Objective Assessment of University Discrimination

Prof. P. A. Robinson, School of Physics, University of Sydney.

robinson@physics.usyd.edu.au

16 November 2020

1. INTRODUCTION

On 25 April 2018 *The Australian* published an article that surveyed the diversity of Australian Vice Chancellors 43 years after the passage of the Sex Discrimination Act (SDA) and Racial Discrimination Act (RDA) in 1975.[1] Of 40 VCs, it noted that 13 were female and one was non-European in ancestry. This was taken to indicate a lack of diversity in this group, but no objective analysis was done to determine just how unlikely such an outcome would be and to what extent it indicates bias in VC selection processes.

The purpose of this document is to show how standard statistical methods can assist discussions of antidiscrimination by providing objective measures of the likelihood of observed appointment outcomes, such as the one above, and estimates of any selection biases that may underlie such outcomes. The aim is to provide objective measures that can be used to illuminate discussions of policy and strategy, and to measure progress. The questions addressed are:

- If certain numbers of people from two mutually exclusive groups (e.g., women and others, Europeans and non-Europeans subject to some definition of European) are observed to have been appointed to a set of positions, what is the likelihood of this outcome?[2]
- What are the margins of error beyond which bias is likely to exist?
- What is the likely level of bias of the selection process?
- What is the likelihood that the selection process is unbiased?

The above issues are illustrated with the above example from *The Australian* and others at senior levels of The University of Sydney. Resulting observations and questions are also briefly noted. To keep the discussion as accessible as possible, the mathematical details are reserved for the accompanying Appendix and have been implemented in a simple-to-use computer program that is intended to be made publicly available. In the present instance, detailed discussion is limited to sexism and racism, but the same approach can be applied to assess discrimination of other types and in other workplaces, as discussed in Sec. 3

2. VC SELECTION

2.1 Sexism in VC selection

How likely was the observed outcome?

We begin by considering the VC example from *The Australian* from the point of view of sexism. The first essential idea is that in an unbiased selection process the sex of appointees should be irrelevant – i.e., it should not affect the outcome.

[1] "The VC club: one third women but light on diversity", S. Powell, *The Australian*, 25/4/18.

[2] Cases that are not mutually exclusive can also be handled but are mathematically more complicated. Nothing is implied here regarding gender identity, sexual orientation, or the broad spectrum of related important issues.

If women constitute 50% of the qualified pool of applicants for Vice Chancellorships, the question of how many are appointed out of 40 VCs is exactly the same as how many heads we expect out of 40 coin tosses – in an unbiased coin neither heads nor tails should be favored. Hence, when 40 VCs are chosen, we expect on average 20 women to be appointed, just as we expect an average of 20 heads out of 40 tosses of an unbiased coin. However, just as with coin tosses, a perfectly fair process can result in some random deviations from this number. So, the question is whether the actual outcome of 13 female VCs implies bias and, if so, how much.

Figure 1(a) shows the expected sex distribution of the outcomes of 40 VC appointments, on the assumption that the pool of qualified applicants is 50% female. (This curve is calculated from the binomial distribution discussed in the Appendix.) We see a strong peak around the ***expected value*** of 20 women, but a significant likelihood of somewhat more or less. There is even a tiny probability that *all* or *none* of the applicants will be female in a totally fair process, but around two thirds of the probability lies in the range 17 – 23 VCs, and 95% of the probability lies between 14 and 26 – this is termed the ***95% confidence interval*** (CI). To determine how likely it is that a fair process would result in 13 or fewer female appointments, we must add up the probabilities of all of these cases – the shaded area in the figure. This gives a chance of 1.9% that 13 or fewer women would be appointed in an *unbiased* selection process – technically, this is termed the ***cumulative distribution function*** (cdf) with cdf(13) = 0.019 in this case.

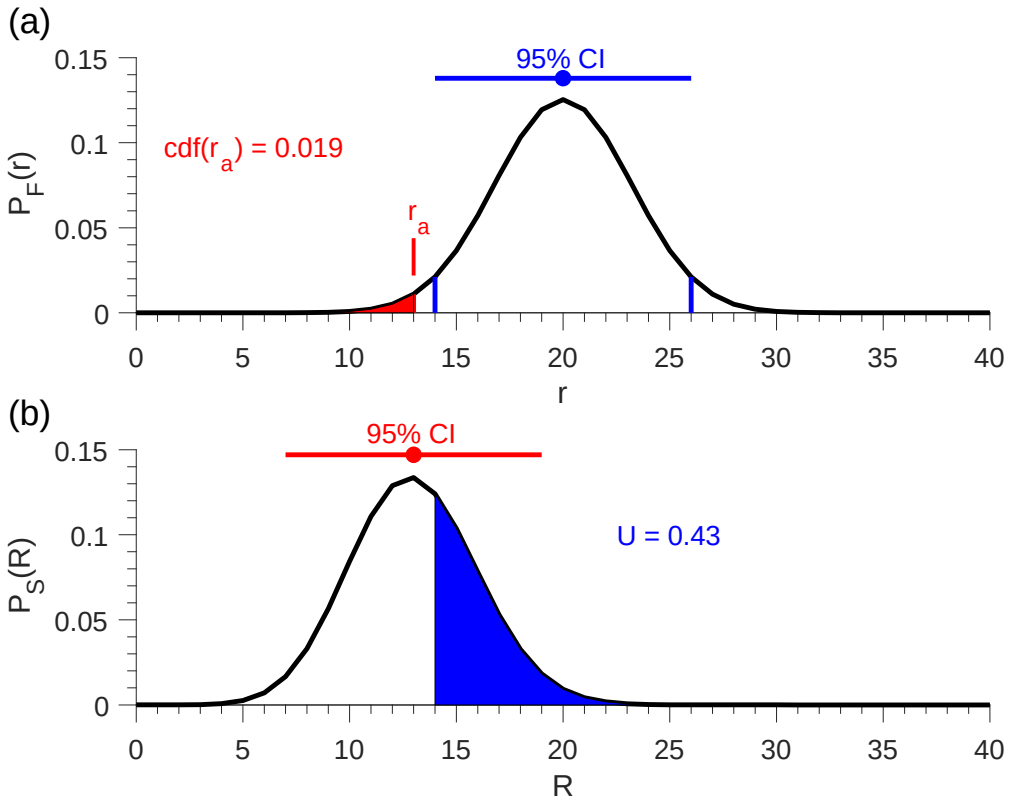

Figure 1. Sexism in Australian VC selection. (a) Probability of different numbers of female appointees out of 40 VCs, assuming women represent 50% of the pool of qualified applicants and the process is unbiased. The actual number (r_a = 13) is indicated by the vertical red line, and the red-shaded region indicates cases where 13 or fewer women are appointed. An unbiased process would yield a result within 95% CI shown by the blue bar and vertical blue lines 95% of the time; the dot indicates the expected number of women appointed, $\bar{r} = nf = 20$. **(b)** Inferred distribution of the numbers *R* of women appointed if the selectors made multiple selections of 40 VCs given that they actually chose 13, also showing the 95% CI with a red bar and dot at r_a. Blue shading indicates the overlap with the 95% CI from Fig. 1(a).

Before proceeding, it is worth stressing a few key points:

i. The proportion of qualified applicants from each of the two groups (women and others in this case) must be estimated. In some cases (e.g., Physiotherapy or Engineering) this proportion will not be 50-50 because of factors that affect the pool of applicants before the selection process takes place (including any search phase) and cannot be attributed to that process.
ii. The fact that there may be some debate over the precise value of the proportions of qualified applicants in point (i) does not mean that nothing can be said. Rather, estimates can be made and the consequences of different assumed values can be explored **objectively**. Any deviations from the proportions in the population at large must be explained and justified before use in the analysis.
iii. The likelihood of a given outcome (1.9% in the present example) is an **objective** probability based solely on the explicit assumption of an unbiased process and estimated proportions of applicants in the qualified pool. It is **not** an estimate of the likelihood that the process is biased – an unbiased process will *occasionally* give highly skewed results. Nor does it answer the question of what to do about bias. We turn to these issues below.

How biased was the process?

One estimate of the disparity in how two groups are viewed by selectors is the **preference ratio**. In the VC case, $r_a = 13$ women were appointed out of an expected $\bar{r} = 20$, which gives a relative proportion of 13/20 = 0.65; the corresponding relative proportion for Others is (40-13)/20 = 1.35, which is approximately twice as high – i.e., the preference ratio is $B = 1.35/0.65$, or about 2.1. Roughly speaking, this implies that VC selectors are roughly twice as likely to appoint a qualified Other (generally a man) as an equally qualified woman when comparing them one-on-one. The preference ratio is a useful indicator, especially if it is very large or very small.[3] However, because we don't expect exactly 20 appointments of women in every set of 40, even in a perfectly fair system, we need a more systematic approach to estimating selection bias.

It is possible to estimate the likely selection bias from the appointments that have actually been made, by using Bayesian statistics (see Appendix). The most likely bias is what is actually observed – i.e., a 2:1 bias against women in the present case – but the probability of other values can be estimated. Figure 1(b) shows the probability that the VC selectors would appoint various numbers R of qualified women if they could make a fresh round of 40 appointments, given their actual appointment of 13 qualified women. An approximate bell curve is seen, with significant tails extending above and below 13 appointments.

The question of selector bias, *whether conscious or unconscious*, can be further addressed by calculating the chance that they would appoint a number of women in the expected 95% CI of 14-26 , given their actual appointment [see Fig. 1(b)]. This gives the ***probability of unbiased selection*** to be $U = 43\%$ or, equivalently, 0.43; i.e., the probability that the process is not sexist by the definition that a fraction U of expected appointment numbers fall within the 95% CI of a fair process. A distribution that matched the expected one exactly would give $U = 95\%$, but an exact match is not normally expected, so somewhat lower values are most likely, even in a perfectly unbiased process. However, numbers much less than 1 do imply bias and $U < 0.1$ should be cause for serious concern, while $U < 0.01$ should provoke urgent action.

[3] A value below 1.0 would imply a bias *in favor* of the group in question (women here), rather than in favor of the Other group.

Summary

The above outcomes imply that women remain underrepresented among VCs relative to their level in the pool of qualified applicants, with a representation just outside the 95% confidence interval. However, because there is a 43% likelihood that the present process would produce a result within that interval, it is probably not highly biased. There is still some way to go to full equality, but the progress since passage of the SDA is encouraging.

2.2 Racism in VC selection.

We now turn to the other issue raised in *The Australian* – racism. To answer the same questions as above we must first estimate the proportion of non-Europeans (by ancestry) in the pool of qualified applicants. Although VC selection processes purport to seek the best applicants from around the globe, the proportion of qualified non-European applicants will not be as high as their proportion (circa 80%) in the world population because of many factors such as unequal access to educational opportunities, language barriers, and low median ages that reduce the fraction of people who could have sufficient experience to be a VC, but it cannot be lower than the proportion in the Australian population (circa 25%)[4]. For simplicity and the sake of argument here, let us suppose that the proportion in the pool of qualified applicants is 40%.

How likely was the observed outcome?

Figure 2(a) shows the probability of different numbers of non-European VCs among 40 appointments if the process is unbiased. The most likely value is $\bar{r}$= 16 with a 95% chance that an unbiased process will deliver between 10 and 22. The actual number is only r_a = 1 and the chance of 0 or 1 being appointed in a nondiscriminatory process is cdf(1) = 3.7 x 10^{-8} (i.e., 1 chance in 27 million).

[4] Australian Human Rights Commission, *Leading for Change: A blueprint for cultural diversity and inclusive leadership revisited* (2018).

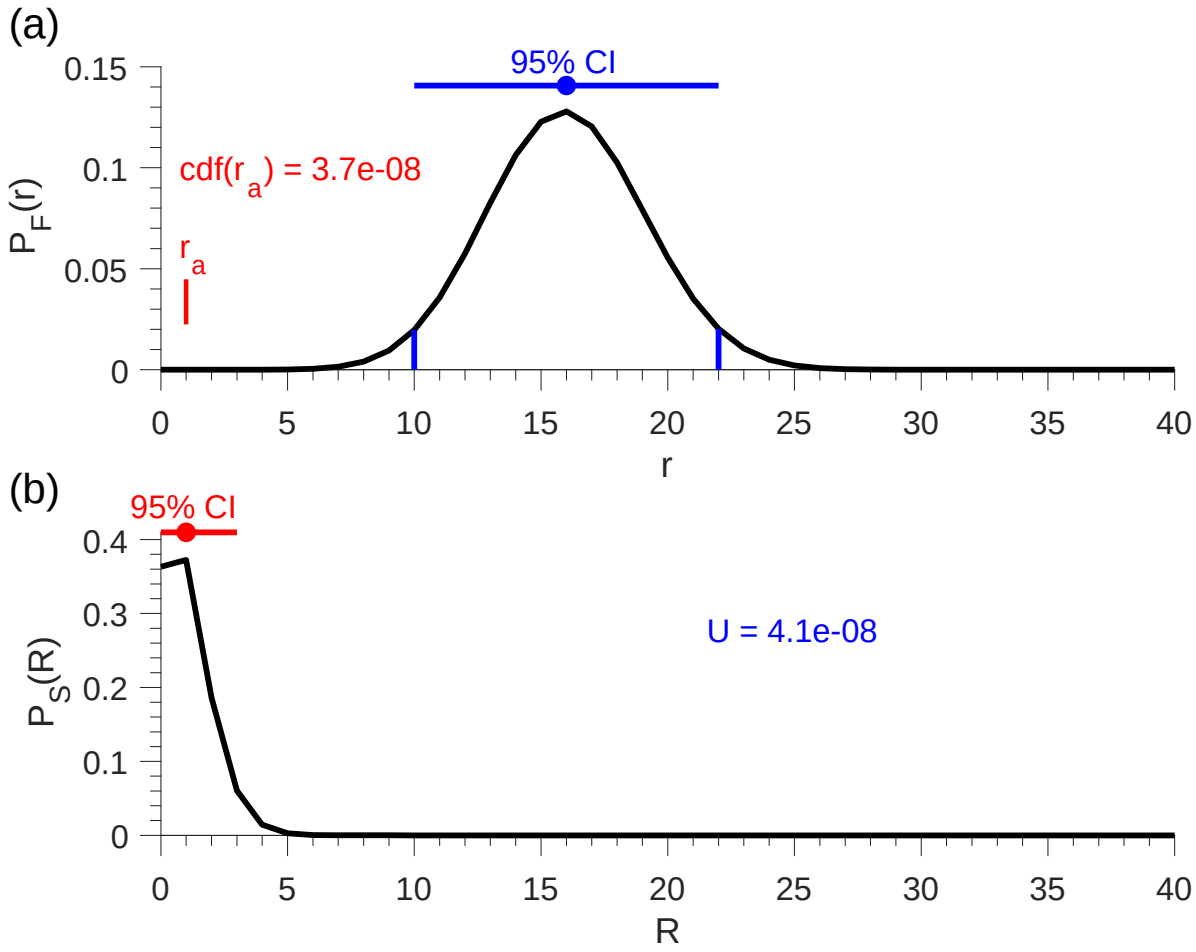

Figure 2. Racism in Australian VC selection. (a) Probability of different numbers of non-European appointees out of 40 VCs, assuming 40% representation in the pool of qualified applicants and that the process is unbiased. The actual number (r_a = 1) is indicated by the vertical red line and the region to its left indicates cases where 0 or 1 non-Europeans are appointed. An unbiased process would yield a result in the 95% CI indicated 95% of the time. **(b)** Inferred distribution of numbers of non-Europeans and Europeans who would be appointed by the selectors if they made multiple selections of 40 VCs. Shading of the overlap with the 95% CI from Fig. 2(a) is too small to be visible at this resolution.

How biased was the process?

With only r_a = 1 non-European appointment out of 40% of the pool of qualified applicants ratio 1/16 = 0.0625 and 39 European appointments out of 60% of the pool ratio 39/24 = 1.625 we see that selectors are B = 1.625/ 0.0625 = 26 times more likely to appoint a qualified individual who is European than an equally qualified individual who is non-European when compared one-on-one. Figure 2b formalizes this by showing the distribution of results expected if this selection process were repeated, given the observed outcome. There is almost no chance that the selectors would appoint more than 3 non-Europeans if they were to make a fresh set of 40 appointments and the probability the process is unbiased is $U = 4.1 \times 10^{-8}$ or 1 in 24 million.

Summary

From these results (summarized in Table 1) there seems to be no alternative but to conclude that the selection processes for Australian VCs are extremely racist and that there has been negligible progress since the passage of the RDA. In fact, the situation is worse than indicated by these figures – only a few per cent of the qualified applicant pool, and roughly 50% of the Australian population, is of Anglo-Celtic/Anglo-Saxon ethnicity,[5] but this group dominates the ranks of VCs, so other Europeans are also discriminated against.

3. UNIVERSITY OF SYDNEY EXECUTIVE, DEANS, AND SENATE

How does the University of Sydney compare with Australian VC selection, especially when filling positions at the highest levels of the organization, which sets the example for conduct by other levels and individuals? Has the University taken effective steps to implement the SDA and RDA? This analysis was done as of August 2020.

We now carry out the same steps as above for the University Executive (UE, including Deans here, 26 people), the Senate (15 people), and for the combination of UE and Senate – 39 people in total because there is an overlap of 2 in the membership of the two bodies. The UE has 12 women and two non-Europeans while the Senate has 6 women and 1 non-European.

As above, we must estimate the proportions of women and non-Europeans in the qualified pool of appointees to the above bodies. We again assume a 50-50 sex ratio; however, because the Senate is locally elected and the non-Academic members of the UE are not selected by such a wide international search process as VCs, we use an estimate of 35% for the fraction of non-Europeans in the pool of qualified applicants – i.e., between the level of 40% for VCs and the domestic proportion of 25% (noting also that the latter proportion in Sydney is higher than 25%).

3.1 Sexism in UE and Senate selection

How likely was the observed outcome?

We begin by considering the combined UE plus Senate (39 people). In this group there are 18 women.

Figure 3(a) shows the expected sex distribution of the outcomes of 39 appointments (including elections in the case of some Senate members), on the assumption that the pool of qualified applicants is f = 50% female. We see a strong peak around the ***expected value*** of $\bar{r}$= 19.5 women, and 95% CI of roughly 14-25. We find a cdf(18) = 37% chance that 18 or fewer women would be appointed in an unbiased process.

[5] *Ibid.*

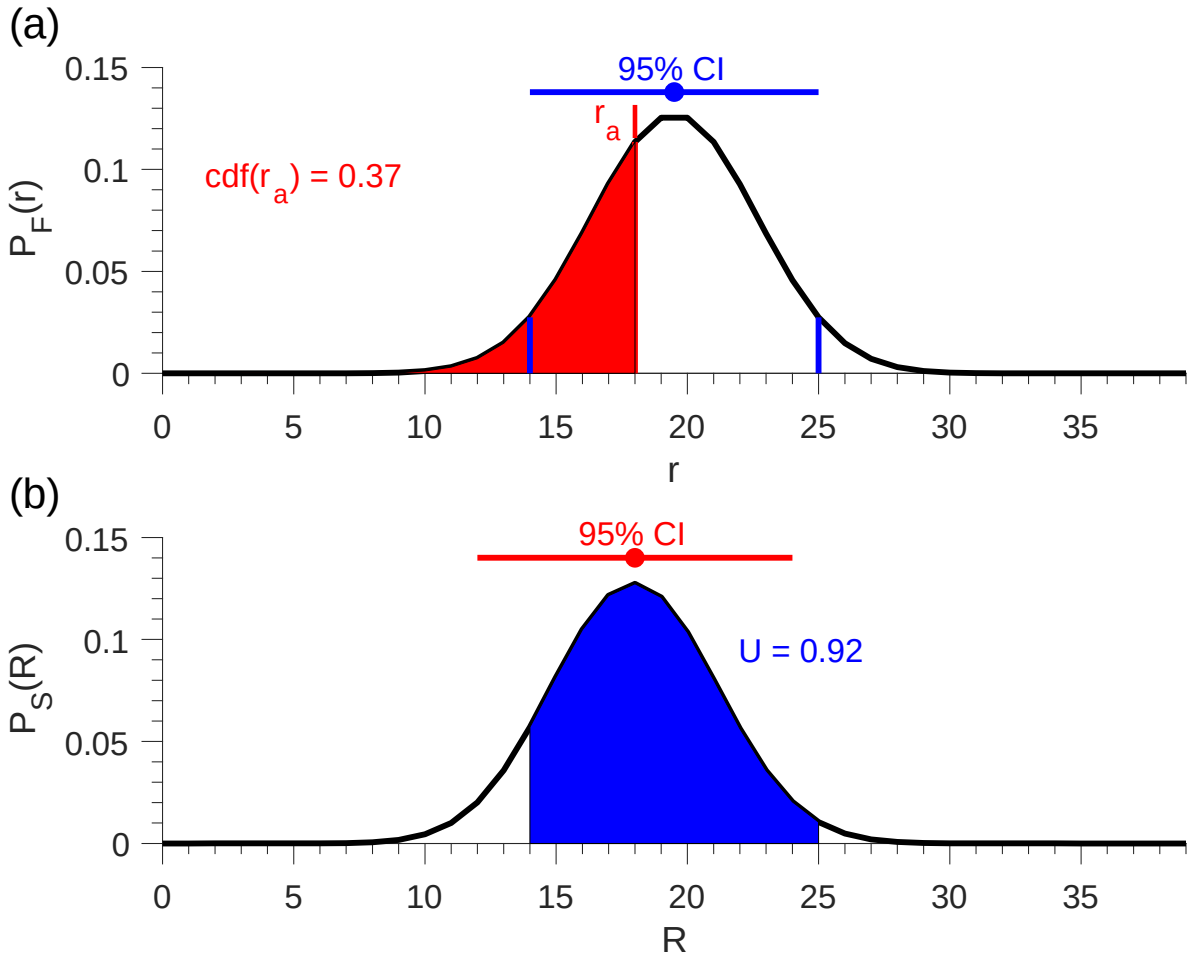

Figure 3. Sexism in University Executive and Senate selection. (a) Probability of different numbers of female appointees out of 39 members, assuming women constitute 50% of the pool of qualified applicants and that the process is unbiased. The actual number (r_a = 18) is indicated by the vertical line, and the shaded region indicates cases where 18 or fewer women are appointed, giving cdf(r_a) = 0.37. An unbiased process would yield a result in the central region indicated 95% of the time. **(b)** Inferred distribution of the number of women who would be appointed if the selectors chose multiple sets of 39 appointees. Shading indicates the overlap with the 95% CI from Fig. 3(a).

How biased was the process?

The data in this case indicate a preference ratio of B = 1.2, which does not indicate strong bias. Similarly, the probability that the process is unbiased is U = 92%, which confirms this conclusion. If the situation is truly unbiased, one would expect B to fluctuate slightly above and below 1 over time.

Summary

The above outcomes imply the representation of women is close to its level in the pool of qualified applicants, with a number well within the 95% confidence interval. There is 92% likelihood that the present process would produce a result within that interval, so it is probably not strongly biased. There is still some way to go to full

equality, but the progress since passage of the SDA shows almost complete success at this level in reaching parity in the combined UE and Senate. Of course, there are many Faculties and other areas of the University in which this is not the case, and the present methods can be used to examine them in a similar way to estimate the effects of external factors and internal processes.

These results are summarized in Table 1, along with those for the UE and Senate separately. Similar conclusions apply for these two groupings.

3.2 Racism in UE and Senate selection

One additional factor must be taken into account in assessing outcomes with regard to racism. As noted early in this document, statistical methods used assume that the characteristics in question are not relevant to the appointments made. In the case of the University executive, the DVC for Indigenous Strategy and Services (DVCISS) is expected to have Indigenous background as a prerequisite for being able to successfully carry out the duties of this position. Hence, this post must be set aside from the statistical analysis, leaving 25 positions in the UE and 38 in the UE-plus-Senate to be analyzed; there is no change to the Senate-only grouping.

How likely was the observed outcome?

We begin by considering the combined UE plus Senate (38 people aside from the DVCISS). In this group there are r_a = 2 non-Europeans.

Figure 4(a) shows the expected distribution of the outcomes of 38 appointments (including elections), on the assumption that the pool of qualified applicants is 35% non-European. We see a strong peak around the expected value of $\bar{r}$ = 13.3, and 95% CI of roughly 8-19. We find a cdf(2) = 1.7×10^{-5} or a one in 57 000 chance that 2 or fewer non-Europeans would be appointed in an unbiased process.

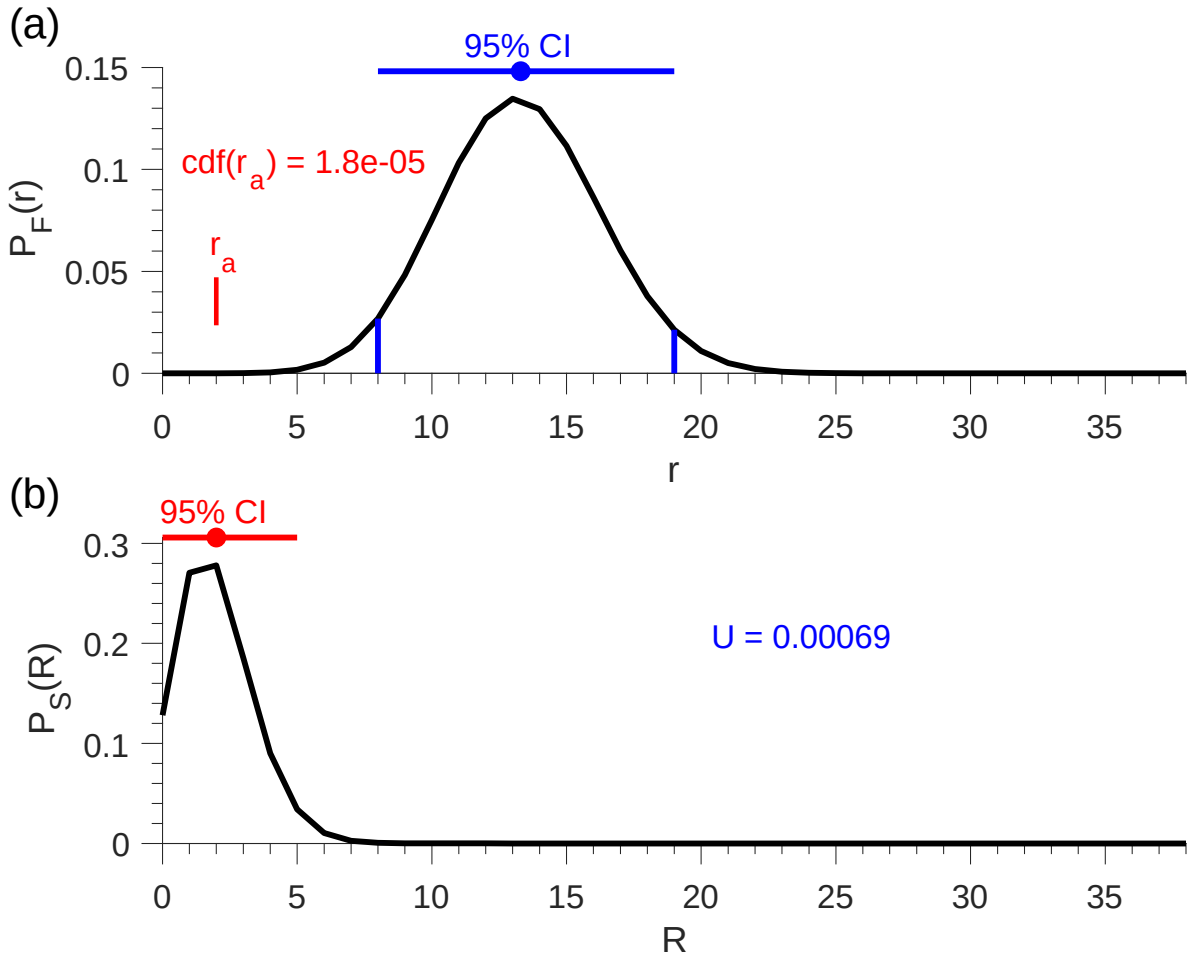

Figure 4. Racism in University Executive and Senate selection. (a) Probability of different numbers of non-European appointees out of 38 members, excluding the DVC-ISS, assuming 35% representation in the pool of qualified applicants and that the process is unbiased. The actual number ($r_a = 2$) is indicated by the vertical red line, and the shaded region indicates cases where 2 or fewer non-Europeans are appointed. An unbiased process would yield a result in the central region indicated 95% of the time. **(b)** Inferred distribution of numbers of non-Europeans the selectors would appoint if they made multiple sets of 38 appointments. Shading to indicate the overlap with the 95% CI from Fig. 4(a) is invisible on this scale near *R = 8.*

How biased was the process?

The data in this case indicate a preference ratio of $B = 9.7$, which indicates that the selectors are nearly 10 times more likely to appoint/elect a European than an equally qualified non-European when considered one-on-one. Similarly, the probability that the process is unbiased is $U = 6.8 \times 10^{-4}$ (i.e., roughly 1 in 1500), which strongly confirms the existence of bias.

Summary

The data from the selection processes for the combined UE and Senate indicate that these involve extreme racism (although not quite as extreme as in VC selection). There has been little progress since the passage of the RDA in 1975.

Similar results are found when one considers the UE and Senate separately, as summarized in Table 1.

4. DISCUSSION

4.1 Sexism

The results of the previous sections are summarized in Table 1. Overall, we see that women remain somewhat underrepresented at VC and University levels but that substantial progress has been made since the passage of the SDA. The University's selection processes do better in this respect than VC selection processes nationwide.

Appointment	*n*	Group	*f (%)*	$\bar{r}$ *=fn*	*CI*	r_a	cdf(r_a)	*B*	f_a *(%)*	*U*
VCs	40	W	50	20	14-26	13	0.019	2.1	32.5	0.43
	40	NE	40	16	10-22	1	3.7×10^{-8}	26	2.5	4.1×10^{-8}
UE	26	W	50	13	8-18	12	0.42	1.2	46	0.96
	25	NE	35	8.8	5-13	1	2.8×10^{-4}	13	4	2.8×10^{-3}
Senate	15	W	50	7.5	4-11	6	0.30	1.5	40	0.91
	15	NE	35	5.2	2-8	1	0.015	7.4	6.7	0.26
UE+Senate	39	W	50	19.5	14-25	18	0.37	1.2	46	0.92
	38	NE	35	13.3	8-19	2	1.8×10^{-5}	9.7	5.3	6.9×10^{-4}

Table 1. Results for the examples in the text: VCs, USyd Executive, USyd Senate, combined UE and Senate, and Physics Professoriate, as shown in the left column. The following columns show the group women (W) or non-Europeans (NE), the assumed percentage *f* of that group in the qualified pool, the number *n* of appointments, the expected number, the 95% confidence interval (*CI*), the actual number r_a appointed, the probability cdf(r_a) of r_a or fewer appointments in an unbiased system, the preference ratio *B,* the actual percentage f_a of appointable group members as judged by the selectors, and the probability *U* that the selection was unbiased.

4.2 Racism

The results demonstrate what could only be termed extreme racism at all levels considered, with massive preference for Europeans (and ethnically Anglo-Celtic persons in particular, as can be demonstrated in the same way) when selectors consider equally qualified candidates one-on-one. The University processes are somewhat better than those at VC level nationwide, but there is still a negligible probability that they are unbiased. Interestingly, the least biased case is that of election to the University Senate,[6] whereas internal appointment processes to the UE are more likely to be biased.

[6] The Senate *U = 0.26* is only this high because of the presence of a single non-European member among 15 and there being a moderate chance that a fresh round of appointments and elections would result in 2, which would be at the bottom of the 95% CI.

4.3 Discussion

Sexism

The previous examples demonstrate that, starting from a poor situation in 1975, sex discrimination has been greatly reduced at the national and institutional level. Much of this appears to have been the result of head-on antisexism awareness-raising at the University level, followed by a raft of measures to remove barriers and to ensure that "closet" sexism did not persist in appointment processes – e.g., by requiring that women be represented on appointment panels.

Many areas within the University still exhibit significant disparities based on sex and other factors. The present method can be used to analyze such areas to assess these imbalances objectively and determine how much of any disparity is due to selection processes. This requires only three pieces of information:

- An estimate of the fraction of qualified applicants who belong to a given target group (e.g., women);
- The total number of appointments; and
- The number of appointments from the target group.

Inserting these 3 numbers into the formulas in the Appendix or the computer program that implements them**,** produces figures like those in this document, plus numbers that parallel one line of Table 1. Use of such an approach removes most subjectivity from the discussion, makes assumptions and approximations explicit, and provides objective measures of bias. Moreover, it enables progress to be tracked and obviates the concern that "*what is not measured is not managed*." The same procedure can be used to investigate other forms of discrimination.

Notably, this can also be done for enrolments and at various stages of progression through to postgraduate education, appointment to the academic staff, and promotion. The key restriction is that cohorts being compared must be sufficiently large to give meaningful statistical results – typically 10-20 at a minimum with at least a few appointments expected from the target group in an unbiased process. University subdivisions and areas to try this on would include:

- Faculties and Schools
- Senior Administrative positions
- Human Resources, particularly its senior ranks
- Long- and short-lists for appointments
- Promotions
- Postgraduate scholarship offers

Racism

Although the RDA became law in 1975 there has been little progress on racism at the Australian VC or University levels. At the University level there has been no head-on antiracism effort at remotely near the same intensity as antisexism; rather, the worthy but oblique goal of encouraging cultural diversity has been pursued. Strangely, despite the effects of the intensely racist processes quantified above being obvious to even a casual observer of a gallery of VCs or members of the UE, no requirement of racial diversity has been imposed on selection committees, where all-Anglo-Celtic examples are still considered perfectly normal and acceptable.

The reasons for the relative lack of action on antiracism may well include:

- University management in 1975 was primarily composed of European men who had European female relatives and might thus have been sympathetic to removing discrimination against them. This was much less likely with regard to non-Europeans, a problem that persists to this day although the management is now roughly balanced between the sexes.
- Then and now, it is human nature not to want to admit that one owes one's position in part to others of equal or greater ability having been excluded by the selection process.
- Those who have benefited from the status quo have little incentive to change it. Encouragingly, the overwhelmingly male management of several decades ago *did* make such changes to implement the SDA, so perhaps it is not too much to hope that the overwhelmingly European management today might act similarly to implement the requirements of the RDA.
- Ignorant biases once used to exclude women (e.g., that "their" thought processes were supposedly not suitable to high-level management, or that "they" did not have the same presumed-essential life experiences as men) persist in analogous forms against non-Europeans (e.g., that "they" might not be sufficiently conditioned into the local ways of doing things, or "they" might have less common accents). In both cases, this simply reduces the available range of expertise, viewpoints, and ideas, and can easily lead to inflexible groupthink and missed opportunities.

Overall, there is an urgent need to quantify and act on racism and other biases of all kinds in the best interests of the University system, which has the benefit that progress can be quantified and advertised. At the University level, the UE needs to lead the way by rectifying its own composition (which could be achieved within a few years at this level, given the number of qualified applicants from around the world, the frequency of appointments, and the typical relatively short duration of postings) and by instituting the same types of University-wide measures as are used against sexism:

- Explicit recognition of the continuing existence and effects of racism, unconscious *and conscious.*
- Admission that the implementation of the RDA has not been given the same attention as that of the SDA.
- Commencing explicit and forceful antiracism, with recognition that encouraging cultural diversity is not enough by itself – after 45 years this approach has made little progress – and still even the University's Mosaic Network top webpage implies that cultural diversity is to do with people who "come to the University from overseas", thus apparently overlooking that locals have a variety of cultures; it also makes no reference to antiracism.
- Recognition of hidden racism – e.g., the widespread preference for graduates of well-known (to the mostly Anglo-Celtic selectors) Universities in wealthy countries, on "quality" grounds that are often spurious (e.g., use of University rankings that depend more on per capita GDP and cultural biases of the rankers than individual-candidate quality); or focusing searches only on a few countries and/or via "old-boy" networks. These factors also bias selection for international scholarships.
- Recognition that there is much overt racism at the University. This should be probed via pointed questions in the anonymous e-surveys that are distributed from time to time, because the fear of reprisals in response to any direct comment is widespread.
- Requirement to document racial diversity in selection processes, just as the male-female ratio is documented, plus a requirement to shortlist realistic candidates of both groups for many positions, or to widen the search process until this is possible.
- Take responsibility at UE, Senate, and HR levels for the failure to make comparable progress on racism as on sexism, identify why, and make changes.
- Implement antiracism measures without delay, starting **immediately** with those that parallel successful antisexism procedures and thus require no additional justification or investigation. After 45 years there can be no legitimate reason to further delay effective action. On the positive side, even though very late, such action would significantly enhance and protect the University's reputation.

General

Overall, these two examples illustrate how much or how little can be achieved, depending on the practical commitment (or lack thereof) to removing specific types of discrimination at the VC-selection and University levels. While considerable progress has been made against sexism one might ask why antiracism has not been given equal priority, and who takes responsibility for this.

The methods discussed here provide specific tools to quantify outcomes and infer discrimination so that biases can be targeted for removal – especially useful are the preference ratio *B* and the probability *U* that a given selection process is unbiased.

Turning to other types of discrimination – based on sexual orientation, disability, gender identity, and other factors, the same methods can be used. The technical limitation, mentioned above, is that the expected and actual numbers of people in the target group need to be large enough to get adequate statistics. Thus, while sex and race discrimination can be assessed at the level of Schools and above, some of the other categories can probably only be assessed at the Faculty level and above. This is not an excuse to avoid action, and the above considerations also apply in these cases.

The author is happy to discuss and/or explain the methods used and welcomes feedback and the pointing out of any errors.

Acknowledgments

Junior staff member(s) and/or student(s) assisted with the code implementation and graphics but do(es) not wish to be identified.