Selection under Uncertainty: Affirmative Action at Shortlisting Stage

LUC BOVENS
London School of Economics and Political Science
L.Bovens@LSE.ac.uk

Choice often proceeds in two stages: We construct a shortlist on the basis of limited and uncertain information about the options and then reduce this uncertainty by examining the shortlist in greater detail. The goal is to do well when making a final choice from the option set. I argue that we cannot realise this goal by constructing a ranking over the options at shortlisting stage which determines of each option whether it is more or less worthy of being included in a shortlist. This is relevant to the 2010 UK Equality Act. The Act requires that shortlists be constructed on grounds of candidate rankings and affirmative action is only permissible for equally qualified candidates. This is misguided: Shortlisting candidates with lower expected qualifications but higher variance may raise the chance of finding an exceptionally strong candidate. If it does, then shortlisting such candidates would make eminent business sense and there is nothing unfair about it. This observation opens up room for including more underrepresented candidates with protected characteristics, as they are more likely to display greater variance in the selector's credence functions at shortlisting stage.

1. Regulations on shortlisting in the 2010 UK Equality Act

Section 159 of the 2010 UK Equality Act states that it is permissible to 'favour' persons with certain 'protected characteristics' (such as age, race, gender etc.) in recruitment and promotion practices when their representation is 'disproportionately low' in a particular sector of employment. Let us call candidates who fit this description 'underrepresented (UR) candidates'. The scope of this favouring is quite restrictive: It is only permissible to favour a UR-candidate X over a non-UR-candidate Y if X is 'as qualified as' Y. This is called 'Positive Action' in the UK Equality Act.

The Government Equality Office has put out a helpful document for employers entitled The Equality Act 2010 – *What Do I Need to Know?* This document clarifies what is meant by the clause 'as qualified as'. Positive action can only come in when both candidates are of 'equal merit'. Hence, it can only be used as a 'tie-breaker' or as a 'tipping point'. It can be invoked at any stage in the selection

Mind, Vol. 125 . 498 . April 2016
doi:10.1093/mind/fzv157 Advance Access publication 21 April 2016

procedure. There is an example in which a company draws up a shortlist of twenty candidates. Suppose that after the first top nineteen candidates are listed, there are two candidates of equal merit next in line, viz. one UR-candidate and one non-UR-candidate. Then we do not need to flip a coin: It is permissible that the twentieth spot on the shortlist simply be offered to the UR-candidate (pp. 5–6).

Contrast this with the 'threshold view' of shortlisting which is not permitted by the equal merit clause. On this view, we ignore all differences in merit above a threshold of qualification and proceed as if all candidates above the threshold were of equal merit. We then favour the UR-candidates above the threshold for shortlist inclusion. As a consequence, some non-UR-candidates who are higher ranked than shortlisted UR-candidates may not be shortlisted.

In parliamentary debates, Baroness Royall of Blaisdon takes position against this threshold view arguing that it violates the equal merit clause:

> Where the assessment process, in whatever form it takes, evaluates one candidate as having scored, say, 95 per cent and another 61 per cent, those candidates cannot be considered as being as qualified as each other to undertake the job. It is immaterial whether the pass mark was set at 60 per cent, 50 per cent or 40 per cent; the clearly superior candidate must always be offered the job. (Hansard – House of Lords Debates, 9 Feb 2010, Column 659 and 2 Mar 2010, Column 1421)

The Baroness points out that recruiting or promoting a less qualified candidate 'would make no business sense' (Hansard – House of Lords Debates, 9 Feb 2010, Column 659). Mark Harper states that appointing female candidates when there are better male candidates is 'not helpful to the cause of equality. That would give equality a bad name and damage the idea of fairness' (Hansard – House of Commons Public Bill Committee on the Equality Bill, 30 June 2009, Column 605).

The Solicitor General does not have much patience for epistemic problems in selection: 'It is not difficult to separate people who are as qualified as each other from those who are not' (Hansard – House of Commons Public Bill Committee on the Equality Bill, 30 June 2009, Column 612). The Equality Act 2010 – *What Do I Need to Know?* echoes this epistemic confidence: At the end stage of the selection procedure, 'all of the relevant factors that the employer will need to know in order to determine whether or not candidates are truly as qualified as each other should have been established' (p. 6).

Anyone who has served on a selection committee knows that there is much uncertainty in our judgment of the qualifications of the candidates.

This uncertainty is even greater at shortlisting stage. If we had a clear view of the qualifications of the candidates at this stage, then we would not need to shortlist—we could just pick the most qualified person.

My question is: What are we permitted to do in the face of uncertainty at shortlisting stage? I will argue that, counter to the Equality Act, it is permissible to pick candidates who are lower ranked over candidates who are higher ranked in our expectation of their quality. This benefits candidates who display greater variance in our assessment and UR-candidates often fit this description. *En route*, we will learn a few things about two-stage selection under uncertainty, that is, selection by isolating a subset of options for more careful investigation.

It is worth noting that the same question holds if we offer multiple candidates probationary contracts with the intention to offer one a permanent contract. Then the list of employees on probationary contracts is like a shortlist and the probationary period functions like an interview, viz. to gather more information before making a permanent hire. This procedure is structurally analogous to shortlisting.

2. The core argument

The Equality Act proceeds on the premise that we can place candidates on a qualifications scale and assess their qualifications in terms of single numbers. Let us assume that such scales are interval scales. They may be continuous or discrete and if discrete they may be more or less fine-grained. They are generated by a test-battery and they assess how good a candidate would be for a firm in a particular role. We let a candidate's 'qualifications' be shorthand for this goodness. Especially at earlier stages, there is uncertainty in the assessment of candidates. For some candidates we may be quite confident that they will perform reasonably well; for other candidates we may fear that they will do poorly, but there is a small chance that they will perform extremely well etc. We can express this uncertainty by means of a credence function. For example, suppose that we have a discrete scale of natural numbers from 0 to 10. A selector might assign credence (or subjective probability) .25 that a candidate is an 8 on the scale, credence .35 that she is a 7, and credence .40 that she is a 6. If a single number is needed to summarise a selector's view of how qualified a candidate is, then we may use the expected qualification of the candidate, in our case, $(.25 \times 8) + (.35 \times 7) + (.40 \times 6) = 6.85$. And we could then rank the candidates relative to their expected qualifications.

Table 1. Credence Functions over Qualifications of Candidates

	Candidates A, B, C	Candidates D, E, F		
Score	8	7	8	9
Credence	1	.20	.70	.10

Selectors on a committee may disagree with each other. I bracket this complication here: There is either a single selector or we have aggregated multiple credence functions so that the committee's judgment of a candidate's qualifications can be expressed in terms of a single credence function.

Suppose that the epistemic confidence of the Solicitor General and in the Equality Act 2010 – *What Do I Need to Know?* is warranted: At the end of a selection procedure, there will be no more uncertainty—the selector's credence function will have variance zero, i.e. the selector will have full credence that the candidate will have one particular score on the scale. We will come to have a ranking of the candidates according to what the selector takes to be their true qualifications. This is unlikely to be the case but we make this assumption for modelling purposes. What we wish to model is that some of the uncertainty that is still present at shortlisting stage will be removed during the interviews. At shortlisting stage, the selector's credence functions over the qualifications of different candidates typically do not have variance zero. This is precisely why we are shortlisting: We need to get a better view of the candidates in order to reduce the variance in our assessment. So when we have m spots on the shortlist, should we rank the candidates according to their expected qualifications and take the top m in the ranking to include in the shortlist? I will show that this way of proceeding would be deeply wrong.

To see this, let us construct a simple case. Suppose that there are six candidates, A through F, for a single job and we need to draw up a shortlist of three candidates. For candidates A, B and C, we have a clear view of their qualifications: We have certainty that their qualifications are at level 8 and so their expected qualifications are 8. For candidates D, E and F, there is much uncertainty: We have credence .20 that their qualifications are at 7, credence .70 that they are at 8 and credence .10 that they are at 9 for each candidate. Hence D, E and F's expected qualifications are at $(.10 \times 9) + (.70 \times 8) + (.20 \times 7) = 7.9$ (See Table 1). All uncertainty will be removed in the interviews of the shortlisted candidates so that we will gain certainty concerning the true qualifications of the candidates at the end of the interviewing process. Who should be shortlisted?

If we insist that we rank the candidates according to their expected qualifications, then A, B and C should be shortlisted. But as an employer, I would not want to be so constrained. I would reason as follows. If I shortlist A, B and C, then the expected qualification of my hire (i.e. the person who will ultimately be selected) is at level 8. If I shortlist D, E and F, then (assuming independence), there is a .271 chance that my hire will be a 9, a .721 chance that she will be an 8 and a .008 chance that she will be a 7.[1] Hence, by shortlisting D, E and F, the expected qualification of my hire is at $(.271 \times 9) + (.721 \times 8) + (.008 \times 7) = 8.26$. So I will shortlist D, E and F, because that choice of a shortlist increases my expectation of the qualifications of my hire.

It makes perfect business sense to shortlist with an eye to procuring the highest expected qualifications of the prospective hire. It does not make business sense to be told that we need to shortlist the candidates who have the highest expected qualifications. This is the core argument. I will now investigate the relevance and scope of this argument.

3. Benefitting UR-candidates

What is clear from our example is that it is not just the first moment of our credence functions (i.e. the expectation) that matters to shortlist inclusion but also the second moment (i.e. the variance). Candidates D, E and F benefit from this greater variance. If UR-candidates typically display a greater variance in the credence functions of selectors before shortlist construction, then our argument will benefit UR-candidates: UR-candidates with lower expectations may trump non-UR-candidates with higher expectations for shortlist inclusion due to the higher variance in the credence functions for UR-candidates.

Why might it be the case that the variance in the credence functions tends to be greater for UR-candidates than for non-UR-candidates? There are two reasons that have to do with the difficulty in assessing CVs of UR-candidates:

[1] To see this, note that the chance that the selector will hire a 9 equals the chance that there is at least one candidate who turns out to be a 9 who will then be picked and hired, i.e. $(1 - (1\text{-}.10)^3) = .271$; the chance that the selector will hire a 7 equals the chance that all candidates are 7s and no candidate with higher qualifications can be hired, i.e. $(.20)^3 = .008$; and the chance that the selector will hire an 8 equals the remaining chance $(1 - (.271 + .008)) = .721$.

(a) *Familiarity.* Non-UR-candidates typically come from the same cultural and educational background as the committee of selectors. For this reason, it is easier to assess the good-making features in their CVs. The committee knows how to read non-UR-candidates' transcripts since they come from the same schools as the candidate. UR-candidates also tend to perform and contribute in ways that are unexpected because they bring in skills from different social environments. For example, they may be proactive and show leadership in different ways than non-UR-candidates. This leads to more uncertainty in assessing the qualifications of UR-candidates.

(b) *Promise.* Non-UR-candidates have typically been exposed to an environment in which they can develop their talents and hone their skills, whereas many UR-candidates have not been exposed to such an environment. The home environment plays a large role in this respect, but also schooling and various social opportunities play a role. Hence we are more likely to be hiring UR-candidates on promise. Hiring on promise is more uncertain than hiring on actual achievements.

Furthermore, there are two reasons that have to do with distorting factors in the selection procedure that disproportionately affect UR-candidates:

c. *Anxiety.* Let there be a stereotype that members of a particular social group tend to lack a particular skill. There is empirical evidence that if members of this group are told that the test is a test which assesses precisely this skill then their average performance drops.[2] Also, if members of this group are asked to reveal their group membership when taking a test for this skill then their average performance drops.[3] This is called the 'stereotype threat'—a kind of anxiety that is responsive to the social expectation that one will perform poorly. This anxiety is a distorting factor in testing which is more prevalent for UR-candidates in general. The need to adjust

[2] Steele and Aronson 1995 observe that African American students who are told that the test is an intelligence test tend to perform worse than when the test is labelled differently.

[3] Hoff and Pandey 2004 find that the test performance of lower caste Indian students is negatively affected by disclosing their name (which is a caste indicator). Danaher and Crandall 2008 find that women do better in math tests when asked to fill in their gender after rather than before the test.

for the possibility of this distorting factor for UR-candidates makes assessment more uncertain.

d. *Implicit Bias.* There is a large literature documenting that selectors will let themselves be influenced by implicit assumptions building on negative stereotypes when evaluating applications.[4] UR-candidates are disadvantaged by such biases for many professional positions. When some selectors succumb to such biases, then this will result in greater uncertainty.

To explain why these distorting factors increase the variance in the assessment of UR-candidates more than in the assessment of non-UR-candidates, let me draw an analogy. Suppose that we assess alcohol consumption by means of the amount of alcohol purchased. We know that there is a distorting factor, viz. home brewing and we estimate that roughly fifty percent of Swedes home brew, whereas only very few Germans do so. Now suppose that we have alcohol purchase data for a Swede named Linnea and for a German named Ute. Then we need to adjust our expectation of alcohol consumption upwards from what the purchase data indicate for Linnea since she may be home brewing. Furthermore the variance of our assessment of the true consumption for Linnea will be greater than for Ute since we do not know whether she is or is not home brewing.

Similarly, we assess qualifications by means of test scores. We know that there is a distorting factor, viz. anxiety and implicit bias, and we estimate that anxiety and implicit bias affects roughly half of the UR-

[4] Corrice 2009 provides a literature overview of some core studies on bias in hiring for faculty and leadership positions. Steinpreis *et al.* 1999 ask academic psychologists whether they would hire or offer tenure to a fictitious candidate with a fictitious CV. They manipulate the gender of the candidate and find that respondents set a higher threshold for women in hiring. King *et al.* 2006 show high- and low-quality fictitious resumés with different ethnicities attached to them and ask whether the candidates would be suitable for high- or low-ranking jobs. The most striking finding is that Asian-Americans tend to be assigned to high-ranking jobs and their resumé quality hardly makes any difference. Bertrand and Mullainathan 2004 send fictitious resumés to help-wanted ads and insert names that are more common for whites and names more common for blacks. White names yield 50 percent more call backs. Carlsson and Rooth 2007 do a similar experiment with Swedish and Arab-Muslim names yielding more call backs for Swedish names. Goldin and Rouse 2000 examine the effect of a screen in auditions on the likelihood that female musicians are hired in orchestras and conclude that the increase in women musicians in orchestras is explained to a significant extent by the introduction of blind auditions in the 70's. Banaji and Greenwald 2013 argue that implicit biases are pervasive in society using data from the Implicit Association Test (IAT). In the IAT subjects make rapid word pairings which indicate the connotations they hold with gender and ethnic groups. Rooth 2010 finds that there is a correlation between Swedish recruiters' IAT scores and their willingness to shortlist Arab-Muslim men.

candidates and hardly any of the non-UR-candidates. We have scores for a UR-candidate Fatima and for a non-UR-candidate John. We need to adjust our expectation of the qualifications upwards from the test scores for Fatima since she may be affected by anxiety and implicit bias. And furthermore, the variance of our assessment of the qualifications of Fatima will be greater than for John since we do not know whether she is so affected.

Why is this the case? Just as we can be quite confident that Ute's alcohol purchases reflect her true consumption, we can be quite confident that John's scores reflect his true qualifications. But for Linnea we need to make an adjustment in our expectation for the fact that she may well be home brewing and, given our uncertainty about this distorting factor, we need to add 'give or take a few litres'. Similarly for Fatima we need to make an adjustment in our expectation for the fact that anxiety or implicit bias may have entered in and, given our uncertainty about these distorting factors, we need to add 'give or take a few points'. It is this 'give or take' that is represented in the greater variance.

Now it is not enough to assume that UR-candidates are *more* affected by anxiety and implicit bias than non-UR-candidates. If all UR-candidates and fifty percent of the non-UR-candidates are so affected, then we make adjustments in our assessment of the true qualifications and the variance in our assessment of John's qualifications will be greater than in our assessment of Fatima's qualifications. Similarly, if all Swedes and half of the Germans home brew, then we make adjustments in our assessment of the true consumption and the variance in our assessment of Ute's true consumption will be greater than in our assessment of Linnea's true consumption. But it is not unreasonable to think that anxiety and implicit bias affect a good number but by no means all UR-candidates, whereas they rarely affect non-UR-candidates. And this would explain the greater variance in our assessment of UR-candidates.

Admittedly, I have not been able to find any empirical studies on the greater variance in the assessment of UR-candidates and have merely provided plausibility arguments for the mechanisms that could bring such a greater variance about. When it comes to familiarity and promise, there are various reasons why selectors who are not from underrepresented groups have difficulty placing non-UR-candidates and it is plausible that this difficulty will be reflected in greater variance in assessment. As to anxiety and implicit bias, there is research showing that UR-candidates tend to be negatively affected in hiring procedures on these grounds and it is plausible to think that this will be reflected in greater variance. Only empirical research can substantiate such plausibility arguments.

4. The holistic nature of shortlisting decisions

Looking at our example involving candidates A through F, it seems that there should be an easy fix. We have measured the goodness of the candidates by means of their expected qualifications. But maybe all that is needed is to add some optimism to our measurement. An optimistic assessor could give a higher score to D, E and F in Table 1 than the expectation of their qualifications. If this score were higher than an 8, then D, E and F would indeed rank higher than A, B and C and we would be justified in shortlisting D, E and F.

There are standard ways of inserting some optimism in the measurement. For example, we could overweight our credence for the better scores and underweight our credence for worse scores. This is how John Quiggin 1993 calculates optimistic *rank-dependent expectations* and how Lara Buchak 2013 calculates risk-loving *risk-weighted expectations*. Or, we could identify candidates with the highest score x such that we have credence of at least .05 that they are at least an x. We can construct a ranking on grounds of these optimistic measures over the set of candidates and include the top m candidates—i.e. the better candidates in the eyes of a more optimistic selector—in the shortlist.

This works in the particular example at hand, but it is mistaken as a general strategy for shortlist construction when the goal is to maximise the expectation of the qualifications of the hire. The reason is that shortlist construction is holistic. There is no ranking of better over worse candidates such that we can just peel off the top m candidates when a set of candidates S is applying. X may be included and Y excluded from the shortlist or vice versa, depending on what other candidates are in the pool and depending on the size of the shortlist. There is no measure of X's and Y's relative goodness that can determine shortlist inclusion or exclusion.

First, shortlist status for the candidates X and Y may depend on the credence functions over the qualifications of the *other* candidates in the candidate set: X may be included and Y excluded from the shortlist when they are contained in a set of candidates S, whereas Y may be included and X excluded when they are contained within a set of candidates S'. Second, shortlist status for X and Y may depend on the size of the shortlist: X may be included and Y excluded from a shortlist with size m, whereas Y may be included and X excluded from a shortlist with size m'. If an employer wishes to maximise the expected qualifications of the hire, then a relative judgment of two candidates as to their eligibility to be placed on the shortlist is always a *holistic* judgment.

Here is an example of different candidate sets S and S'. Define candidate A as before—i.e. there is certainty that she is an 8. For candidate G, we have credence .50 that she is a 7, .40 that she is an 8 and .10 that she is a 9. Let A* and G* be slightly enhanced versions of A and G which will permit us to break ties. For example, suppose that, for small δ, A* is a certain $(8 + \delta)$. (Assume a more fine-grained or a continuous scale for this example.) For small ε, for G* we have credence $(.50 - \varepsilon)$ that she is a 7, .40 that she is an 8, and $(.10 + \varepsilon)$ that she is a 9. Now construct a shortlist of size 2 from the set $S = \{A, G, A^*\}$ and from the set $S' = \{A, G, G^*\}$.

The credence functions in this example are chosen so that we need a certain candidate and an uncertain candidate in shortlists of size 2 constructed from S and S' to maximise expectations. In S, there are two certain candidates and of course we pick the better one of the two for the shortlist. In S', there are two uncertain candidates and again we pick the better one of the two. Hence, the shortlists which provide the highest expectation for the hire are {G, A*} for S and {A, G*} for S'.

G and A are contained in both S and S'. Note that G is included and A is excluded from the shortlist constructed from S and A is included and G is excluded from the shortlist constructed from S'. Hence there can be no measure based on our credence functions which permits us to rank A and G such that this ranking will determine shortlist inclusion. G (rather than A) is included in a shortlist based on S and A (rather than G) is included in a shortlist based on S'. Shortlist inclusion depends not on the relative merits of the two candidates A and G, but on the complete set of candidates that contain A and G.

Here is an example with different shortlist sizes m and m'. Define A, D and E as before (see Table 1). Add candidate H who is probably a genius but might be a phoney—let's say we have credence .90 that H is a 10 and .10 that she is a 0. The set of candidates is $S = \{A, D, E, H\}$. Among all shortlists of size 2, {A, H} provides the highest expectation[5] for the qualifications of the hire whereas among all shortlists of size 3,

[5] Namely the expectation of H plus the chance of H being a phoney times the expectation of A, i.e. $(.90 \times 10) + (.10 \times 8) = 9.80$. Note that {D, H} offers a lower expectation, viz. $(.90 \times 10) + (.10 \times 7.9) = 9.79$.

{D, E, H} provides the highest expectation.[6] The intuitive argument is as follows. In each case, we take the risk of a phoney. With a shortlist of two, we need to shore up the risk of a phoney with a decent certain candidate. With a shortlist of three, we can take the gamble of two uncertain candidates who may be slightly better or slightly worse than just decent. Hence A is included and D and E are excluded from the shortlist of size 2 and vice versa for the shortlist of size 3. The moral is similar: There can be no measure based on our credence functions which permits us to simply rank A, D and E: Shortlist inclusion depends on shortlist size.

Hence, shortlist status for X and Y supervenes on the complete set of information and the nature of the task at hand, viz. the credence functions over all candidates and shortlist size. We cannot just look at our credence function over X's and Y's qualifications and define some measure which determines whether X has more or less of a claim to be on the shortlist than Y on grounds of their relative goodness. Building in optimism in the measurement cannot solve the problem.

One might raise the following objection. There is often substantial attrition while we are working through a shortlist—candidates withdraw or get hired by other companies. For this reason it would be wise to follow the recommendations of the Equality Act and stack the shortlist with candidates that have the greater expected qualifications. This is indeed correct if we are certain that there will be attrition from our shortlist of m candidates to a single candidate. If this is the case then the challenge is basically to come up with m candidates who would all be most fitting for a shortlist of size 1 and, indeed, for the limiting case of a shortlist of size 1, we should pick the candidate with the highest expected qualifications.

But this is a limiting case. In reality attrition will be less severe and less certain. The modelling just becomes more complicated. We will need to assess the attrition rate within a particular market and construct a probability distribution over the random variable with values $i = 0, \ldots, m$ candidates dropping from the shortlist of size m. On the basis of this distribution, we can then determine what an optimal shortlist would be. And indeed the greater the expected attrition, the more our list will come to resemble the list of candidates with the highest expected qualifications.

[6] Viz. the expectation of H plus the chance of H being a phoney times the expectation of the qualification of a hire with a shortlist {D, E}, i.e. $(.90 \times 10) + [.10 \times [((1 - (1 - .10)^2) \times 9) + ((1 - (.20^2 + (1\text{-}(1\text{-}.10)^2))) \times 8) + (.20^2 \times 7)]] = 9.815$. Note that {A, D, H} offers a lower expectation, viz. $(.90 \times 10) + [.10 \times [(.10 \times 9) + (.90 \times 8)]] = 9.81$.

Furthermore we are not helpless in the face of high attrition rates. Bracketing attrition, the expected qualifications of our hire are an increasing function of shortlist size. At some point the marginal gains from increasing the shortlist size are not worth the costs any more. So we can determine optimal shortlist size. In the face of high attrition rates we will simply increase the shortlist size so that after attrition we will still have a reasonable shortlist size, in order to retain the gains in the expected qualifications of the hire.

5. Objectives in hiring

We have assumed so far that a selector wishes to maximise the expectation of the qualifications of the hire. But she may have other objectives. For example, she may be risk averse and run an insurance strategy to make sure that there is at least one decent certain candidate in the shortlist. In this case she is maximising the expected qualification of the hire under the constraint that we should not drop below a particular level. She may be wary of the $.2^3 = .008$ chance of ending up with a 7 when D, E and F are shortlisted in our original example (Table 1). Instead she shortlists {A, D, E} securing at least an 8. The insurance strategy secures that the person hired will be at least an 8. That is correct. The cost of this insurance strategy is that the expectation of the hire drops from 8.26 to 8.19.[7] But the general point remains: The candidates with lower expected qualifications D and E were shortlisted over the candidates with higher expected qualifications B and C.

Alternatively, the selector may maximise the chance that the (or a) best candidate be hired—which is different from maximising the expected qualifications of the candidate who will be hired. In our original example (Table 1), the chance of picking a best candidate is greatest if we pick three uncertain candidates for a three person shortlist; it is lower if we pick one certain and two uncertain candidates; it is still lower if we pick two certain and one uncertain candidate; and it is lowest if we pick three certain candidates.[8] Hence if the selector's

[7] The chance of securing a 9 is $(1 - (1 - .10)^2) = .19$; there is no chance of ending up with a 7; and hence the chance of ending up with an 8 equals $(1 - .19) = .81$. We calculate the expectation: $(.19 \times 9) + (.81 \times 8) = 8.19$.

[8] If we pick three risky candidates {D, E, F}, then we failed to pick a best candidate if all risky candidates turn out to be 7s. Hence the chance of picking a best candidate is $1 - .2^3 = .992$. If we pick one certain and two risky candidates (e.g. {A, D, E}), then we failed to pick a best candidate if the non-picked risky candidate is a 9 and both of the picked risky candidate

objective is to maximise the chance of picking the best candidate then she should pick {D, E, F}. In this particular case maximising expected qualifications and maximising the chance of picking the best candidate make the same recommendations, though this will not always be the case.

This particular objective is highly contested. Frank Jackson 1991 notoriously takes issue with the general strategy of maximising the chance of doing a best action. We can run a case that is similar to Jackson's own counter example.[9] Suppose that there are 26 candidates A, …, Z. A and B are both solid—each have a .50 chance of being an 8 and .50 chance of being a 9. C, …, Z are extremely high risk—each have a .10 chance of being brilliant (i.e. a 10), but a .90 chance of being phoneys (i.e. a 0). Chances are independent. We can construct a shortlist of two candidates. Suppose we have a shortlist {A, B}. The chance of picking a best candidate equals the chance that C through Z are all phoneys, which is quite improbable. It is lower than the chance of picking a best candidate with a shortlist with one solid and one extremely risky candidate, which in turn is lower than the chance of picking a best candidate with a shortlist with two extremely risky candidates.[10] So this objective recommends that we put two extremely risky candidates on our shortlist, rather than two solid candidates or one solid and one extremely risky candidate. It is not a very reasonable objective in selection. It makes sense only if all a selector cares about is

are 7s or 8s, and so the chance of picking a best candidate is $1 - (.1 \times .9^2) = .919$. If we pick two certain candidates and one risky candidate (e.g. {A, B, D}), then we failed to pick a best candidate if at least one of the two non-picked uncertain candidates is a 9 and the picked risky candidate is not a 9, and so the chance of picking a best candidate is $1 - ((1 - .9^2) \times .9) = .829$. If we pick three certain candidates ({A, B, C}), then we picked a best candidate if all risky candidates are 7 or 8s, and so the chance of picking a best candidate is $.9^3 = .729$.

[9] Jackson 1991 (pp. 462–3) considers a case in which Jill, a medical doctor, has a drug A which provides relief for a patient's disease but does not cure it, whereas there is a 50% chance that drug B will cure and drug C will kill the patient and a 50% chance that C will cure and B will kill the patient. Jill should choose A though her chance of doing the best action by choosing A is zero, whereas by choosing either B or C it is .50. My example is similar except that I have built in a shortlisting stage and I have assumed independence between options, since dependence would be quite unrealistic for job candidates (as it is, frankly, for drugs as well).

[10] The chance of picking a best candidate with a shortlist {A, B} is the chance that all C, …, Z are phoneys, i.e. $.9^{24} = .08$. Suppose that we have a shortlist {A, C} (or any combination of a solid and a highly risky candidate). Then the chance of picking a best candidate is the chance that C is brilliant (i.e. .10) plus the chance that A is a better candidate of A and B and that C, …, Z are all phoneys (i.e. $(1 - .5^2) \times .9^{24} = .06$), and so .16. Suppose that we have shortlist {C, D} (or any two highly risky candidates). Then the chance of picking a best candidate is the chance that at least one of C and D are not phoneys, i.e. $1 - .9^2 = .19$.

not appointing a suboptimal candidate and is oblivious to how low she may fall if risky options do not turn out well.

There are many possibilities here. For example, we could model a selector who displays a mixture of the objectives of maximising expected qualifications, securing a threshold and maximising the chance of hiring a best candidate. What I have shown is that the general claim that we should not stack the shortlist with the candidates with the highest expectations is robust given different objectives the selector may be pursuing.

6. Is our shortlisting procedure fair?

Arguments for affirmative action often rest on realising social ideals such as promoting social causes of gender or ethnic equality. My argument does not. It rests on the libertarian ideal that the business of business is business. CEOs have an obligation to shareholders. They have an obligation to make sure that the business flourishes and to do so they should make sure that selectors make the best hires possible. Similarly, other types of organisations (universities, hospitals, governmental services, NGOs etc.) have an obligation to stakeholders imposing the same constraints on selectors. The procedure of shortlisting a candidate with lower expected qualifications may make eminent business sense.

But, one might ask, is it *fair* towards the person with higher expected qualifications to be overlooked for the shortlist?

There is a conception of *fairness* which requires that the best candidate get the job. What would be unfair on this conception is for a selector to knowingly appoint a less qualified person over a more qualified person, since the more qualified person deserves to get the job. This conception does not stand unchallenged, but let us suppose that it could indeed be grounds for complaint.

Even so, it cannot ground complaints from candidates A, B or C. All candidates A, B and C have to go on is that during the selection procedure they were ranked higher than D, E and F on the first moment (i.e. the expectation) of the selector's credence functions over the qualifications of the applicants. This is two steps removed from knowingly appointing a less qualified person over a more qualified person. First, the selector did not *know* A, B or C to be better candidates—for all she knew, D, E and F could have been 9s and she did indeed give some non-zero credence to them being 9s. And second, the preference was given during the selection procedure and not at the final hiring stage. The selector will point out that D, E and F were put on the shortlist during the procedure precisely because she was aiming to appoint a more qualified rather than a less qualified person for the position at the final stage, which is in line with the conception of fairness that our interlocutor appeals to.

7. Conclusion

I developed an argument about two-stage choice under uncertainty and applied it to shortlist construction in hiring, showing that the prescriptions concerning positive action of the 2010 UK Equality Act are untenable.

Let there be a set of options (say, bottles of wine, places to live etc.). We are allowed to pick one and we want to do well on grounds of our picking. There is uncertainty about the goodness of the options. We can focus on a subset of options for further investigation. How do we go about this? What I have shown is that we cannot construct a ranking that determines of each option whether it is more or less worthy of being included in this subset. We can do no better than make holistic judgments about where we should place our energy for further information gathering, depending on the complete set of options and subset size. And this holds whatever our objective is—to maximise expectations, with or without a threshold, or to maximise the chances of picking a best option.

If this is the case, then we shouldn't let the UK Equality Act 2010 tell us to construct a ranking over potential candidates on some goodness measure and stack the shortlist with the top m candidates from a set of applicants S. This is not how shortlist construction works.

Furthermore, whatever the objective, if our concern is to make a strong hire, then this tends to favour candidates at shortlisting stage who display greater variance in the selector's credence function. There is a range of reasons why UR-candidates are more likely to display greater variance. Hence my argument may favour the inclusion of UR-candidates in a shortlist even if we have lower expectations for them. Furthermore, the argument also holds if we make temporary hires with the aim to retain the best candidates, since the problem is structurally analogous.

I have provided a strict business reason for shortlisting and hiring practices that will often be tantamount to favouring UR-candidates with lower expected qualifications. Furthermore, non-UR-candidates with higher expected qualifications who were overlooked in this way have no grounds to complain that they were treated unfairly.[11]

References

Banaji, M. and A. G. Greenwald 2013: *Blind spot: Hidden biases of good people*. New York: Delacourt.

Bertrand, Marianne and Sendhil Mullainathan 2004: 'Are Emily and Greg more employable than Lakisha and Jamal? A field experiment on labor market discrimination'. *American Economic Review*, 94, pp. 991–1013.

Buchak, L. 2013: *Risk and Rationality*. Oxford: Oxford University Press.

Carlsson, Magnus and Rooth Dan-Olof 2007: 'Evidence of ethnic discrimination in the Swedish labor market using experimental data'. *European Association of Labour Economists 18th annual conference CERGE-EI, Prague, Czech Republic 21-23 September 2006*, 14, pp. 716–29.

Commons Debates. Public Bill Committee. Equality Bill. 30 June 2009: Column 612. http://www.publications.parliament.uk/pa/cm200809/cmpublic/equality/090630/pm/90630s10.htm.

Corrice, April 2009: 'Unconscious bias in faculty and leadership recruitment: A literature review'. *Analysis in Brief. Association of American Medical Colleges*, 9(2), pp. 1–2.

Danaher, K and C.S. Crandall 2008: 'Stereotype threat in applied settings re-examined'. *Journal of Applied Social Psychology*, 38, pp. 1639–55.

Equality Act 2010, Section 159. <http://www.legislation.gov.uk/ukpga/2010/15/section/159>, accessed 9 June 2015.

Equality Act 2010 – *What Do I Need to Know? A Quick Start Guide to Using Positive Action in Recruitment and Promotion*. Government Equalities Office. Jan 2011 <https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/85014/positive-action-recruitment.pdf>, accessed 9 June 2015.

[11] Acknowledgements: I am grateful to Jason Alexander, Anna Bartsch, Richard Bradley, Veselin Karadotchev, Alexander Kirshner, Alex Marcoci, Johanna Thoma, Jane von Rabenau and Alex Voorhoeve for discussion and comments. My research was partly supported by a Laurance S. Rockefeller Visiting fellowship in the University Center for Human Values in Princeton University.

Goldin, Claudia and Cecilia Rouse 2000: ‘Orchestrating Impartiality: The Impact of “Blind” Auditions on Female Musicians’. *American Economic Review*, 90, pp. 715–41.
Hansard – House of Lords Debates. 9 Feb 2010, Column 659 <http://www.publications.parliament.uk/pa/ld200910/ldhansrd/text/100209-0009.htm>, accessed 9 June 2015.
Hansard – House of Lords Debates. 2 Mar 2010, Column 1421 <http://www.publications.parliament.uk/pa/ld200910/ldhansrd/text/100302-0015.htm>, accessed 9 June 2015.
Hansard – House of Commons Public Bill Committee on the Equality Bill. 30 June 2009, Column 605. <http://www.publications.parliament.uk/pa/cm200809/cmpublic/equality/090630/pm/90630s09.htm>, accessed 9 June 2015.
Hansard – House of Commons Public Bill Committee on the Equality Bill. 30 June 2009, Column 612. <http://www.publications.parliament.uk/pa/cm200809/cmpublic/equality/090630/pm/90630s10.htm>, accessed 9 June 2015.
Hoff, Karla Ruth and Pandey Priyanka 2004: *Belief systems and durable inequalities: an experimental investigation of Indian caste*. World Bank Publications, Vol. 3351.
Jackson, Frank 1991: ‘Decision-Theoretic Consequentialism and the Nearest and Dearest Objection’. *Ethics*, 101, pp. 461–82.
King, E. B., S. A. Mendoza, J. M. Madera, M. R. Hebl and J. L. Knight 2006: ‘What’s in a Name? A Multiracial Investigation of the Role of Occupational Stereotypes in Selection Decisions’. *Journal of Applied Social Psychology*, 36, pp. 1145–59.
Rooth, Dan-Olof 2010: ‘Automatic Associations and Discrimination in Hiring: Real World Evidence’. *Labor Economics*, 17, pp. 523–34.
Quiggin, John 1993: *Generalized Expected Utility Theory. The Rank-Dependent Model*. Boston: Kluwer.
Steele, Claude M., and Joshua Aronson 1995: ‘Stereotype threat and the intellectual test performance of African Americans’. *Journal of personality and social psychology*, 69, pp. 797–811.
Steinpreis, Rhea E., Katie Anders and Dawn Ritzke 1999: ‘The impact of gender on the review of the curricula vitae of job applicants and tenure candidates: a national empirical study’. *Sex Roles*, 41, pp. 509–28.
Williams, Bernard 1973: ‘Ethical Consistency’. In *Problems of the Self: Philosophical Papers* 1956–72. Cambridge: Cambridge University Press. An earlier version of the essay appeared in *Proceedings of the Aristotelian Society*, 39, 1965, pp. 103–24.

