Women, Men, and Pólya Urns

Underrepresentation at equal talent in the absence of discrimination*

Laura Caron, Alessandra Casella and Victoria Mooers

February 28, 2026

Preliminary and Incomplete

Abstract

In a world where two groups have equal distributions of talent, where candidates are objectively and accurately evaluated, and no discrimination occurs, the underrepresentation of one group in prestigious positions is nonetheless highly sticky. The point applies broadly, but we study it using women and men as our running example. Suppose the sample of women candidates for a selective position is smaller. Then, at equal distribution of talent, the most qualified candidate is likely to be a man, mirroring the larger number of men candidates. If future samples of candidates respond as expected–increasing if the selection came from the sample, increasing less if it did not–men's higher probability of success will persist. We capture this process with a well-known stochastic dynamic process: the Pólya urn. The model delivers a surprising number of useful empirical benchmarks: long-term persistence of unequal outcomes, and dispersion and non-monotonicities in the representation of women across employers and time. The streamlined model allows us to compare different policy interventions. A simple app (https://caron.shinyapps.io/Women-Men-Polya-Urns/) encourages readers to run their own experiments. Two robust results are that all affirmative action interventions, no matter how temporary, improve the long-term representation of women; all affirmative action interventions, no matter how protracted, self-correct any initial decline in the quality of the hired candidates. Requiring employers to interview an equal number of candidates from each sample of applicants can only be effective if the interviewed candidates are chosen randomly, ignoring their (observable) talent.

*We are grateful for comments from seminar participants at Columbia, PSE, Universita' Cattolica di Milano, and the 2026 ADRES Conference at Universite' Paris-1. We are particularly grateful to Jean Francois Laslier, who originally directed us to the literature on Pólya urns.

1 Introduction

A parable

Anna is a very distinguished senior economist, engaged in an influential research program with stellar publications; she is an effective teacher and a generous collaborator. Marie suggested her name when discussing potential senior hires. "Excellent idea!" was the committee's reaction: "Anna is creative and very productive, top publications, willing to move, and in a subfield where women are underrepresented!" The committee returned to Marie a week later. "Now, Anna is an extremely strong candidate. Really very impressive. [Pause]. However, you must admit that Bob is even stronger." And Bob received the offer. Let's accept Bob's dazzling achievements as indeed superior. What puzzled Marie was not the specific outcome, but the realization that she had already witnessed the same process, with the same outcome, too many times.

Stories very similar to Anna and Bob's occur frequently in many academic departments, typically with the same ending. There is much welcome attention to underrepresented candidates, but when the discussion turns to actions and to offers, underrepresented groups remain, well, underrepresented. And yet, there was no discrimination in the decision.

The example of women in male-dominated academic fields stands more broadly for the hiring or nomination of members of underrepresented groups in selective environments.[1] The goal of this work is to show that the ending we just described is built into the very fact of underrepresentation, and yet works to perpetuate it, even when all groups are treated equally. In a world where two groups have equal distributions of talent, where candidates are objectively evaluated, there are no information frictions, and no discrimination occurs, underrepresentation by itself is highly sticky. The key insight is that, if groups have historically had different access to positions of power, it is likely true that candidacies for such positions remain unbalanced. But if the difference in the number of candidates from two different group is large, for any candidate in the smaller group it is almost always possible to find a candidate in the larger group that dominates her, at equal distribution of talent. Or, stating the same idea more precisely, if two samples of different sizes are drawn from the same distribution and their respective highest order statistics are compared, the highest realization is more likely to belong to the larger sample. And if the sizes of the future samples respond to the realized selection in the expected direction- increasing if the selection came from the sample, increasing less if it did not-the hysteresis can be very strong.

We model this simple point and study its long-term implications exploiting the statistical properties of Pólya urns. A Pólya urn is an urn containing a given number of balls of different colors, two colors in our parable: w_0 initial white balls and m_0 initial mauve balls. At each step, one or more balls are randomly drawn, their colors are observed, and the balls are replaced in the urn; in addition, k new balls are added to the urn, with colors depending

[1] For concreteness, we use "men" and "women" throughout the paper, but the two labels are not meant to be read only literally. "Women" could refer more broadly to any underrepresented group, and "men" to the converse.

on the colors of the balls that were drawn. The process continues forever, and the question typically studied is the limit composition of the urn. In the simplest case, a single ball is drawn and a single additional ball of the same color is added. The fraction of white balls converges to a random limit, a random variable with a Beta distribution of parameters w_0 and m_0. In other words, if we replicate the process many times always starting from the same initial conditions (w_0, m_0), each time, the proportion of white balls will eventually stabilize to a limit; the limit itself will differ across replications but the mean limit will come to equal the initial fraction of white balls, $w_0/(w_0 + m_0)$.

Note that if the initial urn contains w_0 white balls and m_0 mauve balls, the probability that a randomly drawn ball is white $(w_0/(w_0 + m_0))$ equals the probability that the highest order statistics in two samples W and M, of sizes w_0 and $m + 0$ and drawn from the same distribution, belongs to sample W. If this model represents hiring and the evolutions of the samples of candidates from which hiring occurs, in expectation there is perfect hysteresis: on average, the probability of selecting a woman never changes, and if the initial sample of women candidates is small, relative to the initial sample of men, such probability is expected to remain small. Forever. With equal talent and no discrimination.

Groups who historically have been oppressed or subordinated remain underrepresented in positions of power or influence. Focusing on women, only about 10% of CEO's are women in OECD and EU countries (OECD, 2025); 11% of Fortune 500 companies' CEO's are women (a fraction that stalled at just above 10%, in 2023 and 2024). In universities, women representation among faculty falls with seniority and prestige. In the US, across all non-medical fields and ranks, women are 45% of full-time faculty and 37% of full professors (AAUP, 2025), but Harvard, Stanford and MIT, for example, all report about 25% of women among senior faculty [refs.]. In Economics, the 2024 share of women full professor was 18% in departments with doctoral programs and 14% for top-10 departments (AEA, 2024).

Explaining such persistence is at the core of our understanding of inequality and justice and shapes the policies our societies enact. As Brooks (2015) says, "libraries have been written" documenting that inequality can be persistent because relative positions in society tend to reproduce themselves, even in the absence of active discrimination. And yet, the Chief Justice of the US Supreme Court can still assert: "The way to stop discrimination on the basis of race is to stop discriminating on the basis of race." (Roberts, 2007).[2] Roberts' position is reflected in recent Supreme Court rulings in the US and in the broader backlash against affirmative action.[3]

This paper illustrates a very intuitive mechanism contradicting Roberts' statement. In our stylized world, there are neither biases nor informational frictions, hiring is fully meritocratic, and the majority and minority group have identical distributions of talent. The results we obtain are useful benchmarks about the evolution of underrepresentation in the absence of typical channels of discrimination.

The result described earlier –the forever-smaller chances of success for the minority– is weakened under more general assumptions, but the long-term stickiness of underrepresenta-

[2]Parents Involved in Cmty. Schs. v. Seattle Sch. Dist. No. 1, 551 U.S. 701, 748 (2007). The statement in fact did not originate with Roberts. See the discussion in Brooks (2015).

[3]See the landmark 2023 decision in Students for Fair Admissions v. Harvard, 600 U.S. 181, and its companion case against the University of North Carolina, outlawing preferential admissions in colleges on the basis of race.

tion at the aggregate level is very robust. In addition, the stochasticity of the process yields large disparities across individual employers, and marked non-monotonicities over time for each employer, without any underlying structural difference. Different paths are random and idiosyncratic, but the source of the aggregate inequality remains active. Note how the persistence of the minority's underrepresentation can easily contribute to the propagation of prejudices. Because men are hired with higher frequency, by the identical statistical reasoning described so far, they are also likely to be the most talented among those candidates who have been hired. Any prior (mistaken) belief in their superiority is easily confirmed by empirical observation.

The model has two components: the inequality in the size of the initial samples, which leads to the unequal probability of success, and, through the role-model effect, the dynamic link that perpetuates the imbalance. The link between the probability of success and the size of the sample is an immediate statistical observation, although often neglected. It is explained with great clarity in Ma (2020), a blog entry on prejudice against women chess players (https://en.chessbase.com/post/what-gender-gap-in-chess), and yet the comments to the post reveal how counterintuitive it remains to many readers. The second mechanism is the role-model effect: the increase in participation after observing the success of a member of one's own group. There is a large empirical literature, especially with regard to gender, and robust results in its favor.[4] Among theoretical studies, Athey et al. (2000), Chung (2000), and Müller-Itten and Öry (2017) build sophisticated dynamic models and discuss policy recommendations. These are however complex models, not easy to manipulate. The contribution of the present paper is to capture the same forces in a streamlined and very well studied framework: the statistical model of Pólya urns. The elegant simplicity of Pólya urns and the wealth of existing results about their properties lend transparency to the forces we are highlighting. They allow us to study the effects of different parametrizations, to propose and simulate a multiplicity of policy interventions, and to build a simple application (https://caron.shinyapps.io/Women- Men-Polya-Urns/) that lets the reader experiment with different assumptions and different interventions.

Our model does not offer an efficiency rationale for policies intended to equalize the representation of women. Without any intervention, the process selects the most qualified candidate at each step, and with unbounded populations and continuous distributions of talent, there is no decline in expected quality as one gender continues to be over-represented. This said, women's lower probability of success has a purely historical root, and in the presence of equal distributions of talent defies plausible notions of fairness.[5]

[4]Remaining with chess, De Sousa and Niederle (2002) discuss how the requirement of at least one woman player in French competitive chess teams lead to a large increase in the number and quality of women competitive players. Other studies document that women are more likely to choose a STEM major in college if they have been exposed to women mentors (Porter and Serra, 2020; Dennehy and Dasgupta, 2017; Carrell, Page, and West, 2010) and similar results have been shown for PhD students (Gaule and Piacentini, 2018). In Economics, there is a positive correlation between the number of women PhD's across programs and the number of female faculty (Boustan and Langan, 2019), and between the grades of undergraduate students and the gender of the professor in programs with strong gender imbalances (Griffith, 2014). (See AEA (2021) for more references. We expect the role-model effect to be stronger in environments one's gender is particularly salient. In line with this hypothesis, Kofoed and mcGovney (2019) document a marked increase in women cadets choosing traditionally male-dominated occupations at the West Point military academy after having been mentored by a female officer.

[5]In real applications, over-representation is likely to be accompanied by increasing scarcity of talent and by the discouragement of minority group members. Our goal, however, is not to investigate inefficiencies

Across all interventions we study, we find two strong results. All affirmative action policies, no matter how temporary, improve the long-term representation of women; all affirmative action policies, no matter how protracted, self-correct the initial decline, if any, in the quality of the hired candidates.

We consider two broad classes of policies: interventions on hires, which introduce a pro-women bias when a candidate is selected; and interventions on candidacies, which attempt to correct the imbalances in the samples of candidates. The latter interventions attack directly the source of women's disadvantage, but are of more difficult implementation. The former type modifies the probability of success in women's favor, and its influence is then transmitted through time via the role-model effect, the law-of-motion of the system. Because our underlying model selects the most able candidate by assumption, any binding intervention on hires must reduce the quality of the chosen candidate by construction. The effect however is only temporary, regardless of the duration of the policy. A pro-women bias leads to an increase in the number of women candidates and works towards equalizing the distributions of talent in the samples of candidates, reflecting the equal distributions in the population. As a result, the probability that the most qualified candidate is in fact a woman increases, triggering a virtuous cycle that eventually makes the affirmative action policy redundant.

Among interventions on hires, we compare two policies in more detail: Quota policies that mandate a specified fraction of women hires, and Rank policies that mandate that a woman be hired if ranked within the top k candidates. When the imbalance in the number of women and men candidates is not large, the latter policies are more desirable: rank policies can then generate both more women hires and higher expected hires quality. However, when the imbalance in the samples of candidates is large, quota policies have larger and faster impact, and thus can be interrupted sooner. The analysis of interventions on candidacies is still in progress. One immediate result however is worth mentioning. A commonly recommended policy is for the employer to conduct an equal number of interviews with women and men candidates, regardless of the difference in the size of the two samples. Our model, and our motivating parable, show that the policy will typically be ineffective. It can have an equalizing effect only if the candidates to be interviewed are chosen randomly within each sample, as opposed to being selected on the basis of their observable talent. The intervention would be quite inefficient, and to our knowledge is never implemented in these terms.

Affirmative action interventions are effective in the model because favoring the selection of women increase women's participation among candidates without affecting negatively their qualifications. Note the distinction between the qualifications of the winners (which in our model decline in the short-term) and the qualifications of future candidates. The empirical literature on affirmative action, whether focused on gender or on race, addresses primarily the former, while we focus on the latter.[6] Some works, however, document protracted effects of affirmative action after the policy stops binding (Matsa and Miller, 2011;

but to highlight the force of persistence in a stripped-down world of minimal assumptions.

[6]See for example the large body of work on affirmative action and higher education, as in Arcidiacono and Lovenheim (2016) or Bleemer (2021), and the references therein.

Miller and Segal, 2012; Miller, 2017)[7] The longer-term impacts are in line with our model but may reflect both supply and demand effects. Experimental tests on tournaments are valuable in isolating cleanly the origin of the effect: in response to a policy favoring the initially disadvantaged group, Schotter and Weigelt (1992) and Calsamiglia et al. (2013) find increases in total effort; Niederle et al. (2013) find increases in women's participation to the tournament that cannot be explained solely by the higher probability of winning.

Finally, the last component of our approach is the Pólya urn model. Pólya urns have been used in countless applications across fields. In Economics, their most influential applications have been to modelling self-confirming processes in location choices and technological innovation (Arthur et al., 1987; Arthur 1989, and 1994; David, 1985). In a similar spirit, Pólya urns have been a useful technical tool in evolutionary approaches to industrial policy (see for example Dosi et al., 1994, and Stepanova et al., 2019). In the literature on discrimination and persistent inequality, Pólya urns have been evoked as a powerful rhetorical device (see for example Roithmayr, 2014, and the discussion in Brooks 2015) but to our knowledge there have been few precise technical results with this application in mind, and the model has not been exploited to yield precision to intuitive conjectures. In particular, Pólya urns have not been applied to the study and comparison of different affirmative action interventions.

When the mathematics of Pólya urns has been exploited, the focus has been on the limit of the process. Although we share the interest on limit properties, our numerical simulations also inform us about the process' dynamic path. From an empirical point of view, that allows us to find results on the speed with which different interventions can affect minority representation, and on the changes in such a speed as the process evolves. Empirical observations that the progress of women towards equal representation appears to have slowed in more recent times, or that the effects of women mentors in the workplace seem to have dampened can be rationalized through our model (England, Levine, and Mishel 2020, Kurtulus and Tomaskovic-Devey 2012).

In what follows, we begin by describing the basic model and its properties, as applied to our problem, and a few natural extensions (Section 2). Section 3 discusses possible interventions, distinguishing between interventions on hires (Section 3.1) and interventions on candidacies (Section 3.2). The Appendix collects some mathematical proofs.

2 The Model

We begin by describing the simplest model, first referring to our hiring scenario, and then discussing how it can be captured by the statistical model of Pólya urns. Once the logic is clear, generalizations will be easy to formalize.

At all t, with $t \in \{0, 1, 2, ...\}$, a set of candidates N_t, of size n_t, compete for a single, selective position. Each candidate i is characterized by aptitude a_i, fully observable, and the position goes to the candidate with highest aptitude. The set N_t consists of both women (denoted by W_t, with $|W_t| \equiv w_t$) and men (denoted by M_t, with $|M_t| \equiv m_t$). W_t and M_t are two samples composed of independent draws from two underlying populations, women

[7] Bertrand et al. (2019) qualify the results in Matsa and Miller (2011).

(W), and men (M), each formed by a continuum of individuals with varying aptitude a, described by the same continuous distribution $F(a)$, common to both populations. At time $t+1$, individual s_{t+1} with $a_{(n)} = Max\{a_i | i \in W_t \cup M_t\}$ is selected. We denote as *samples* of candidates the two samples drawn from each population, women and men, and by *set* of applicants the union of the two samples, or the set N_t.

At time $t = 0$, the majority of candidates are men: $w_0 < m_0$. At each successive hiring cycle, two new samples of women and men candidates are realized, W_t and M_t, again composed of independent draws from the two populations. The size of the two samples is influenced by the identity of the individual last selected for the position: observing the success of a member of one's own group reinforces the desire to compete. Denoting by Δw_{t+1} (Δm_{t+1}) the difference $w_{t+1} - w_t$ $(m_{t+1} - m_t)$, we assume that $\Delta w_{t+1} > \Delta m_{t+1}$ if $s_{t+1} \in W_t$, and $\Delta w_{t+1} < \Delta m_{t+1}$ if $s_{t+1} \in M_t$: the sample from which the successful candidate was chosen grows more than the other sample. The process then continues: given $\{W_{t+1}, M_{t+1}\}$, the highest aptitude candidate is selected at $t+2$, etc.. We are interested in the probability with which the successful candidate belongs to one or the other population, and in the evolution of this probability over time.

The dynamics of the model have two ingredients: (1) the selection of the successful candidate, and (2) the law of motion of the samples over time. Given the maintained assumptions of a common aptitude distribution $F(a)$, fully observable characteristics, and no discrimination, the logic is transparent. Consider first the selection phase. Because the two samples are independent draws from populations with the same $F(a)$ distribution, each candidate has the same ex ante probability of being the highest a_i draw, a probability equal to $1/(w_t + m_t)$. Hence: $\Pr(s_{t+1} \in W_t) = w_t/(w_t + m_t)$, and $\Pr(s_{t+1} \in M_t) = m_t/(w_t + m_t)$. At $t = 0$, $w_0 < m_0$; hence $\Pr(s_1 \in W_0) < \Pr(s_1 \in M_0)$: in the absence of discrimination and with identical aptitude distribution, women start out by being chosen less frequently than men, simply as a result of their smaller numerical representation among the set of candidates, and the more so the more skewed are the sizes of the two samples–the smaller is w_0/m_0. Along the dynamic path, any role-model effect, symmetric across the two groups, will work to sustain the difference in sample sizes in the future and propagate the lower chances of a woman being selected. The initial numerical difference in the samples of candidates translates into long-term persistence of underrepresentation, at equal distribution of aptitude and no discrimination.

This very simple model maps directly into a statistical model of Pólya urns. The selection of the successful candidate and the role model effect that drive the dynamic path of the samples of applicants are mirrored in Pólya urns' classic assumptions. A large number of generalizations have been studied and several can be relevant for our problem[8], but let us begin from the simplest scenario.

Suppose that after the successful candidate is selected, the new sample of candidates from population W (M) is larger by one unit if the position was given to a woman (man) and remains of equal size otherwise: $n_{t+1} = n_t + 1$, with $\{w_{t+1} = w_t + 1,\ m_{t+1} = m_t\}$ if $s_{t+1} \in W_t$, and $\{m_{t+1} = m_t + 1,\ w_{t+1} = w_t\}$ if $s_{t+1} \in M_t$. We want to know how the probability of selecting a woman evolves over time. The model corresponds to the classic

[8]Useful surveys are Johnson and Kotz (1977), Pemantle (2007), Mahmoud (2008).

two-color Pólya urn with a single draw and a single addition. (Eggenberger and Pólya, 1923; Pólya, 1931): an urn contains balls of two colors; at each step, a single ball is extracted, its color is observed, and the ball is replaced in the urn together with a single additional ball of the same color. Note that the number of balls in the urn always grows by one unit.

Suppose the two colors are white (W) and mauve (M). If the urn at t contains w_t and m_t balls of each color, the probability of drawing a white ball is $w_t/(w_t+m_t)$, corresponding to the probability that the highest aptitude candidate is a woman in our model of hiring. The law of motion of the balls in the urn mirrors the role-model effect described above.[9]

The question studied is the composition of the urn over time, and its limit properties, or, in our application, the evolution of the probability of selecting a woman, and its limit properties. In the remainder of this section, we summarize results from the Pólya urns literature, rephrased in terms of our hiring model.[10] We interpret a single urn as a single employer, and a large number of urns with the same initial conditions (w_0 and m_0) and the same law of motions as a market.[11]

2.1 Lessons from Pólya urns

Define r_t as the share of women among all applicants after the tth hiring cycle: $r_t \equiv w_t/(w_t + m_t)$. As noted, r_t corresponds to the probability that the successful candidate is a woman at step $t+1$. Then:

Result 1. *The probability that the successful candidate is a woman is a martingale:* $E(r_{t+1}|r_t) = r_t$.

Note that the martingale property implies $E[r_t] = r_0 = w_0/(w_0 + m_0)$ for all t: the probability that a woman is successful is never expected to change. If $w_0 < m_0$, we expect women to be *forever* selected less frequently than men, even though aptitude is equally distributed in the two populations and the selection is transparent and meritocratic.

The stickiness of the probability of selecting a woman extends to the fraction of women chosen over time, and thus of women's representation in the selective position. Denoting by $\widehat{w}_t$ the total number of women selected after the tth hire, and by ρ_t the share of women in the set of selected candidates at t, we have:[12] $E(\rho_t) = r_0$. At equal talent and with fair hirings,

[9] In the hiring model, the two samples of applicants are newly drawn at every t, ensuring that each candidate has the same ex ante probability of having the highest aptitude–something that would not be true if candidates not hired in the previous samples remained among the applicants. As in the academic market, new candidates come with a vintage. In the Pólya urn model, existing balls remain in the urn and an additional one is added. However, the balls do not have additional attributes, beyond their color, and thus can reflect the fresh draws of candidates. The random draw of a ball from the urn implies that each has equal probability of being drawn (the logical parallel to having the same probability of being of highest aptitude in the hiring model).

[10] These are known results and proofs are omitted, unless they are very short and can help intuition. See the Appendix.

[11] The reader should keep in mind, however, the lack of competition across urns and the separate evolution of each urn's composition (or applicant samples). Alternatively, one could interpret each urn as a separate labor market, accommodating these features more naturally but requiring an extension to multiple hires (which, as we will show, is possible). At the end, the best choice depends on the application. In an academic frame, for example, we can think of an urn as a department considering its junior faculty for promotion to tenure, or as the academic job market in a specific discipline, starting with equal imbalances in applications but diverging over time.

[12] See Remark 1 in the Appendix.

for any foreseeable future, the expected fraction of women among all hired candidates equals the share of women among candidates at time 0, and this remains a minority for all $w_0 < m_0$.

Similarly, the limit behavior of ρ_t is pinned down by the limit behavior of r_t. Indeed:

Result 2 (Eggenberger and Pólya, 1923; Pólya, 1931). *As* $t \to \infty$, $r_t \overset{a.s.}{\to} z \sim Beta(w_0, m_0)$ and $\rho_t \overset{a.s.}{\to} z \sim Beta(w_0, m_0)$.

This result, justly famous in mathematical statistics, can be usefully decomposed into three separate observations.[13] First, it says that a limit exists: the probability of selecting a woman converges to a stable limit. However, it also says that the limit z is a random variable and thus will differ across employers. Replicating the experiment across many different employers, starting from the same initial conditions (w_0, m_0) and applying exactly the same rules, over a very large number of hiring cycles the probability that women's success rate stabilizes goes to 1 for all hiring firms, but the probability that two firms stabilize to the same rate is zero. This is an important remark. Randomness dictates which limit emerges, and different success rates for women candidates across different employers need not be imputed to underlying differences but can simply result from the stochasticity of the dynamic path. And yet, and this is the third observation, when a large number of employers are observed, order reappears: the limit distribution of the frequency of women hires exists and has a known parametric form, a Beta distribution with parameters (w_0, m_0). In the aggregate a structure emerges.[14] We have phrased these observations in terms of the probability of selecting a woman, but identical remarks, and identical limit distribution, apply to ρ_t, the fraction of women among successful candidates.

The existence of a limit depends on the fixed absolute increase in the set of applicants after each hire–one unit, in the example we are studying here, but any fixed absolute increase would also lead to convergence. As the sample grows, the increase eventually becomes infinitesimal. Applied to our problem, this corresponds to the very plausible assumption that the role-model impact of recent hires declines in importance as a field matures: recent hires matter when a sector or profession is new and small; over time, their influence declines and the role-model effect becomes dominated by the average composition of its labor force, as opposed to the marginal changes of the last hiring cycle.

It is helpful to visualize these results. We have built a simple application which simulates the dynamic paths of multiple contemporaneous employers (multiple Pólya urns): (https://caron.shinyapps.io/Women-Men-Polya-Urns/). It is designed to be flexible, and we encourage the reader to experiment with different parameter values and, as we will discuss, different assumptions and different interventions.

Figure 1 reports the results of simulating 100 successive hiring cycles over 100 different employers, where each line is a separate employer. As the theory predict, the mean probability of selecting a woman is constant and equal to $w_0/(w_0 + m_0)$. The same expected value applies to the proportion of women in the stock of all hires: $E[\rho_t]$ in the right-hand

[13] Cohen (1973) is an unusually engaging description of the properties of the Pólya-Eggenberger urn.

[14] The moments of the limit distribution are pinned down by the initial sizes of the two samples, w_0 and m_0: $E(z) = r_0 = w_0/(w_0 + m_0)$, and $Var(z) = (w_0 m_0)/\left((w_0 + m_0)^2 (w_0 + m_0 + 1)\right)$. The larger is the initial set of applicants, the smaller is the increment represented by the fixed increase after each hiring, and the more concentrated the limit distribution. When the initial set is smallest $(w_0 = 1, m_0 = 1)$, the dispersion of the limit distribution is maximal: the distribution of the limit z is uniform.

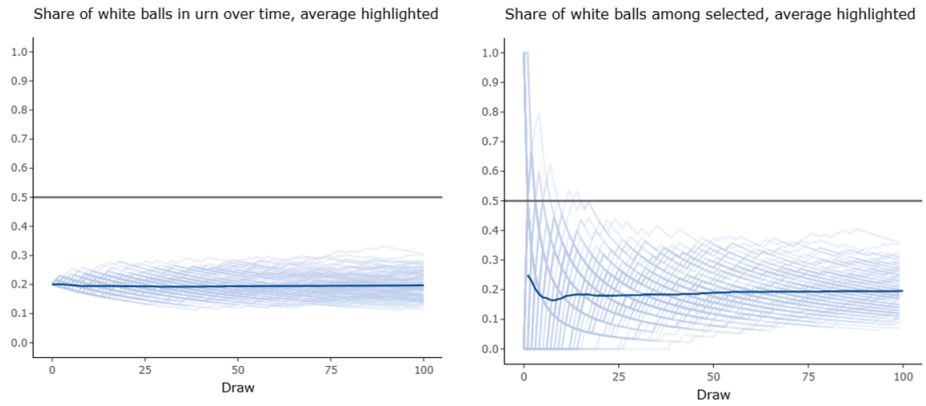


Figure 1: *Selection of candidates as draws from Pólya urns.* 100 urns; 100 successive draws; initial samples $w_0 = 10$; $m_0 = 40$. Left panel: probability of hiring a W at each draw. Right panel: share of W among selected. Averages highlighted: $E(r_t = E(\rho_t) = 0.25$ at all t.

side panel, and is again constant for all t.[15]

As noted in footnote 14, the moments of the limit distribution are pinned down by the initial sizes of the two samples of candidates: the larger the initial set of applicants, the more concentrated the limit distribution. Figure 2 again reports simulations of 100 successive hiring cycles over 100 different employers but in this case for three markets starting at different initial conditions. Although $r_0 = 0.2$ in all three markets, $n_0 = 200$ in the blue market, $n_0 = 50$ in the red market, and $n_0 = 20$ in the yellow market. The three markets have equal means, but different dispersions.

Because the model is well-known, we do not insist on all its details. But note its richness. Because each market's evolution is stochastic, it is typically not monotonic: for the same employer, increases and declines in the share of women candidates or hires can follow each other. Figure 3, zooming in on a few urns, distinguished by their colors, makes that transparent.

In addition, especially when the total number of candidates is small, random realizations can have large influence on future paths, and small initial differences across employers become amplified over time: different employers can evolve quite differently. And yet, over all employers mean values are predictable: the highlighted averages are very stable. Mean values reflect the initial imbalance in the samples of candidates and thus, without intervention, the model could not deliver a long term majority of women hires, or indeed, parity. Although the model is stochastic and the stochasticity allows it to generate realistic instability at the individual employer level, which group remains underrepresented is not random.

The martingale property that induces the model's hysteresis is robust to generalizations

[15] In all that follows, when reporting numerical simulations we will use the default case of $W_0 = 10$; $M_0 = 40$. These values blunt the importance of the first few draws and mimic the 2024 share of women among associate professors in top 10 US Economics programs (21%, according to AEA, CSWEP, 2024). Precise numbers should not be given weight; their role is to anchor comparative results when looking at different scenarios.

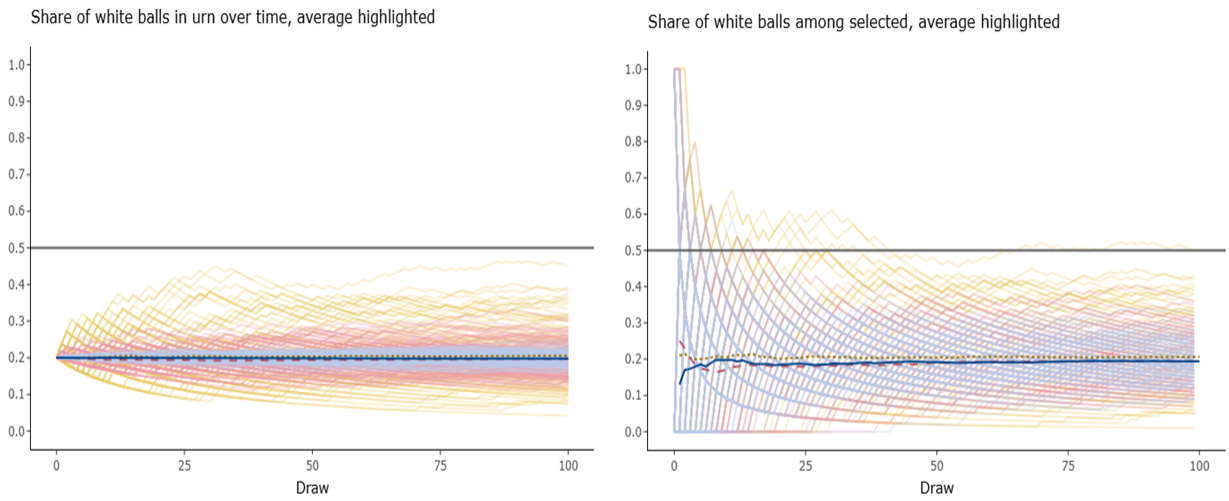


Figure 2: *Thin and Thick Markets.* Left: Share of white balls in urns. Right: Share of white balls among selected. Blue: $w_0 = 40, m_0 = 160$, Red: $w_0 = 10, m_0 = 40$, Yellow: $w_0 = 4, m_0 = 16$. Averages highlighted.

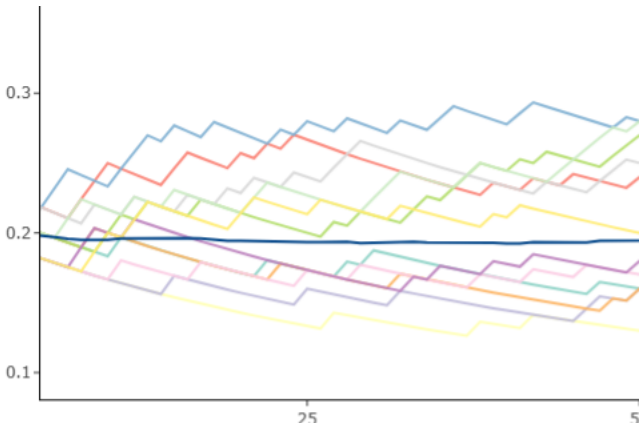


Figure 3: *Zoom on non-monotonicity.* 10 urns, 50 draws; $w_0 = 10, m_0 = 40$ for all.

that apply quite naturally to our story. For example, the case of a single hire per period and a single addition to the set of candidates over time can be extended to multiple contemporaneous hires and an increase of multiple units to the set of candidates. Suppose v candidates are hired at each cycle–the v best candidates among the set of all candidates. Maintaining our focus on simple models, suppose that each hire then results in one additional candidate of the same gender in the next cycle's set of candidates. The set of candidates always increases by v additional units at each cycle.[16]

The model conserves the martingale property of the simpler Pólya urn with a single addition:

Result 3. (Chen and Kuba, 2013; Kuba and Mahmoud, 2017). *Suppose every hiring cycle consists of v hires of the v most able candidates. Each hire results in one additional candidate of the same gender. Then: $E[r_{t+1}|r_t] = r_t$, and thus $E[r_t] = r_0$ for all t.*

The expected probability of hiring a woman remains constant over time.[17]

A different interesting variation is studying a role model effect that mirrors not the latest hire but the fraction of women in the stock of successful candidates, $\rho_t \equiv \hat{w}_t/t$. Returning to the reference case of a single hire and a single addition, suppose that, following a hire at t, the set of candidates always grows by one unit, an additional woman with probability ρ_t, or an additional man with probability $1 - \rho_t$.[18]

The process continues to exhibit the same stickiness. As we show in the Appendix, we find:[19]

Result 4. *Suppose the single most able candidate is hired at every hiring cycle t. Each hire results in one additional W candidate with probability ρ_t, and one additional M candidate with probability $1 - \rho_t$. Then: $E(r_t) = r_0 \ \forall t$ and $E(\rho_t) = r_0 \ \forall t$.*

Result 4 is not unexpected: the evolution of the stock of successful candidates and the evolution of the candidates samples are closely tied, and a role model effect that focuses on the share of past successes cannot overcome the sluggishness built into the system.

At least at first sight, more unexpected is a contrasting result from another variant of the basic model. Suppose that, as in Result 4, we allow for stochasticity in the law of motion of the candidates samples, but hold such stochasticity constant. Maintaining the assumption of a single hire and a single addition, suppose that the additional candidate mirrors the gender of the person hired with probability $p < 1$ and is instead of the other gender with

[16]The literature has studied two alternative models, corresponding to whether the v draws take place without or with replacement. (In both models, the additions to the urn occur, all together, after all draws). The first model, which we adopt here, fits a labor market where the set of candidates applies simultaneously to multiple jobs–for example for a position in v academic institutions–and the top v candidates are hired. We note in passing that the model then fits better the interpretation of each urn as a separate labor market, now now filling v positions.

[17]As for limit results, Chen and Kuba (2013) establish that, in the setting we are discussing, there exists a random variable x such that $lim_{t\to\infty} E(r_t) = x$, and $E(x) = r_0$, but the limit distribution is neither Beta nor gaussian.

[18]More explicitly, after a woman is hired at t, $w_t = w_{t-1} + 1$ with probability $\frac{\hat{w}_{t-1}+1}{t}$, and $w_t = w_{t-1}$ otherwise; after a man is hired at t, $w_t = w_{t-1} + 1$ with probability $\frac{\hat{w}_{t-1}}{t}$, and $w_t = w_{t-1}$ otherwise. Note that the last selected hire affects the law of motion through its influence on ρ_t.

[19]Characterizing the limit distribution is complicated because the changes in the samples of candidates are stochastic and the probabilities are time-dependent. However, the results in Pemantle (1990), applied to our model, imply that the limit share of women converges a.s. to a random variable that has no atoms.

probability $1-p$. We find:[20]

Result 5. (Bagchi and Pal, 1985, Chen and Kuba, 2013, Morcrette and Mahmoud, 2018, among others). *Suppose the single most able candidate is hired at every hiring cycle. Each hire results in one additional candidate of the same gender with probability $p<1$, and of the other gender with probability $1-p$. Then: (i) For all $p<1$ and $r_t<1/2$, $E(r_{t+1}|r_t)>r_t$; (ii) For all $p\in[1/2,1)$, $E\left(\lim_{t\to\infty} r_t\right)=E\left(\lim_{t\to\infty}\rho_t\right)=1/2$.*

If the role-model effect exists but is less than fully predictable, the martingale property ceases to hold, and equality is reestablished in the limit. Note that the source of the convergence is the stochasticity in the response to an M candidate hire. Following such a hire, we are assuming that the M sample of candidates grows by 1 unit with probability p, and the W sample grows by 1 unit with probability $1-p$. We know that M hires are more frequent, and thus this factor more than compensates for the $1-p$ decline in the W's role model effect following a W hire, relative to the deterministic case. Shifting the same probability of an increase in W candidates from a less to a more likely event helps long term equality.

The observation invites two comments. First, it calls attention to role model effects in response to *both* W and M hires, something that deserves more attention in the literature.[21] Second, it is important to keep in mind that the result is asymptotic. Convergence remains very slow. Consider for example 100 urns, with $r_0=0.2$. If $p=0.7$, our app tells us that after 100 draws, the average probability of hiring a woman, as well as the share of women in the cumulative number of hires, is still below 35 percent; it reaches 45 percent after 1,000 draws and rises only to 47 percent after 2,000. The case $p=0.5$ is an interesting reference because it corresponds to the full absence of role model effects, and thus the lack of the dynamic transmission from current to future underrepresentation. With $p=0.5$, it takes about 230 draws to reach $r=0.45$ and about 425 draws to reach $r=0.47$. But the initial disadvantage created by the numerical inferiority of the sample of women candidates is never fully overcome. With $p=0.5$, the exact probability distributions of the evolving samples of candidates are simple binomials.[22] After 1,000 draws, the probability that the sample of women candidates is strictly smaller than the men's sample ($\Pr(w_{1000}<m_{1000})$) is 82 percent; after 10,000 draws it is still 61 percent. At best, the disadvantage of the minority group becomes less and less relevant as the market expands and the initial disparity pales. But it never goes fully away.

Figure 4 shows three markets with $p=\{0.5,0.7,1\}$, each with 100 employers and 1,000 draws. The strong inertia of the system and the progressively slower improvement as the curves, on average, approach the asymptote at $E\left(\lim_{t\to\infty} r_t\right)=1/2$ are very evident.

The literature on Pólya urns is very rich and although it mostly focuses on limit results,

[20]The first claim is proven in the Appendix; for the second, see the references, which also establish: For all $p\in(1/2,3/4)$: $\frac{w_t-t/2}{\sqrt{t}}\xrightarrow{\mathcal{D}}\mathcal{N}(0,\frac{(2p-1)^2}{4(3-4p)})$; $\frac{\widehat{w}_t-t/2}{\sqrt{t}}\xrightarrow{\mathcal{D}}\mathcal{N}(0,\sigma^2)$.

[21][Suppose the role model differed across the two genders: following a W hire, $w_{t+1}=w_t+1$ with probability p_w (and $m_{t+1}=m_t+1$ otherwise); following a M hire, $m_{t+1}=m_t+1$ with probability p_m (and $w_{t+1}=w_t+1$ otherwise. A plausible conjecture is that $E\left(\lim_{t\to\infty} r_t\right)$, if it exists, would equal 1 if $p_w>p_m$, and 0 otherwise. However, this remains to be verified.]

[22]$\Pr(w_{s+t}=s+k)=\binom{t}{k}(1/2)^t$. If $p\neq 0.5$, and $p<1$, the problem becomes much more difficult because of the need to keep track of the exact timing of each W or M hire. We are not aware of a characterization of the precise distribution.

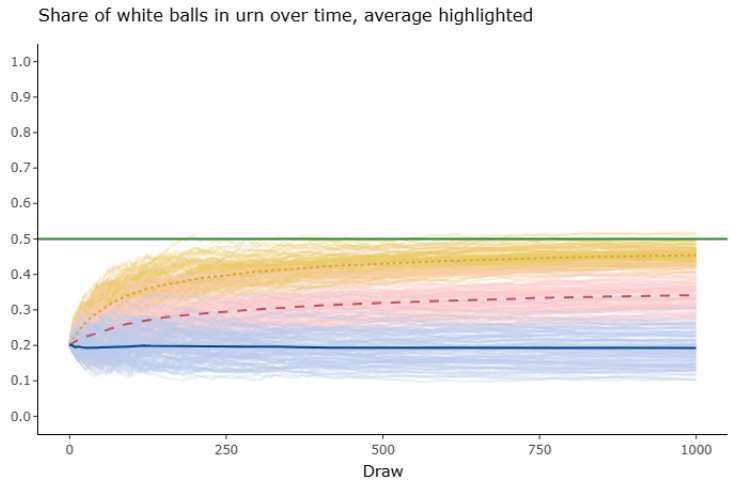


Figure 4: *Stochastic Additions.* Number of white balls in urns, $w_0 = 10; m_0 = 40$, 100 urns per market, 1,000 draws. Blue market: $p = 1$, Red: $p = 0.9$, Yellow: $p = 0.7$. Averages highlighted.

it can help to evaluate the robustness of our model.[23] The results reported so far, however, give us a sufficient basis to study possible policy interventions.

3 Interventions

In our model, only fairness considerations call for more equal representation. With continuous distributions of talent over unbounded populations, as long as the selection is unbiased and talent is observable, there is no efficiency reason to increase women's presence either among candidates or in hiring. We take such fairness motivation as a given: we suppose that even someone who believed in the assumptions of the model would agree that, with inequality stemming from past history and equal distributions of talent for the two populations, a target of equal representation is desirable. Our goal in this section is to clarify how to obtain it (in finite time) and to compare short and long-term effects of alternative policies.

The objective is to increase the representation of women among hires and among candidates with minimal decline in the selected candidates' aptitude. We compare two broad classes of interventions. The first class operates at the hiring stage and includes standard affirmative action policies: for example, favoring a woman candidate if in the top j of candidates, or hiring at least one woman for each k hires. The second class, closer to the underlying logic of the model but more difficult to implement, operates directly on the set of candidates, increasing the probability that the sample of women candidates grows, regardless of hires.

We begin by studying the first class of interventions, first deriving results that are common to all such policies, and then highlighting quantitative differences in their effects through

[23] For example, Aoudia and Perron (2012) and Crimaldi et al. (2022) discuss urn models that could be applied to multiple hirings where both the number of hirings and the role model effects are allowed to be stochastic and time dependent.

our app. We return to the simplest set-up, where at any hiring cycle a single candidate is hired, the set of candidates increases by one unit, and such unit mirrors the gender of the last hired candidate with probability 1.[24]

3.1 Policies targeting hires

We begin by a definition:

Definition. *We call a policy an "intervention on hires" if it satisfies the following conditions:*

1. *Hires are biased in favor of women:*

$$Pr(s_{t+1} \in W_t) > r_t$$

2. *Within each sample of candidates, the selection remains meritocratic: a woman (man) hired has highest aptitude among all women (men) candidates*:

$$\text{If } s_{t+1} \in W_t \Rightarrow a_{s_{t+1}} = Max\{a_i | i \in W_t\}$$

$$\text{If } s_{t+1} \in M_t \Rightarrow a_{s_{t+1}} = Max\{a_i | i \in M_t\}$$

We denote by $\mathcal{H}$ the set of such policies, and by H one such policy.

All policies in $\mathcal{H}$ share some common properties. We summarize them in the following Results.

Result 6. *Consider a policy $H \in \mathcal{H}$ enacted at $\underline{t}$.*

(i) For all H and all $t > \underline{t}$:

$$\Pr(Max\{a_i | i \in W_t\} = Max\{a_i | i \in N_t\}) > \Pr(Max\{a_i | i \in W_{t-1}\} = Max\{a_i | i \in N_{t-1}\}).$$

(ii) For all H, there exists $t' > \underline{t}$ such that for all $t \geq t'$:

$$\Pr(a_{s_{t+1}} = Max\{a_i | i \in N_t\}) > \Pr(a_{s_t} = Max\{a_i | i \in N_{t-1}\}) :$$

Both claims arise immediately from the logic of the model. Claim (i) states that the probability that the best woman candidate is the best candidate overall increases monotonically over time. The result must hold because by definition any policy H increases the expected fraction of hires who are women, and thus, via the role-model effect, the ratio of women to men in the set of candidates. Since candidates are independent draws from the same distribution of talent, it then follows that the probability that the best candidate overall is a woman must increase monotonically throughout the duration of the policy.

Claim (ii) is particularly interesting and less intuitive. It says that, after the policy has been implemented for some time, the probability that the selected candidate is the best

[24] We conjecture that many of the qualitative results below hold for arbitrary $p > 1/2$, i.e. with a probabilistic role model effect. For now, however, this is just a conjecture.

overall must increase monotonically over time, and thus must tend to 1: any decline in the aptitude of the selected candidate induced by the policy must be self-correcting. Absent any policy intervention, our model always delivers, by assumption, the most qualified hire. Thus, by construction, any policy that biases hiring must reduce the expected qualification of the person hired. However, claim (ii) states that any such decline must be temporary. The claim is not immediately obvious because it combines two different components: the monotonic expected increase in the aptitude of any hired candidate favored by the policy (a woman, in our case), following claim (i), and the varying probability of hiring the best woman or the best man, which depends on the policy and on t.

Consider for example an extreme quota policy, such that only women are hired until the samples of candidates are equalized. The probability that the selected candidate is the highest aptitude applicant overall jumps downward at the start of the intervention and remains below 1 until the samples are equalized, but must increase monotonically throughout the duration of the policy. Less immediate is the answer under different H policies, policies that increase the chances that a woman is hired, but do not guarantee it. Consider for example the policy "hire a woman if in the top j candidates" (with $j > 1$). The probability that the best candidate is hired corresponds to the probability that either the best candidate is a woman, or that there is no woman among the top j candidates:

$$\Pr(a_{s_{t+1}} = Max\{a_i | i \in N_t\}) = \frac{w_t}{n_t} + \prod_{z=0}^{j-1} \left(\frac{m_t - z}{n_t - z} \right) =$$

$$= r_t + \prod_{z=0}^{j-1} \left(1 - r_t \frac{n_t}{n_t - j} \right)$$

For small enough j and large n_t:

$$\Pr(a_{s_{t+1}} = Max\{a_i | i \in N_t\}) \approx r_t + (1 - r_t)^j \tag{1}$$

For large n_t and j not too high, equation 1 shows transparently the non-monotonic trend in the probability that the selected candidate has highest aptitude overall, first declining, during some initial interval after implementation of the policy, and then self-correcting. Intuitively, when the fraction of women candidates is small, the policy has little bite, and the best candidate is chosen with high frequency; when it is large, the policy does bite but the chances that the best woman candidate is the best candidate overall are high as well. The probability that the best candidate is chosen reaches a minimum when the fraction of women candidate has intermediate values.

The self-correction in the probability that the selected candidate is the highest aptitude applicant, while the affirmative action policy is implemented, is an important feature of the model, reminding us of the true origin of the disadvantage of the underrepresented group. The feedback loop between hires and the role-model effect also deliver an immediate result about the long-term effect of the policy:

Result 7. *Under any* $H \in \mathcal{H}$, $\lim_{t\to\infty} r_{t,H} = 1$, *and* $\lim_{t\to\infty} \rho_{t,H} = 1$.

Proof. Dropping the subscripts H for ease of notation, and recalling that s_{t+1} denotes the successful candidate among the set n_t, the selection criterion implies:

$$E(r_{t+1}|r_t) = \left(\frac{w_t+1}{n_t+1}\right) Pr(s_{t+1} \in W_t|H) + \left(\frac{w_t}{n_t+1}\right)(1 - Pr(s_{t+1} \in W_t|H) =$$

$$= \left(\frac{n_t}{n_t+1}\right)\left[\left(r_t + \frac{1}{n_t}\right)(Pr(s_{t+1} \in W_t|H) + r_t\,(1 - Pr(s_{t+1} \in W_t|H)\right] =$$

$$= \left(\frac{n_t}{n_t+1}\right)\left[r_t + \frac{(Pr(s_{t+1} \in W_t|H)}{n_t}\right] > \left(\frac{n_t}{n_t+1}\right)\left(r_t + \frac{r_t}{n_t}\right) = r_t$$

Or $E(r_{t+1}|r_t) > r_t$. But r_t has a ceiling at 1, for all t, and thus $E(r_{t+1}|r_t) > r_t$ implies $E\left(\lim_{t\to\infty} r_t\right) = 1$. But then $\lim_{t\to\infty} r_t = 1$, which then implies $\lim_{t\to\infty} \rho_t = 1$. □.

Result 7 has a strong implication: if the policy is designed for fairness reasons and hiring favors women candidates, then the policy must be temporary only. The logic is transparent: the disproportionate hiring of women reduces the relative candidacy of men, leading to a decline in their aptitude, relative to woman candidates, and reinforcing the selection of women on meritocratic grounds. In the limit, only women are hired and only women apply. The spare model makes the dynamics very stark, but captures transparently a leading concern about affirmative action. The same model that teaches that affirmative action policies are needed to reach more equal representation and fairness, also tells us that such policies must eventually stop.

Corollary to Result 7. Any policy $H \in \mathcal{H}$ enacted for fairness reasons must be temporary.

Denote then by $\bar{t}_H$ a time $t(H)$ such that $r_{\bar{t}_H} = 1/2$. We want to study the effect of a temporary policy H lasting from some $\underline{t}$ to some $t' \in (\underline{t}, \bar{t}_H]$. The crucial question now is whether gains from temporary policies will fade. The answer is negative.

Result 8. *Consider a policy $H \in \mathcal{H}$ enacted from $\underline{t}$ to some $t' \in (\underline{t}, \bar{t}_H]$, and an alternative scenario, distinguished by the index a, in which the policy is not enacted. For any $t > t'$, $E(r_{t,H}) > E(r_{t,a})$, and $E(\rho_{t,H}) > E(\rho_{t,a})$.*[25]

Any policy in $\mathcal{H}$, no matter how temporary, increases the long term shares of women among candidates and among hires. By increasing the number of selected women, the policy activates a feed-back process working through the role-model effect, resulting in more women candidates and thus in a higher probability that some are highly qualified, an effect that persists after the policy is stopped. A virtuous circle works to equalize representation.[26]

[25]The proof is immediate. By construction, for all $t > t'$, $E(\widehat{w}_{t,H}) > E(\widehat{w}_{t,a})$ and thus, via the workings of the role model effect, $E(w_{t,H}) > E(w_{t,a})$. Thus $E(\rho_{t,H}) > E(\rho_{t,a})$ and $E(r_{t,H}) > E(r_{t,a})$.

[26]The model's maintained assumption is that each candidate is an independent draw from the aptitude distribution $F(a)$. The positive effect of a biased hiring policy on the aptitude of minority candidates would not follow if applications from the minority group responded to the policy by increasing draws from the lower part of the support. But note that if this logic applies, it also means that, absent the policy, minority applications are more likely to be drawn from the upper tail of the distribution, relative to applications from the majority group. Explaining systematic underrepresentation then becomes problematic, in the absence of discrimination.

In fact, were the policy to stop in each market at $\bar{t}_H$, we know from Results 1 and 2 that as hiring cycles continue, $E(r_t|\bar{t}_H) = 1/2$ for all $t > \bar{t}_H$ for each employer, converging to a random limit with expected value 1/2.[27]

3.1.1 Policies targeting hires. Two examples

Can we say more about different policies that target hires? In this section we consider two examples: a quota policy mandating a minimal specified fraction of women hires, and a rank-based policy that favors a woman candidate if among a small set of most qualified candidates. We denote by Q_k the Quota policy "hire at least one w every k hires", and by R_j the Rank policy "hire a w if in the top j candidates". In all cases, the candidate hired is the top candidate in the cohort who satisfies the policy constraint–if a woman is hired, it is the best-ranked woman; if a man is hired, it is the best ranked overall.

The Qk quota policy is defined over k hires, or, in the scenario we are considering with a single hire per period, k periods, while the Rj policy is defined over a single hire and a single period. To make the two policies comparable, we evaluate both over k periods, and denote by Rj^k the Rj policy adopted in each of k contiguous periods. We study the policies' expected impact on the number of women hired, the evolution of the sample of women candidates, and the overall ranking of the successful candidates. In line with the previous discussion, we consider policies applied only as long as $r < 1/2$.

To gain some intuition on the relative workings of Quota and Rank policies, we focus on two extreme cases. $R2^2$ ("hire a woman if among the two top candidates") is the least aggressive Rank policy. $Q2$ ("hire at least one woman every two hires"), on the other hand, is the most aggressive Quota policy if we impose equal representation as a pragmatic upper bound, even in the short run, to any policy bias in hirings.[28]

We assume that the quota policy is implemented with the goal of optimizing the expected average rank of the 2-period hires, subject to the quota constraint. For example, under Q2, if the best woman candidate is drawn with an overall rank worse than the expected rank of a woman in the next period, the employer will wait to fulfill the quota until the next period.[29]

We denote by $E_t(\Delta\widehat{w}_2|H)$ the expected number, at t, of women hired between t and $t+2$, conditional on policy H, and by $E_t(\overline{\phi}_2|H)$ the expected average rank of candidates hired at $t+1$ and $t+2$, where each candidate's rank is defined relative to the candidate's cohort. We find:

[27] Note that across different employers $\bar{t}_H$ would vary, and thus so would $n_{\bar{t}_H}$ and the variance of the random limit.

[28] $Q2$ is clearly more aggressive than any Qk, with $k > 2$. Quota policies could take the more general form "hire at least z women every k hires. If we constrain $z/k \leq 1/2$, then $Q2$ corresponds to imposing $z/k = 1/2$ and the shortest possible time horizon, without allowing any carryover across periods. For example, over four successive hiring periods, the quota must be satisfied over periods 1 and 2, and then over periods 3 and 4, etc., as opposed to granting the employer a longer horizon. Constraints of this type work to limit the chances of non-compliance–for example, cap-and-trade environmental policies typically do not allow borrowing from future permits allocations.

[29] We ignore discounting and focus on relative talent. Note that in our model, the sample of candidates increases over time, and thus expected absolute talent increases over time for any fixed order statistics. We concentrate on rank both because focusing on absolute aptitude pushes us towards more parametric restrictions on the distribution of talent $F(a)$, and because affirmative action policies are particularly problematic when they select candidates that are ranked worse than other available candidates, a concern with rank rather than absolute aptitude.

Result 9. *For all n, There exists $\underline{r}(n) < 2/5$ such that for all $r \in (\underline{r}(n), 1/2]$, $E_t(\Delta\widehat{w}_2|R_2^2) > E_t(\Delta\widehat{w}_2|Q2)$, and for all $r < \underline{r}(n)$, $E_t(\Delta\widehat{w}_2|R_2^2) < E_t(\Delta\widehat{w}_2|Q2)$.* (The proof is in the Appendix).

In words, there exists a threshold value for r, $\underline{r}(n)$, such that when the share of women candidates, although a minority, is larger than $\underline{r}(n)$, the Rank policy leads to a larger expected number of women hires than the Quota policy, while the reverse holds if $r < \underline{r}(n)$. The threshold $\underline{r}(n)$ depends on n, but for all n it is smaller than $2/5$. For any size of the set of candidates, $R2^2$ is expected to have a larger impact on equalizing hires than $Q2$ as long as $w_t \geq (2/3)m_t$. Because the inequality holds when the policies are compared at equal candidate sets, it will continue to hold, and in fact be strengthened, as long as the policies are maintained and the sets of candidates diverge via the workings of the role model effect.

Finally, if we restrict attention to Qk and Rj^k policies as described above, as noted earlier, $R2^2$ is the least aggressive of the Rank policies, and $Q2$ the most aggressive Quota, implying that at $r > \underline{r}(n)$ any Rank policy leads to higher expected women hires than any Quota policy. We state this observation as a Corollary to Result 9.

Corollary to Result 9. *For all n, there exists $\underline{r}(n) < 2/5$ such that for all $r \in (\underline{r}(n), 1/2]$ any Rank policy R_j^k leads to higher expected women hires than any Quota policy Q_k.*

The comparison between Rank and Quota policies is important because explicit quotas are illegal in most situations. Rank policies maintain a selectivity criterion over all candidates, not only within each sample, and thus are easier to defend. When the minority represents a very small fraction of the set of candidates, however, Rank policies have to be relaxed substantially to reach the effectiveness of Quota policies.[30] What is less expected is that this ceases to be true when the minority, still remaining a minority, is better represented among candidates. In such cases, consideration of candidates that are "close to the best" is a sufficient intervention. In our model, Rank policies can (and should) take the place of Quota policies before full equality is reached.

Besides efficacy in increasing minority hiring, a second important criterion is the impact of the policy on the talent of who is hired. As discussed, under our assumptions, a short-term decline when the affirmative action intervention is enacted is inevitable, but is temporary only. Can the comparison of Quota versus Rank policies tell us more? Whenever a minority candidate is not favored, the candidate hired is the best overall. It then follows automatically that more aggressive interventions must be associated with higher declines in selectivity at the start. Two questions remain open. How do Quota and Rank policies compare at equal expected effectiveness on hirings? And given that more aggressive policies will be less lasting, how do different interventions affect the selectivity of the pool of hired candidates over a longer horizon? Before turning to our app, the two policy examples of $Q2$ and $R2^2$ give us an initial answer to the first question. We find:

Result 10. *If $E_t(\Delta\widehat{w}_2|R_2^2) = E_t(\Delta\widehat{w}_2|Q2)$, then $E_t(\overline{\phi}_2|R2^2) < E_t(\overline{\phi}_2|Q2)$. At equal*

[30] It is always possible to design a Rank policy that matches a Quota policy, but for any given Quota policy, obtaining equal expected minority hires through a Rank policy requires lower selectivity the lower is r, eventually making the supposedly easier defense of Rank policies doubtful. [Examples]

expected numbers of women hires, $R2^2$ is expected to be a more selective policy than $Q2$. (The proof is in the Appendix.)

Contrary to quotas, Rank policies set a floor on the aptitude of any hired candidate. As a result, at comparable effectiveness, they tend to be more selective.

Our app can help us lend some concreteness to these results. Figures 5 and 6 compare $R2$ and $Q2$ at different values of r: $r = 0.2 < \underline{r}$ (Figure 5) or $r = 0.4$, just above $\underline{r}$ (Figure 6). In both figures, the interventions are initiated after 50 draws and are interrupted for any individual employer as soon as the set of candidates reaches $r_i = 0.5$.[31] The role of the policies in redressing the stickiness of hires over the first 50 draws is very clear. In the longer term, both policies equalize both the flow and the stock of hires (both the probability that a new hire is a woman or a man, and the fraction of total hires that are of either gender). Predictably, the length of time required to reach equality, average over all urns, depends on the initial conditions, and is longer when r_0 is lower.

As Figure 5 shows, when women are a small minority in the set of candidates ($r < \underline{r}$), the Quota policy $Q2$ is more effective at rebalancing hires and candidacies in the short term. It causes a larger decline in the selectivity of hires than the Rank policy, as measured by the fraction of hired candidates who had highest aptitude among their cohort, but the effect is temporary. Because the policy is more effective, it also ends earlier, and the rebound in selectivity also takes place earlier: by draw 500, $Q2$ is associated with a more highly selected stock of hires than $R2^2$.[32] The impacts of $Q2$ and $R2^2$ are instead very closely comparable when $r_0 = 0.4$, just above $\underline{r}$ (Figure 6). $R2$ is barely more effective at increasing the fraction of women hired, and causes a hardly perceptible decline in the selectivity of the stock of hires.[33]

Note however that the selectivity criterion reported in the figures (the fraction of all hired who were ranked highest in their cohort) underplays the difference between Quotas and Rank policies: under the latter, candidates who are not ranked highest are hired, as under Quotas, but their rank is by construction close to the top, which is not necessarily true under Quotas. Figure 7 reports the average rank of hired candidates under the two policies with different initial conditions: $r_0 = 0.2$ in the left panel, and $r_0 = .4$ in the right panel. (The current figure reports each hire, as opposed to averaging over two periods. Hence the spikes.) Recall that a higher number means a higher order statistics and hence lower ranking (and note that the scale of the two figures is different). The difference between the two policies is very clear, and equally clear is the impact of more/less unbalanced starting conditions as well as the earlier ending of the policy under $Q2$ when r_0 is low (but not at $r_0 = 0.4$). The simulations give visual evidence to the theoretical results: when the minority's disadvantage is not too large, Rank policies are preferable to Quotas: they can

[31] Stopping the policy for each employer as soon as the target is reached, as opposed to stopping it when the average across all employers reaches the target, results in lower aggregate minority hires but seems a more reasonable implementation. Note also that if the target is $r = 1/2$ (i.e. equal probability of success), as opposed to $\rho = 1/2$ (or equal experience of success), there will be times t and employers i, still bound by the policy but with $\rho_{i,t} > 1/2$: $r_{i,t}$ has more inertia than $\rho_{i,t}$ because of the initial imbalance in the candidate samples.

[32] Note also the lower variance in hirings and candidacies across markets under $Q2$, relative to $R2^2$.

[33] Figures 9 and 10 in the Appendix report simulations of the two policy interventions under the same parameters when the role model is probabilistic, with $p = 0.7$. The qualitative comparisons between $Q2$ and $R2^2$ are preserved, but the difference between the two policies is predictably reduced.

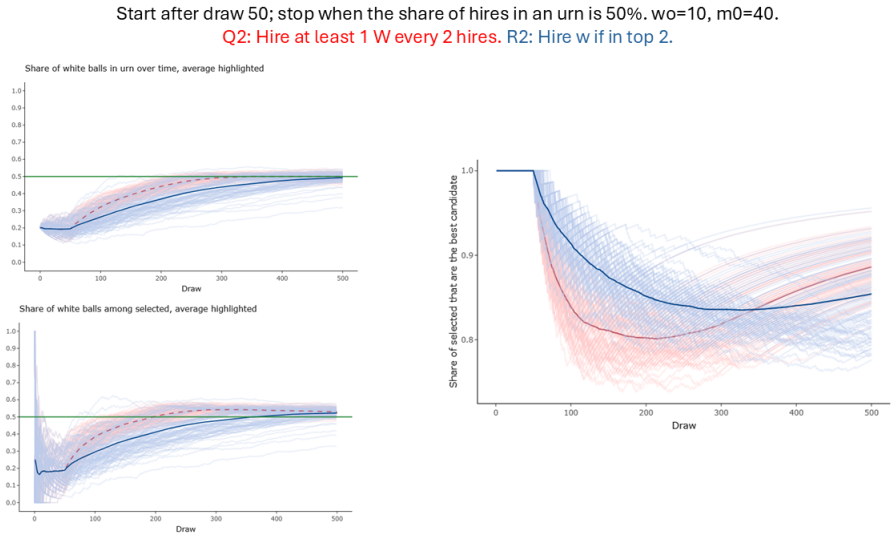


Figure 5: *Comparing Q2 (red) and R2 (blue).* $w_0 = 10$, $m_0 = 40$. The policy starts after 50 draws and ends for each employer the first time $r_i = 0.5$.

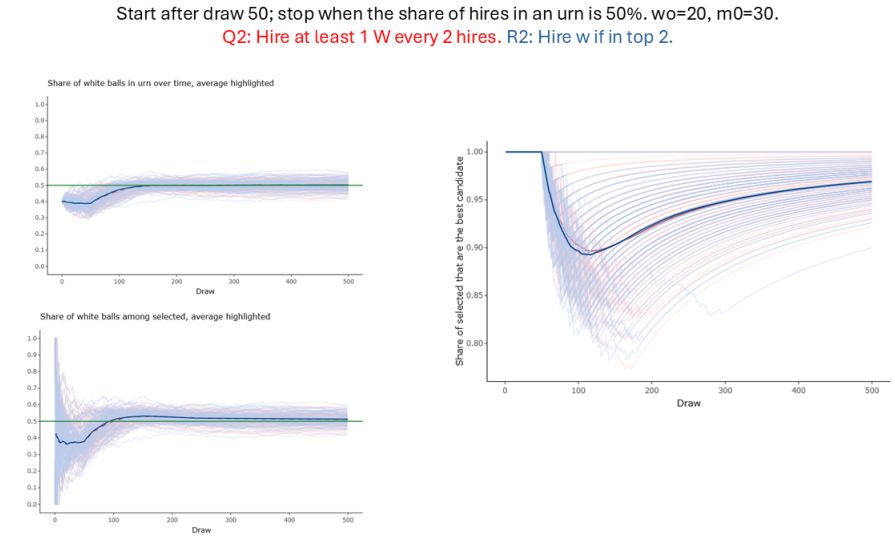


Figure 6: *Comparing Q2 (red) and R2 (blue).* $w_0 = 20$, $m_0 = 30$. The policy starts after 50 draws and ends for each employer the first time $r_i = 0.5$.

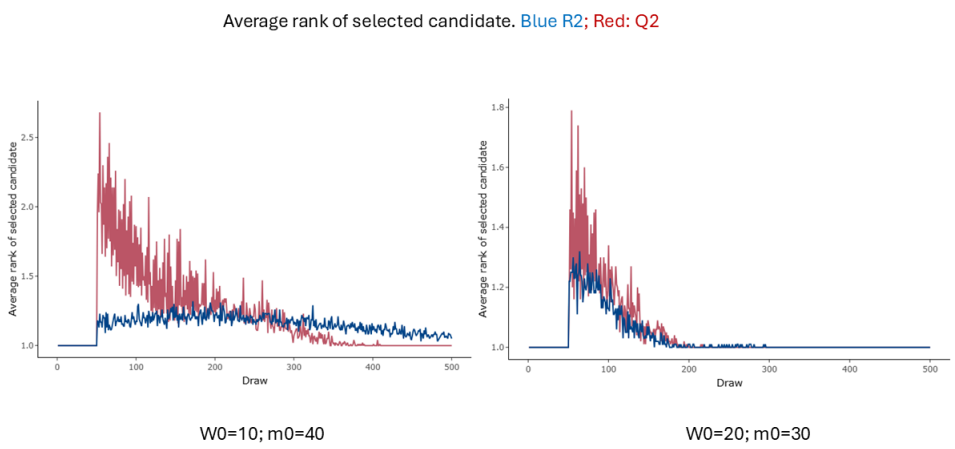


Figure 7: *Comparing Q2 (red) and R2 (blue).* Average rank of hired candidate. Left: $w_0 = 10$, $m_0 = 40$. Right: $w_0 = 20$, $m_0 = 30$. The policy starts after 50 draws and ends for each employer the first time $r_i = 0.5$.

be more selective and yet have comparable or superior effects on minority hirings; when the disadvantage is large, a trade-off appears: selective Rank policies will lead to lower minority hirings and longer affirmative action interventions.

Figures 9, 10, and 11 in the Appendix report simulations of the two policy interventions under the same parameters when the role model is probabilistic, with $p = 0.7$. Although we see some indication that $Q2$ continues to lead to more women hires than $R2^2$ at low r (while the opposite holds for r close to or above $\overline{r}$), the difference between the two policies is sharply reduced. Indeed, taking also into account the difference in selectivity, it becomes difficult to claim an advantage for the Quota policy, even at low r. More work is needed to understand and verify the robustness of these regularities but, if confirmed, would support relying on Rank policies more widely.

An additional question worth studying is whether these results continue to hold, qualitatively, when each employer hires not one, but multiple candidates, as in Result 3 of Section 2.1: at every hiring cycle, each employer hires the top v candidates from that cycle's set of candidates. Beyond realism, one advantage of this scenario is that Quota and Rank policies both can be defined over a single hiring cycle. Consider the following version of the model. The role model effect is deterministic: next period set of candidates increases by v, and each hire results in one additional candidate of the same gender. Quota policies fix the minimum number of hires, among the v hires, that must be women; Rank policies fix the number of top candidates z, $z > v$, such that any woman in z is hired, under the constraint that the total number of women hires cannot be larger than v.

We explore the relative effectiveness of the two policies in a simple example. Consider two very mild versions of the policies: a Quota policy establishing that at least 1 of the v hires must be a woman (which, by coherence with previous notation, we denote by Qv) and

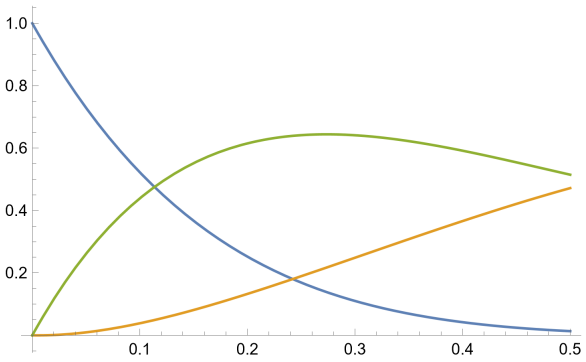


Figure 8: Comparing $Pr(Qv > R_v(v+1)$ (blue), $Pr(Qv < R_v(v+1)$ (orange), and $Pr(Qv = R_v(v+1)$ (green) as function of r. $n = 100$, $v = 5$.

a Rank policy that requires hiring a woman if ranked $v+1$ unless all top v candidates are women (denoted $R_v(v+1)$). Note that, over a single cycle of v hires, the two policies can differ at most by 1 hire.

Qv hires 1 additional woman if the top ranked woman has rank higher than $v+1$. $R_v(v+1)$ hires 1 additional woman if $s < v$ of the top v candidates are women, and the candidate ranked $v+1$ is a woman.

Using $Pr(R_v(v+1) > Qv$ to denote the probability of an additional woman hire under the Rank policy (and correspondingly for the other cases), we can write:

$$Pr(Qv > R_v(v+1)) = \binom{m}{v+1} \Big/ \binom{n}{v+1}$$

$$Pr(R_v(v+1) > Qv) = \sum_{s=1}^{v-1} \left[\binom{w}{s} \Big/ \binom{n-v+s}{s} \right] \left[\binom{m}{v-s} \Big/ \binom{n}{v-s} \right] \left(\frac{w-s}{n-v} \right)$$

$$Pr(R_v(v+1) = Qv) = 1 - Pr(Qv > R_v(v+1)) - Pr(R_v(v+1) > Qv)$$

It is immediately clear from the first equation that, for given n, the probability that Qv yields more women hires than $R_v(v+1)$ is increasing in m or, equivalently, decreasing in r. Figure 8 makes the point graphically.

As in the case of individual hires, the Quota policy is more effective (and the Rank policy has very little bite) when r is low; the opposite is true when r is closer to 0.5.[34]

3.2 Policies targeting candidates [In progress]

The entire dynamics of the model are driven by the initial imbalance in the samples of candidates. Policies that operate on the samples of candidates directly seem then naturally preferable. For example, if role model effects were inverted–the set of candidates increases by

[34] With the parameters used in the figure, the relative effectiveness of the two policies changes sign at r just below 0.25. The crossing point is barely sensitive to n (it converges to 0.25 as n approaches infinity), but moves significantly with v, falling as v rises. Because the policies we are considering are only one possible example, we do not attempt to reach broader analytical results.

a man whenever a woman is hired, and by a woman whenever a man is hired[35]– convergence to equality would be rapid. The challenge here is to devise policies that can realistically affect candidacies, a practical problem that recruiters are very well aware of.

In the literature on Pólya urns, recent papers study models with unbalanced reinforcement rules: rules where, for example, two balls are drawn at each cycle, and two balls are added, but at least one of the two additional balls must be w. Or rules such that drawing a w ball leads to adding x new w balls; drawing a m ball, leads to adding y new m balls, with $x \neq y$.[36] We are investigating the applicability of these approaches to our problem. But note again the difficulty of practical implementation.

One intervention that is feasible, commonly recommended and indeed, to our experience, at least occasionally followed is to recommend or require that hiring interviews following candidacies include an equal number of women and men candidates. This is not unrelated to the parable with which we started this paper, and as in the parable, it is typically ineffective. The simple Pólya model clarifies the reason: given the samples of candidates, it is natural to interview the top ranked applicants. And by the logic we have followed throughout this analysis, if the candidate samples are unbalanced, the top applicants from the larger sample will dominate, in expectation, the top applicants from the smaller sample. Interviewing an equal number of candidates from the two samples will not correct the problem. Unless, that is, the interviews are assigned randomly. Writing at the end of the academic job market in Economics, it is difficult not to comment that random interviews would spare everybody a lot of time, and yes, they would be fair. They would also, however, be highly inefficient.

4 Appendix

4.1 Proofs

Result 1. *The probability that the successful candidate is a woman is a martingale:* $E(r_{t+1}|r_t) = r_t$.

Proof. The proof is immediate, once we keep in mind that $w_t + m_t$ evolves deterministically and r_t is the probability the following draw is a woman. Hence:

$$E[r_{t+1}|r_t] = \frac{(w_t+1)r_t + w_t(1-r_t)}{w_t+m_t+1} = \frac{w_t+r_t}{w_t+m_t+1}$$

But:

$$w_t = r_t(w_t+m_t)$$

Hence:

$$E[r_{t+1}|r_t] = \frac{r_t(w_t+m_t+1)}{w_t+m_t+1} = r_t.\square$$

Remark 1 $E(\rho_t) = r_0$.

[35]What the literature calls opposite reinforcement. See also the example in Brooks (2015)

[36]The rules can be deterministic or probabilistic. See for example Crimaldi et al. (2022); Aguech, Lasmar, and Selmi (2019); Lasmar, Mailler, and Selmi (2021).

Proof. Note that, by the law of motion of the sample of applicants, $\widehat{w}_t = w_t - w_0$, and $w_t = r_t(w_0 + m_0 + t)$. Hence:

$$\rho_t = \frac{\widehat{w}_t}{t} = r_t(\frac{w_0 + m_0 + t}{t}) - \frac{w_0}{t} = r_t(1 + \frac{w_0 + m_0}{t}) - \frac{w_0}{t} \tag{2}$$

It then follows that:

$$E(\rho_t) = E(r_t)(1 + \frac{w_0 + m_0}{t}) - \frac{w_0}{t} = r_0 + \frac{w_0}{w_0 + m_0}\frac{w_0 + m_0}{t} - \frac{w_0}{t} = r_0.\square$$

Equation 2 also makes clear that the limit behavior of ρ_t is pinned down by the limit behavior of r_t.

Result 4. *Suppose the single most able candidate is hired at every hiring cycle t. Each hire results in one additional W candidate with probability ρ_t, and one additional M candidate with probability $1 - \rho_t$. Then: $E(r_t) = r_0 \ \forall t$ and $E(\rho_t) = r_0 \ \forall t$.*

Proof. When the dynamics of the samples of candidates depend on the stock of hire, it is important to be very clear about timing. Recall that at each time t, the process proceeds in two steps. In the first step, one of the candidates is selected; this determines the number of women in the total stock of candidates at t, $\widehat{w}_t$,and their ratio among all hires, ρ_t. In the second step, a new set of candidates emerges, formed by w_t women and m_t men, and such that the ratio of women is r_t. By the laws of motion of the system:

$$E(\rho_t|\rho_{t-1}, r_{t-1}) = \frac{\rho_{t-1}(t-1) + r_{t-1}}{t}$$
$$E(r_t|\rho_t, r_{t-1}) = \frac{r_{t-1}(n_t - 1) + \rho_t}{n_t}$$

or:

$$E(\rho_t|\rho_{t-1}, r_{t-1}) = \left(\frac{t-1}{t}\right)\rho_{t-1} + \frac{1}{t}r_{t-1}$$
$$E(r_t|\rho_t, r_{t-1}) = \left(\frac{n_t - 1}{n_t}\right)r_{t-1} + \frac{1}{n_t}\rho_t \tag{3}$$

Hence $E(\rho_1) = E(r_1) = r_0$, and similarly, $E(\rho_2) = (1/2)\left(E(\rho_1) + E(r_1)\right) = r_0$. Thus, by induction, $E(\rho_t) = r_0$ for all t, and $E(r_t) = r_0 \ \forall t$. $\square$.

Result 5. *(i) For all $p < 1$ and $r_t < 1/2$, $E(r_{t+1}|r_t) > r_t$; (ii) For all $p \in [1/2, 1)$, $E\left(\lim_{t\to\infty} r_t\right) = E\left(\lim_{t\to\infty} \rho_t\right) = 1/2$.*

Proof of Result 5 (i).

Given $Pr(s_{t+1} \in W_t) = w_t/n_t = r_t$, it follows that:

$$
\begin{aligned}
r_{t+1} &= r_t\left(p\frac{w_t+1}{n_t+1}+(1-p)\frac{w_t}{n_t+1}\right)+(1-r_t)\left(p\frac{w_t}{n_t+1}+(1-p)\frac{w_t+1}{n_t+1}\right)\\
&= r_t\left(\frac{w_t}{n_t+1}\right)+(1-r_t)\left(\frac{w_t}{n_t+1}+\frac{1-p}{n_t+1}\right)\\
&= \frac{w_t}{n_t+1}+\left(\frac{r_tp+(1-r_t)(1-p)}{n_t+1}\right)\\
&= r_t+\left(\frac{r_t(2p-2)+(1-p)}{n_t+1}\right)\\
&= r_t+\frac{(1-p)(1-2r_t)}{n_t+1} > r_t \ \forall p<1 \text{ and } r_t<1/2.
\end{aligned}
$$

Result 9. *For all n, There exists $\underline{r}(n) < 2/5$ such that for all $r \in (\underline{r}(n), 1/2]$, $E_t(\Delta\widehat{w}_2|R_2^2) > E_t(\Delta\widehat{w}_2|Q2)$, and for all $r < \underline{r}(n)$, $E_t(\Delta\widehat{w}_2|R_2^2) < E_t(\Delta\widehat{w}_2|Q2)$.* (The proof is in the appendix).

Proof. The proof proceeds in stages.

1. We begin by considering $Q2$. Given the constraint that one of the two future hires must be a woman, the question is how the employer proceeds in the first period. For ease of notation, we denote the two hiring periods by 1 and 2, and by $n_1 = m_1 + w_1$ the set of candidates entering hiring period 1. In period 1, the employer hires a woman if the best woman ranks not higher[37] than the expected rank of the highest aptitude woman at 2, when the period 1 hire is a man (and thus when $m_2 = m_1 + 1$). Denote such expected rank by $E(\phi_{w,2}|(m_1+1))$. Note that it corresponds to 1 plus the expected number of draws of mauve balls before the first white ball is drawn when the number of mauve balls is m_1+1, the number of white balls is w_1 and the total number of balls is $m_1+1+w_1 = n_1+1$. Thus:

$$E(\phi_{w,2}|(m_1+1)) = 1+\sum_{s=1}^{m_1+1} s\left(Pr(X=0|s)-Pr(X=0|s+1)\right)$$

where $Pr(X=0|s)$ the probability of no successes (no w) in the first s draws. Or:

$$Pr(X=0|s) = \binom{m}{s}/\binom{w+m}{s}$$

and correspondingly for $Pr(X=0|s+1)$.

Simplifying the binomial terms:[38]

$$E(\phi_{w,2}|(m_1+1)) = 1+\frac{1+m_1}{1-m_1+n_1}$$

or:

$$E(\phi_{w,2}|(m_1+1)) = \left(\frac{2+n_1}{1+r_1n_1}\right)$$

[37] Recall that high rank means higher order statistics, or lower aptitude.

[38] See for example the steps in https://mikespivey.wordpress.com/2018/09/

It then follows that $E(\phi_{w,2}|(m_1+1)) \in [2,3)$ for all $r_1 \in [1/3, 1/2]$. Over this range of r_1 values, the employer under $Q2$ hires a woman in period 1 if the woman has rank 1 or 2. Thus, if $r_1 \in [1/3, 1/2]$, the first period hiring rule is identical under $Q2$ and $R2^2$.

2. Consider then $r_1 \in [1/3, 1/2]$. The two policies differ in period 2. Call $\phi_{w,1}$ the rank of the highest aptitude woman in period 1. There are two possibilities: (i) If $\phi_{w,1} \in \{1,2\}$, then under $Q2$ a woman is hired in period 2 only if $\phi_{w,2}|m_1 = 1$. Under $R2^2$, on the other hand, a woman is hired at 2 if $\phi_{w,2}|m_1 \in \{1,2\}$. Thus, if $\phi_{w,1} \in \{1,2\}$ and $\phi_{w,2}|m_1 = 2$, $R2^2$ hires one more woman than $Q2$ over the 2-period horizon. This occurs with probability:

$$p_1(\frac{m_1}{n_1+1}\frac{w_1+1}{n_1}) = p_1\left[(1-r_1)\left(\frac{r_1 n_1+1}{n_1+1}\right)\right]$$

where p_1 is the probability a woman is hired in period 1, or:

$$p_1 = \frac{w_1}{n_1} + \frac{m_1}{n_1}(\frac{w_1}{n_1-1}) = r_1\left[1+(1-r_1)\frac{n_1}{n_1-1}\right]$$

(ii) If $\phi_{w,1} > 2$, then under both $Q2$ and $R2^2$, the first period hire is a man. In period 2, $Q2$ always hires a woman, while $R2^2$ hires a woman only if $\phi_{w,2}|(m_1+1) \in \{1,2\}$. Thus $Q2$ hires one more woman than $R2$ with probability:

$$(1-p_1)\left(\frac{m_1+1}{n_1+1}\right)\left(\frac{m_1}{n_1}\right) = (1-p_1)(1-r_1)\left[1-\left(\frac{r_1 n_1}{n_1+1}\right)\right]$$

Considering both scenarios, it then follows that $R2^2$ is expected to lead to more women hires if:

$$p_1(r_1 n_1+1) > (1-p_1)[n_1(1-r_1)+1]$$

Substituting p_1 from above, the condition becomes:

$$\left(\frac{1-r_1}{n_1^2-1}\right)[1-2r_1-n_1(n_1+n_1(r_1-3)r_1+2(r_1-1)r_1)] > 0$$

Expression 4.1 is increasing in r_1 everywhere and crosses 0 at some $r_1 = \underline{r}_1(n_1)$.[39] $\underline{r}_1(n_1)$ is strictly increasing in n_1, with $\underline{r}_1(n_1) > 1/3$ for all $n_1 \geq 5$ and $\lim_{n_1\to\infty}\underline{r}_1(n_1) = 0.38 < 2/5$. It follows that for any n_1 and $r_1 \in [1/3, 1/2)$ there exists $\underline{r_1}(n_1) \in (1/3, 2/5)$ such that $E_t(\Delta\widehat{w}_2|R_2^2) > E_t(\Delta\widehat{w}_2|Q2)$ for all $r_1 > \underline{r_1}$, $E_t(\Delta\widehat{w}_2|R_2^2) < E_t(\Delta\widehat{w}_2|Q2)$ for all $r_1 \in [1/3, \underline{r_1})$, and $E_t(\Delta\widehat{w}_2|R_2^2) = E_t(\Delta\widehat{w}_2|Q2)$ at $r_1 = \underline{r_1}$.

3. Consider now $r_1 < 1/3$. Hiring decisions in period 1 may now differ between the two policies. There exists a $\tilde{\phi}_{w,1}(r_1) > 2$ such that $Q2$ hires a woman in period 1 while $R2^2$ does not.
(i) Suppose first that either $\phi_{w,1} \in \{1,2\}$, or $\phi_{w,1} > \tilde{\phi}_{w,1}$. In these two cases, the two policies lead to the same hiring choice in period 1, and exactly the same comparisons dis-

[39] $\underline{r}_1(n) = -\left(2-n_1(2+3n_1)+\sqrt{4+n_1^2(-4+n_1(4+5n_1))}\right)/(2n(2+n))$.

cussed above apply. Because expression 4.1 is everywhere increasing in r_1 and equals 0 at $\underline{r_1}(n_1) > 1/3$, it follows that, conditional on $\phi_{w,1} \in \{1,2\}$, or $\phi_{w,1} > \tilde{\phi}_{w,1}$ and $r_1 < 1/3$, $E_t(\Delta\widehat{w}_2|R2_2^2) < E_t(\Delta\widehat{w}_2|Q2)$. (ii) Suppose then $\phi_{w,1} = \tilde{\phi}_{w,1}$. In this case, in period 1 $Q2$ hires a woman candidate while $R2^2$ does not; in period 2, $Q2$ hires a second woman candidate if $\phi_{w,2}|(m_1)) = 1$; $R2^2$ hires a woman if $\phi_{w,2}|(m_1+1)) \in \{1,2\}$. Thus, conditional on $\phi_{w,1} = \tilde{\phi}_{w,1}$, $E_t(\Delta\widehat{w}_2|R2_2^2) < E_t(\Delta\widehat{w}_2|Q2)$. It then follows that $E_t(\Delta\widehat{w}_2|R2_2^2) < E_t(\Delta\widehat{w}_2|Q2)$ for all $r_1 < 1/3$. But we established earlier that $E_t(\Delta\widehat{w}_2|R2_2^2) < E_t(\Delta\widehat{w}_2|Q2)$ for all $r_1 \in [1/3, \underline{r_1})$. Hence $E_t(\Delta\widehat{w}_2|R2_2^2) < E_t(\Delta\widehat{w}_2|Q2)$ for all $r_1 < \underline{r_1}$. □

Result 10. *For all $n \geq 5$, if $E_t(\Delta\widehat{w}_2|R2_2^2) = E_t(\Delta\widehat{w}_2|Q2)$, then $E_t(\overline{\phi}_2|R2^2) < E_t(\overline{\phi}_2|Q2)$. At equal expected numbers of women hires, $R2^2$ is expected to be a more selective policy than $Q2$.*

Proof. To evaluate the two policies at equal expected women hires, by Result 9 we need to evaluate them at $r = \underline{r_1}(n_1) \in (1/3, 1/2)$ for all $n_1 \geq 5$. Hence we know that, for all $n_1 \geq 5$ under both policies a woman is hired in period 1 if $\phi_{w,1} \in \{1,2\}$. In period 2, $R2^2$ again hires a woman if $\phi_{w,2} \in \{1,2\}$; $Q2$'s hire depends on the outcome at period 1: if a man was hired, then $Q2$ always hires a woman in period 2; if a woman was hired, then in period 2 $Q2$ only hires a woman if $\phi_{w,2} = 1$. It then follows that:

$$E_t(\overline{\phi}_2|Q2) = \left[1\frac{w_1}{n_1} + 2\left(\frac{m_1}{n_1}\right)\left(\frac{w_1}{n_1-1}\right)\right] + 1\left(\frac{w_1+1}{n_1+1}\right) + \left[1 - \frac{w_1}{n_1} - \left(\frac{m_1}{n_1}\right)\left(\frac{w_1}{n_1-1}\right)\right]\left(\frac{2+n_1}{1+r_1n_1}\right)$$

and, conditional on hiring:

$$E_t(\overline{\phi}_2|R2^2) = \left[1\frac{w_1}{n_1} + 2\left(\frac{m_1}{n_1}\right)\left(\frac{w_1}{n_1-1}\right)\right] + \left[1\left(\frac{w_1+1}{n_1+1}\right) + 2\left(\frac{m_1}{n_1+1}\right)\left(\frac{w_1+1}{n_1}\right)\right]$$

Substituting $w_1 = r_1n_1$ and $m_1 = (1-r_1)n_1$ and simplifying, we find:

$$E_t(\overline{\phi}_2|R2^2) < E_t(\overline{\phi}_2|Q2) \Longleftrightarrow 2\left(\frac{r_1n_1+1}{n_1+1}\right) < \left(1 - \frac{r_1n_1}{n_1-1}\right)\left(\frac{2+n_1}{1+rn_1}\right)$$

Setting $r_1 = \underline{r}_1(n_1)$ and again simplifying, we find that the inequality is satisfied at $\underline{r}_1(n_1)$ if and only if the following condition holds:

$$F(n_1) \equiv 2 + n_1(2+5n_1) - 3\sqrt{4 + n_1^2(n_1(4+5n_1)-4)} < 0$$

$F(n_1)$ is everywhere decreasing in n_1, equals -0.613 at $n_1 = 5$ and limits to $(5-3\sqrt{5})/2$ as n_1 goes to infinity. Hence, for any $n_1 \geq 5$, $F(n_1) < 0$ and thus $E_t(\overline{\phi}_2|R2^2) < E_t(\overline{\phi}_2|Q2)$ at $r_1 = \underline{r}_1(n_1)$. Since, by Result 9, $r_1 = \underline{r}_1(n_1)$ is the only value of r_1 at which $E_t(\Delta\widehat{w}_2|R2_2^2) = E_t(\Delta\widehat{w}_2|Q2)$, Result 10 is proven. □

4.2 Policies affecting hires: comparing $Q2$ and $R2^2$ when the role model effect is probabilistic

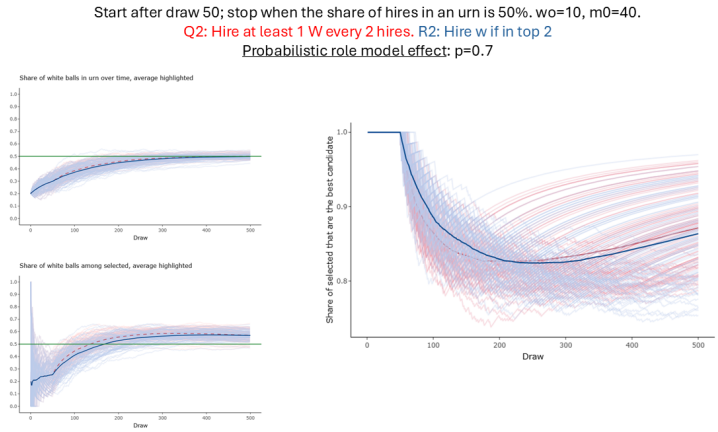


Figure 9: $Q2$ *(red) and* $R2^2$ *(blue),* $w_0 = 10$, $m_0 = 40$. *The policy starts after 50 draws and ends for each urn the first time* $r_i = 0.5$.

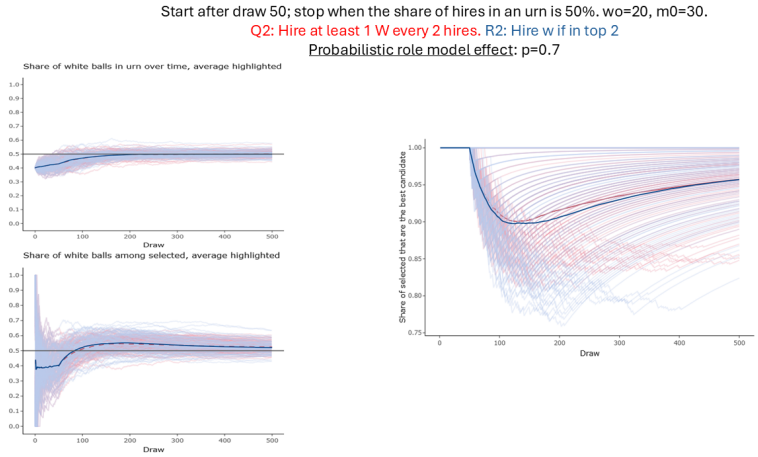


Figure 10: $Q2$ *(red) and* $R2^2$ *(blue),* $w_0 = 20$, $m_0 = 30$. *The policy starts after 50 draws and ends for each urn the first time* $r_i = 0.5$.

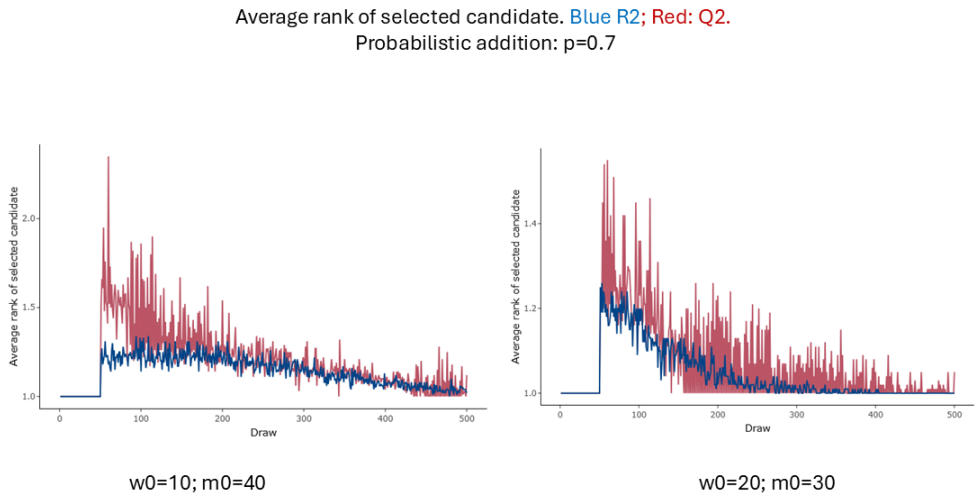


Figure 11: *Comparing Q2 (red) and R2 (blue).* Average rank of hired candidate; probabilistic role model effect ($p = 0.7$). Left: $w_0 = 10$, $m_0 = 40$. Right: $w_0 = 20$, $m_0 = 30$. The policy starts after 50 draws and ends for each urn the first time $r_i = 0.5$.

5 References

AAUP, 2025, Data Snapshot: Women Faculty and Faculty of Color, Fall 2023, based on IPEDS data.
https://www.aaup.org/academe/issues/fall-2025/data-snapshot-women-faculty-and-faculty-color-fall-2023

Aguech, R., N., Lasmar and O. Selmi, 2019, "A generalized urn with multiple drawing and random addition", *Annals of the Institute of Statistical Mathematics* 71, 389-408.

American Economic Association: CSWEP, 2024, *The 2024 Report of the Committee on the Status of Women in the Economics Profession,*
https://www.aeaweb.org/content/file?id=22178

American Economic Association: CSWEP, 2021, *Annotated Bibliography on Research Related to Women in the Economics Profession*,
https://www.aeaweb.org/about-aea/committees/cswep/survey/related-literature, compiled by D. Ginther and J. Stefanov.

Aoudia, D. A., and F. Perron, 2012, "A New Randomized Polya Urn Process", *Applied Mathematics*, 3, 2118-2122.

Arthur, B., Yu. M. Ermoliev and Yu. M. Kaniovski, 1987, Path-dependent processes and the emergence of macro-structure, *European Journal of Operational Research*, 30, 294-303.

Arthur, B., 1989, "Competing Technologies, Increasing Returns, and Lock-In by Historical Events", *Economic Journal*, 99, 116-131.

Arthur, B., 1994, *Increasing Returns and Path Dependence in the Economy*, University

of Michigan Press: Ann Arbor, Mi.

Athey, S., C. Avery, and P. Zemsky, 2000, "Mentoring and Diversity", *American Economic Review*, 90, 765-786.

Bagchi, A. and Pal, A., 1985, "Asymptotic normality in the generalized Pólya-Eggenberger urn model with applications to computer data structures", *SIAM Journal on Algebraic and Discrete Methods*, 6, 394–405.

Boustan, L and A. Langan, 2019, "Variation in Women's Success across PhD Programs in Economics" *Journal of Economic Perspectives*, 33, 23-42.

Brooks, R. 2015, The Banality of Racial Inequality (a review of *Reproducing Racism: How Everyday Choices Lock in White Advantage* by Daria Roithmayr, New York: NYU Press, 2014), *Yale Law Journal*, 124, 2626-2662.

Calsamiglia, C., Franke and P. Rey-Biel, 2013, "The Incentive Effects of Affirmative Action in Real-Effort Tournament", *Journal of Public Economics*, 98, 15-31.

Carrell, S. E., M. E. Page, and J. E. West, 2010, "Sex and Science: How Professor Gender Perpetuates the Gender Gap", *The Quarterly Journal of Economics*, 125, 1101-1144.

Chen, M-R. and M. Kuba, 2013, "On Generalized Pölya Urn Models", *Journal of Applied Probability*, 50, 1169-1186.

Chung, K.-S., 2000, "Role Models and Arguments for Affirmative Action", *The American Economic Review*, 90: 640-648.

Cohen, J., 1976, "Irreproducible Results and the Breeding of Pigs (Or Nondegenerate Limit RandomVariables in Biology)", *BioScience*, 26, 391-394.

Crimaldi, I., P.-Y. Louis and Ida G. Minelli, 2022, "An Urn Model with Random Multiple Drawing and Random Addition", *Stochastic Processes and their Applications*, 147, 270-299.

David, P. 1985, "Clio and the Economics of QWERTY", *American Economic Review*, 75, 332-337.

Dennehy, T. and N. Dasgupta, 2017. "Female Peer Mentors Early in College Increase Women's Positive Academic Experiences and Retention in Engineering", *Proceedings of the National Academy of Sciences* 114: 5964-69.

Dosi, G,. Y. Ermoliev and Y. Kaniovski, 1994, "Generalized Urn Schemes and Technological Dynamics", *Journal of Mathematical Economics* 23, 1-19.

Eggenberger, F. and Pólya, G., 1923, "Über die statistik verketteter vorgänge", *Zeitschrift für Angewandte Mathematik und Mechanik*, 3, 279-289.

England, P., A. Levine, and E. Mishel, 2020, "Progress toward gender equality in the United States has slowed or stalled", *Proceedings of the National Academy of Sciences*, 117, 6990-6997.

Freedman, D., 1965, "Bernard Friedman's urn", *The Annals of Mathematical Statistics*, 36, 956–970.

Friedman, B., 1949, "A simple urn model", *Communications of Pure and Applied Mathematics*, 2, 59–70.

Gaule, P., and M. Piacentini, 2018, "An advisor like me? Advisor gender and postgraduate careers in science", *Research Policy,* 47, 805-813.

Griffith, A., 2014, "Faculty Gender in the College Classroom: Does It Matter for Achievement and Major Choice?", *Southern Economic Journal*, 81, 211-231.

Kofoed, M. and E. McGovney, 2019, "The Effect of Same-Gender or Same-Race Role Models on Occupation Choice: Evidence from Randomly Assigned Mentors at West Point", *Journal of Human Resources*, 54, 430-467.

Kuba, M. and H. M. Mahmoud, 2017, "Two-color balanced affine urn models with multiple drawings", *Advances in Applied Mathematics*, 90, 1-26.

Kurtulus, F. A., and D. Tomaskovic-Devey, 2012, "Do Female Top Managers Help Women to Advance? A Panel Study Using EEO-1 Records", *The Annals of the American Academy of Political and Social Science,* 639, 173-197.

Johnson, N. and Kotz, S., 1977, *Urn models and Their Applications: An Approach to Modern Discrete Probability Theory*, Wiley: New York.

Lasmar, N., C. Mailler and O. Selmi, 2021, "Multiple drawing multi-colour urns by stochastic approximation", arXiv:1611.09090v6 [math.PR].

Mahmoud, H. M., 2008, *Pólya Urn Models*, Chapman and Hall Texts in Statistical Science Series, Taylor & Francis Group: Boca Raton, FL.

Matsa, D. and A. R. Miller, 2011, "Chipping away at the Glass Ceiling: Gender Spillovers in Corporate Leadership", *American Economic Review*, 101, 635-639.

Miller, A. R., and C. Segal, 2012, "Does Temporary Affirmative Action Produce Persistent Effects? A Study of Black and Female Employment in Law Enforcement." *The Review of Economics and Statistics*, 94, 1107-1125.

Miller, C.. 2017, "The Persistent Effect of Temporary Affirmative Action", *American Economic Journal: Applied Economics*, 9, 152-190.

Morcrette, B. and H. M. Mahmoud, 2018, "Exactly Solvable Balanced Tenable Urns with Random Entries via the Analytic Methodology", 23rd International Meeting on Probabilistic, Combinatorial, and Asymptotic Methods in the Analysis of Algorithms (AofA'12), Jun 2012, Montreal, Canada. pp.219–232. arXiv:1207.4948v1 [math.CO].

Müller-Itten, M. and A. Öry, 2022, "Mentoring and the Dynamics of Affirmative Action", *American Economic Journal: Economic Policy*, 14, 402-444.

Niederle, M, C. Segal and L. Vesterlund, 2013, "How Costly Is Diversity? Affirmative Action in Light of Gender Differences in Competitiveness", *Management Science*, 59, 1-16.

OECD, 2025, *Gender Equality in a Changing World: Taking Stock and Moving Forward, Gender Equality at Work*, OECD Publishing, Paris, https://doi.org/10.1787/e808086f-en.

Pemantle, R., 1990, "A Time-Dependent Version of Polya's Urn", *Journal of Theoretical Probability*, 3, 627-637.

Pemantle, R., 2007, "A survey of random processes with reinforcement", *Probability Surveys*, 4, 1-79.

Pólya, G., 1931, "Sur quelques points de la théorie des probabilités", *Annales de l'Institut Henri Poincaré*, 1, 117-161.

Porter, C., and D. Serra, 2020, "Gender Differences in the Choice of Major: The Importance of Female Role Models", *American Economic Journal: Applied Economics*, 12, 226-254.

Schotter, A. and K. Weigelt, 1992, "Asymmetric Tournaments, Equal Opportunity Laws, and Affirmative Action: Some Experimental Results", *Quarterly Journal of Economics*, 107, 511-539.

Smythe, R., 1996, "Central limit theorems for urn models", *Stochastic Processes and Their Applications*, 65, 115–137.

Spivey, Mike, ongoing, "A Math Blog: A Narrow Margin": https://mikespivey.wordpress.com/.

Stepanova, E., G. Dosi and A. Moneta, 2019, "Dynamic Increasing Returns and Innovation Diffusion: bringing Polya Urn processes to the empirical data", available at SSRN: https://ssrn.com/abstract=5314575.