Algorithmic Monoculture and Social Welfare

The rise of algorithms used to shape societal choices has been accompanied by concerns over monoculture—the notion that choices and preferences will become homogeneous in the face of algorithmic curation. One of many canonical articulations of this concern was expressed in the New York Times by Farhad Manjoo [2019], who wrote, “Despite the barrage of choice, more of us are enjoying more of the same songs, movies and TV shows.” Because of algorithmic curation, trained on collective social feedback [Salganik et al. 2006], our choices are converging.

When we move from the influence of algorithms on media consumption and entertainment to their influence on high-stakes screening decisions about whom to offer a job or whom to offer a loan, the concerns about algorithmic monoculture become even starker. Even if algorithms are more accurate on a case-by-case basis, a world in which everyone uses the same algorithm is susceptible to correlated failures when the algorithm finds itself in adverse conditions. This type of concern invokes an analogy to agriculture, where monoculture makes crops susceptible to the attack of a single pathogen [Power and Follett 1987]; the analogy has become a mainstay of the computer security literature [Birman and Schneider 2009], and it has recently become a source of concern about screening decisions for jobs or loans as well. Discussing the post-recession financial system, Citron and Pasquale [2014] write, “Like monocultural-farming technology vulnerable to one unanticipated bug, the converging methods of credit assessment failed spectacularly when macroeconomic conditions changed.”

The narrative around algorithmic monoculture thus suggests a tradeoff: in “normal” conditions, a more accurate algorithm will improve the average quality of screening decisions, but when conditions change through an unexpected shock, the results can be dramatically worse. But is this tradeoff genuine? In the absence of shocks, does monocultural convergence on a single, more accurate screening algorithm necessarily lead to better average outcomes?
In this chapter, we show that algorithmic monoculture poses risks even in the absence of shocks. We investigate a model involving minimal assumptions, in which two competing firms can either use their own independent heuristics to perform screening decisions or they can use a more accurate algorithm that is accessible to both of them. (Again, we think of screening job applicants or loan applicants as a motivating scenario.) We find that even though it would be rational for each firm in isolation to adopt the algorithm, it is possible for the use of the algorithm by both firms to result in decisions that are worse on average. This in turn leads, in the language of game theory, to a type of “Braess’s paradox” [Braess 1968] for screening algorithms: the introduction of a more accurate algorithm can drive the firms into a unique equilibrium that is worse for society than the one that was present before the algorithm existed.

Note that the harm here is to overall performance. Another common concern about algorithmic monoculture in screening decisions is the harm it can cause to specific individuals: if all employers or lenders use the same algorithm for their screening decisions, then particular applicants might find themselves locked out of the market when this shared algorithm doesn’t like their application for some reason. While this is clearly also a significant concern, our results show that it would be a mistake to view the harm to particular applicants as necessarily balanced against the gains in overall accuracy—rather, it is possible for algorithmic monoculture to cause harm not just to particular applicants but also to the average quality of decisions as well.

Our results thus have a counterintuitive flavor to them: if an algorithm is clearly more accurate than the alternatives when one entity uses it, why does the accuracy become worse than the alternatives when multiple entities use it? The analysis relies on deriving some novel probabilistic properties of rankings, establishing that when we are constructing a ranking from a probability distribution representing a “noisy” version of a true ordering, we can sometimes achieve less error through an incremental construction of the ranking—building it one element at a time—than we can by constructing it in a single draw from the distribution. We now set up the basic model, and then frame the probabilistic questions that underpin its analysis.

6.1 Algorithmic Hiring as a Case Study

To instantiate the ideas introduced thus far, we’ll focus on the case of algorithmic hiring, where recruiters make decisions based in part on scores or recommendations provided by data-driven algorithms. In this setting, we’ll propose and analyze a stylized model of algorithmic hiring with which we can begin to investigate the effects of algorithmic monoculture.
Informally, we can think of a simplified hiring process as follows: rank all of the candidates and select the first available one. We suppose that each firm has two options to form this ranking: either develop their own, private ranking (which we will refer to as using a “human evaluator”) or use an algorithmically produced ranking. We assume that there is a single vendor of algorithmic rankings, so all firms choosing to use the algorithm receive the same ranking. The firms proceed sequentially, each hiring their favorite remaining candidate according to the ranking they’re using—human-generated or algorithmic. Thus, we can frame the effects of monoculture as follows: are firms better off using the more accurate, common algorithm or should they instead employ their own less accurate, but private, evaluations?

In what follows, we’ll introduce a formal model of evaluation and selection, using it to analyze a setting in which firms seek to hire candidates.

### 6.1.1 Modeling Ranking

More formally, we model the $n$ candidates as having intrinsic values $x_1, \ldots, x_n$, where any employer would derive utility $x_i$ from hiring candidate $i$. Throughout the chapter, we assume without loss of generality that $x_1 > x_2 > \cdots > x_n$. These values, however, are unknown to the employer; instead, they must use some noisy procedure to rank the candidates. We model such a procedure as a randomized mechanism $\mathcal{R}$ that takes in the true candidate values and draws a permutation $\pi$ over those candidates from some distribution. Our main results hold for families of distributions over permutations as defined below:

**Definition 6.1** Noisy permutation family

A noisy permutation family $\mathcal{F}_\theta$ is a family of distributions over permutations that satisfies the following conditions for any $\theta > 0$ and set of candidates $x$:

1. **(Differentiability)** For any permutation $\pi$, $\Pr[\pi] = \pi$ is continuous and differentiable in $\theta$.

2. **(Asymptotic optimality)** For the true ranking $\pi^*$, $\lim_{\theta \to \infty} \Pr[\pi^*] = 1$.

3. **(Monotonicity)** For any (possibly empty) $S \subset x$, let $\pi^{(-S)}$ be the partial ranking produced by removing the items in $S$ from $\pi$. Let $\pi_1^{(-S)}$ denote the value of the top-ranked candidate according to $\pi^{(-S)}$. For any $\theta' > \theta$,

$$\mathbb{E}_{\mathcal{F}_{\theta'}}[\pi_1^{(-S)}] \geq \mathbb{E}_{\mathcal{F}_{\theta}}[\pi_1^{(-S)}]. \quad (6.1)$$

Moreover, for $S = \emptyset$, (6.1) holds with strict inequality.

$\theta$ serves as an “accuracy parameter”: for large $\theta$, the noisy ranking converges to the true ranking over candidates. The monotonicity condition states that a higher
value of $\theta$ leads to a better first choice, even if some of the candidates are removed after ranking. Removal after ranking (as opposed to before) is important because some of the ranking models we will consider later do not satisfy Independence of Irrelevant Alternatives. Examples of noisy permutation families include Random Utility Models (RUMs) [Thurstone 1927] and the Mallows Model [Mallows 1957], both of which we will discuss in detail later.

As an objective function to evaluate the effects of different approaches to ranking and selection, we'll consider each individual employer's utility as well as the sum of employers' utilities. We think of this latter sum as the social welfare since it represents the total quality of the applicants who are hired by any firm. (For example, if all firms deterministically used the correct ranking, then the top applicants would be the ones hired, leading to the highest possible social welfare.)

### 6.1.2 Modeling Selection

Each firm in our model has access to the same underlying pool of $n$ candidates, which they rank using a randomized mechanism $\mathcal{R}$ to get a permutation $\pi$ as described above. Then, in a random order, each firm hires the highest-ranked remaining candidate according to their ranking. Thus, if two firms both rank candidate $i$ first, only one of them can hire $i$; the other must hire the next available candidate according to their ranking. In our model, candidates automatically accept the offer they get from a firm. For the sake of simplicity, throughout this chapter we restrict ourselves to the case where there are two firms hiring one candidate each, although our model readily generalizes to more complex cases.

As described earlier, each firm can choose to use either a private “human evaluator” or an algorithmically generated ranking as its randomized mechanism $\mathcal{R}$. We assume that both candidate mechanisms come from a noisy permutation family $F_\theta$, with differing values of the accuracy parameter $\theta$: human evaluators all have the same accuracy $\theta_H$, and the algorithm has accuracy $\theta_A$. However, while the human evaluator produces a ranking independent of any other firm, the algorithmically generated ranking is identical for all firms who choose to use it. In other words, if two firms choose to use the algorithmically generated ranking, they will both receive the same permutation $\pi$.

The choice of which ranking mechanism to use leads to a game-theoretic setting: both firms know the accuracy parameters of the human evaluators ($\theta_H$) and the algorithm ($\theta_A$), and they must decide whether to use a human evaluator or the algorithm. This choice introduces a subtlety: for many ranking models, a firm's rational behavior depends not only on the accuracy of the ranking mechanism but also on the underlying candidate values $x_1, \ldots, x_n$. Thus, to fully specify a firm's behavior, we assume that $x_1, \ldots, x_n$ are drawn from a known joint distribution $\mathcal{D}$. 
Our main results will hold for any $D$, meaning they apply even when the candidate values (but not their identities) are deterministically known.

### 6.1.3 Stating the Main Result

Our main result is a pair of intuitive conditions under which a Braess’s Paradox-style result occurs—in other words, conditions under which there are accuracy parameters for which both firms rationally choose to use the algorithmic ranking, but social welfare (and each individual firm’s utility) would be higher if both firms used independent human evaluators. Recall that the two firms hire in a random order. For a permutation $\pi$, let $\pi_i$ denote the value of the $i$th-ranked candidate according to $\pi$.

We first state the two conditions, and then the theorem based on them.

**Definition 6.2 Preference for the first position**

A candidate distribution $D$ and noisy permutation family $F_\theta$ exhibits a preference for the first position if for all $\theta > 0$, if $\pi, \sigma \sim F_\theta$,

$$\mathbb{E}[\pi_1 - \pi_2 \mid \pi_1 \neq \pi_2] > 0.$$ 

In other words, for any $\theta > 0$, suppose we draw two permutations $\pi$ and $\sigma$ independently from $F_\theta$, and suppose that the first-ranked candidates differ in $\pi$ and $\sigma$. Then the expected value of the first-ranked candidate in $\pi$ is strictly greater than the expected value of the second-ranked candidate in $\pi$.

**Definition 6.3 Preference for weaker competition**

A candidate distribution $D$ and noisy permutation family $F_\theta$ exhibits a preference for weaker competition if the following holds: for all $\theta_1 > \theta_2$, $\sigma \sim F_{\theta_1}$ and $\pi, \tau \sim F_{\theta_2}$,

$$\mathbb{E}[\pi_1(\cdot)_{\{\pi_1\}}] < \mathbb{E}[\pi_1(\cdot)_{\{\tau_1\}}].$$

Intuitively, suppose we have a higher accuracy parameter $\theta_1$ and a lower accuracy parameter $\theta_2 < \theta_1$; we draw a permutation $\pi$ from $F_{\theta_1}$; and we then derive two permutations from $\pi: \pi(\cdot)_{\{\pi_1\}}$ obtained by deleting the first-ranked element of a permutation $\sigma$ drawn from the more accurate distribution $F_{\theta_1}$, and $\pi(\cdot)_{\{\tau_1\}}$ obtained by deleting the first-ranked element of a permutation $\tau$ drawn from the less accurate distribution $F_{\theta_2}$.

Then the expected value of the first-ranked candidate in $\pi(\cdot)_{\{\pi_1\}}$ is strictly greater than the expected value of the first-ranked candidate in $\pi(\cdot)_{\{\tau_1\}}$—that is, when a random candidate is removed from $\pi$, the best remaining candidate is better in expectation when the randomly removed candidate is chosen based on a noisier ranking.
Using these two conditions, we can state our theorem.

**Theorem 6.1** Suppose that a given candidate distribution $D$ and noisy permutation family $F_\theta$ satisfy Definitions 6.2 (preference for the first position) and 6.3 (preference for weaker competition).

Then, for any $\theta_H$, there exists $\theta_A > \theta_H$ such that using the algorithmic ranking is a strictly dominant strategy for both firms, but social welfare would be higher if both firms used human evaluators.

### 6.1.4 A Preference for Independence

Before we prove Theorem 6.1, we provide some intuition for the two conditions in Definitions 6.2 and 6.3. The second condition essentially says that it is better to have a worse competitor: the firm randomly selected to hire second is better off if the firm that hires first uses a less accurate ranking (in this case, a human evaluator instead of the algorithmic ranking).

The first condition states that when two identically distributed permutations disagree on their first element, the first-ranked candidate according to either permutation is still better, in expectation, than the second-ranked candidate according to either permutation. In what follows, we'll demonstrate that this condition implies that firms in our model rationally prefer to make decisions using independent (but equally accurate) rankings.

To do so, we need to introduce some notation. Recall that the two firms hire in a random order. Given a candidate distribution $D$, let $U_s(\theta_A, \theta_H)$ denote the expected utility of the first firm to hire a candidate when using ranking $s$, where $s \in \{A, H\}$ is either the algorithmic ranking or the ranking generated by a human evaluator, respectively. Similarly, let $U_{s_1,s_2}(\theta_A, \theta_H)$ be the expected utility of the second firm to hire given that the first firm used strategy $s_1$ and the second firm uses strategy $s_2$, where again $s_1, s_2 \in \{A, H\}$. Finally, let $\pi_1, \sigma \sim F_\theta$.

In what follows, we will show that for any $\theta$,

$$
\mathbb{E} [\pi_1 - \pi_2 \mid \pi_1 \neq \sigma] > 0 \iff U_{AH}(\theta, \theta) > U_{AA}(\theta, \theta).
$$

In other words, whenever a ranking model meets Definition 6.2, the firm chosen to select second will prefer to use an independent ranking mechanism from its competitor, given that the ranking mechanisms are equally accurate.

First, we can write

$$
U_{AH}(\theta_A, \theta_H) = \mathbb{E} \left[ \pi_1 \cdot 1_{[\pi_1 \neq \sigma]} + \pi_2 \cdot 1_{[\pi_1 = \sigma]} \right]
$$

$$
U_{AA}(\theta_A, \theta_H) = \mathbb{E} [\sigma_2]
$$

$$
= \mathbb{E} \left[ \sigma_2 \cdot 1_{[\pi_1 \neq \sigma]} + \sigma_2 \cdot 1_{[\pi_1 = \sigma]} \right]
$$
Thus,
\[
U_{AH}(\theta_A, \theta_H) - U_{AA}(\theta_A, \theta_H) = \mathbb{E} \left[ (\pi_1 - \sigma_2) \cdot 1_{[\pi_1 = \sigma_1]} + (\pi_2 - \sigma_2) \cdot 1_{[\pi_1 = \sigma_1]} \right],
\]
Conditioned on either \(\pi_1 = \sigma_1\) or \(\pi_1 \neq \sigma_1\), \(\pi_2\) and \(\sigma_2\) are identically distributed and therefore have equal expectations. As a result,
\[
U_{AH}(\theta_A, \theta_H) - U_{AA}(\theta_A, \theta_H) = \mathbb{E} \left[ (\pi_1 - \pi_2) \cdot 1_{[\pi_1 = \sigma_1]} \right],
\]
which implies (6.2). Thus, whenever a ranking model meets Definition 6.2, firms rationally prefer independent assessments, all else equal.

To provide some intuition for what this preference for independence entails, consider a setting where a hiring committee seeks to hire two candidates. They meet, produce a ranking \(\sigma\), and hire \(\sigma_1\) (the best candidate according to \(\sigma\)). Suppose they have the option to either hire \(\sigma_2\) or reconvene the next day to form an independent ranking \(\pi\) and hire the best remaining candidate according to \(\pi\) which option should they choose? It’s not immediately clear why one option should be better than the other. However, whenever Definition 6.2 is met, the committee should prefer to reconvene and make their second hire according to a new ranking \(\pi\). After proving Theorem 6.1, we will provide natural ranking models that meet Definition 6.2, implying that under these ranking models independent re-ranking can be beneficial.

6.1.5 Proving Theorem 6.1

With this intuition, we are ready to prove Theorem 6.1.

**Proof of Theorem 6.1.** For given values of \(\theta_A\) and \(\theta_H\), using the algorithmic ranking is a strictly dominant strategy as long as
\[
U_{A}(\theta_A, \theta_H) + U_{AH}(\theta_A, \theta_H) > U_{H}(\theta_A, \theta_H) + U_{AH}(\theta_A, \theta_H) \quad (6.4)
\]
\[
U_{A}(\theta_A, \theta_H) + U_{HA}(\theta_A, \theta_H) > U_{H}(\theta_A, \theta_H) + U_{HH}(\theta_A, \theta_H) \quad (6.5)
\]
Note that (6.5) is always true for \(\theta_A > \theta_H\) by the monotonicity assumption on \(F_\theta\): 
\(U_{A}(\theta_A, \theta_H) \geq U_{H}(\theta_A, \theta_H)\) because a more accurate ranking produces a top-ranked candidate with higher expected value, and 
\(U_{HA}(\theta_A, \theta_H) \geq U_{HH}(\theta_A, \theta_H)\) because this holds even conditioned on removing any candidate from the pool (in this case, the candidate randomly selected by the firm that hires first). Crucially, in (6.5), the first firm’s random selection is independent from the second firm’s selection; the same logic could not be used to argue that (6.4) always holds for \(\theta_A \geq \theta_H\). Moreover, when \(\theta_A > \theta_H\), 
\(U_{A}(\theta_A, \theta_H) > U_{H}(\theta_A, \theta_H)\) by the monotonicity assumption, meaning (6.5) holds.
Let $W_{s_1s_2}(\theta_A, \theta_H)$ denote social welfare when the two firms employ strategies $s_1, s_2 \in \{A, H\}$. Then, when both firms use the algorithmic ranking, social welfare is

$$W_{AA}(\theta_A, \theta_H) = U_A(\theta_A, \theta_H) + U_{AA}(\theta_A, \theta_H).$$

By (6.2), Definition 6.2 implies that for any $\theta$, $U_{AA}(\theta, \theta) < U_{AH}(\theta, \theta)$, implying

$$U_A(\theta_H, \theta_H) + U_{AA}(\theta_H, \theta_H) < U_H(\theta_H, \theta_H) + U_{AH}(\theta_H, \theta_H).$$

However, by the optimality assumption on $F_\theta$ in Definition 6.1, for sufficiently large $\hat{\theta}_A$,

$$U_A(\hat{\theta}_A, \theta_H) + U_{AA}(\hat{\theta}_A, \theta_H) > U_H(\hat{\theta}_A, \theta_H) + U_{AH}(\hat{\theta}_A, \theta_H).$$

Note that $U_{s_1}(\theta_A, \theta_H)$ and $U_{s_1s_2}(\theta_A, \theta_H)$ are continuous with respect to $\theta_A$ for any $s_1, s_2 \in \{A, H\}$ since they are expectations over discrete distributions with probabilities that are by assumption differentiable with respect to $\theta_A$. Therefore, by the Differentiability assumption on $F_\theta$ from Definition 6.1, there is some $\theta_A^* > \theta_H$ such that

$$U_A(\theta_A^*, \theta_H) + U_{AA}(\theta_A^*, \theta_H) = U_H(\theta_A^*, \theta_H) + U_{AH}(\theta_A^*, \theta_H),$$

that is, given that its competitor uses the algorithmic ranking, a firm is indifferent between the two strategies. For such $\theta_A^*$, using the algorithmic ranking is still a weakly dominant strategy. By definition of $W_{AA}$,

$$W_{AA}(\theta_A^*, \theta_H) = U_H(\theta_A^*, \theta_H) + U_{AH}(\theta_A^*, \theta_H).$$

If both firms had instead used human evaluators, social welfare would be

$$W_{HH}(\theta_A^*, \theta_H) = U_H(\theta_A^*, \theta_H) + U_{HH}(\theta_A^*, \theta_H).$$

By Definition 6.3, for $\sigma \sim F_{\theta_A}$ and $\pi, \tau \sim F_{\theta_H}$,

$$\mathbb{E}\left[\pi_1^{(-[\sigma_1])}\right] < \mathbb{E}\left[\pi_1^{(-[\tau_1])}\right].$$

Note that

$$U_{AH}(\theta_A^*, \theta_H) = \mathbb{E}\left[\pi_1^{(-[\sigma_1])}\right]$$

$$U_{HH}(\theta_A^*, \theta_H) = \mathbb{E}\left[\pi_1^{(-[\tau_1])}\right]$$
Thus, Definition 6.3 implies that for \( \theta_A' > \theta_H, U_{HH}(\theta_A^*, \theta_H) > U_{AH}(\theta_A^*, \theta_H) \). As a result for \( \theta_A' > \theta_H \), using the algorithmic ranking is a weakly dominant strategy, but

\[
W_{HH}(\theta_A^*, \theta_H) = U_H(\theta_A^*, \theta_H) + U_{HH}(\theta_A^*, \theta_H) > U_H(\theta_A^*, \theta_H) + U_{AH}(\theta_A^*, \theta_H) = U_A(\theta_A^*, \theta_H) + U_A(\theta_A^*, \theta_H) = W_{AA}(\theta_A^*, \theta_H),
\]

meaning social welfare would have been higher had both firms used human evaluators.

We can show that this effect persists for a value \( \theta_A' \) such that using the algorithmic ranking is a strictly dominant strategy. Intuitively, this is simply by slightly increasing \( \theta_A' \) so the algorithmic ranking is strictly dominant. For fixed \( \theta_H \), define

\[
\begin{align*}
    f(\theta_A) &= U_A(\theta_A, \theta_H) + U_{AA}(\theta_A, \theta_H) \\
g(\theta_A) &= U_H(\theta_A, \theta_H) + U_{AH}(\theta_A, \theta_H) \\
h(\theta_A) &= U_H(\theta_A, \theta_H) + U_{HH}(\theta_A, \theta_H)
\end{align*}
\]

Because (6.5) always holds for \( \theta_A > \theta_H \), it suffices to show that there exists \( \theta_A' \) such that \( g(\theta_A') < f(\theta_A') < h(\theta_A') \). This is because \( g(\theta_A') < f(\theta_A') \) is equivalent to (6.4) and \( f(\theta_A') < h(\theta_A') \) is equivalent to \( W_{AA}(\theta_A^*, \theta_H) < W_{HH}(\theta_A^*, \theta_H) \).

First, note that \( h(\theta_A) \) is a constant, and by Definition 6.3, \( g(\theta_A) < h(\theta_A) \) for all \( \theta_A > \theta_H \). By the optimality assumption of Definition 6.1, there exists sufficiently large \( \theta_A \) such that \( f(\theta_A) > g(\theta_A) \). Recall that by definition of \( \theta_A^*, f(\theta_A^*) = g(\theta_A^*) \). Both \( f \) and \( g \) are continuous by the Differentiability assumption in Definition 6.1. Thus, there must exist some \( \theta_A^* > \theta_A^* \) such that \( g(\theta_A^*) < f(\theta_A^*) < h(\theta_A^*) \). This means that for \( \theta_A' \), using the algorithmic ranking is a strictly dominant strategy, but social welfare would still be larger if both firms used human evaluators.

### 6.2 Instantiating with Ranking Models

Thus far, we have described a general set of conditions under which algorithmic monoculture can lead to a reduction in social welfare. Under which ranking models do these conditions hold? In the remainder of this chapter, we instantiate the model with two well-studied ranking models: RUMs [Thurstone 1927] and the Mallows Model [Mallows 1957]. While RUMs do not always satisfy Definitions 6.2 and 6.3, they do under some realistic parameterizations, regardless of the candidate distribution \( D \). Under the Mallows Model, Definitions 6.2 and 6.3 are always met, meaning that for any candidate distribution \( D \) and human evaluator accuracy
\( \theta_H \), there exists an accuracy parameter \( \theta_A \) such that a common algorithmic ranking with accuracy \( \theta_A \) decreases social welfare.

### 6.2.1 Random Utility Models

In RUMs, the underlying candidate values \( x_i \) are perturbed by independent and identically distributed noise \( \varepsilon_i \sim \mathcal{E} \), and the perturbed values are ranked to produce \( \pi \). Originally conceived in the psychology literature [Thurstone 1927], this model has been well-studied over nearly a century [Daniels 1950, Block and Marschak 1960, Manski 1977, Yellott Jr. 1977, Strauss 1979, Joe 2000], including more recently in the computer science and machine learning literature [Azari Soufiani et al. 2012, 2013, Ragain and Ugander 2016, Zhao et al. 2018, Makhijani and Ugander 2019].

First, we must define a family of RUMs that satisfies the conditions of Definition 6.1. Assume without loss of generality that the noise distribution \( \mathcal{E} \) has unit variance. Then, consider the family of RUMs parameterized by \( \psi \) in which candidates are ranked according to \( x_i + \frac{\varepsilon_i}{\sqrt{\psi}} \). By this definition, the standard deviation of the noise for a particular value of \( \psi \) is simply \( 1/\sqrt{\psi} \). Intuitively, larger values of \( \psi \) reduce the effect of the noise, making the ranking more accurate. In Theorem F.1 in Section F.1, we show as long as the noise distribution \( \mathcal{E} \) has positive support on \((-\infty, \infty)\), this definition of \( \mathcal{F}_\psi \) meets the differentiability, asymptotic optimality, and monotonicity conditions in Definition 6.1. For distributions with finite support, many of our results can be generalized by relaxing strict inequalities in Definition 6.1 and Theorem 6.1 to weak inequalities.

Because RUMs are notoriously difficult to work with analytically, we restrict ourselves to the case where \( n = 3 \), that is, there are three candidates. Under this restriction, we can show that for Gaussian and Laplacian noise distributions, Definitions 6.2 and 6.3—the two conditions of Theorem 6.1—are met, regardless of the candidate distribution \( \mathcal{D} \). We defer the proof to Section F.3.

**Theorem 6.2** Let \( \mathcal{F}_\psi \) be the family of RUMs with either Gaussian or Laplacian noise with standard deviation \( 1/\psi \). Then, for any candidate distribution \( \mathcal{D} \) over three candidates, the conditions of Theorem 6.1 are satisfied.

It might be tempting to generalize Theorem 6.2 to other distributions and more candidates; however, certain noise and candidate distributions violate the conditions of Theorem 6.1. Even for three-candidate RUMs, there exist distributions for which each of the conditions is violated; we provide such examples in Section F.2.

Moreover, while Gaussian and Laplacian distributions provably meet Definitions 6.2 and 6.3 with only three candidates, this doesn’t necessarily extend to larger candidate sets. Figure 6.1 shows that Definition 6.2 can be violated under
Figure 6.1 \(U_{AH}(\theta, \theta) - U_{AA}(\theta, \theta)\) for three noise models with \(n\) candidates whose utilities are drawn from a uniform distribution with unit variance for \(n = 3\), \(n = 5\), and \(n = 15\). Note that for \(n = 15\), \(U_{AH}(\theta, \theta) - U_{AA}(\theta, \theta) < 0\) for Laplacian noise, meaning Definition 6.2 is not met.

a particular candidate distribution \(D\) for Laplacian noise with 15 candidates. This challenges the intuition that independence is preferable—under some conditions, it can actually be better in expectation for a firm to use the same algorithmic ranking as its competitor, even if an independent human evaluator is equally accurate overall. Unlike Theorem 6.2, which applies for any candidate distribution \(D\), certain noise models may violate Definition 6.2 only for particular \(D\). It is an open question as to whether Theorem 6.2 can be extended to larger numbers of candidates under Gaussian noise.

Finally, there exist noise distributions that violate Definition 6.2 for any candidate distribution \(D\). In particular, the RUM family defined by the Gumbel distribution is well-known to be equivalent to the Plackett–Luce model of ranking, which is generated by sequentially selecting candidate \(i\) with probability

\[
\frac{\exp(\beta x_i)}{\sum_{j \in S} \exp(\beta x_j)},
\]

where \(S\) is the set of remaining candidates [Luce 1959, Block and Marschak 1960]. Under the Plackett–Luce model, for any \(\theta\), \(U_{AH}(\theta, \theta) = U_{AA}(\theta, \theta)\). To see this, suppose the firm that hires first selects candidate \(i^*\). Then, the firm that hires second gets each candidate \(i\) with probability given by (6.7) with \(S = \{1, \ldots, n\}\setminus i^*\). As a result, by (6.3), if \(\pi, \sigma \sim F_\theta\),

\[
\mathbb{E}[\pi_1 - \pi_2 \mid \pi_1 \neq \sigma_1] = 0
\]
for any candidate distribution \( D \), meaning the Plackett–Luce model never meets Definition 6.2. Thus, under the Plackett–Luce model, monoculture has no effect—the optimal strategy is always to use the best available ranking regardless of competitors’ strategies.

Given the analytic intractability of most RUMs, it might appear that testing the conditions of Theorem 6.1, especially for a particular noise and candidate distributions, may not be possible; however, they can be efficiently tested via simulation: as long as the noise distribution \( E \) and the candidate distribution \( D \) can be sampled from, it is possible to test whether the conditions of Theorem 6.1 are satisfied. Thus, even if the conditions of Theorem 6.1 are not met for every candidate distribution \( D \), it is possible to efficiently determine whether they are met for any particular \( D \).

It is also interesting to ask about the magnitude of the negative impact produced by monoculture. Our model allows for the qualities of candidates to be either positive or negative (capturing the fact that a worker’s productivity can be either more or less than their cost to the firm in wages); using this, we can construct instances of the model in which the optimal social welfare is positive but the welfare under the (unique) monocultural equilibrium implied by Theorem 6.1 is negative. This is a strong type of negative result, in which suboptimality reverses the sign of the objective function, and it means that in general we cannot compare the optimum and equilibrium by taking a ratio of two non-negative quantities, as is standard in Price of Anarchy results. However, as a future direction, it would be interesting to explore such Price of Anarchy bounds in special cases of the problem where structural assumptions on the input are sufficient to guarantee that the welfare at both the social optimum and the equilibrium are non-negative. As one simple example, if the qualities for three candidates are drawn independently from a uniform distribution centered at 0, and the noise distribution is Gaussian, then there exist parameters \( \theta_A > \theta_H \) such that expected social welfare at the equilibrium where both firms use the algorithmic ranking is non-negative, and approximately 4% less than it would be had both firms used human evaluators instead.

### 6.2.2 The Mallows Model

The Mallows Model also appears frequently in the ranking literature [Lu and Boutilier 2011, Das and Li 2014] and is much more analytically tractable than RUMs. Under the Mallows Model, the likelihood of a permutation is related to its distance from the true ranking \( \pi^* \):

\[
\Pr[\pi] = \frac{1}{Z} \phi^{-d(\pi, \pi^*)},
\]  

(6.8)
where $Z$ is a normalizing constant. In this model, $\phi > 1$ is the accuracy parameter: the larger $\phi$ is, the more likely the ranking procedure is to output a ranking $\pi$ that is close to the true ranking $r$. To instantiate this model, we need a notion of distance $d(\cdot, \cdot)$ over permutations. For this, we’ll use Kendall tau distance, another standard notion in the literature, which is simply the number of pairs of elements in $\pi$ that are incorrectly ordered [Kendall 1938]. In Section F.4, we verify that the family of distributions $\mathcal{F}_\theta$ given by the Mallows Model satisfies Definition 6.1, defining $\theta = \phi - 1$ (for consistency, so $\theta$ is well-defined on $(0, \infty)$).

In contrast to RUMs, the Mallows Model always satisfies the conditions of Theorem 6.1 for any candidate distribution $D$, which we prove in Section F.5.

**Theorem 6.3** Let $\mathcal{F}_\theta$ be the family of Mallows Model distributions with parameter $\theta = \phi - 1$. Then, for any candidate distribution $D$, the conditions of Theorem 6.1 are satisfied.

![Figure 6.2](image_url)

**Figure 6.2** Regions for different equilibria. When human evaluators are more accurate than the algorithm, both firms decide to employ humans (HH). When the algorithm is significantly more accurate, both firms use the algorithm (AA). When the algorithm is slightly more accurate than human evaluators, two possible equilibria exist: (1) one firm uses the algorithm and the other employs a human (AH), or (2) both decide whether to use the algorithm with some probability $p$. The shaded portion of the green AA region depicts where social welfare is smaller at the AA equilibrium than it would be if both firms used human evaluators.
Figure 6.2 characterizes firms’ rational behavior at equilibrium in the \((\theta_H, \theta_A)\) plane under the Mallows Model. The decrease in social welfare found in Theorem 6.3 is depicted by the shaded portion of the green region labeled AA, where social welfare would be higher if both firms used human evaluators.

While the result of Theorem 6.3 is certainly stronger than that of Theorem 6.2, in that it applies to all instances of the Mallows Model without restrictions, it should be interpreted with some caution. The Mallows Model does not depend on the underlying candidate values, so according to this model monoculture can produce arbitrarily large negative effects. While insensitivity to candidate values may not necessarily be reasonable in practice, our results hold for any candidate distribution \(D\). Thus, to the extent that the Mallows Model can reasonably approximate ranking in particular contexts, our results imply that monoculture can have negative welfare effects.

### Models with Multiple Firms

Our main focus in this chapter has been on models with two competing firms. However, it is also interesting to consider the case of more than two firms; we will see that the complex and sometimes counterintuitive effects that we found in the two-firm case are further enriched by additional phenomena. Primarily, we will present the result of computational experiments with the model, exposing some fundamental structural properties in the multi-firm problem for which a formal analysis remains an intriguing open problem. For concreteness, we will focus on a model in which rankings are drawn from the Mallows model. As before, each firm must choose to order candidates according to either an independent, human-produced ranking or an algorithmic ranking common to all firms who choose it. These rankings come from instances of the Mallows model with accuracy parameters \(\phi_H\) and \(\phi_A\), respectively, as defined in (6.8).

**Braess’s Paradox for \(k > 2\) firms.** First, we ask whether the Braess’s Paradox effect persists with \(k > 2\) firms. We find that it is possible to construct instances of the problem with \(k > 2\) for which Braess’s Paradox occurs—using an algorithmic evaluation is a dominant strategy, but social welfare would be higher if all firms used human evaluators instead. Under the Mallows Model, suppose \(n = 4, k = 3, \phi_A = 2, \phi_H = 1.75\), and candidate qualities are drawn from a uniform distribution on \([0, 1]\). We find via computation that at equilibrium, each firm will rationally decide to use the algorithmic evaluator and experience utility \(\approx 0.551\), but if all firms instead used human evaluators, they would experience utility \(\approx 0.552\). Thus, Braess’s Paradox can occur for \(k > 2\) firms. Proving a generalization of Theorem 6.3, to show that
Braess’s Paradox can occur for any candidate distribution $D$ and any value of $\phi_H$ for $k > 2$ firms remains an open question.

**Sequential decision-making.** Since the equilibrium behaviors we are studying take place in a model where firms make decisions in a random order, a crucial first step is to characterize firms’ optimal behavior when making decisions *sequentially*—that is, when firms hire in a fixed, known order as opposed to a random order. In this context, consider the rational behavior of each firm: given a distribution over candidate values, which ranking should each firm use? Clearly, the first firm to make a selection should use the more accurate ranking mechanism; however, as shown previously, subsequent firms’ decisions are less clear-cut.

For a fixed number of firms, number of candidates, and distribution over candidate values, we can explore the firms’ optimal strategies over the possible space of $(\phi_H, \phi_A)$ values.

An optimal choice of strategies for the $k$ firms moving sequentially can be written as a sequence of length $k$ made up of the symbols $A$ and $H$; the $i$th term in the sequence is equal to $A$ if the $i$th firm to move sequentially uses the algorithm as its optimal strategy (given the choices of the previous $i - 1$ firms), and it is equal to $H$ if the $i$th firm uses an independent human evaluation. We can therefore represent the choice of optimal strategies, as the parameters $(\phi_H, \phi_A)$ vary, by a labeling of the $(\phi_H, \phi_A)$-plane: we label each point $(\phi_H, \phi_A)$ with the length-$k$ sequence that specifies the optimal sequence of strategies.

We can make the following initial formal observation about these optimal sequences:

**Theorem 6.4** When $\phi_H \geq \phi_A$, one optimal sequence is for all firms to choose $H$. When $\phi_H > \phi_A$, the unique optimal sequence is for all firms to choose $H$.

We prove this formally in Section F.6.1, but we provide a sketch here. When $\phi_H \geq \phi_A$, the first firm to move in sequence will simply use the more accurate strategy, and hence will choose $H$. Now, proceeding by induction, suppose that the first $i$ firms have all chosen $H$, and consider the $(i + 1)$st firm to move in sequence. Regardless of whether this firm chooses $A$ or $H$, it will be making a selection that is independent of the previous $i$ selections, and hence it is optimal for it to choose $H$ as well. Hence, by induction, it is an optimal solution for all firms to choose $H$ when $\phi_H \geq \phi_A$. (This argument, slightly adapted, also directly establishes that it is uniquely optimal for all firms to choose $H$ when $\phi_H > \phi_A$.)

Beyond this observation, if we wish to extend to the case when $\phi_A > \phi_H$, the mathematical analysis of this multi-firm model remains an open question; but it is possible to determine optimal strategies computationally for each choice of $(\phi_H, \phi_A)$, and then to look at how these strategies vary over the $(\phi_H, \phi_A)$-plane.
Figure 6.3 Regions for different optimal strategy profiles, where each strategy profile is a sequence of “A” and “H” representing the optimal strategies of each firm sequentially. For this plot, there are five firms \((k = 5)\) and six candidates \((n = 6)\) whose values are drawn from a uniform distribution. Note that when \(\phi_A\) is much larger than \(\phi_H\), all firms use the algorithmic ranking, but when \(\phi_A\) is only slightly larger than \(\phi_H\), only the first firm uses the algorithmic ranking.

Figure 6.3 shows the result of doing this—producing a labeling of the \((\phi_H, \phi_A)\)-plane as described above—for \(k = 5\) firms and \(n = 6\) candidates, with the values of the candidates drawn from a uniform distribution.

We observe a number of interesting phenomena from this labeling of the plane. First, the region where \(\phi_H \geq \phi_A\) is labeled with the all-H sequence, reflecting the argument above; for the half-plane \(\phi_A > \phi_H\), on the other hand, all optimal sequences begin with A since it is always optimal for the first firm to use the more accurate method. The labeling of the half-plane \(\phi_A > \phi_H\) becomes quite complex; in principle, any sequence over the binary alphabet \(\{A, H\}\) that begins with \(A\) could be possible, and in fact we see that all \(2^4 = 16\) of these sequences appear as labels in some portion of the plane. This means that the sequential choice of optimal strategies for the firms can display arbitrary non-monotonicities in the choice of algorithmic or human decisions, with firms alternating between them;
for example, even after the first firm chooses $A$ and the second chooses $H$, the third may choose $A$ or $H$ depending on the values $(\phi_H, \phi_A)$.

The boundaries of the regions labeled by different optimal sequences are similarly complex; some of the regions (such as $AAHHH$) appear to be bounded, while others (such as $AHHA$ and $AHHAH$) appear to only emerge for sufficiently large values of $\phi_H$.

Perhaps the most intriguing observation about the arrangement of regions is the following. Suppose we think of the sequences of symbols over $\{A, H\}$ as binary representations of numbers, with $A$ corresponding to the binary digit 1 and $H$ corresponding to the binary digit 0. (Thus, for example, $AAHHH$ would correspond to the number $16 + 8 + 4 = 28$, while $AHHA$ would correspond to the number $16 + 4 + 1 = 21$.) The observation is then the following: if we choose any vertical line $\phi_H = x$ (for a fixed $x$), and we follow it upward in the plane, we encounter regions in increasing order of the numbers corresponding to their labels, in this binary representation. (First $HHHHH$, then $AHHHH$, then $AHHHA$, then $AHHAH$, and so forth.)

We do not know a proof for this fact, or how generally it holds, but we can verify it computationally for the regions of the $(\phi_H, \phi_A)$-plane mapped out in Figure 6.3, as well as similar computational experiments not shown here for other choices of $k$ and $n$. This binary-counter property suggests a rich body of additional structure to the optimal strategies in the $k$-firm case, and we leave it as an open question to analyze this structure mathematically.

6.4 Conclusion

Concerns about monoculture in the use of algorithms have focused on the danger of unexpected, correlated shocks, and on the harm to particular individuals who may fare poorly under the algorithm’s decision. Our work here shows that concerns about algorithmic monoculture are in a sense more fundamental, in that it is possible for monoculture to cause decisions of globally lower average quality, even in the absence of shocks. In addition to telling us something about the pervasiveness of the phenomenon, it also suggests that it might be difficult to notice its negative effects even while they’re occurring—these effects can persist at low levels even without a shock-like disruption to call our attention to them. Our results also make clear that algorithmic monoculture in decision-making doesn’t always lead to adverse outcomes; rather, we given natural conditions under which adverse outcomes become possible, and show that these conditions hold in a wide range of standard models.

Our results suggest a number of natural directions for further work. To begin with, we have noted earlier in the chapter that it would be interesting to give more
comprehensive quantitative bounds on the magnitude of monoculture's possible negative effects in decisions such as hiring—how much worse can the quality of candidates be when selected with an equilibrium strategy involving shared algorithms than with a socially optimal one? In formulating such questions, it will be important to take into account how the noise model for rankings relates to the numerical qualities of the candidates.

We have also focused here on the case of two firms and a single shared algorithm that is available to both. It would be natural to consider generalizations involving more firms and potentially more algorithms as well. With more algorithms, we might see solutions in which firms cluster around different algorithms of varying accuracies, as they balance the level of accuracy and the amount of correlation in their decisions. It would also be interesting to explore the ways in which correlations in firms’ decisions can be decomposed into constituent parts, such as the use of standardized tests that form input features for algorithms, and how quantifying these forms of correlation might help firms assess their decisions.

Finally, it will be interesting to consider how these types of results apply to further domains. While the analysis presented here illustrates the consequences of monoculture as applied to algorithmic hiring, our findings have potential implications in a broader range of settings. Algorithmic monoculture not only leads to a lack of heterogeneity in decision-making; by allowing valuable options to slip through the cracks—be they job candidates, potential hit songs, or budding entrepreneurs—it reduces total social welfare, even when the individual decisions are more accurate on a case-by-case basis. These concerns extend beyond the use of algorithms; whenever decision-makers rely on identical or highly correlated evaluations, they miss out on hidden gems, and in this way diminish the overall quality of their decisions.