to reduce implicit bias. And one of the best ways to do this is exposure to counter-stereotypical exemplars: people who are from stigmatized groups but who don't fit the stereotypes of these groups. We need, then, to increase the representation of members of stigmatized groups in philosophy. If we can do this to the point where we break down the stereotypical associations between philosophy and maleness, philosophy and whiteness, and so on, we may finally be in a position to make accurate judgments of quality. Until that time, all of these exercises must be taken with a substantial grain of salt.

Finally, we need to discuss these issues as a profession, and exchange and build upon ideas about how to improve these exercises. They look to be a permanent part of the landscape for us, so we should think about how we can best use them to improve the profession. To do this, however, requires first squarely facing up to the problems that they pose. My hope is that this paper can be a useful early step in this process.

Anonymous. 2009. "A Dearth of Woman Philosophers", New York Times October 2: < http://ideas.blogs.nytimes.com/2009/10/02/a-dearth-of-women-philosophers/>.

Antony, L. "Different Voices or Perfect Storm—Why Are There So Few Women in Philosophy?", Unpublished MS.

Baggini, J. 2011. "The Long Road to Equality", The Philosophers' Magazine 53: < http://www.philosophypress.co.uk/?p=1881>.

Banaji, M. and Greenwald, A. 1995. "Implicit gender Stereotyping in Judgments of Fame", Attitudes and Social Cognition, 68: 2, 181–198.

Beebee, H. forthcoming. 'Women and Deviance in Philosophy, forthcoming in *Women in Philosophy: What Needs to Change?* Edited by Fiona Jenkins and Katrina Hutchison. Available at <http://www. bpa.ac.uk/resources/women/>.

Blair, I. 2002. "The Malleability of Automatic Stereotypes and Prejudice", Personality and Social Psychology Review, 3: 242–261.

Blanton, H. and Jaccard, J. 2006. "Arbitrary Metrics in Psychology", American Psychologist 61: 1, 27–41.

Bradley, Clare. (1984) "Sex bias in the evaluation of students." British Journal of Social Psychology, Vol 23(2), Jun 1984, 147–153.

Bradley, C. 1993. "Sex Bias in Student Assessment Overlooked?", Assessment and Evaluation in Higher Education, 18: 1, 3–8.

Budden, A., Tregenza, T., Aarssen, L., Koricheva, J., Leimu, R. and Lortie, C. 2008. "Double-Blind Review Favours Increased Representation of Female Authors", Trends in Ecology and Evolution 23:1, 4–6.

Buckwalter , W. and Stich, S. 2010. "Gender and Philosophical Intuition: Why Are There So Few Women in Philosophy?", talk given at Society for Philosophy and Psychology, Portland, OR.

Calhoun, C. 2009. "The Undergraduate Pipeline Problem", Hypatia 24: 2, 216–223.

Correll, S. and Benard, S. 2007. "Getting a Job: Is There a Motherhood Penalty? ", American Journal of Sociology 112: 1297–1338.

Goddard, E. 2008. "Improving the Position of Women in the Philosophy Profession",
<http://aap.org.au/women/reports/index.html#Ausreports>
Haslanger, S. 2008. "Changing the Ideology and Culture of Philosophy: Not By Reason (Alone)", Hypatia 23: 2, 210–223.

Jost, J.T., Rudman, L.A., Blair, I.V., Carney, D., Dasgupta, N., Glaser, J. & Hardin, C.D. (2009). The existence of implicit bias is beyond reasonable doubt: A refutation of ideological and methodological objections and executive summary of ten studies that no manager should ignore. Research in Organizational Behavior.
Kang, J. and Banaji, M. 2006. "Fair Measures: A Behavioral Realist Revision of 'Affirmative action'", California Law Review 94: 1063–1118.
Lane, K. A., Kang, J. and Banaji, M. 2007. "Implicit Social Cognition and Law", Annual Review of Law and Social Science 3: 427–51.
Lee, C. and Schunn, C. 2010. "Philosophy Journal Practices and Opportunities for Bias", APA Newsletter on Feminism and Philosophy.
Lee, C. and Schunn, C. 2011. "Social Biases and Solutions for Procedural Objectivity", Hypatia 26:2, 352–373.
Lee, C, 2012. "Implicit Bias in Peer Review: Empirical and Philosophical Quandaries". Draft paper, presented at Implicit Bias and Philosophy workshop, April 2012.

Leiter, B. 2011 "PGR Welcome and Instructions"

Levy, S., Stroessner, S., and Dweck, C. 1998. "Stereotype Formation and
        Endorsement: The Role of Implicit Theories", Journal of Personality
        and Social Psychology 74: 6, 1421–1436.

Merton, Robert K. 1996. The Matthew Effect, II (1988). In *On Social
Structure and Science*, edited by P. Sztompka. Chicago, IL: The University
of Chicago Press.

Misra, J., Lundquist, J. H., Holmes, E. and Agiomavritis, S.'The Ivory Ceiling of
Service Work',Academe Online January-February 2011:
<http://www.aaup.org/AAUP/pubsres/academe/2011/JF/Feat/misr. htm>.

Moskowitz, G.B., & Li, P. (2011). Egalitarian goals trigger stereotype inhibition: A
proactive form of stereotype control. Journal of Experimental Social Psychology.

Nosek, B., Banaji, M., and Greenwald, J. 2002. "Math = male, me =
        female, therefore math ≠ me." Journal of Personality and Social
        Psychology, 83, 44–59.

Peters, Douglas P., and Stephen J. Ceci. 1982. Peer-Review Practices of
        Psychological Journals: The Fate of Published Articles, Submitted Again.
        Behavioral and Brain Sciences 5:187-255.

Sadker, M. and Sadker, D. 1984. Failing at Fairness: How America's
        Schools Cheat Girls, New York: Touchstone Press

Sandström, U., Hällsten, M. 2008. "Persistent nepotism in peer-review",
        Scientometrics 74: 175–189.

Saul, Jennifer. "Implicit Bias, Stereotype Threat and Women in Philosophy",
        fortcoming in Jenkins and Hutchinson, Women in Philosophy: What
        Needs to Change.

Schwitzgebel, E. 2010. "On Being Good at Seeming Smart",
http://schwitzsplinters.blogspot.com/2010/03/on-being-good-at-
seeming-smart.html

Steele, C. 2010. Whistling Vivaldi: And Other Clues to How Stereotypes Affect Us, Norton.

Steinpreis, R., Anders, K., and Ritzke, D. 1999. "The Impact of Gender on the Review of the Curricula Vitae of Job Applicants and Tenure Candidates: A National Empirical Study", Sex Roles, 41: 7/8, 509–528.

Stewart, B. D. and Payne, B. K. 2008. "Bringing Automatic Stereotyping Under Control: Implementation Intentions as Efficient Means of Thought Control", Personality and Social Psychology Bulletin, 34: 1332–1345.

Uhlmann, E. and Cohen, G. 2007. ""I think it, therefore it's true": Effects of self-perceived objectivity on hiring discrimination", Organizational Behavior and Human Decision Processes, 207–223.

Valian, V. 1999. Why So Slow? The Advancement of Women. Cambridge, MA: MIT Press.

Valian, V. 2005. "Beyond Gender Schemas: Improving the Advancement of Women in Academia", Hypatia 20:3, 198–213.

Webb, T., Sheeran, P. and Pepper, J. 2011. "Gaining Control Over Responses to Implicit Attitude Tests: Implementation Intentions Engender Fast Responses on Attitude-Incongruent Trials", British Journal of Social Psychology 51: 1, 13–32.

Wenneras, C. and Wold, A. 1997. "Nepotism and Sexism in Peer Review", Nature 341–343.

Women in Science and Engineering Leadership Institute. 2006. "Reviewing Applicants: Research on Bias and Assumptions", <wiseli.engr.wisc.edu/docs/BiasBrochure_2ndEd.pdf>.