



Algorithmic monoculture and social welfare

Jon Kleinberg^a and Manish Raghavan^{a,1}

^aDepartment of Computer Science, Cornell University, Ithaca, NY 14853

Edited by Matthew O. Jackson, Stanford University, Stanford, CA, and approved February 3, 2021 (received for review August 31, 2020)

As algorithms are increasingly applied to screen applicants for high-stakes decisions in employment, lending, and other domains, concerns have been raised about the effects of algorithmic monoculture, in which many decision-makers all rely on the same algorithm. This concern invokes analogies to agriculture, where a monocultural system runs the risk of severe harm from unexpected shocks. Here, we show that the dangers of algorithmic monoculture run much deeper, in that monocultural convergence on a single algorithm by a group of decision-making agents, even when the algorithm is more accurate for any one agent in isolation, can reduce the overall quality of the decisions being made by the full collection of agents. Unexpected shocks are therefore not needed to expose the risks of monoculture; it can hurt accuracy even under “normal” operations and even for algorithms that are more accurate when used by only a single decision-maker. Our results rely on minimal assumptions and involve the development of a probabilistic framework for analyzing systems that use multiple noisy estimates of a set of alternatives.

monoculture | ranking | random utility model | algorithmic decision-making

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 (1), 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 (2), 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 to whom to offer a job or to 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 (3); the analogy has become a mainstay of the computer-security literature (4), and it has recently become a source of concern about screening decisions for jobs or loans as well. Discussing the postrecession financial system, Citron and Pasquale (5) 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 trade-off: 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 trade-off genuine? In the absence of shocks, does monocultural convergence on a single, more accurate screening algorithm necessarily lead to better average outcomes?

In this work, 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’ paradox” (6) 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 probabilistic properties of rankings, establishing

Significance

Algorithmic monoculture is a growing concern in the use of algorithms for high-stakes screening decisions in areas such as employment and lending. If many firms use the same algorithm, even if it is more accurate than the alternatives, the resulting “monoculture” may be susceptible to correlated failures, much as a monocultural system is in biological settings. To investigate this concern, we develop a model of selection under monoculture. We find that even without any assumption of shocks or correlated failures—i.e., under “normal operations”—the quality of decisions may decrease when multiple firms use the same algorithm. Thus, the introduction of a more accurate algorithm may decrease social welfare—a kind of “Braess’ paradox” for algorithmic decision-making.

Author contributions: J.K. and M.R. designed research, performed research, contributed new reagents/analytic tools, analyzed data, and wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission.

Published under the [PNAS license](#).

¹To whom correspondence may be addressed. Email: manish@cs.cornell.edu.

This article contains supporting information online at <https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2018340118/-/DCSupplemental>.

Published May 25, 2021.

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.

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 in a random order, each hiring their favorite remaining candidate according to the ranking they’re using—human-generated or algorithmic (see Fig. 1 for an example). 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.

Modeling Ranking. More formally, we model the n candidates as having intrinsic values x_1, \dots, x_n , where any employer would derive utility x_i from hiring candidate i . Throughout the paper, we assume without loss of generality that $x_1 > x_2 > \dots > x_n$. 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 π over those candidates from some distribution. Our main results hold for families of distributions over permutations as defined below:

Definition 1 (Noisy Permutation Family): A noisy permutation family \mathcal{F}_θ is a family of distributions over permutations that satisfies the following conditions for any $\theta > 0$ and set of candidates \mathbf{x} :

1. **Differentiability:** For any permutation π , $\Pr_{\mathcal{F}_\theta}[\pi]$ is continuous and differentiable in θ .

2. **Asymptotic optimality:** For the true ranking π^* , $\lim_{\theta \rightarrow \infty} \Pr_{\mathcal{F}_\theta}[\pi^*] = 1$.
3. **Monotonicity:** For any (possibly empty) $S \subset \mathbf{x}$, let $\pi^{(-S)}$ be the partial ranking produced by removing the items in S from π . Let $\pi_1^{(-S)}$ denote the value of the top-ranked candidate according to $\pi^{(-S)}$. For any $\theta' > \theta$,

$$\mathbb{E}_{\mathcal{F}_{\theta'}}[\pi_1^{(-S)}] \geq \mathbb{E}_{\mathcal{F}_\theta}[\pi_1^{(-S)}]. \quad [1]$$

Moreover, for $S = \emptyset$, Eq. 1 holds with strict inequality.

θ serves as an “accuracy parameter”: For large θ , the noisy ranking converges to the true ranking over candidates. The monotonicity condition states that a higher value of θ 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) (7) and the Mallows Model (8), 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.)

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 π , 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 paper, 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 \mathcal{F}_θ , with differing values of the accuracy parameter θ : Human evaluators all have the same accuracy θ_H , and the algorithm has accuracy θ_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 π .

The choice of which ranking mechanism to use leads to a game-theoretic setting: Both firms know the accuracy parameters of the human evaluators (θ_H) and the algorithm (θ_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, \dots, x_n . Thus, to fully specify a firm’s behavior, we assume that x_1, \dots, x_n are drawn from a known joint distribution \mathcal{D} . Our main results will hold for any \mathcal{D} , meaning that they apply even when the candidate values (but not their identities) are deterministically known.

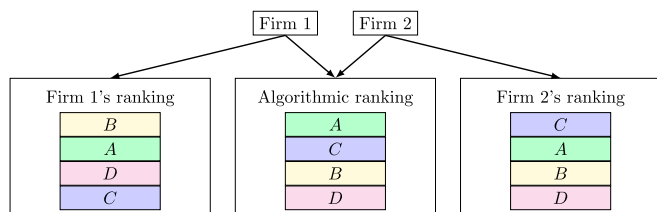


Fig. 1. Each firm has the choice to use either their own private ranking or the common algorithmic ranking to order the n candidates. In a random order, each firm hires the highest-ranked available candidate according to the ranking they chose. For example, if firm 1 uses their private ranking and firm 2 uses the algorithmic ranking, then firm 1 hires candidate B , and firm 2 hires candidate A . If both firms use the algorithmic ranking, then the firm randomly selected to hire first hires candidate A , and the firm randomly selected to hire second hires candidate C .

Stating the Main Result. Our main result is a pair of intuitive conditions under which a Braess' 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 π , let π_i denote the value of the i th-ranked candidate according to π .

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

Definition 2 (Preference for the First Position): A candidate distribution \mathcal{D} and noisy permutation family \mathcal{F}_θ exhibits a preference for the first position if for all $\theta > 0$, if $\pi, \sigma \sim \mathcal{F}_\theta$,

$$\mathbb{E}[\pi_1 - \pi_2 \mid \pi_1 \neq \sigma_1] > 0.$$

In other words, for any $\theta > 0$, suppose we draw two permutations π and σ independently from \mathcal{F}_θ , and suppose that the first-ranked candidates differ in π and σ . Then, the expected value of the first-ranked candidate in π is strictly greater than the expected value of the second-ranked candidate in π .

Definition 3 (Preference for Weaker Competition): A candidate distribution \mathcal{D} and noisy permutation family \mathcal{F}_θ , exhibits a preference for weaker competition if the following holds: For all $\theta_1 > \theta_2$, $\sigma \sim \mathcal{F}_{\theta_1}$ and $\pi, \tau \sim \mathcal{F}_{\theta_2}$,

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

Intuitively, suppose we have a higher accuracy parameter θ_1 and a lower accuracy parameter $\theta_2 < \theta_1$; we draw a permutation π from \mathcal{F}_{θ_2} ; and we then derive two permutations from π : $\pi^{(-\{\sigma_1\})}$ obtained by deleting the first-ranked element of a permutation σ drawn from the more accurate distribution \mathcal{F}_{θ_1} , and $\pi^{(-\{\tau_1\})}$ obtained by deleting the first-ranked element of a permutation τ drawn from the less accurate distribution \mathcal{F}_{θ_2} .

Then, the expected value of the first-ranked candidate in $\pi^{(-\{\tau_1\})}$ is strictly greater than the expected value of the first-ranked candidate in $\pi^{(-\{\sigma_1\})}$ —that is, when a random candidate is removed from π , 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 1. Suppose that a given candidate distribution \mathcal{D} and noisy permutation family \mathcal{F}_θ satisfy Definition 2 (preference for the first position) and Definition 3 (preference for weaker competition).

Then, for any θ_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.

A Preference for Independence. Before we prove Theorem 1, we provide some intuition for the two conditions in Definitions 2 and 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

\mathcal{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 $\pi, \sigma \sim \mathcal{F}_\theta$.

In what follows, we will show that for any θ ,

$$\mathbb{E}[\pi_1 - \pi_2 \mid \pi_1 \neq \sigma_1] > 0 \iff U_{AH}(\theta, \theta) > U_{AA}(\theta, \theta). \quad [2]$$

In other words, whenever a ranking model meets Definition 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

$$\begin{aligned} U_{AH}(\theta_A, \theta_H) &= \mathbb{E}[\pi_1 \cdot \mathbb{1}_{\pi_1 \neq \sigma_1} + \pi_2 \cdot \mathbb{1}_{\pi_1 = \sigma_1}] \\ U_{AA}(\theta_A, \theta_H) &= \mathbb{E}[\sigma_2] \\ &= \mathbb{E}[\sigma_2 \cdot \mathbb{1}_{\pi_1 \neq \sigma_1} + \sigma_2 \cdot \mathbb{1}_{\pi_1 = \sigma_1}]. \end{aligned}$$

Thus,

$$\begin{aligned} U_{AH}(\theta_A, \theta_H) - U_{AA}(\theta_A, \theta_H) &= \mathbb{E}[(\pi_1 - \sigma_2) \cdot \mathbb{1}_{\pi_1 \neq \sigma_1} + (\pi_2 - \sigma_2) \cdot \mathbb{1}_{\pi_1 = \sigma_1}]. \end{aligned}$$

Conditioned on either $\pi_1 = \sigma_1$ or $\pi_1 \neq \sigma_1$, π_2 and σ_2 are identically distributed and, therefore, have equal expectations. As a result,

$$U_{AH}(\theta_A, \theta_H) - U_{AA}(\theta_A, \theta_H) = \mathbb{E}[(\pi_1 - \pi_2) \cdot \mathbb{1}_{\pi_1 \neq \sigma_1}], \quad [3]$$

which implies Eq. 2. Thus, whenever a ranking model meets Definition 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 σ , and hire σ_1 (the best candidate according to σ). Suppose they have the option to either hire σ_2 or reconvene the next day to form an independent ranking π and hire the best remaining candidate according to π ; which option should they choose? It's not immediately clear why one option should be better than the other. However, whenever Definition 2 is met, the committee should prefer to reconvene and make their second hire according to a new ranking π . After proving Theorem 1, we will provide natural ranking models that meet Definition 2, implying that under these ranking models, independent reranking can be beneficial.

Proving Theorem 1. With this intuition, we are ready to prove Theorem 1.

Proof of Theorem 1. For given values of θ_A and θ_H , using the algorithmic ranking is a strictly dominant strategy as long as

$$U_A(\theta_A, \theta_H) + U_{AA}(\theta_A, \theta_H) > U_H(\theta_A, \theta_H) + U_{AH}(\theta_A, \theta_H), \quad [4]$$

$$U_A(\theta_A, \theta_H) + U_{HA}(\theta_A, \theta_H) > U_H(\theta_A, \theta_H) + U_{HH}(\theta_A, \theta_H). \quad [5]$$

Note that Eq. 5 is always true for $\theta_A > \theta_H$ by the monotonicity assumption on \mathcal{F}_θ : $U_A(\theta_A, \theta_H) \geq U_H(\theta_A, \theta_H)$ 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 Eq. 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 Eq. 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 Eq. 5 holds.

Let $W_{s_1 s_2}(\theta_A, \theta_H)$ 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 Eq. 2, Definition 2 implies that for any θ , $U_{AA}(\theta, \theta) < U_{AH}(\theta, \theta)$, 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 \mathcal{F}_θ in Definition 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 $U_{s_1 s_2}(\theta_A, \theta_H)$ are continuous with respect to θ_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 θ_A . Therefore, by the differentiability assumption on \mathcal{F}_θ from Definition 1, there is some $\theta_A^* > \theta_H$ such that

$$U_A(\theta_A^*, \theta_H) + U_{AA}(\theta_A^*, \theta_H) = U_H(\theta_A^*, \theta_H) + U_{AH}(\theta_A^*, \theta_H), \quad [6]$$

i.e., given that its competitor uses the algorithmic ranking, a firm is indifferent between the two strategies. For such θ_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 3, for $\sigma \sim \mathcal{F}_{\theta_{A^*}}$ and $\pi, \tau \sim \mathcal{F}_{\theta_H}$,

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

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 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

$$\begin{aligned} W_{HH}(\theta_{A^*}, \theta_H) &= U_H(\theta_{A^*}, \theta_H) + U_{HH}(\theta_{A^*}, \theta_H) \\ &> U_H(\theta_{A^*}, \theta_H) + U_{AH}(\theta_{A^*}, \theta_H) \\ &= U_A(\theta_{A^*}, \theta_H) + U_{AA}(\theta_{A^*}, \theta_H) \\ &= W_{AA}(\theta_{A^*}, \theta_H), \end{aligned}$$

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

We can show that this effect persists for a value θ'_A , such that using the algorithmic ranking is a strictly dominant strategy. Intuitively, this is simply by slightly increasing θ_A^* so the algorithmic ranking is strictly dominant. For fixed θ_H , define

$$\begin{aligned} f(\theta_A) &= U_A(\theta_A, \theta_H) + U_{AA}(\theta_A, \theta_H) \\ g(\theta_A) &= U_H(\theta_A, \theta_H) + U_{AH}(\theta_A, \theta_H) \\ h(\theta_A) &= U_H(\theta_A, \theta_H) + U_{HH}(\theta_A, \theta_H). \end{aligned}$$

Because Eq. 5 always holds for $\theta_A > \theta_H$, it suffices to show that there exists θ'_A such that $g(\theta'_A) < f(\theta'_A) < h(\theta'_A)$. This is because $g(\theta'_A) < f(\theta'_A)$ is equivalent to Eq. 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 3, $g(\theta_A) < h(\theta_A)$ for all $\theta_A > \theta_H$. By the optimality assumption of Definition 1, there exists sufficiently large $\hat{\theta}_A$ such that $f(\hat{\theta}_A) > g(\hat{\theta}_A)$. Recall that by definition of θ_A^* , $f(\theta_A^*) = g(\theta_A^*)$. Both f and g are continuous by the differentiability assumption in Definition 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 θ'_A , using the algorithmic ranking is a strictly dominant strategy, but social welfare would still be larger if both firms used human evaluators.

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 paper, we instantiate the model with two well-studied ranking models: RUMs (7) and the Mallows Model (8). While RUMs do not always satisfy Definitions 2 and 3, they do under some realistic parameterizations, regardless of the candidate distribution \mathcal{D} . Under the Mallows Model, Definitions 2 and 3 are always met, meaning that for any candidate distribution \mathcal{D} and human evaluator accuracy θ_H , there exists an accuracy parameter θ_A such that a common algorithmic ranking with accuracy θ_A decreases social welfare.

RUMs. 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 π . Originally conceived in the psychology literature (7), this model has been well-studied over nearly a century, (9–14), including more recently in the computer science and machine-learning literature (15–19).

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

Because RUMs are notoriously difficult to work with analytically, we restrict ourselves to the case where $n = 3$ —i.e., there are three candidates. Under this restriction, we can show that for Gaussian and Laplacian noise distributions, Definition 2 and 3—the two conditions of Theorem 1—are met, regardless of the candidate distribution \mathcal{D} . We defer the proof to *SI Appendix*.

Theorem 2. *Let \mathcal{F}_θ be the family of RUMs with either Gaussian or Laplacian noise with SD $1/\theta$. Then, for any candidate distribution \mathcal{D} over three candidates, the conditions of Theorem 1 are satisfied.*

It might be tempting to generalize Theorem 2 to other distributions and more candidates; however, certain noise and candidate distributions violate the conditions of Theorem 1. Even for three-candidate RUMs, there exist distributions for which each of the conditions is violated; see *SI Appendix* for examples.

Moreover, while Gaussian and Laplacian distributions provably meet Definitions 2 and 3 with only three candidates, this doesn't necessarily extend to larger candidate sets. Fig. 2 shows that Definition 2 can be violated under a particular candidate

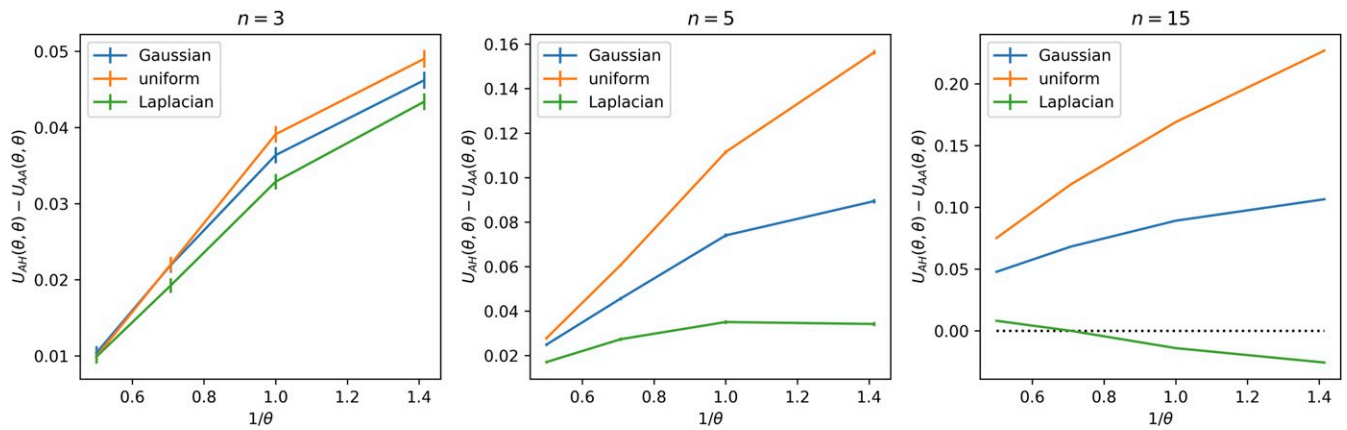


Fig. 2. $U_{AH}(\theta, \theta) - U_{AA}(\theta, \theta)$ for three noise models with n candidates whose utilities are drawn from a uniform distribution with unit variance for $n = 3$, $n = 5$, and $n = 15$. Note that for $n = 15$, $U_{AH}(\theta, \theta) - U_{AA}(\theta, \theta) < 0$ for Laplacian noise, meaning Definition 2 is not met.

distribution \mathcal{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 2, which applies for any candidate distribution \mathcal{D} , certain noise models may violate Definition 2 only for particular \mathcal{D} . It is an open question as to whether Theorem 2 can be extended to larger numbers of candidates under Gaussian noise.

Finally, there exist noise distributions that violate Definition 2 for any candidate distribution \mathcal{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

$$\frac{\exp(\theta x_i)}{\sum_{j \in S} \exp(\theta x_j)}, \quad [7]$$

where S is the set of remaining candidates (10, 20). Under the Plackett–Luce model, for any θ , $U_{AH}(\theta, \theta) = U_{AA}(\theta, \theta)$. 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 Eq. 7 with $S = \{1, \dots, n\} \setminus i^*$. As a result, by Eq. 3, if $\pi, \sigma \sim \mathcal{F}_\theta$,

$$\mathbb{E}[\pi_1 - \pi_2 \mid \pi_1 \neq \sigma_1] = 0,$$

for any candidate distribution \mathcal{D} , meaning the Plackett–Luce model never meets Definition 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 1, especially for particular noise and candidate distributions, may not be possible; however, they can be efficiently tested via simulation: As long as the noise distribution \mathcal{E} and the candidate distribution \mathcal{D} can be sampled from, it is possible to test whether the conditions of Theorem 1 are satisfied. Thus, even if the conditions of Theorem 1 are not met for every candidate distribution \mathcal{D} , it is possible to efficiently determine whether they are met for any particular \mathcal{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 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 nonnegative 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 nonnegative. As one simple example, if the qualities for three candidates are drawn independently from a uniform distribution centered at zero, 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 nonnegative and approximately

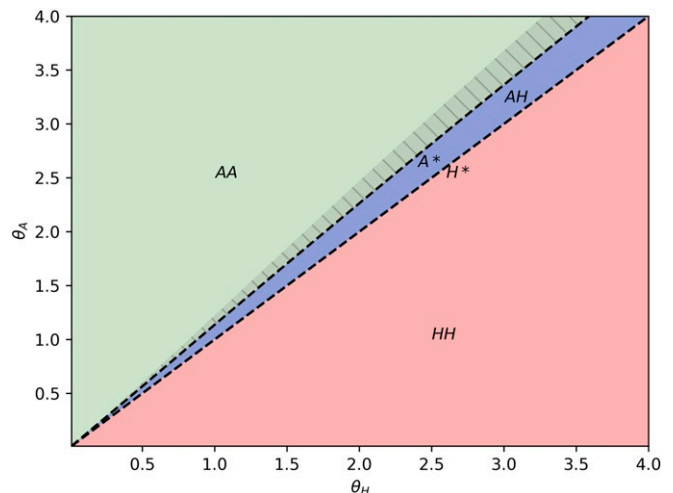


Fig. 3. Regions for different equilibria. When human evaluators are more accurate than the algorithm, both firms decide to employ humans (HH). When the algorithm is significantly more accurate, both firms use the algorithm (AA). When the algorithm is slightly more accurate than human evaluators, two 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.

4% less than it would be had both firms used human evaluators instead.

The Mallows Model. The Mallows Model also appears frequently in the ranking literature (21, 22) 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 π^* :

$$\Pr[\pi] = \frac{1}{Z} \phi^{-d(\pi, \pi^*)}, \quad [8]$$

where Z is a normalizing constant. In this model, $\phi > 1$ is the accuracy parameter: The larger ϕ is, the more likely the ranking procedure is to output a ranking π that is close to the true ranking π^* . 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 π that are incorrectly ordered (23). In *SI Appendix*, we verify that the family of distributions \mathcal{F}_θ given by the Mallows Model satisfies Definition 1, defining $\theta = \phi - 1$ (for consistency, so θ is well-defined on $(0, \infty)$).

In contrast to RUMs, the Mallows Model always satisfies the conditions of Theorem 1 for any candidate distribution \mathcal{D} , which we prove in *SI Appendix*.

Theorem 3. *Let \mathcal{F}_θ be the family of Mallows Model distributions with parameter $\theta = \phi - 1$. Then, for any candidate distribution \mathcal{D} , the conditions of Theorem 1 are satisfied.*

Fig. 3 characterizes firms' rational behavior at equilibrium in the (θ_H, θ_A) plane under the Mallows Model. The decrease in social welfare found in Theorem 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 3 is certainly stronger than that of Theorem 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 \mathcal{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.

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 give natural conditions under which such 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 paper 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.

Materials and Methods

The results in Figs. 2 and 3 were obtained by computational methods. The computational results for RUMs in Fig. 2 were performed via simulation, taking the average of 10,000,000 trials and reporting 95% CIs. Candidate utilities were drawn from a uniform distribution on $[-\sqrt{3}, \sqrt{3}]$, which is symmetric and has unit variance. The computational results for the Mallows Model in Fig. 3 were obtained symbolically by using the sympy module in Python. Candidate utilities were drawn from a uniform distribution on $[0, 1]$.

Data Availability. There are no data underlying this work.

ACKNOWLEDGMENTS. This work has been supported in part by a Simons Investigator Award, a Vannevar Bush Faculty Fellowship, a Multidisciplinary University Research Initiative grant, Air Force Office of Scientific Research Grant FA9550-19-1-0183, an NSF Graduate Research Fellowship, and grants from the Army Research Office and the MacArthur Foundation.

1. F. Manjoo, This summer stinks. But at least we've got 'Old Town Road.' *The New York Times*, 7 August 2019. <https://www.nytimes.com/2019/08/07/opinion/old-town-road.html>.
2. M. J. Salganik, P. S. Dodds, D. J. Watts, Experimental study of inequality and unpredictability in an artificial cultural market. *Science* **311**, 854–856 (2006).
3. J. Power, R. Follett, Monoculture. *Sci. Am.* **256**, 78–87 (1987).
4. K. P. Birman, F. B. Schneider, The monoculture risk put into context. *IEEE Secur. Priv.* **7**, 14–17 (2009).
5. D. K. Citron, F. Pasquale, The scored society: Due process for automated predictions. *Wash. Law Rev.* **89**, 1 (2014).
6. D. Braess, Über ein paradoxon aus der verkehrsplanung. *Unternehmensforschung* **12**, 258–268 (1968).
7. L. L. Thurstone, A law of comparative judgment. *Psychol. Rev.* **34**, 273 (1927).

8. C. L. Mallows, Non-null ranking models. I. *Biometrika* **44**, 114–130 (1957).
9. H. Daniels, Rank correlation and population models. *J. Roy. Stat. Soc. B* **12**, 171–191 (1950).
10. H. Block, J. Marschak, "Random orderings and stochastic theories of responses" in *Contributions to Probability and Statistics*, I. Olkin, S. G. Ghurye, W. Hoeffding, W. G. Madow, H. B. Mann, Eds. (Stanford University Press, Stanford, CA, 1960), pp. 97–132.
11. H. Joe, Inequalities for random utility models, with applications to ranking and subset choice data. *Methodol. Comput. Appl. Probab.* **2**, 359–372 (2000).
12. J. I. Yellott, Jr, The relationship between Luce's choice axiom, Thurstone's theory of comparative judgment, and the double exponential distribution. *J. Math. Psychol.* **15**, 109–144 (1977).
13. C. F. Manski, The structure of random utility models. *Theor. Decis.* **8**, 229 (1977).

14. D. Strauss, Some results on random utility models. *J. Math. Psychol.* **20**, 35–52 (1979).
15. H. Azari Soufiani, D. C. Parkes, L. Xia, "Random utility theory for social choice" in *NIPS'12: Proceedings of the 25th International Conference on Neural Information Processing Systems*, F. Pereira, C. J. C. Burges, L. Bottou, K. Q. Weinberger, Eds. (Curran Associates, Red Hook, NY, 2012), pp. 126–134.
16. H. Azari Soufiani, H. Diao, Z. Lai, D. C. Parkes, "Generalized random utility models with multiple types" in *NIPS'13: Proceedings of the 26th International Conference on Neural Information Processing Systems*, C. J. C. Burges, L. Bottou, M. Welling, Z. Ghahramani, K. O. Weinberger, Eds. (Curran Associates, Red Hook, NY, 2013), pp. 73–81.
17. S. Ragain, J. Ugander, "Pairwise choice Markov chains" in *NIPS'16: Proceedings of the 30th International Conference on Neural Information Processing Systems*, D. D. Lee, U. von Luxburg, R. Garnett, M. Sugiyama, I. Guyon, Eds. (Curran Associates, Red Hook, NY, 2016), pp. 3198–3206.
18. Z. Zhao, T. Villamil, L. Xia, "Learning mixtures of random utility models" in *Thirty-Second AAAI Conference on Artificial Intelligence*, S. A. McIlraith, K. Q. Weinberger, Eds. (AAAI Press, Palo Alto, CA, 2018), pp. 4530–4538.
19. R. Makhijani, J. Ugander, "Parametric models for intransitivity in pairwise rankings" in *WWW'19: The World Wide Web Conference*, L. Liu, R. White, Eds. (Association for Computing Machinery, New York, NY, 2019), pp. 3056–3062.
20. R. D. Luce, *Individual Choice Behavior: A Theoretical Analysis* (Wiley, New York, NY, 1959).
21. S. Das, Z. Li, "The role of common and private signals in two-sided matching with interviews" in *WINE 2014: International Conference on Web and Internet Economics*, T. Y. Liu, Q. Qi, Y. Ye, Eds. (Lecture Notes in Computer Science, Springer, Cham, Switzerland, 2014), vol. 8877, pp. 492–497.
22. T. Lu, C. Boutilier, "Learning Mallows models with pairwise preferences" in *ICML'11: Proceedings of the 28th International Conference on Machine Learning*, L. Getoor, T. Scheffer, Eds. (Omnipress, Madison, WI, 2011), pp. 145–152.
23. M. G. Kendall, A new measure of rank correlation. *Biometrika* **30**, 81–93 (1938).