

Against Candidate Quality*

Liam Kofi Bright[†]

January 27, 2017

Abstract

Academics often make judgements about the relative merit or quality of scholars. These judgements play an important role in various aspects of academic life, most noticeably during the hiring process. Using conceptual apparatus drawn from the work of Sally Haslanger, I argue that the procedures we use to make such judgements are unable to justify the institutional role that estimations of candidate quality currently play.

1 Introduction

Candidates for academic positions are supposed to be sortable by quality. Knowing their quality as candidates is taken to justify certain inferences about them – for instance, inferences concerning probability of offering intelligent opinions during discussion and collaborations, whether or not they regularly produce high quality work, whether or not they will earn the department they are associated with greater renown by their presence, and so on. In virtue of

*Thanks for helpful comments to Helen de Cruz, Olúfémí O. Táíwò, Yena Lee, Megan Fairchild, Tim Kenyon, Kevin Zollman, Jessica Wilson, Elizabeth Barnes, Remco Heesen, Eric Schwitzgebel, Eric Schliesser, Cailin O'Connor, Jennifer Saul, Yuzuko Nakamura, Daniel Malinsky, Danielle Wenner, Bryce Huebner, Carole Lee, and Zina B. Ward. This work partially supported by NSF grant SES 1254291.

[†]Department of Philosophy; Baker Hall 161; Carnegie Mellon University, Pittsburgh, PA 15213-3890, USA. Email: lbright@andrew.cmu.edu .

justifying these inferences, attempting to discern a candidate's quality is taken to be an important part of academic job searches or funding competitions. These decisions being made throughout the academy, in turn, culminate in an inegalitarian distribution of prestige and resources among academics.

In this paper I will submit potential analyses of the concept of 'quality' as applied to academic candidates to critical scrutiny. Both realist and instrumentalist analyses of the concept shall be examined. I shall make a very weak request of these analyses. I require only that it be feasible to, supposing the analysis in question to capture the nature of 'quality', know ourselves to be doing better than chance at picking out which candidates are higher quality than others. Even facing only this weak requirement, all analyses shall be found wanting. If 'quality' or 'merit' is analysed in the manners to be suggested, academics presently have no way of knowing they are reliably tracking its presence, comparative degree, or absence in the candidates they consider. Further, some of these analyses, if applied in guiding hiring or funding decisions, may lead to biases against members of certain already disadvantaged demographic categories.

My central claim, therefore, is that the prestige hierarchies and inegalitarian distributions of resources in the academy cannot be justified by appeal to the fact that we are a meritocracy that favours the highest quality academics. We have no way of knowing if we are even roughly tracking any such alleged meritocratic hierarchy of candidates. I illustrate how difficulties arise due to the widespread false belief that we can differentiate among candidates by quality through an examination of the ADVANCE programme. My arguments will suggest that it is failing to live up to its potential as a means of promoting diversity in science because of the role that notions of 'quality' play in how it is organised.

I am aware that readers are apt to be sceptical of my thesis that we cannot know if we are tracking quality when assessing candidates. Many will have experience of assessing candidates for job positions who seemed to be clearly

differentiable by quality, will have debated whether or not Affirmative Action policies (or their local equivalent) are enacted at the cost of lowering the average quality of scientist we can expect to see employed, will have felt the keen sting of insecurity at the recognition that many of their peers are better scholars than they are. I ask the reader to at least temporarily set aside such scepticism, and bear in mind the following point: a central role the concept of ‘quality’ plays in academia is to justify the existence of a social hierarchy among academics. For book length discussions of the empirical evidence that ‘quality’ plays this role, see (Cole and Cole 1973), or Lamont’s observations of the way grant committees choose among potential projects and researchers to fund (Lamont 2009, pgs. 168-170). Given that this is the role ‘quality’ plays in academia, even those who are sceptical about my central thesis have reason to read on.

For, even where one thinks a social hierarchy is just, and placed on sound metaphysical and conceptual foundations, it still behoves a philosopher to do the work of examining those foundations to ensure as much. What is more, in this case we should not be so complacent, since there is already evidence that belief in ‘excellence’ is worrying correlated with demographic exclusion (Leslie et al. 2015). If I am right in the arguments that follow this may well be the result of a systematic problem in how concepts of ‘quality’ are understood and used to organise academic life. Hence, even if still at the end of this essay you are committed to retaining a notion of quality, I hope to convince you that there are important and as yet barely addressed questions of how to provide an analysis that is applicable in a fairly organised academy.

2 Method

I will make use of a conceptual distinction drawn from Sally Haslanger’s work in *Resisting Reality*, introducing some new terminology below. Haslanger outlines a distinction between the ‘*operative concept*’, which is determined by how people apply (or withhold) the concept in question, and the ‘*manifest concept*’, which is how people think, or would describe, themselves to be using

the concept (Haslanger 2012, pg. 92). These things may not be extensionally equivalent. To take a made up example, suppose that the manifest concept for ‘profound person’ is ‘Person who utters deep truths or insights concerning the nature of reality’, whereas the operative concept may be better described as ‘Person perceived to be high of socio-economic-status who confidently asserts enigmatic or unintelligible things’. If there are some people who have working class accents but who are given to uttering deep truths, they will then fall under the manifest but not the operative concept.

I now introduce some terminology for the sakes of clarifying my present project. An *analysis of a manifest concept* is a set of conditions such that in a high enough proportion of cases if people in the target audience agree the conditions are (not) satisfied they will agree that the concept should (not) be applied. What counts as a high enough proportion of cases may be determined by context, and nothing in this essay will turn on the specifics of the choice made there. An *analysis of an operative concept* is a procedure specified such that for a given set of input conditions it outlines a series of steps that should be taken that terminate in a ‘Yes/No’ decision as to whether or not the concept should be applied. Or, if the requirement of a ‘Yes/No’ decision is too strong, perhaps ‘Yes/No/Undecided’ or ‘Yes with degree of confidence r ’ or some other weaker requirement. Again, nothing in this paper will depend upon the specifics of how discerning one wants the operative analysis to be.

Note that the point of introducing the idea of an analysis of an operative concept is for it to be relevant in explaining or justifying the behaviour of those making use of the associated manifest concept. For this reason I will concentrate below on operative concepts that we could deploy given our present levels of theoretical or technical development. If our analysis of an operative concept required us to compute an uncomputable function, observe what is going on in far distant regions of space, or wait until the end of time before offering a response, then we have given an analysis of the operative concept that can neither explain present behaviour nor serve as a workable proposal for reformed behaviour. These examples are meant to make clear that I am taking

our present technological and theoretical limitations to account in deciding what is or is not implementable. Perhaps in the future we shall easily be able to travel to far off regions of space, or predict with a high degree of confidence what shall happen at the end of time. None the less, such things could not presently describe or guide people's behaviour in deploying a concept. To give an example relevant to our purposes: any analyses of the operative concept of academic quality which required us to see how a person's work fits into the completed science of the domain they work in must be rejected, since even if such an analysis were attractive on some grounds it could not describe what is underlying our present practice, and so would not be relevant to this essay.

In order to evaluate academic's deployment of the concept of 'quality' I will consider three questions one may ask concerning the relationship between analyses of operative and manifest concepts. Those three questions are;

Question 1. Do they correlate with sufficient regularity?

Question 2. Can the operative concept be implemented?

Question 3. Can we evince that our operative concept has a greater than chance probability of being correlated with our manifest concept?

I will argue that in order to play its role in justifying a hierarchy an analysis of the manifest and operative concepts of 'quality' must be able to answer all three questions with 'Yes'. First, however, I illustrate an example of analyses of manifest and operative concepts that could answer all questions with 'Yes'. The manifest concept of quality for a computer processor is its being able to run programs fast and to be reliable / not break down. The operative concept of a computer processor's quality would be to examine its price and some simpler measurements, such as series number (e.g. i7 is better than i5) and its listed speed, and then to declare the computer of sufficient quality when enough of these indicators are to your satisfaction. Sometimes these manifest and operative concepts diverge in which processors they pick out as great, but overall they correlate fairly well (so may answer (1) with 'Yes'), it's easy to implement the operative concept by examining specifications (answering (2)

with ‘Yes’), and people can review the actual performance of processors using benchmarks to test how well the operative concepts are tracking the manifest one (thus answering (3) with ‘Yes’).

Each question will be explained in more detail, starting with question (1). This should be answered ‘Yes’ if in a sufficiently high proportion of cases people in the target audience will agree that the operative concept is being (in)correctly deployed only when the manifest concept has (not) been satisfied. In other words, in a great enough proportion of cases, people align in their judgements of when the operative and manifest concepts should or should not be deployed. Otherwise question (1) should be answered ‘No’. I take it that ideally manifest and operative concepts would be perfectly correlated in the above sense. However, given various difficulties one encounters in the practice of inquiry, the best that can usually be hoped for is that we at least correlate with some tolerable degree of reliability. Nothing in this essay will turn on what threshold is chosen for the desired level of reliability, and for rhetorical effect I will proceed as if nothing more than better-than-chance correlation is required for the relationship between operative and manifest analyses of ‘quality’.

Question (2) should be answered ‘Yes’ if and only if the procedure specified by the operative concept is such that human beings could actually carry it out. Otherwise it should be answered ‘No’. The desirability of this being answered with ‘Yes’ is motivated by the earlier discussion of the point of introducing the notion of an operative concept.

Question (3) should be answered ‘Yes’ if and only if it is possible to prove, or evince, that our operative concept will be better than randomly correlated – in the sense of question (1) – with the manifest concept. That saying ‘Yes’ to question (3) is desirable is a presupposition of this paper, and so is worth making explicit just for the sakes of exposing it to critique. However, it is also defensible on more positive normative and epistemic grounds. If we cannot evince that we are doing better than chance in matching our operative to our

manifest analysis then we are in the awkward position of not really knowing what we are doing when we evaluate candidate quality. If we cannot answer question (3) with ‘Yes’ then we are unable to demonstrate to rejected candidates that the procedure we used to decide they were not the best candidate available fairly evaluated their quality, or even lines up with the relevant notion of ‘quality’ at all. It is, or ought be, rarely considered satisfactory when those with the power to distribute resources can do nothing more to justify their decisions than assert ‘We know what we’re doing’, and I see no reason to make academia an exception to the prohibition on such credulous deference to elites.

These questions are especially interesting in the context of decisions about the distribution of resources and opportunities in academia, which is the use of ‘quality’ I have in mind. There may be contexts in which we can afford to be happy with analyses of properties that do not allow us to reliably identify their presence in objects or events of concern, or do not allow us to attain evidence that we are reliable. However, this is not so if the presence or absence of that property is taken to either guide policy or justify social hierarchies. In such cases it is important that policy makers themselves are able to recognise the presence, absence, or comparative degree, of the property in question in the domain in question; and to demonstrate that they have this capacity. This may take the form of recognising properties that only correlate with, or indirectly indicate the presence of, the property of interest. One can allow that they will be imperfect even in this, and only a certain degree of confidence or reliability can be expected. However, policy makers must have some procedure for recognising the property of interest. If they cannot recognise the property, or provide evidence that they are recognising it, there is no reason to ascribe the presence or absence of the property in question any causal role in explaining their decision, or normative role in justifying their decision. Note, also, that I am not concerned with decisions about who is the highest quality ‘overall’ or in any sense beyond that which is relevant to particular, local decisions; I take it that what is at issue is who is the highest quality candidate for a given role or opportunity. Since the alleged presence or absence (or comparative degree)

of ‘quality’ is taken to guide and justify such hiring decisions in academia, members of hiring or grant committees must be able to outline procedures which answer each of the above three questions with ‘Yes’.

3 Realism

The first analyses of the manifest concept of candidate quality I consider shall be realist analyses. Under these analyses the quality candidates have is some metaphysically real property inhering in the candidates, the presence of which we may hope to measure by various means. The first such analysis would be some sort of direct realism about quality. On this analysis of the manifest concept a candidate’s quality would be something we can directly observe on encountering them and their work, without the need for any intervening proxies or complex analyses. A candidate’s quality is perhaps something like Moorean Goodness. That is to say, candidate quality may be simple, unanalysable, and only detectable by means of directly considering questions like: how good is candidate A as compared to candidate B? (Soames 2003, pg. 43) This is perhaps what people have in mind when they say things like ‘This candidate’s high quality simply shines forth to any fair minded observer’. The operative concept to accompany this, then, would be nothing more than instruction to take a look at the candidate’s CV, and perhaps witness a few talks. It is possible, for instance, that this is what was meant in (Steinpreis et al. 1999, pg.523), where the authors suggest that academics are able to discern quality by just looking at CVs.

This manifest-operative concept pair should be rejected. First, there is simply no evidence that academics have the required ability to discern quality. Indeed, there is evidence that academics’ evaluations of the quality of published manuscripts correlates poorly with other intuitive metrics of quality (Eyre-Walker and Stolezki 2013), which bodes ill for academics’ ability to evaluate authors of manuscripts. It is not that evaluations made without explicit reasoning, of the sort the Moorean Goodness view relies upon, are never reliable. Where there exists some feedback mechanism which allows people to

realise they have gone wrong in specific judgements people are able to become reliable in making such judgements (Railton 2014). However, evaluations of candidate quality are not such cases. We have no means of distinguishing between a candidate who was promising but then whose quality changed over time, and a candidate whose quality we were mistaken about (Dennett 1993, c.f. Intuition Pump #7). As such, there is no feedback mechanism which allows us to tell when we were right or wrong, even granting the Moorean picture and where we are able to observe the subsequent performance of the candidate. This analysis thus leaves us with no reason to suppose we may answer question (2) with ‘yes’.

Second, the history of thought about race and IQ should give us ample reason to be suspicious of shadowy properties that nobody can quite clearly define but which we are sure can be discerned, and once discerned justify social hierarchies. This worry is supported by the evidence that where there is ambiguity in the metrics that are supposed to determine quality, implicit biases favouring socially dominant groups are liable to have greater effect (Dovidio and Gaertner 2000) (Zheng Forthcoming). Further, beliefs that candidates must have a certain brilliance – which may or may not be the same thing as quality, but is plausibly the sort of thing people have in mind here – correlates with exclusion of certain marginalised demographic groups (Leslie et al. 2015), and likely actively promotes such exclusion (Yap 2014). While perhaps such concepts can, on examination, be put to good social and scientific use, our default assumption should be wariness and suspicion of such claims, rather than confidence in our own abilities to detect and make use of them.

A second realist analysis of candidate quality would be to treat it as expertise in the sense favoured by social epistemologists. That is to say, according to this analysis candidate A is of higher quality than candidate B if and only if, within some specified domain of interest, A is more likely to answer questions correctly than B (Goldman 2003, c.f. pg.268). Note that there may be a colloquial sense of ‘expertise’ which is more expansive than the social epistemologists’, but I consider the social epistemologists’ sense because it is well defined, and the

more colloquial sense of ‘expertise’ risks being no clearer than the initial notion of ‘quality’ or ‘merit’ we wanted analysed. Further, this analysis has the virtue of simplicity, and may even match something like what is in mind when panels assess candidates during some (not all) funding competitions. Finally, this analysis has received some previous attention in the literature: David Lewis once considered this, ultimately to set it aside, as a measure of the quality of candidates for academic appointments (Lewis 1989).

However, this analysis will not do even as an analysis of the manifest concept of candidate quality for the majority of theoretical and practical purposes we put the concept to. For instance, in many research roles a high quality candidate is supposed to be able to pick interesting projects for their next research topic. It does not automatically make a candidate of lower quality if they pick an area of research which is of greater theoretical interest but which is harder to produce results about. Besides being able to pick fruitful lines of research, we also expect from academic candidates high productivity, collegiality, successful teaching, ability to give good talks, ability to successfully compete for grants, leadership and service skills, skill at training/advising grad students. This analysis is unable to answer question (1) with ‘yes’, since, in general, it is not enough to be able to recite a higher-than-average proportion of true rather than false beliefs regarding some topic(s); we expect more than that from a high quality academic.

A third realist analysis one might give is that candidate quality is an underlying property possessed by individuals, which is indicated by the various sorts of observables which hiring committees attempt to pay attention to, such as number and calibre of publications, teaching evaluations, regularity of conference attendance, reports of their helpfulness as a colleague, etc. Under this analysis, candidate quality is an unobservable cause which may take various values (e.g. Low, Medium, High), and it a cause of the value of various observables (e.g. number of papers, place of publication, etc). Our task in constructing an operative concept is to come up with a way of estimating the value of candidate quality given our observations of the variable values

it causes. If such a property existed and we were able to reliably track the degree to which somebody possessed it, this would (perhaps) justify some of our hiring, promotion, and grant giving behaviours; it is hence plausible that something like this is what academics take candidate quality to be when they make use of the concept in their day to day professional lives.

This manifest concept of quality makes the problem of candidate selection into one of estimating the value of an unobserved variable. Given normal scientific practice, to carry this out with any degree of reliability we should require a model of some sort that told academics what they could infer about the value of the unobserved ‘quality’ variable in light of what they observe of the candidates publication behaviour, demeanor at conferences, peer reputation, etc. My first argument against this manifest-operative concept pair is therefore that we do not presently have anywhere near enough theoretical understanding of the sociology of science or psychology of inquiry to construct such a model. As such, for all we know we are not in a position to answer question (2) with ‘Yes’. My second claim is that if we could get lucky and chance upon such a model, we are not presently in a position to give any evidence as to whether or not it is the correct model. As such, we are not in a position to answer question (3) with ‘Yes’. In short: this manifest concept is itself so difficult to spell out that it renders us unable to provide an operative concept that could describe or justify our behaviour while answering all questions ‘Yes’.

Attempts to analyse candidate quality sociologically often use citation count (Cole and Cole 1973), publication count¹, or impact factor (Simonton 2004) as a proxy for quality, and then explore the relationship between this variable and other variables. Such models are not intended to be examples of the sort of model we have in mind, but rather assume that these observables can serve as a proxy for quality because there is some prior unarticulated reason to suppose their observables will systematically vary as level of quality varies. In the terms of §2, these are attempts to offer an analysis of an operative concept

¹See, for instance <http://genderandprestige.blogspot.com/> - last accessed 09/04/2015

for a manifest concept that has not itself been analysed. Some early attempts at making use of an underlying notion of ‘quality’ causing behaviours (in this case paper quality causing citation patterns) have been made (Wang 2014), but as yet they are nowhere near developed enough to form a full analysis of candidate quality. If we were to precisify this manifest concept, we would have to start nearly from scratch. Nobody is presently in possession of such a model, and as such nobody is in possession of an adequate manifest concept of this second realist sort.

Further, there are considerable difficulties in constructing such a model. Suppose we believed that we should observe number of publications and number of citations, awards, or other visible signs of peer recognition, when estimating a candidate’s quality. If these are our variables, we must contend with the following facts. There is good reason to think that the number of publications one attains can be affected by the candidate’s gender (van Arensbergen et al. 2012). There is good reason to think that the number of citations or awards one receives can be affected by small initial success followed by iterated processes of cumulative advantage (Merton 1968), sheer good luck (Heesen Forthcoming), and also by the ethnic backgrounds of memberships of coauthorship teams (Simonton 2004, ch.5) (Freeman and Huang 2014). It is also far from settled in bibliometrics how precisely to count citations (Pendlebury 2009). Further, both of these measures will be heavily affected by the observed (Kuhn 1996, pg. 151) and theoretically predicted (Kummerfeld and Zollman 2015) conservatism in science, such that high quality but novel candidates would be punished by finding it hard to publish and difficult to attain recognition. I assume these latter features are all confounders rather than themselves indicators of quality. Or, even if they are not, we presently lack any theory which would allow us to discern relevant as opposed to confounding instances of the above phenomena. In summary: we are aware that a variety of factors make a causal difference to how a candidate performs according to these quality metrics, but we are presently unable to tell what is signal, and what is noise. We don’t know what the required model for explicating this manifest concept of ‘quality’ would look like, and we don’t know how to begin

constructing one.

Suppose, however, that somehow some department were to strike upon a method of estimating the value of an unobserved quality metric without a model in hand, and even given the present poor state of theoretical development in the sociology of science. Even were that possible, the answer to (3) should be ‘No’ for this proposed manifest – operative concept pair. For lack of any model of ‘quality’, there is presently no way of giving any evidence whatsoever that the operative procedure is better than simply guessing who is of the highest quality. Hence the manifest concept of quality as possessing some not directly observable feature which manifests itself via various proxies leaves us unable to construct, or show ourselves to be already operating with, an adequate operative concept.

Personal correspondence suggests to me that many academics strongly believe the factors such as publication record and ability to converse intelligently correlate with some worthwhile notion of quality. In particular, I suspect this confidence may be so strong that people are willing to say that although we cannot answer question (3) with ‘Yes’ for any operative concept here, still we are so well assured that we are answering (2) with ‘Yes’ that we may go on in confidence while hoping to improve our modelling capacities as social science progresses. To give an instance of this attitude, a source referred to above conceded that publication count was a ‘crude’ measure of merit, but then said

However, to try and argue that publication count is not a major indicator of academic merit is not a reasonable position to take.²

However, when it comes to the organisation of social life and the distribution of resources, strong intuitions that something like the status quo is on the right track are best ignored. The nature of ideology is to inculcate in those who live under it a sense that things are more or less as they ought to be (Marx and Engels 1978, pg. 736) (Mills 2007) (Stanley 2015, ch. 5). Cognitive psychology attests to various “just world” biases we have, some of which will tend

²<http://genderandprestige.blogspot.com/> - last accessed 09/04/2015

to lead us to believe that when apparently innocent people suffer (perhaps, for instance, when deserving candidates are passed over for employment) that something must justify their suffering (Hafer and Begue 2005, pgs. 148 - 150 contain review and friendly critique of this literature). We should therefore expect such pro-status-quo intuitions whatever the status-quo, and so should avoid deploying arguments based on these intuitions as a form of justifying social hierarchies if we are concerned to avoid entrenching unjust status-quos. The matter can be put in stark terms: those presently at the top of academic hierarchies have no better ability to demonstrate that it is their greater possession of some underlying ‘quality’ that separates them from candidates who are rejected than Europe’s feudal aristocracy had to demonstrate that it was their possession of some underlying ‘nobility’ that separated them from the peasantry.

It is worth being explicit about what I do and do not claim to have shown. I do not claim to have shown that it is simply impossible to provide an adequate model of the operation of ‘merit’ or ‘quality’ along the lines suggested by this realist analysis. I claim that we have not yet done so and are not yet near to being able to do so; whatever explains or justifies the way we presently distribute resources in the academy, it cannot be that we are reliably tracking quality understood in this sense. Were we to just so happen to be able to track this property without an explicit model, by good luck or keen intuition (though the arguments against the Moorean analysis speak against this latter possibility), without the explicit model, or even some crude approximate to a fully explicit model, we are not in a position to answer question (3) with a ‘Yes’. Even if I am overly pessimistic in thinking it beyond our present capabilities to construct such a model, I take it that it is less controversial that none of the difficulties in constructing such a model and using it to form or evaluate operative analyses have yet been faced. Hence even if the underlying cause analysis of candidate quality is correct, the present way in which we reason about candidate quality is highly slapdash and theoretically inadequate.

Three varieties of realism have thus been examined and found wanting. Candidate quality is not identical to expertise, because displays of expertise are only a proper subset of what a good academic is supposed to be able to do. Candidate quality is not some sort of underlying property we have access to via proxies, because we have no reason to think that we could reliably estimate the state of any such underlying property. Nor should we be direct realists about candidate quality, or at least we should err on the side of rejecting this possibility, because we have reason to believe we are bad at making such judgements, and evidence that shaky judgements concerning such properties can serve to justify social hierarchies based on race or gender. Realist analyses of candidate quality hence lead us to make or endorse judgements that are false, pernicious, or both.

4 Constructivism

I next consider constructivist analyses of the manifest concept of quality. According to this sort of analysis, there are a variety of properties and virtues we are interested in that candidates may possess - average impact factor of published articles, teaching skill, regular publication habits, collegiality, etc. Candidate quality is not some additional property besides all these, for instance an underlying cause, which may correlate with them or manifest itself through them. Rather, candidate quality is simply an amalgamation we construct out of them, whose institutional purpose is to record a summary of a candidate's overall performance when taking all things into consideration. Being a high quality candidate is, on this analysis, something like the property of being the winner of an election. There is no property of being an election winner which a given candidate has that manifests itself, or has some propensity to manifest itself, in all the individual voting profiles of election participants. We do not try and choose electoral procedures with the hope of facilitating this property manifesting itself, or ensuring that those who have it actually take office. Rather, to say somebody is an election winner records a fact about the candidate's relationship to each individual voting profile as compared to their rival candidates relationship to those profiles, in light of whatever electoral

procedure is in place.

When we consider how a (close enough to) co-extensive operative concept may be constructed, it is significant that no convention has ever been established regarding the manner in which the amalgamation of candidate virtues is to proceed. One result of this is that the ‘etc’ which ended the list of candidate virtues is misleading. There is no agreed upon way of deciding which properties of a candidate should be taken into account for any given job. For instance, there is widespread and principled disagreement as to whether a candidate’s race, gender, or demographic background should count towards their quality. Some would say ‘yes’ (Harding 1993), others would say ‘no’ (Cole and Cole 1973, e.g. ch.1). More generally, sociological study reveals disagreement between scientific sub-communities about what constitutes quality (Hemlin 1993), and scientometricians have not achieved consensus on what constitutes quality even within a given field (Bucholz 1995). Until agreement is reached, we will be unable to amalgamate virtues for the simple reason that we do not know what virtues to amalgamate. Since it is often not the case that departments or funding bodies have explicitly stated the virtues they will be paying attention to, we will hence presently not be able to evince whether any operative analyses satisfies the desideratum, or can answer any of the questions ‘Yes’. Even if on reflection academics would endorse a constructivist analysis, therefore, they would very often not presently be in a position to justify their own behaviour according to the criteria of §2.

It is worth noting that this problem would also afflict any realist version of this proposal, along the lines of Boyd’s homeostatic properties (Boyd 1989). Without any agreement about what properties count towards ‘quality’ we are unable to develop methods of tracking it. As noted in the previous section, for my purposes in this paper I am not committed to the claim that no such homeostatic property cluster exists. Rather, I am claiming that even if such a property does exist, we are not yet able to track it in a way that allows us to answer question (3) with ‘Yes’, because we do not even agree what we are meant to be tracking, let alone whether or not we are successful in tracking it.

Suppose, though, that we had agreed, at least locally for a given job or award, about which properties of the candidates to take into account, and how each candidate ranked according to the respective virtues. The key point to be made is that, even with all these matters settled, when we amalgamate these virtues we must still decide upon some means of weighting them. Does being a good teacher count as much for a candidate's overall quality as being a good colleague? If not, how are they weighted respective to each other? Once again, there is no agreed upon convention regarding these matters, and there is no reason for optimism in thinking that this can be reached. It is plausible that the nature of a given position or award will at least provide some guidance as to which of a candidate's properties are functionally relevant. But the overall ranking by quality one will form of candidates will be sensitive to very fine distinctions between possible ways of weighting the properties against each other. There is no presently agreed upon way of using job requirements to decide whether (e.g.) teaching should be 2 times as highly weighted as collegiality or 2.5 times. What is more, it is hard to imagine how any decision context could decide what weightings different qualities should be given to such a level of exactness as to settle all gradations of difference that may affect overall ranking. But some weighting system must be decided upon for any operative analysis of this manifest concept.

It is especially important to note the distinction between what I do and do not claim to have shown. I am not here defending Humean inductive scepticism. I do not, therefore, claim to have shown that we cannot infer anything about future performance regarding publication rate, citation count, teaching ability, etc, from information about a candidate's past performance according to those metrics. Rather, my point is that to make use of this information about candidates to decide who should receive whatever benefit is being conferred, we must compress the information we get from these various measures into a single ordering. Or, at least, we must devise some sort of choice function which operates on profiles of such information. If we want to do this in a way which satisfies the requirements laid down in §2, we must, first, be able to specify what information was taken into account, and, second,

how the various measures were weighted. Unless and until we adopt some conventions specifying these matters, we have not yet constructed a concept of candidate quality that can justify its use in the maintenance of a social hierarchy.

5 Illustration

The ADVANCE programme is run by the National Science Foundation. Its aim is to improve the recruitment and retention of women and members of racial minority groups in academic science. As part of this programme, a number of guidebooks and resources have been developed for assisting departments, hiring committees, funding agencies (etc) deal with candidates with differing demographic categories in a more equitable manner. Many of their materials explicitly refer to notions of ‘excellence’³ As such it seems like a good test ground for much of my claims here. In this brief section I will first argue that, contrary to the self-description of some of the reports associated with ADVANCE, there is no reason to think that what they in fact achieve is the promotion of ‘quality’. I will then offer some reasons to be concerned about the difference between their self-description and their actual practice, and in particular raise worries about how this difference might hinder their ability to improve the recruitment and retention of under-represented groups in academic science. As I will make clear, I think that even given the criticisms I am putting forward there is important work for the ADVANCE programme to do, and I single it out for critique in the hope that it will help carry that work out if these concerns about the notion of ‘quality’ are considered and acted upon.

The materials ADVANCE circulate suggest they have something like either the realist ‘hidden variable’ analysis in mind or the constructivist concept (or its Boydean realist equivalent) in mind. For instance, one Candidate Evaluation Tool developed by researchers at Michigan University is designed to make

³E.g. http://diversity.arizona.edu/sites/diversity/files/evaluation_annotated_bibliography.pdf - last accessed 26/3/2015

people on hiring committees have to be explicit about the factors they took into account when rating potential applicants for a position.⁴ The hope is that forced to be explicit in their reasoning, it will be harder for (un)conscious biases to play an unnoticed role in determining who is offered what position. However, the judgements members of hiring committees are asked to be explicit about are things like ‘Potential for scholarly impact’, ‘Potential for research productivity’, ‘Fit with Department’s Priorities’, etc. These appear to be dispositional properties, and it is highly likely that these in particular were picked out because it is thought that they were either indicators of some underlying quality property, or are the stuff out of which quality is formed in the Boydean or constructivist sense. In this the Candidate Evaluation Tool is fairly representative, as, for instance, other ADVANCE documents promote various measures to ensure that evaluations are done according to explicit criteria.^{5 6}

I claim, however, that despite the self-described purpose of such materials in promoting the identification of excellence or ensuring well-functioning meritocracy, they are better understood as promoting equitable treatment. That we should not believe them to be promoting the identification of excellence follows from my previous arguments. The notion of excellence the ADVANCE programme members seem to have in mind fall under the categories I have already offered arguments against. However, this should not be read as a claim that the ADVANCE programme is failing to achieve anything worthwhile. They are for the most part following well known procedures for reducing the role of implicit bias in evaluations, and so it would be actively surprising if this was simply ineffectual in the peculiar context of evaluating academic candidates. Consistent with the argument of this paper, we can understand the ADVANCE programme as ensuring that whatever standards are applied to members of privileged demographic groups are also applied to mem-

⁴‘Candidate Evaluation Tool’ Document downloadable at http://sitemaker.umich.edu/advance/stride_committee - last accessed 26/3/2015

⁵http://diversity.arizona.edu/sites/diversity/files/evaluation2_bestpractices.pdf - last accessed 26/3/2015

⁶http://sish.fiu.edu/initiatives/advance-grant/stride-committee-2012/resources/ib_strategies_033012.pdf - last accessed 26/3/2015

bers of stigmatised demographic groups. We do not know these standards are tracking any notion of quality. But even without any belief in ‘quality’ many would hold that demographic information such as race or gender should not influence the sort of hiring or resource allocation decisions that ADVANCE monitors or advises on. For instance, even for a job so low skilled as to be performable by almost all members of a population, and where applicant quality is hence not a real issue in hiring decisions, it could still be found objectionable if hiring practices favoured members of privileged demographic groups. There is reason to think that whatever standards are applied in scientific hiring decisions are applied inconsistently, and to the detriment of members of certain demographic groups. Therefore there is reason to support programmes like ADVANCE even if they do not deliver on their stated goals of identifying excellence but rather promote equitable treatment.

However, while ensuring that candidates are considered by equal standards is a worthwhile project, it is worth noting the limits imposed on ADVANCE in even this regard due to the focus on ‘quality’. For ADVANCE’s materials work only to ensure synchronic equality of standards. That is to say, by making people be explicit about the considerations they are using to evaluate candidates with respect to each other, they are working towards ensuring that all those candidates in the same applicant pool are treated equally. But this is no reason to assume there is diachronic consistency in standards. If it were the case, as I suspect those involved in the ADVANCE project believe, that there was some underlying quality they are directing academics towards basing their judgements on, we would be in a position to suppose that by fastening evaluation to considerations relevant to that quality year after year academics were judging all those who apply by the same standard. According to the previous arguments we have no reason to believe academics are being directed towards any such property. The possibility of diachronic inequality of standards may itself be felt to be unfair. It should be considered especially worrying in light of evidence that where there is ambiguity about what considerations matter, hiring committees will shift the weighting attached to different values so as to favour applicants who are men (Steinpreis et al. 1999, pg.523) (Uhlmann and

Cohen 2006). Hence without the background presumption of an underlying and constant quality variable being tracked, we cannot rule out the possibility that the relative importance given to the considerations the ADVANCE programme highlights will be shifted in a gender biased fashion.

6 Conclusion

My arguments suggest that candidate quality is not expertise in some domain, nor an unobserved property tracked by proxies, nor an observable property we directly observe, nor a conventionally established construction. I have no argument to suggest that these properties exhaust the options. The typical dialectic in analytic philosophy suggests that further conceptual possibilities can be generated at will, and I would not like to place road blocks to inquiry by declaring it *a priori* that the concept of candidate quality must forever remain unanalysed or empty. My conclusion will hence be limited to: at present I see no reason to believe that candidates for academic positions can be ordered or compared in terms of their overall quality, and for that reason I am unconvinced that we presently organise ourselves according to any metric of quality.

Setting aside the failed attempt at an expertise based analysis, the reasons realist analyses failed differed from the reasons constructivist analyses failed. Realist analyses failed because we do not have the ability to (show that we can) reliably track the property they claim grounds our decisions. Constructivist analyses failed because to put them into practice we need to make communal decisions which we have not yet faced, and where it is presently unclear what would count as making these decisions well. This difference in cause of failure is worth paying attention to for anybody who wishes to explicate the concept of ‘quality’ so as to avoid the objections outlined here. In particular, this difference suggests that a constructivist analysis may be an easier foundation to work from if one wishes to devise a workable manifest-operative concept pair for evaluating candidate quality. Although we do not know how to evaluate proposed constructions, we can at least construct them. Whereas it is not clear

how to even begin producing a model of some hidden or directly perceived ‘quality’ candidates are alleged to have, let alone how to evaluate proposed models

This essay is just one of a number that raise concerns about attempts to aggregate various epistemic judgements into a decision about which theory is the ‘overall best’. It has been argued that there is no best ontology of science but rather our practical goals in carrying out the inquiry may dictate using theories with different ontological commitments on different occasions (Danks 2015), that sometimes there is no neutral decision procedure for deciding between competing paradigms (Kuhn 1996, pgs.199-200), that combining assessments of scientific virtues to arrive at an all-things-considered most virtuous theory may actually be formally impossible (Okasha 2011), that attempts to amalgamate peer reviews into a ‘Yes/No’ funding decision sometimes leads to pernicious bias (Lee 2015). The present essay should be considered a natural extension of this tradition, and by extending it raises the possibility that there is some underlying fact about how judgement aggregation works in the context of science that challenges received ideas about scientific theorising. If there is reason to doubt we can amalgamate virtues so as to pick an overall best theory in the relatively sanitised context of scientific theory choice, how much more fraught with difficulty and danger must it be to engage in the social project of trying to amalgamate judgements so as to decide on an overall best theorist.

Candidate quality is just one part of a wider hierarchical social structure which governs the organisation of science. It plays a key role in our decisions about how to distribute the rewards and responsibilities associated with academic life. If what I maintain in this essay is correct, it has managed to play this role in deciding how we organise ourselves despite being ill suited to this task. I hope this essay prompts the realisation that, first, other aspects of how we in academia now organise ourselves may be without justification once this support is removed, and that, second, if so central a notion as candidate quality turns out to be ill founded, then other aspects of the governing ideology of academia ought also to be critically examined and rejected if found wanting.

References

- Richard Boyd. What realism implies and what it does not. *Dialectica*, 43(1-2): 5 – 29, 1989.
- Klaus Bucholz. Criteria for an analysis of scientific quality. *Scientometrics*, 32(2):195 – 218, 1995.
- Jonathan Cole and Stephan Cole. *Social Stratification In Science*. The University of Chicago Press, 1973.
- David Danks. Goal dependence in (scientific) ontology. *Synthese*, 39(7857):1 – 16, 2015.
- Daniel Dennett. Quining qualia. In Alvin I. Goldman, editor, *Readings in Philosophy and Cognitive Science*. MIT Press, 1993.
- John F. Dovidio and Samuel L. Gaertner. Aversive racism and selection decisions: 1989 and 1999. *Psychological Science*, 11(4):315 – 319, 2000.
- Adam Eyre-Walker and Nina Stolezki. The assessment of science: The relative merits of post-publication review, the impact factor, and the number of citations. *PLOS Biology*, DOI: 10.1371/journal.pbio.1001675, 2013.
- Richard B. Freeman and Wei Huang. Collaborating with people like me: Ethnic co-authorship within the us. In Sarah Turner and William Kerr, editors, *High Skill Immigration*. University of Chicago Press, 2014.
- Alvin I. Goldman. *Knowledge in a Social World*. Oxford Univeristy Press, 2003.
- Carolyn L. Hafer and Laurent Begue. Experimental research on just-world theory: Problems, developments, and future challenges. *Psychological Bulletin*, 131(1):128 – 167, 2005.
- Sandra Harding. Rethinking standpoint epistemology: What is “strong objectivity”? In Linda Martin Alcoff and Elizabeth Potter, editors, *Feminist Epistemology*. Routledge, 1993.

- Sally Haslanger. *Resisting Reality: Social Construction and Social Critique*. Oxford University Press, 2012.
- Remco Heesen. Academic superstars: Competent or lucky? *Synthese*, Forthcoming.
- Sven Hemlin. Scientific quality in the eyes of the scientist: a questionnaire study. *Scientometrics*, 27(1):3 – 18, 1993.
- Thomas Kuhn. *The Structure of Scientific Revolutions*. University of Chicago Press, 1996.
- Erich Kummerfeld and Kevin J. Zollman. Conservatism and the scientific state of nature. *British Journal for the Philosophy of Science*, doi: 10.1093/bjps/axv013, 2015.
- Michele Lamont. *How Professors Think*. Harvard University Press, 2009.
- Carole Lee. Commensuration bias in peer review. *Philosophy of Science*, 82(5):1272–1283, 2015.
- S. J. Leslie, A. Cimpian, M. Meyer, and E. Freeland. Expectations of brilliance underlie gender distributions across academic disciplines. *Science*, 347(6219):262 – 265, 2015.
- David K. Lewis. Academic appointments: Why ignore the advantage of being right? In *Ormond Papers*, pages 187–200. University of Melbourne Press, 1989.
- Karl Marx and Friedrich Engels. On morality. In Robert C. Tucker, editor, *The Marx-Engels Reader*. W. W. Norton & Company, 1978.
- Robert K. Merton. The Matthew effect in science. *Science*, 159(3810):55 – 63, 1968.
- Charles Mills. White ignorance. In Shannon Sullivan and Nancy Tuana, editors, *Race and Epistemologies of Ignorance*. State University of New York Press, 2007.

- Samir Okasha. Theory choice and social choice: Kuhn versus arrow. *Mind*, 120(477):83 – 115, 2011.
- David A. Pendlebury. The use and misuse of journal metrics and other citation indicators. *Archivum Immunologiae et Therapiae Experimentalis*, 57(1):1 – 11, 2009.
- Peter Railton. The affective dog and its rational tale: Intuition and attunement. *Ethics*, 124(4):813–859, 2014.
- Dean Keith Simonton. *Creativity in Science: Chance, Logic, Genius, and Zeitgeist*. Cambridge University Press, 2004.
- Scott Soames. *Philosophical Analysis in the Twentieth Century*, volume 1. Princeton University Press, 2003.
- Jason Stanley. *How Propoganda Works*. Princeton University Press, 2015.
- Rhea E. Steinpreis, Katie A. Anders, and Dawn Ritzke. 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, 1999.
- Eric Luis Uhlmann and Geoffrey L. Cohen. Constructed criteria: Redefining merit to justify descrimination. *Psychological Science*, 16(6):474 – 480, 2006.
- Pleun van Arensbergen, Inge van der Weijden, and Peter van den Besselaar. Gender differences in scientific productivity: a persisting phenomenon? *Scientometrics*, 93(3):857 – 868, 2012.
- Jian Wang. Unpacking the matthew effect in citations. *Journal of Infometrics*, 8(1):329 – 339, 2014.
- Audrey Yap. Teaching logic, fighting stereotypes. Manuscript, 2014.
- Robin Zheng. Attributability, accountability, and implicit bias. In Jennifer Saul and Michael Brownstein, editors, *Implicit Bias and Philosophy: Volume 2 – Moral Responsibility, Structural Injustice, and Ethics*. Oxford Univeristy Press, Forthcoming.