Response to David Wallace

Brian Bruya

In responding to David Wallace, I'll lay out the sections of my argument from my article first (highlighting the sections that Wallace addresses), along with a quick summary of my main conclusions, and then review his complaints one by one, assessing along the way how much, if at all, they affect the overall argument. Here are the section headings in my article:

> Flaws in PGR
> A. Selection Bias I: Sampling Methods
> B. Selection Bias II: Expert Opinion and Overall Ranking
> C. Selection Bias III: Underrepresentation of Methodological Continentalists
> D. Methodological Flaw I: Misapplication of the Expert Committee
> E. Methodological Flaw II: Area Dilution
> F.1. PGR Results Demonstrate a Bias against History
> F.2. PGR Demonstrates a Bias Against Other.
>
> Suggestions for improving PGR
> 1. Use a Random Sample for the Evaluator Pool
> 2. Use Mathematically Aggregated Specialty Scores to Calculate Overall Ranking
> 3. Allow Evaluators to Evaluate Only One Specialty
> 4. Revise the List of Specialties
> 5. Offer a Special Score to Indicate Comprehensive Balance Within Programs

Summary of conclusions: The PGR employs a deeply flawed sampling method that provides no generalizable information about the quality of the programs being evaluated. The evaluators evaluate mainly outside of their areas of expertise, and so the rationale for using "experts" and for thereby excluding all other potential evaluators is void. This exclusiveness is slanted toward the area of Analytic M&E and away from non-Analytic History, and very far away from Other. As a result, specialties in areas that the PGR marginalizes or excludes outright are at a disadvantage when any program appeals to its PGR rank in making hiring decisions. Fields that the PGR marginalizes will be discounted, and fields that the PGR excludes will be excluded from consideration. Assuming this practice is occurring in the profession to a significant degree, the PGR is damaging the profession by unnecessarily and unproductively narrowing the scope of the field at exactly the time when the rest of academia and the world are globalizing.

I'll use Wallace's section numbering and headings from his article to help readers locate quotations from it.

1. Introduction

Wallace says:

> I don't find the qualitative parts of the criticisms persuasive (for the most part they seem to assume a straw man – that the PGR is intended as an opinion poll rather than an expert review – and then draw conclusions from that).

Wallace seems to be saying that the PGR is an expert review and that I treat it as an opinion poll. I entertain Brian Leiter's claim that the PGR is an expert review and show that it is not. I don't see where the straw man is in that.

Assessment: Wallace targets my argument in section B on expert opinion and the overall ranking, impugning my argument without support. Since no arguments were challenged on their merits, all arguments remain unaffected.

2. Hidden reallocation of subject speciality areas to groups

Wallace says, "[Bruya's classification method] appears nowhere in the main paper, which is entirely silent on the reclassification" (p. 1).

This is not correct. I begin looking closely at area classification in section E, on p. 669. I immediately provide a footnote (number 15) referring the reader to Appendix 2, where my methods are explained in full. I originally had it all, beginning right there on the same page, in footnote 15, but the journal said it was too long for a footnote, so we decided to move it to an appendix. The case is the same with the other appendix, by the way. To immediately refer the reader from a footnote to an explanation in the appendix seems perfectly reasonable to me. It is not hidden if I point the reader directly to it.

It would be great if Professor Wallace would take the same scrupulousness with respect to the placement of methodological explanations in my article and apply it also to the PGR, where it is not a matter of placement but a matter of absence! Why is the Methods & Criteria section of the PGR so silent on its actual methods and criteria, and why doesn't this silence bother Professor Wallace?
[At various places, I can't help but express indignation at the high level of energy that Wallace devotes to critiquing my article compared to the absence of energy he directs at obvious and not so obvious flaws in the PGR. I set these remarks off by putting them in italics. Anyone coming to this post with an open mind will be well-served, I think, by these reminders.]

Wallace says:

> Several of Bruya's arguments rely centrally on this reclassification. Firstly, Bruya's "egregious explosion" critique of M&E is based on the fact that there are 15 specialty areas in M&E, as against 6 in Value, 9 in History, and 3 in Other. Once the PGR's original classification is restored, there are only 7 areas in M&E, 8 in Science, 6 in Value, 9 in History, and 3 in Other.

Wallace's statement is misleading because it takes my claim out of context. Here is the full statement in the article (in section F.2, demonstration of bias against Other):

> Under M&E, not only is there a specialty called "General Philosophy of Science," there are also specialties identified as "Philosophy of Physics," "Philosophy of Biology," "Philosophy of Social Science," and "Philosophy of Cognitive Science"—all of equal standing in the ontology. Similarly, there are not just "Philosophical Logic" and "Philosophy of Mathematics" but also "Mathematical Logic." These seem to indicate an egregious explosion of M&E specialties, given that there is not even a distinct category for bioethics, environmental ethics, philosophy of education, existentialism, hermeneutics, Indian philosophy, Buddhist philosophy, Islamic philosophy, African philosophy, or Latin American philosophy, among many other possibilities. All of these latter are either lumped under other categories, such as twentieth-century Continental, or not given any recognition at all. (674)

Perhaps, Wallace would be satisfied if I changed the language to "egregious explosion of Analytic specialties." Either way, it doesn't affect the overall argument. All of the excluded specialties are still excluded.

Here is my note 21, which follows very soon after:

> It is worth comparing the PGR's list of philosophical specialties to those put out in a survey from the American Philosophical Association (2013), the largest society of philosophers in the United States. As I've already remarked, the PGR has the following number of specialties in each area: M&E—15, value—6, history—9, other—3. The survey by the APA was sent out by the executive director (Amy Ferrer) following the Eastern Division annual meeting (the largest annual meeting of philosophers in the United States) in order to evaluate the success of the meeting and how welcoming the climate was for underrepresented groups. In the demographic section of the survey, sixty philosophical specialties are listed. Compare this to the PGR's thirty-three and you begin to see indications of exclusivity in the PGR. Using the PGR's own way of grouping specialties into

areas, and standardized as described in Appendix 2, the APA's grouping would look like this: M&E—11, value—11, history—20, other—18. The differences are dramatic.

Here is Wallace's second argument on the point that my reclassification of the specialties is flawed:

> Bruya's "dependence ratios" (table 1) purport to show that evaluation of non-M&E areas is dependent on M&E, in that a high fraction of assessors in other areas co-assess in M&E. This effect is drastically reduced when the PGR's classification is restored:

Table 1: recalculated degrees of dependence between areas (cf Bruya table 1)					
Field	M&E	Value	History	Science	Other
Degree of dependence on M&E		0.20	0.31	0.58	0.19
Degree of dependence on Value	0.12		0.2	0.11	0.63
Degree of dependence on History	0.20	0.21		0.18	0.19
Degree of dependence on Science	0.24	0.074	0.12		0.25
Degree of dependence on Other	0.019	0.11	0.03	0.061	
Overall degree of dependence	0.51	0.51	0.55	0.72	0.93

> The overall levels of dependence for M&E, Value, and History are now about the same, all around 50%. The only subject-specific level of dependences above 50% are now the dependence of Science on M&E and the dependence of Other on Value.
>
> To repeat: it's not crazy to imagine a critique of the PGR based on the fact that it gives much too much attention to philosophy of science, and/or that philosophy of science is disguised M&E. But to conceal all this in an appendix and to bake that critique silently into the main analysis is very troubling.

Wallace did a good job reproducing my calculations using the original data, but he left an important column. Here is my re-do along his lines (let's call it Table 2):

A: Area	B: No. of Times Area Listed with Another Area	C: No. of Times Area Listed in Total	D: Degree of Dependence: Ratio of A:B	E: No. of Times Area Listed with M&E	F: No. of Times Area Listed with Value	G: No. of Times Area Listed with History	H: No. of Times Area Listed with Science	I: No. of Time Listed with C
M&E	81	157	0.52		17	25	30	2
Value	48	94	0.51	17		16	7	9
History	55	100	0.55	25	16		3	3
Science	48	66	0.73	30	7	3		2
Other	15	16	0.94	2	9	3	6	

And here is my original table for reference (let's call it Table 1):

A: Area	B: No. of Times Area Listed with Another Area	C: No. of Times Area Listed in Total	D: Degree of Dependence: Ratio of A:B	E: No. of Times Area Listed with M&E	F: No. of Times Area Listed with Value	G: No. of Times Area Listed with History	H: No. of Times Area Listed with Other
M&E	61	185	0.33		25	37	5
Value	48	94	0.51	25		20	10

History	55	100	0.55	37	20		3
Other	15	17	0.88	5	10	3	

Let's first remind ourselves of my original argument (in section E on area dilution):

> These results [in Table 1 above] demonstrate that for the most part M&E functions as an independent area while also influencing the areas of value and history, thereby implicitly compromising the independence of value and history in the rankings. Overall, this means that M&E has a disproportionate influence on the results. We see, then, that not only do we begin with a bias in which MC is underrepresented, as I demonstrated in section C, but now it is compounded by being diluted by M&E. Likewise, value is being diluted, and the poor area of other has virtually no independence at all.

So the claim I'm making is that there is a danger of the presuppositions and methodology of one area to dilute the distinct presuppositions and methodology of another area if there is a significant imbalance toward one area. Let's look at column D in Tables 1 and 2. Wallace notes that the number for M&E changed dramatically, going from largely independent to more or less just as dependent as Value and History. Notice, however, in Column E that it is still M&E (even this slimmed down version) that is co-occurring with both Value and History—so it still seems to be having its influence, unless you think that Value and History are somehow ganging up to water down M&E. Which is more likely? Suppose there is someone who does both M&E and Heidegger. In doing M&E, this person will either be doing it from an analytic perspective (say, Quinean or Davidsonian) or be doing it from a Heideggarian perspective. These ways of doing M&E are radically different. So when you look at the co-occurrence of History and M&E in the table, which is more likely—that Continental methodologies are influencing Analytic methodologies or the other way around? I've already shown in section C how History (as methodological Continentalism) is devalued in the PGR—and many other critics of the PGR have shown the same (citations of other critiques are given in the article). The point of this section of the article is to demonstrate bias on top of bias. I think the argument still stands, but even if it doesn't, it doesn't affect the overall thrust of the article—namely, that there is an exclusivity in the way the PGR is constructed that, when used to build departments, marginalizes and locks out certain perspectives in the field.

Wallace says that it is "very troubling" that I put the discussion of area delineations in an appendix. What about the lack of Indian Philosophy, existentialism, hermeneutics, philosophy of education, environmental philosophy, Latin American philosophy, African philosophy, Confucianism, Daoism, Japanese philosophy, Buddhist philosophy, etc. in the PGR and in PGR-ranked philosophy Ph.D. programs more generally—does Wallace find this troubling, too? If so, why not spend some time thinking about why this might be and working to rectify it?

Assessment: Wallace accuses me of hiding a methodological explanation, when in fact it is in plain sight. He then targets a minor point (egregious explosion of M&E specialties) in my section F.2 (bias against Other) by unpacking the M&E categorization. I demonstrate that his quibble does not affect any facet of my argument regarding bias against Other in the PGR. Wallace goes on to use the same unpacking of M&E to target my arguments in section E (area dilution). The point he makes about the change in the level of influence from M&E on History and Value might reduce one's level of confidence that the area of M&E is unduly influencing the area of History, but I show how confidence can be restored. Even if I were to concede this minor point, it does not negate my other evidence for bias against History in section C (or the bias against History that so many others have shown elsewhere), nor does it affect the fact that Other is still diluted away to nothing. All other arguments are also unaffected.

3. "Aggregated rank" for departments actually tracks PGR rank closely

Wallace says:

> Bruya calculates an "aggregated rank" for each department by simply summing their ranks in the specialty ratings. The method of aggregation is simply to sum the marks in each specialty area. This is problematic for three reasons:
>
> 1) it assumes that the PGR categories ought to be seen as of equal weight, something the PGR itself never claims. (This is a recurring issue in Bruya 2015.) Philosophy of physics, for instance, is a niche subject in almost all institutions, and a fairly high rank can be achieved by just having a couple of well-regarded people in the field; regarding it as of equal weight to (say) Ethics is hard to motivate. (This seems to be driving Notre Dame's very high rank on Bruya's analysis). □

Here is my question in this section (Section F.1, PGR results demonstrate a bias against History): if the PGR purports to be an "expert ranking," why are the evaluators for the overall ranking evaluating mostly outside of their areas of expertise? This is a methodological issue that the PGR does not address. One point I make in creating an alternative ranking based on specialty scores is to show that "the rankings of the PGR give the illusion of numerical precision" (p. 671), and so I offer an alternative to show that quite different results can be achieved through a ranking that is no more questionable in its assumptions than the overall ranking, and the results show that the overall ranking exhibits a bias against History. Wallace doesn't dispute this.

What's more, my equal weighting solution receives a more nuanced argument than Wallace acknowledges. Here is how I put it:

> There are any number of ways that the specialty rankings could be aggregated, but the obvious and simplest way would be to simply sum them for each program.
>
> A related improvement would be to rank individual professors, rather than entire programs. If a department has three philosophers working in epistemology who all deserve top scores in the specialty of epistemology, that program's score in epistemology should be three times higher than the score of a program that has only one epistemologist who deserves a top score. Granted, this would give a numerical advantage to larger programs, but this is a genuine advantage of larger programs, not an artifact of the poll's methodology. [Note 26: It is not so clear, however, that two specialists who rate a score of 3 are better than one specialist who rates a score of 5. This problem could be remedied by adjusting the evaluation scale. On the other hand, given that professors have a tendency to go on leave, having one 3 and one absent 3 would be better for a student than having an absent 5.] It seems obvious that a set of students can learn much more from three specialists in a field than they can from just one, all else being equal. There are two added benefits to this method. The first is that it would encourage growth in programs without discouraging comprehensiveness. Second, if a program could approach its university administration with a data-driven argument that growing the program would improve the program (in substance and in its ranking), that would be to the good. (681)

I don't work out the solution entirely, myself, but I think I go far enough to show that it is a better, fairer, more reliable method. I think a representative committee of the profession should work on refining it. I note further in my article that Jennifer Saul, in her own peer-reviewed article, recommends something similar.

Wallace says:

> 2) It confines the strength of an area to 5/165 of the total rank. The only way to get a high score on Bruya's analysis is to be strong across a very large number of disciplines; exceptional strength in a smaller number of large subject areas contributes little. □

I consider this a strength of this method, not a limitation, as I explain, in section F.2 (PGR shows a bias against Other). I think that comprehensiveness in a program is good for students. Many philosophers in other programs must agree, as this and terms like it are often used on program websites to promote

programs. This isn't to say that programs shouldn't have niches, but comprehensiveness has value, too, and this is not reflected in the PGR.

Wallace says:

> 3) It assumes that the PGR marks are aggregative, so that having three speciality areas ranked 2.0 or two ranked 3.0 is better than having one speciality area ranked 5.0. On the PGR scale 5.0 means "distinguished" while 2.0 means "adequate" and 3.0 "good", all of which are qualitative – and the scale is capped at 5 even for large and super-distinguished research groups, so this looks prima facie implausible. (This, and the previous point, seems to be driving Rutgers' comparatively low rank on Bruya's analysis.) □

I do consider this point. To repeat (note 26, p. 681):

> It is not so clear, however, that two specialists who rate a score of 3 are better than one specialist who rates a score of 5. This problem could be remedied by adjusting the evaluation scale. On the other hand, given that professors have a tendency to go on leave, having one 3 and one absent 3 would be better for a student than having an absent 5.

I guess I don't see the prima facie implausibility that Wallace sees.

Wallace goes on to say:

> But for all that, the striking fact about Bruya's alternative ranking is that it tracks the PGR rank very closely. Rounding Bruya's ranking to the same accuracy as the PGR (which only distinguishes c. 30 possible scores, from 2.0 to 4.8)…
>
> The correlation between PGR score and Bruya score is 94%; the standard error (average difference) is 4 ranks. Bruya describes this as "quite large" but it's pretty trivial given the spread of ranks of the PGR overall. That would seem to render moot most of Bruya's complaints about the aggregate rank, even if Bruya rank was methodologically sound. In other words, the effects of supposed assessor bias really aren't doing much to the overall rank. Add to that the really fairly low R-factor for history dependence and the result really doesn't look very significant.
> Though to repeat: Bruya rank isn't methodologically sound. A possible reply is that PGR rank isn't methodologically sound either. That isn't a legitimate response. "My interlocutor's method is unsound, so I will use an unsound method too" isn't okay. (In addition, the question of PGR soundness is a matter of judgment and argument; the problems with Bruya rank are straightforwardly mathematical.)

Wallace is not speaking to the point I am making in the article. I am *not* suggesting that everything about the PGR stay the same and that we simply switch from the overall rank to the summed specialty rank, which is what Wallace seems to be attributing to me. The most important thing is to use a more representative sample of evaluators. I think it is also important to revise the list of specialties. Then allow evaluators to evaluate in one specialty only. When these are done, sum the scores. If all of this is done, it will be a very different, more fair, more useful, more legitimate instrument. Wallace shrugs off the fact that in my table Notre Dame jumps from 18th place to 4th, and Rutgers falls from 2nd to 8th. Those seem like significant differences to me.

Wallace seems to be admitting the possibility that the PGR ranking is methodologically unsound, but that doesn't bother him. Why not?

Assessment: Wallace turns his attention to section F.1 (PGR results demonstrate a bias against History), and misconstrues my method of aggregating specialty scores to demonstrate bias against History as a stand-alone proposal for reforming the PGR. That proposal comes later and is more nuanced than Wallace allows. If that one part were all that I suggest, his counter-argument would stand. But that one part depends on other concomitant reforms and allows for related alternatives which together change the situation

entirely. Because Wallace does not address any of these other suggestions for reform or the alternatives, my suggestion for reform remains unaffected, as do all of my other arguments.

4. Concern about distribution of speciality areas

Wallace says:

> Bruya uses the APA's division of specialities as a target level for the spread of PGR evaluators. This seems to mischaracterise the rationale for categories both in the APA and in the PGR. In each case (one assumes) the reason is because that division is useful for whatever purpose is being served by the categoriser, and for neither the APA nor the PGR is the purpose to divide the philosophy demographic evenly.
>
> Philosophy of physics, for instance, is clearly smaller than ethics, but the PGR treats each as a special case because for prospective graduate students (the clearly-identified target of the speciality ratings), if they want to specialise in philosophy of physics they have fairly bespoke faculty needs. That carries no implication that philosophy of physics is comparable in size to ethics!

I don't understand the complaint here. I don't use "the APA's division of specialties as a target level for the spread of PGR evaluators." I bring up the APA's division as an alternative, to demonstrate that the PGR's method is not the only way to do it and that it is excluding a great number of specialties.

Wallace suggests that the PGR has a rationale for carving up the areas in a certain way, for selecting some specialties for inclusion and, more importantly, for excluding others. I would like to see that rationale. Does it not disturb Wallace that it doesn't exist?

> Bruya writes that "[t]here is no way to say for sure which way of slicing up specialities is most representative of philosophy in the United States, but let's just say that the APA – the largest body of philosophers in the United State [sic] – has it more correct."
>
> The APA division strikes me as so unreliable a guide to the demographics of the profession (does Bruya really think as many people do philosophy of biology as do Ethics? That only one philosopher in 60 has a Metaphysics primary AOS?)

Wallace is referring to my note 21 that runs from p. 674-675. I say:

> Using the PGR's own way of grouping specialties into areas, and standardized as described in Appendix 2, the APA's grouping would look like this: M&E—11, value—11, history—20, other—18. The differences are dramatic. No longer is M&E the dominant area; instead, history and other dominate, while M&E and value are equally sized minorities.

Wallace is saying that we shouldn't assume an equal distribution of scholars across each category. Point taken, but this doesn't alter the fact that so many specialties are excluded from the PGR, which is what I am driving at in this part of my argument.

I don't understand what Wallace means about the reliability of a guide to the demographics of the profession. My main issue here is not the number of people in a field but the lack of acknowledgement of the existence of the fields themselves. The PGR fails to acknowledge a great number of fields, and the APA survey points these fields out.

Wallace says:

> That if it was the best we could do, we'd just have to accept that we can't know the demographics of the profession. But in fact, we can do better. Passing over the fact that the PGR is not confined to the US (I write from Oxford!) I note that rough data on the spread of philosophers is fairly readily available, so there is no need to fall back on the flawed method of categories. For instance, the Philosophy Documentation Center collects data from university websites and departments, and from individuals, and on that basis compiles a list of how many philosophers are in each AOS. This is clearly a very crude indicator, as people can and often do have more than one AOS, but it at least bears some methodological relation to actual demographics.
> On that basis (excluding the catch-all category of "modern philosophy") we get the following:

Category	% of AOS in PDC data	% of PGR categories	% of APA categories
M&E	27%	21%	Est. 9%[2]
Value	30%	18%	18%
Science	9%	24%	Est. 9%
History	27%	27%	33%
Other	8%	9%	9%

> As it happens, the PGR categories track the demographics moderately well, much better than the APA categories (not that either is intended to track the demographics, and not that the fine details of the PDC-derived demographics should be taken too seriously).

Notice that Wallace is now changing the subject in a subtle but important way. He has moved from the demographics of specialties to the demographics of areas. I demonstrate that a great number of specialties are excluded from the PGR, and Wallace shows that the PGR tracks the demographics of areas "moderately well." But what about specialties? Wallace doesn't speak to the point and instead distracts by putting forth related but ultimately irrelevant information.

I visited the site Wallace mentions but can't tell how he derives his figures, and Wallace provides no URL or description of his method. I have been able to find a list of specialties on the site. I count 372. Wow! Did Wallace find the number of philosophers working in each of these 372 specialties and then slot the specialties into the PGR areas? If so, what did he use as a guide for doing the slotting? Where, for example would Adam Smith, Taoism, American Philosophy, and Romanian philosophy fall? When we do this kind of slotting, should we try to get into Leiter's head and do it in a way that we think he would do it, or should we do it in a way that each of us individually thinks it should be done. I, for example would prefer to slot Taoism under M&E. Would other people do that? This issue is the reason, as I recommend in my fourth recommendation (revise the list of specialties, p. 672), that a representative committee should be formed for the purpose of settling specialty and area demarcations. Wallace ignores this argument.

Assessment: Wallace again addresses my argument in F.2 (PGR demonstrates a bias against Other). He focuses on some minor details about demographics of the profession, purporting to show that the PGR tracks the demographics "moderately well." This is interesting preliminary information but beside my main point in this section that so many philosophical specialties are excluded from the PGR, a point that Wallace does not contest. If no one were working in any of the fields, their exclusion would be of no consequence, which is what Wallace seems to be suggesting, but the fields *are* populated, and so his criticism is moot. The result is that the PGR does demonstrate a bias against Other. All other arguments also remain unaffected.

5. Concerns about differing mean in speciality areas

Wallace says:

> Bruya's Table 4 (p.675) and associated discussion compares the "speciality ratios using actual PGR scores" to the "speciality ratings by highest possible scores", and adduces bias from them ("a

compelling result" – p.675). It's a little difficult to interpret what is going on mathematically here, but I think what it comes down to is that the mean score in M&E (including Science) is higher than in other areas. Recovering mean scores (as ratios of the average) by reverse-engineering Bruya's data, we get:

Category	Mean speciality score compared to mean across all subjects
M&E (including science)	+17%
Value	+ 5%
History	-15%
Other	-45%

Bruya claims, without further argument, that "[i]f programs on average have higher scores in a particular area, that means that evaluators are recognizing scholars in that area more often than in other areas." This doesn't follow from the data.

This is a nicely done table, and it is revealing that Wallace does not question the results. Instead, he turns to a statement that I make in which I say that the cause for these results is the disproportionate number of evaluators in particular specialties. Let's look at my argument:

> What if we calculate this same set of ratios for each program in the PGR and create an average over all the programs for each area? Doing that, we get the results in the third row of Table 4, which show us that on average programs ranked in the PGR are scoring disproportionately high in M&E, slightly disproportionately high in value, disproportionately low in history, and very disproportionately low in other.
>
> In other words, M&E, the most independent of all areas (as demonstrated above), is overrepresented by eight percentage points in its total score with respect to an "ideally balanced" score, whereas the comparatively dependent area of history is underrepresented by four percentage points—amounting to a twelve-percentage-point difference with respect to those two areas. Value looks to be slightly well-off, being overrepresented by one percentage point, but this is deceptive because it is a dependent area, like history, and tied to M&E evaluators. Most salient is that the already scarce and dependent area of other drops 45 percent from the "ideal balance," to account for a mere 5 percent of all specialty scores.
>
> The takeaway from this is that in programs that make the rankings of the PGR, there is an imbalance in program scores tilted toward M&E and away from history, and especially away from other. If evaluators are recognizing history disproportionately less than other fields and if history is more dependent as an area, then it is corroborating evidence of the relative lack of importance of history specialties in the PGR. The same goes even more so for other. (676-677)

My main point is that the area of other (the title of this subsection is "PGR Demonstrates a Bias Against Other") is marginalized to the point of irrelevance in the PGR. As I say, Wallace doesn't contend this point and instead focuses on the cause, but not on the cause of exclusivity, on the more general cause of why programs are, on average, scoring disproportionately in relation to an idealized score designed as a metric of area balance within programs. Of course, the general cause should also account for the case of Other, but Wallace's account seems to do it less successfully than my own account. Again, my focus is to explain why the score of other is off by 45%. Wallace's reasoning doesn't explicitly consider the case of Other. This complaint of mine against Wallace cuts both ways, of course. For both of us, our account of the cause must account for the disproportionality in all areas. He offers four alternatives in what he says is ascending order of likelihood. Here they are:

1) Bruya's suggestion: bias in favour of M&E/science

> 2) Work in M&E/science, on average, is somewhat better than in other areas in contemporary philosophy. (It is not an a priori truth that at each instant in time, each area in philosophy is doing equally well.) □
>
> 3) The PGR only names departments that score at least 3+ on the speciality ranking. If there are a large number of reasonably good Ethics departments (scoring 3-4) but comparatively fewer in Philosophy of Language, philosophy of language could obtain a higher average speciality score as a consequence, but that spread doesn't obviously mean that Ethics is doing worse (there may well be a number of departments rated 2 or 1 in philosophy of language that didn't show up; equally there might just be more Ethics concentrations period). This distribution within subject areas could readily have been analysed; not doing so seems a serious methodological lacuna in the critique. □
>
> 4) There is no prima facie reason to expect different groups of assessors to be systematically scaling in the same way – and it doesn't really matter if they are, given that the point of the speciality rankings is to rank-order within a speciality. Only if the PGR used an aggregative formula to work out overall rankings would it necessary to assume comparability of numerical rankings. (For that reason, an aggregative approach probably should aggregate ranks, not scores.) □

Let's start with number 4, the most likely, according to him. Hmm, no cause is given. It's just a statement about what is not a cause.

Let's move back to number 3, then. I'm not sure I fully understand Wallace's argument, but it seems to hinge on lower-rated specialty scores not being counted in the specialty rankings. He seems, however, to be comparing actual area scores to actual area scores, not actual area scores to the idealized scores, so I don't see the relevance. I do take exception, however, to his accusation that my not making the analysis he is making amounts to a "serious methodological lacuna" in my critique. *Would he also believe that the very same lack of analysis in the PGR's "Methods and Criteria" section also amounts to a serious methodological lacuna? If so, why doesn't he raise that objection?*

Number 2. If I understand him correctly, Wallace is saying that most scholars working in M&E are, on account of the fantastic work they do, more visible across specialties than people working in other fields, who don't do fantastic work and are therefore less visible. This seems unlikely to me. Superstars are visible across specialties, scholars like Peter Singer and Daniel Dennett, but there are very few of those. I think that folks in Chinese and comparative philosophy do pretty fantastic work, but for some reason it hasn't caught the eyes of the PGR evaluators. Why would that be? I suggest that it is because there are an insufficient number of evaluators specializing in Chinese and comparative philosophy in the PGR evaluator pool.

This brings us back to number 1, my proposal. Here is my argument:

> Evaluators are very rarely recognizing expertise in other. Does this mean that in philosophy programs across the country there are few experts working in the specialties of other? Not necessarily. It means that the PGR evaluators are rarely identifying them. For example, suppose that there were no PGR evaluators who had any specialty of other as a field of expertise. Suppose further that there were ten Ph.D. programs consisting of fifteen professors each with extremely high expertise only in specialties that the PGR classifies as other (or doesn't classify at all) and with no expertise in the other three PGR areas. None of these programs would even make the PGR rankings—not because they didn't have philosophical expertise but because the slate of evaluators in the PGR wouldn't be diverse enough to be able to identify them. This extreme hypothetical demonstrates the importance of having a slate of evaluators that is both diverse enough and balanced enough to equitably rate philosophers in all specialties of philosophy. Any imbalance in evaluator specialties with respect to the broader population will necessarily be reflected in the rankings, undercounting the specialties that are not adequately represented. (677)

Wallace fails to address this argument. Taken together with my other arguments for bias against History, it accounts for all cases of disproportionality, whereas Wallace's alternatives do not.

Assessment: Wallace again addresses my argument in section F.2 (PGR demonstrates a bias against Other), this time by attempting to account for the cause of the difference in mean specialty areas compared to "ideal" scores. He ignores my argument and proffers his own, which does not adequately account for the case of Other—the main target in my discussion. Wallace also seems to concede the point of exclusivity, quibbling only over the cause. So my main point in that section stands, as does the overall conclusion that the PGR is biased against Other. All other arguments remain unaffected.

[6.] Conclusion

Wallace's says:

> A large fraction of the "data-driven" part of Bruya's paper is open to severe criticism on methodological grounds, quite apart from one's assessment of the more qualitative issues. The severity of the criticisms are such that it's very hard to see the paper passing peer review in any journal of the quantitative sciences. Indeed, the first and most severe criticism in this note – the silence, in the main part of the paper, about the reclassification of the PGR "Science" category as M&E – would in other contexts be troublingly close to academic malpractice.

This so-called severe criticism has been addressed and shown to be a non-issue. Further, there is no silence on my part about classification, as I show, and is readily apparent to anyone who looks.

Wallace concentrates his critique on three sections of my article: E. Methodological Flaw II: Area Dilution; F.1. PGR results demonstrate a bias against History; and F.2. PGR Demonstrates a Bias Against Other. As I demonstrate, none of Wallace's complaints affect any of my main arguments of these sections, nor do they threaten any of the arguments in the other four sections, nor my recommendations for reform. The conclusions still stands the PGR has a deeply flawed sampling method and that the evaluators in the overall ranking are mostly working outside of their specialties. The conclusion that the PGR is exclusive in regard to who it allows to evaluate and in regard to which specialties it allows in for evaluation also still stands. Most importantly, the conclusion still stands that when the PGR is taken into consideration in making hiring decisions, it has a detrimental affect on the profession overall.

Wallace's contention that my paper would not pass peer-review in a journal of the quantitative sciences is at best an unsupported conjecture.

Why, by the way, does Wallace not wonder if the PGR, itself, would pass peer-review in a journal of the quantitative sciences? It is published on the website of an academic press and makes quantitative claims. Has it gone through a rigorous review process? I actually asked Blackwell a long while back and was not dignified with a response. Why is Wallace not as concerned about this as he is about my placing a methodological issue in an appendix and directing readers to it from an appropriately placed footnote?

In working through Wallace's critique of my article, I am disheartened. I make an argument about the role of the PGR in philosophy, demonstrating that it is damaging the profession, and Wallace, instead of entertaining any of the arguments that speak to this, hones in on niggling issues that in the end do not pan out in any significant way.

Now that Professor Wallace has examined some of my arguments in detail, I hope he will turn his considerable acumen toward the PGR and ask some equally probing questions, beginning with what follows.

Here is a set of questions that everyone interested in critiquing my article should ask themselves first:

1. Is it true that some philosophy Ph.D. programs in the U.S. use the PGR to build their programs? By this, I mean, does anyone in the hiring process ever ask "How will hiring this person affect our rank in the PGR?" (Alternatively, a program could use a PGR rank to substantiate its status in the profession to administrators or aspire to do so.) If the answer to this question is: "No, no significant number of programs use the PGR to build their programs," then my article is of little value, and there is no point in even critiquing it, except to say that it relies on this one flawed assumption. If, however, the answer is "yes," then the next question must be asked:

2. Is it true that hiring in any one specialty over other specialties can affect one's overall ranking, all else being equal? I show decisively, and others have also shown, that it does.

If both 1 and 2 are true, then there must be structural flaws in the PGR which will lead to artificial and damaging imbalances in the field over the long run, as programs pursue specialists in fields that will raise their PGR rank over specialists in fields that won't, or won't do it as well. [Perhaps someone who works on game theory could try to model this process to see where it will lead.] I ask in my article and try to answer the question of what those structural flaws are. The existence and deleterious effects of these flaws shout out for a response to a third question:

3. How can the PGR be reformed to be more inclusive?

All programs that want to raise or maintain their rank in the PGR have a very strong incentive to hire in fields already well-represented in the PGR and a very strong negative incentive not to take risks on marginalized fields. Over time, this will mean that marginalized fields will become more marginalized until disappearing altogether. This is exactly what is happening to Chinese philosophy. Fifteen years ago, there were specialists at Berkeley, Stanford, and Michigan. Instead of that number growing in other elite schools as multiculturalism and globalization and diversity have been spreading steadily through universities, now only one half-timer remains at Berkeley. The situation for Chinese philosophy is not unique. All non-Western philosophies are largely non-existent in PGR-ranked philosophy Ph.D. programs. There are many other marginalized fields as well (as I note in my article, and as others have noted) that face the same fate over time. Maybe all the blame can't be put at the feet of the PGR, but if any of the blame can, it is a reason in favor of reforming the PGR. In my article, I offer five suggestions for reform. Instead of focusing narrow critiques on minor parts of my article, let's open a discussion on all possible methods of reform.

Before closing, allow me to mention several more things (for those who have the patience to continue reading).

1. I appreciate the attention the article is receiving and hope that it contributes to rectifying the situation, but I can't keep up the pace of responding to blog posts. I suggest that anyone who has further criticisms of my critique write them up and get them published. Maybe we can start a little cottage industry examining the pros and cons of the PGR.

2. On that note, I would like to draw attention to another peer-reviewed critique of the PGR: Jennifer Saul's excellent "Ranking Exercises in Philosophy and Implicit Bias" (*Journal of Social Philosophy*, v. 43, n. 3). Here is a thought experiment that she puts forward:

> The practical effects of this [implicit bias in the PGR] are even more disturbing [than the under-rating of women philosophers in the PGR]. 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 this, imagine that you are 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 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 are 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 the Gourmet Report surveys. And when the survey participants are ranking whole departments and *his* name is one among many dimly remembered names, it is likely to be judged more famous than *her* name would be. The instruction from your dean, then, amounts to an instruction to discriminate against members of stigmatized groups. (268)

I was inspired to publish my piece after reading Saul's article. Her piece also offers a critique and valuable potential solutions. Implicit bias is another angle from which the PGR board of advisors should look at reforming the PGR. Take measures to reduce its influence as much as possible. Saul doesn't see anything being done about reforming the PGR in substantial ways, and neither do I, but both of us have suggestions on how to do it. Maybe someone of influence can open a conversation to discuss them.

3. There is one very large potential sticking point to the prospect of reforming the PGR. Let's say there is a discussion of solutions to reforming the PGR, but none can be agreed on as workable, or none actually gets implemented. Then what? Then, it should be scrapped altogether for the damage it is doing to the profession. Gregory Wheeler argues that the PGR is not reformable at all and so recommends that it be abolished now.

4. I would like to draw your attention also to the many other critiques of the PGR on sites across the blogosphere, as listed in my bibliography. These also were a great inspiration to me and were what prompted Leiter to defend the PGR on his blog in the first place, defenses that I take issue with in my article. Saul and I are not the only ones who have noticed and detailed significant problems in the PGR.

5. I've seen comments saying something to the effect "If you don't like the PGR, create your own equivalent." This is like telling a member of an excluded minority, "If you don't like our exclusive white, heterosexual country club, go create your own." No attempt to actually recognize the systemic roots of the issue, let alone rectify the situation. Actually, the analogy of the country club is not off the mark in the sense that country clubs function (or used to function) as gatekeepers to the upper echelons of society. The PGR functions this way, too. This is why its exclusivity is so problematic. On a closely related topic, a colleague in the social sciences suggested to me as I was writing this article that I should refer to the literature on gatekeepers in academia. I didn't do that, but perhaps that is another angle that someone would like to take up and pursue for a future article.

6. I would also like to direct readers' attention to the reason I wrote this article in the first place. My "Appearance and Reality" article started out as an appendix to the article "The Tacit Rejection of Multiculturalism in Philosophy Ph.D. programs" in order to support my claims there before eventually outgrowing that role. I would love to see a blog discussion of the arguments in the "Tacit Rejection" article, especially my argument for the value of pluralism. The "Tacit Rejection" article was also originally intended as an appendix—to the book *The Philosophical Challenge from China* (MIT Press, 2015)—yes, I'm shamelessly plugging my own work, but it is relevant. I would love to also see some discussion of the Introduction to that book and even of its contents and their relation to the prospect of widening the scope of analytic philosophy into non-Western resources.

7. Some people seem to look at the field of philosophy as if it were an economy, with rules of supply and demand governing hiring decisions, and various programs creating niches for

themselves within this economy. This is an attractive way to view the field because it allows for autonomy within all programs under the assumption that when all these autonomous programs pursue their own self-interest, the entire profession will flourish. When the PGR is introduced into this economy, however, it creates a situation analogous to the classic tragedy of the commons. Suddenly there is a limited resource that everyone pursues at the expense of the greater good. The purpose of my article is to turn our attention to this greater good and see how the PGR can be reformed in order to better serve it.

8. In section 5 above, a hypothetical department is discussed, one that has a high concentration of scholars in specialties that the PGR does not recognize. How plausible is the existence of such a department, and could such a department, if found, act as a test case for inclusiveness in the PGR? The PGR purports to represent the entire profession of philosophy. One of the main underlying assumptions of the PGR, in its role as a useful tool for prospective graduate students, is that there is a strong correlation between the quality of faculty in a program and the job placement rate when graduates come out of that program. If we can find a program that has a high job placement rate but does not make the PGR rankings, that would suggest a blind spot in the PGR evaluator pool. I know of just such a program: the University of Hawai'i. According to data on the UH philosophy website, UH's initial tenure-track placement rate for graduates since 2004 is 49%. According to a report put out by PhilosophyNews (using data from the year 2000), the figure of 49% places UH exactly even with Princeton University's philosophy program, and above the University of Chicago, University of Michigan, MIT, Carnegie Mellon, and Stanford, all well-ranked PGR programs. [The initial tenure-track placement statistic does not take into account other placements such as post-docs, visiting positions, lectureships, non-academic jobs, later placement, or the quality of the school, but for many graduates faced with adjuncting or outright unemployment, placement into a tenure-track position is the gold standard]. So if UH is placing graduates into tenure-track positions at a rate that exceeds most PGR-ranked programs, by the underlying rationale of the PGR (helping students choose programs that will get them jobs across the entire spectrum of philosophy), then UH should rank also.

The fact that UH doesn't register even a blip on the overall PGR rankings points to a blind spot in the evaluator pool. What is the pool lacking that UH has? Here is a list of the number of UH faculty working in various specialties, drawing from the specific listings on their faculty webpage and allowing for multiple specialties:

Aesthetics 2
American 2
Ancient Greek 1
Buddhist 1
Chinese 3
Comparative 4
Confucianism 1
Continental 1
Daoism 1
Environmental 1
Epistemology 2
Ethics 3
Feminism 2
Hermeneutics 1
Historical ontology 1
Indian 2
Islamic 1
Japanese 2
Mathematical logic 1
Metaphysics 2
Modern 1
Moral Psychology 1

Neo-Confucianism 1
Phenomenology 1
Philosophy for Children 1
Philosophy of Language 1
Philosophy of Law 2
Philosophy of Religion 1
Political 2
Process philosophy 1
Renaissance 1

Since most UH graduates are placed into positions in non-Western fields, the blind spot is obvious—specialties in non-Western philosophy, which are almost entirely absent in the PGR evaluator pool.

I happen to be a graduate of UH and so am not entirely unbiased on this issue, but the facts are the facts, and my association with UH doesn't change its value as a test case for assessing the inclusiveness of the PGR, using the PGR's own standards. Remember, the PGR does purport to represent the entire profession, not just part of it. Surely there are other test cases as well, and concerned scholars, like David Wallace and the PGR's own board, should look into these to see how the PGR can be improved and live up to its own aspirations for providing reliable information to *all* prospective philosophy graduate students.

[I know, making this claim opens me up to all kinds of sour grapes arguments: "Bruya's test case is a laughing stock: If his alma mater doesn't make the ranking, then the ranking must be flawed. That's what they all say!" But it is a good test case—for the reasons given.]

9. Finally, the PGR should look to the U.S. Fulbright program as a model of inclusiveness. Fulbright is one of the most prestigious fellowship programs in the U.S., but it doesn't draw it's awardees or its evaluators only from the most elite universities. It draws from community colleges right up through Ivy League schools, always attempting to retain a balance. If the PGR really aspires to represent the entire profession of philosophy by providing generalizable conclusions about what philosophers think about faculty quality in Ph.D. programs, or generalizable information about all philosophy Ph.D. programs that will place students into jobs, or generalizable information about job placement prospects into schools, then it should strive for a similar kind of inclusiveness.