Online Appendix for "The Managerial Effects of Algorithmic Fairness Activism"

by Bo Cowgill, Fabrizio Dell'Acqua and Sandra Matz

Tables and Figures

TABLE A1—**DESCRIPTIVE STATISTICS: SUBJECTS**

	Op-Ed	Veneer
Male	0.49	0.52
Female	0.49	0.45
Other Gender	0.02	0.02
Latinx	0.10	0.07
White	0.81	0.85
Black	0.07	0.05
Asian	0.05	0.04
Other Ethnicity	0.06	0.06
AI Decisions	0.49	0.40
Prepared by Educ.	0.75	0.71
Knows ML	0.46	0.45

Notes: All reported variables are binary. Subjects self identified their gender as "Male", "Female" or "Other". They selected their ethnic background, and whether they are Latinx. "AI Decisions" captures subjects who report working (or potentially working) in roles where they make decisions like those in our surveys. "Prepared by education" indicates whether subjects feel their education has prepared them well enough for this type of decisions. "Knows ML" takes the value 1 if subjects know "a great deal", "a lot", or "a moderate amount" about machine learning and predictive modeling, and 0 otherwise. Subjects were recruited from Prolific Academic, which is evaluated in Peer et al. (2017).

TABLE A2—STATUS QUO CONDITIONS AND DEMOGRAPHICS: STUDY 1 ("COUNTERFACTUAL" AND "FATALISM" ACTIVISM)

	Positive	Rec (Scale)	Rec (Y/N)	Lawsuit	Damage Size
Algorithm Fix	.47***	.45***	1.3e-14	-.7***	-.34***
	(.057)	(.051)	(.)	(.061)	(.044)
Status Quo Shown	.52***	.65***	.57***	-.43***	-.44***
	(.1)	(.1)	(.094)	(.095)	(.099)
Status Quo Shown (only if clicked)	.46***	.63***	.54***	-.38***	-.42***
	(.11)	(.099)	(.096)	(.092)	(.098)
Fix × Status Quo Shown	-.17**	-.26***	-2.2e-14	.19**	.12*
	(.072)	(.063)	(.)	(.084)	(.061)
Fix × Status Quo Shown (only if clicked)	-.19**	-.3***	-1.9e-14	.2**	.091
	(.076)	(.062)	(.)	(.082)	(.061)
Female	-.16**	-.11	-.15**	.11	.29***
	(.07)	(.072)	(.071)	(.066)	(.071)
Black	-.25*	-.22*	-.048	.18	.22*
	(.14)	(.13)	(.13)	(.11)	(.13)
Asian	-.17	-.027	-.084	-.026	-.16
	(.13)	(.13)	(.16)	(.14)	(.15)
Other Ethnicity	-.33***	-.26*	-.21	.083	-.086
	(.12)	(.14)	(.16)	(.14)	(.15)
Political Conservatism (Standardized)	.094**	.082**	.026	-.067**	-.079**
	(.038)	(.037)	(.036)	(.034)	(.038)
AI Decisions	.12*	.12*	.078	.025	.019
	(.069)	(.072)	(.075)	(.067)	(.074)
Prepared by Educ.	.13	.14*	.099	-.12*	-.12
	(.08)	(.083)	(.082)	(.075)	(.08)
Knows ML	.0055	.0028	-.056	.062	-.097
	(.07)	(.074)	(.074)	(.068)	(.076)
Observations	1,992	1,992	1,992	1,992	1,992
R^2	.1	.12	.12	.12	.097
p-value: Status Quo (SQ) Shown + Fix × SQ Shown=0	.0005	.00018	2.4e-09	.024	.0023
p-value: SQ Optional + Fix × SQ Opt=0	.0075	.0018	3.0e-08	.097	.0017
p-value: SQ Shown = SQ Opt	.48	.85	.74	.57	.85
p-value: Fix × SQ Shown = Fix × SQ Opt	.75	.37	1	.86	.69
p-value: SQ Shown+Fix × SQ Shown = SQ Opt+Fix × SQ Opt	.32	.47	.74	.5	.95

Notes: This table contains regression specifications described in Section B. The exact wording of survey questions used as outcome variables is in Section D. "Status Quo Shown (only if clicked)" takes the value 1 if subjects have the option to view the status quo with an extra click, and choose to click, 0 otherwise.

TABLE A3—STATUS QUO CONDITIONS AND DEMOGRAPHICS: STUDY 1 (“SCIENTIFIC VENEER”)

	Positive	Rec (Scale)	Rec (Y/N)	Lawsuit	Damage Size
Algorithm Fix	.34***	.32***	-2.9e-15	-.51***	-.23***
	(.055)	(.047)	(2.9e-09)	(.057)	(.042)
Status Quo Shown	.35***	.55***	.51***	-.21**	-.16*
	(.086)	(.086)	(.081)	(.085)	(.086)
Status Quo Shown (only if clicked)	.28***	.57***	.57***	-.27***	-.21**
	(.088)	(.087)	(.08)	(.084)	(.092)
Fix × Status Quo Shown	-.11	-.16**	3.0e-15	.087	.03
	(.072)	(.064)	(2.3e-09)	(.079)	(.056)
Fix × Status Quo Shown (only if clicked)	-.043	-.14**	5.5e-15***	.089	.023
	(.074)	(.062)	(1.5e-15)	(.078)	(.062)
Female	-.05	-.016	-.034	.0037	.1
	(.063)	(.067)	(.067)	(.066)	(.075)
Black	-.55***	-.42***	-.28**	-.032	-.048
	(.11)	(.12)	(.12)	(.17)	(.19)
Asian	-.17	-.12	-.2	-.053	.21*
	(.11)	(.12)	(.15)	(.14)	(.12)
Other Ethnicity	-.53***	-.43**	-.19	.2	.14
	(.15)	(.17)	(.15)	(.15)	(.18)
Political Conservatism (Standardized)	.16***	.13***	.091***	-.15***	-.14***
	(.032)	(.033)	(.034)	(.033)	(.04)
AI Decisions	.019	-.0022	-.058	-.0055	-.028
	(.064)	(.067)	(.069)	(.068)	(.077)
Prepared by Educ.	.17***	.08	.062	-.2***	-.16**
	(.064)	(.068)	(.073)	(.067)	(.077)
Knows ML	.12*	.15**	.13*	-.027	-.098
	(.065)	(.067)	(.068)	(.069)	(.078)
Observations	1,984	1,984	1,984	1,984	1,984
R^2	.11	.11	.086	.094	.063
p-value: Status Quo (SQ) Shown + Fix × SQ Shown=0	.0038	8.7e-06	9.4e-10	.19	.16
p-value: SQ Optional + Fix × SQ Opt=0	.0032	5.5e-07	3.5e-12	.043	.058
p-value: SQ Shown = SQ Opt	.41	.81	.4	.45	.55
p-value: Fix × SQ Shown = Fix × SQ Opt	.31	.76	1	.98	.91
p-value: SQ Shown+Fix × SQ Shown = SQ Opt+Fix × SQ Opt	.99	.63	.4	.48	.53

Notes: This table contains regression specifications described in Section B. The exact wording of survey questions used as outcome variables is in Section D. “Status Quo Shown (only if clicked)” takes the value 1 if subjects have the option to view the status quo with an extra click, and choose to click, 0 otherwise.

TABLE A4— **EFFECTS OF ACTIVISM: STUDY 1 ("COUNTERFACTUAL" AND "FATALISM" ACTIVISM)**

Panel A: Effects on Adoption Decisions

	Positive	Rec (Scale)	Rec (Y/N)	Lawsuit	Damage Size
Hiring	-.088	-.055	2.0e-17	-.04	-.28***
	(.082)	(.085)	(.)	(.087)	(.079)
Algorithm Fix	.37***	.27***	.019	-.61***	-.28***
	(.04)	(.034)	(.034)	(.045)	(.034)
Counterfactual Op-Ed	.24**	.19*	.14**	-.12	-.14
	(.11)	(.11)	(.054)	(.12)	(.11)
AI Fatalism Op-Ed	-.16	-.28***	-.079	.19*	.061
	(.11)	(.11)	(.063)	(.11)	(.094)
Fix× Counterfactual	-.042	-.0064	-.28**	.16*	.041
	(.078)	(.061)	(.11)	(.086)	(.075)
Fix× AI Fatalism	-.044	.012	.16	.1	.012
	(.075)	(.062)	(.13)	(.077)	(.058)
Fixed Effects	Subject	Subject	Subject	Subject	Subject
Observations	1,992	1,992	1,992	1,992	1,992
R^2	.68	.72	.76	.63	.73
p-value: Counterfactual (CF)=Fatalism (FAT)	.00061	.00003	.00079	.0057	.051
p-value: Fix×CF=Fix×FAT	.98	.82	.00079	.59	.74
p-value: CF+Fix×CF=FAT+Fix×FAT	.00016	.000021	.00079	.031	.11

Panel B: Effect on the Use of Counterfactual Information:

	Rec (Scale)	Rec (Scale)	Damage Size	Damage Size
Status Quo Seen (Instrumented by Op-Ed)	1.2***	.69***	-.84***	-.35*
	(.28)	(.2)	(.32)	(.21)
Instrument	CF Op-Ed	FAT Op-Ed	CF Op-Ed	FAT Op-Ed
Observations	1,992	1,992	1,992	1,992
1st Stage F-Stat	86	187	86	187

Notes: This table contains regression specifications described in Section B. The exact wording of survey questions used as outcome variables is in Section D. Insofar as the op-eds influence choices directly rather than through the examining of the status quo, these instruments may exhibit exclusion restriction limits. Our treatments to always/never/optionally show the status quo measure allow us to measure these direct effects.

TABLE A5—EFFECTS OF ACTIVISM: STUDY 2 (“SCIENTIFIC VENEER”)

Panel A: Effects on Adoption Decisions

	Positive	Rec (Scale)	Rec (Y/N)	Lawsuit	Damage Size
Hiring	-.13	-.23**	-1.3e-17	.0022	-.17*
	(.096)	(.094)	(.)	(.091)	(.093)
Algorithm Fix	.35***	.27***	.049	-.52***	-.24***
	(.04)	(.036)	(.044)	(.043)	(.036)
Scientific Veneer	.036	.093	-.21***	-.17	-.036
	(.16)	(.15)	(.079)	(.15)	(.13)
Pro-AI Argument	1.1***	1.1***	.17**	-.86***	-.48***
	(.13)	(.14)	(.071)	(.14)	(.13)
Scientific Veneer × Pro AI	-.019	-.072	.27**	.24	-.037
	(.21)	(.2)	(.11)	(.21)	(.19)
Fix × Scientific Veneer	.078	-.046	.43***	.0013	-.0084
	(.1)	(.08)	(.16)	(.1)	(.07)
Fix × Pro-AI	-.29***	-.28***	-.34**	.24**	.033
	(.073)	(.062)	(.14)	(.11)	(.07)
Fix × Scientific Veneer × Pro-AI	-.084	.23*	-.54**	.049	.1
	(.14)	(.12)	(.23)	(.16)	(.11)
Fixed Effects	Subject	Subject	Subject	Subject	Subject
Observations	1,984	1,984	1,984	1,984	1,984
R^2	.62	.67	.63	.62	.69
p-value: Veneer (Ven) +Ven×ProAI = 0	.9	.87	.49	.66	.58
p-value: Ven×Fix +Ven×ProAI×Fix = 0	.96	.027	.49	.69	.29
p-value: Ven+Ven×Fix +Ven×ProAI+Ven×ProAI×Fix = 0	.93	.11	.49	.36	.89

Panel B: Effect on the Use of Counterfactual Information:

	Views Status Quo
Scientific Veneer	-.39
	(.26)
Pro-AI Argument	-.45*
	(.26)
Scientific Veneer × Pro AI	.77*
	(.43)
Observations	334
R^2	.013

Notes: This table contains regression specifications described in Section B. The exact wording of survey questions used as outcome variables is in Section D.

Regression Specifications

B1. Study 1:"Counterfactual" and "Fatalism"

$$Y_{i,t,fix} = \beta[1(t=1) + 1(fix=1) + Hiring + FE_i + CF_{i,t} + FAT_{i,t} + CFxFix_{i,t,fix} + FATxFix_{i,t,fix}] + \varepsilon \tag{B1}$$

Where:

- $Y_{i,t,fix}$ refers to the response to question Y individual i on case t regarding before/after the 6-month effort to fix. We run separate regressions for each of the five questions categories.
- i indexes subjects (approximately 500 each for Study 1 and 2).
- $t \in \{1,2$ indexes the first or second order. $\beta_{1(t=1)}$ is a fixed effect for the first period.
- $fix \in 0,1$ differentiates answers to the questions before ($fix = 0$) or after ($fix = 1$) the six-months of dedicated effort. $\beta_{1(fix=1)}$ is a fixed effect for after the fix.
- $Hiring$ is equal to 1 for the hiring business case.
- FE_i refers to the subject-level fixed effect. Note that this subsumes the effects of a) the order of the two cases, as well as b) whether the status quo details were shown, hidden or optional.
- CF refers to the "counterfactual" condition. This is 1 for the subjects reading the counterfactual in the second case, and 0 for everyone in the first case.
- FAT refers to the "fatalism" condition. This is 1 for the subjects reading the fatalism in the second case, and 0 for everyone in the first case.
- Note that no op-ed is the excluded condition.
- $CFxFix$ is equal to $CF \times 1(fix = 1)$
- $FATxFix$ is equal to $FAT \times 1(fix = 1)$
- ε: Standard errors are clustered by subject.

B2. Study 2: "Scientific Veneer"

$$Y_{i,t,fix} = \beta[1(t=1) + 1(fix=1) + Hiring + FE_i + ProAi_{i,t} + SciVen_{i,t} + ProAixSciVen_{it} + ProAixFix_{i,t} + SciVenxFix_{i,t} + ProAixSciVenxFix_{it}] + \varepsilon \tag{B2}$$

Where:

- $Y_{i,t,fix}$ refers to the response to question Y individual i on case t regarding before/after the 6-month effort to fix. We run separate regressions for each of the five questions categories.
- i indexes subjects (approximately 500 each for Study 1 and 2).
- $t \in \{1,2$ indexes the first or second order. $\beta_{1(t=1)}$ is a fixed effect for the first period.
- $fix \in 0,1$ differentiates answers to the questions before ($fix = 0$) or after ($fix = 1$) the six-months of dedicated effort. $\beta_{1(fix=1)}$ is a fixed effect for after the fix.
- $Hiring$ is equal to 1 for the hiring business case.

- FE_i refers to the subject-level fixed effect. Note that this subsumes the effects of a) the order of the two cases, as well as b) whether the status quo details were shown, hidden or optional.
- *ProAi* refers to the condition in which a positive argument about AI ethics are made. This is 1 for the subjects reading the positive opinion in the second business case, and 0 for everyone in the first case.
- *SciVen* refers to whether the expert applied scientific veneer.. This is 1 for the subjects seeing scientific veneer in the second case, and 0 for everyone in the first case.
- *ProAixSciVen* refers to subjects who see pro-AI arguments with scientific veneer. This is 1 for the subjects reading these arguments in the second case, and 0 for everyone in the first case.
- *ProAixFix* is equal to $ProAi \times 1(fix = 1)$
- *SciVen* is equal to $SciVen \times 1(fix = 1)$
- *ProAixSciVenxFix* is equal to $ProAi \times 1(fix = 1)$
- Note that anti-AI, excluded is the omitted condition.
- ε: Standard errors are clustered by subject.

Experimental Materials: Case Studies

C1. Text of Hiring-Related Business Case

The focus in this business case is on one of the world's top technology companies, headquartered in California (think of Facebook, Alphabet, Apple...). The case refers to it as "Toptech Co.".

Please imagine you are an engineering executive for Toptech Co. The organization you lead at Toptech implements software behind Toptech's consumer-facing products, digital infrastructure and advertising revenue.

As part of your job, you need to recruit and hire workers. Hiring qualified technology workers is difficult and high-stakes. Hiring quality people can make an enormous difference in product quality, market leadership and profitability. Many technology companies have reported shortages of qualified workers.

Toptech gets high volumes of applications from aspiring software engineers from around the world. Because of this high volume, it is cost-prohibitive to hire recruiters to read each job application with great care. As a result, a team at Toptech was assigned to work with your team to develop software to read the text of job applications and score them.

The technology looks at Toptech's historical hiring and job performance records. This data contains information about what job candidates and workers perform well on the job, who doesn't, and what each candidates' personal characteristics, experiences and qualifications were on the job application

Using this data, a team of statistical and machine learning experts has built a sophisticated model of what candidates are is likely to succeed using this data. The model is truly sophisticated; it allows for many different "paths" to being labeled a likely success. In data from the past five years, the model is very accurate at predicting which candidates will succeed or fail, even as circumstances have changed in the tech industry and at Toptech.

Advocates of the program claim that it improves the objectivity and consistency of evaluation. However, critics are concerned that the algorithm may propagate biases in the historical data. For example, critics inside the company point to the fact that the software's recommendations for who to hire in technical roles are approximately 71% men, and 29% women.

Your questions today are about whether your organization should adopt this technology.

C2. Text of Lending-Related Business Case

For this scenario, imagine you are a finance executive at one of the largest banks in the country (think of Bank of America, Wells Fargo, Citigroup...). The case will refer to it as ”Financial Co.” The organization you lead at Financial oversees consumer lending products and services.

As part of your job, you need to oversee a process for making decisions about small business lending. This is a critical part of Financial's strategy.

Financial gets high volumes of applications from small businesses. Because of this high volume, it is costly to hire loan officers to read each loan application with great care. As a result, a team at Financial was assigned to work with your team to develop software to read the content of loan applications and score them.

The technology looks at Financial's historical data about loan performance. This contains data about what borrowers historically paid back their loans, who didn't and what all applicants' loan characteristics are.

Using this data, a team of statistical and machine learning experts has built a sophisticated model of what borrowers are likely to pay back using this data. The model is truly sophisticated; it allows for many different "paths" to being labeled a likely success. In data from the past five years, the model is very accurate at predicting which borrowers will pay back or not, even as circumstances have changed in the economy.

Advocates of the program claim that it improves the objectivity and consistency of loan evaluation. However, critics are concerned that the algorithm may propagate biases in the historical data. For example, critics inside the company point to the fact that the software's recommendations charge higher and more expensive interest rates – about 3.2 higher basis points – to Latinx and African-American borrowers, compared to white borrowers.

Your questions today are about whether your organization should adopt this technology.

Survey Questions

- How negative or positive is the impact of this technology? (very negative - very positive)
- What is the probability that a fairness issue is alleged and becomes a problem (through lawsuits or PR)? (very low - very high)
- If a problem is alleged, how damaging would that be? (not at all damaging - very damaging)
- To what extent do you recommend using this algorithm rather than the status quo processes? (Do not recommend - Strongly recommend)
- If you were forced to make a decision on the future use of the algorithm now, what would you do? (Use the status quo methods (don't switch to the algorithm) OR Switch to adopt this algorithm).

[pagebreak]
Now suppose that the team agrees to spend the next six months making an effort to address fairness issues. Please re-evaluate the proposal based on how you expect the algorithm to perform after six additional months of your team attempting to address fairness issues.

Experimental Materials: "Scientific Veneer" Interventions

E1. Expert Introductions

For our veneer treatment, the following introduction was provided to the expert:

To help your team make a decision, your company has hired an expert to offer an assessment of the technology.

The expert, Taylor, has a PhD in Physics from UC Berkeley. Taylor has been employed as a third-party evaluator of digital technology for the past two years, working on questions like those you're facing.

On the next page you will read Taylor's summary and analysis of the technology and decision you are facing.

For our non-veneer treatment, the following introduction was provided to the expert:

To help your team make a decision, your company has hired an expert to offer an assessment of the technology.

The expert, Taylor, has a PhD in Sociology from UC Berkeley. Taylor has been employed as a third-party evaluator of digital technology for the past two years, working on questions like those you're facing.

On the next page you will read Taylor's summary and analysis of the technology and decision you are facing.

E2. Pro-AI Argument Without Scientific Veneer

Evaluation and Recommendation

I have analyzed the methods and data of this technology, and I recommend **in favor of** adopting it. Below I have outlined my main set of findings and reasons for this recommendation.

Main Arguments

1. After job qualifications are accounted for, the algorithm's recommendations are uncorrelated with demographics like race and gender.
2. My analysis suggests that algorithm makes decisions by maximizing the measurable, legitimate business objectives that your company has identified.
3. I have assessed whether the algorithm uses gender or race variables, and discovered it does not use these variables in any way.
4. Even a truly biased algorithm might show positive results (or better) by sheer chance. My analysis suggests this is highly unlikely to be responsible for my conclusions.

By using objective data and measurable outcomes, the algorithm removes human biases and arbitrary factors. In addition, as your company uses this algorithm through multiple rounds, the benefits will accumulate even higher. As you select better applicants, this algorithm will learn from these applicants' performances and optimize future decisions based on this learning.

E3. Pro-AI Argument With Scientific Veneer

Evaluation and Recommendation

I have analyzed the methods and data of this technology, and I recommend **in favor of** adopting it. Below I have outlined my main set of findings and reasons for this recommendation.

Main Argument	**Quantitative Expression**
1. After job qualifications are accounted for, the algorithm's recommendations are uncorrelated with demographics like race and gender.	1. Job qualifications are expressed by q. I measured that $\rho(\hat{y}\|q, g) \approx 0$ and $\rho(\hat{y}\|q, r) \approx 0$, for all genders $g \in G$ and races $r \in R$.
2. My analysis suggests that algorithm makes decisions by maximizing the measurable, legitimate business objectives that your company has identified.	2. The algorithm was designed to maximize the expression $\int_{x \in X} u(d, x)\hat{p}(x)dx$. u is the business objective and $\hat{p}(x)$ is the probability of facing each possibility. Our analysis suggests it successfully maximizes this.
3. I have assessed whether the algorithm uses gender or race variables, and discovered it does not use these variables in any way.	3. In the scoring algorithm, the weights associated with demographic variables are β (gender) and τ (race). We find that $\beta = 0$ and $\tau = 0$ for this algorithm $f(\beta, \tau; ...)$.
4. Even a truly biased algorithm might show positive results (or better) by sheer chance. My analysis suggests this is highly unlikely to be responsible for my conclusions.	4. $p < 0.00001$ (***)

By using objective data and measurable outcomes, the algorithm removes human biases and arbitrary factors. In addition, as your company uses this algorithm through multiple rounds, the benefits will accumulate even higher. As you select better applicants, this algorithm will learn from these applicants' performances and optimize future decisions based on this learning.

E4. Anti-AI Argument Without Scientific Veneer

Evaluation and Recommendation

I have analyzed the methods and data of this technology, and I recommend **against** adopting it. Below I have outlined my main set of findings and reasons for this recommendation.

Main Arguments

1. The algorithm's recommendations are highly correlated with being white and male.
2. Although the algorithm maximizing the **measurable** business objectives, many important objectives are harder to measure. Maximizing **only** the measurable ones leads to poor decision-making overall.
3. Although the algorithm does not directly utilize race or gender, it does use other characteristics correlated with race and gender. The algorithm's internal processes put very high "weight" on these characteristics.
4. Even a truly biased algorithm might show negative results (or better) by sheer chance. My analysis suggests this is highly unlikely to be responsible for my conclusions.

By using data from historical observations, the algorithm learns to replicate this historical bias – embedding it into future decisions and entrenching it. As your company uses this algorithm through multiple rounds, bias will accumulate even more. As you select applicants in a more biased way, this algorithm will learn from these applicants' selections and further entrench bias.

E5. Anti-AI Argument With Scientific Veneer

Evaluation and Recommendation

I have analyzed the methods and data of this technology, and I recommend **against** adopting it. Below I have outlined my main set of findings and reasons for this recommendation.

Main Argument	**Quantitative Expression**
1. The algorithm's recommendations are highly correlated with being white and male.	1. White and male are coded as $w = 1$ and $m = 1$ and zero otherwise. My measurements show $\rho(\hat{y}, w)$ and $\rho(\hat{y}, w)$ are both above thresholds.
2. Although the algorithm maximizing the **measurable** business objectives, many important objectives are harder to measure. Maximizing **only** the measurable ones leads to poor decision-making overall.	2. Your company should be maximizing the sum of i) $\int_{x \in X} p(x)u(d, x)dx$ and ii) $\int_{x \in X} p(x)m(d, x)dx$ where m is measurable objectives and u is harder-to-measure.
3. Although the algorithm does not directly utilize race or gender, it does use other characteristics correlated with race and gender. The algorithm's internal processes put very high "weight" on these characteristics.	3. In the scoring algorithm, the weights associated with demographic **proxy** variables are β (gender proxies) and τ (race proxies). We find that β and τ for this algorithm $f(\beta, \tau; ...)$ are both above thresholds.
4. Even a truly biased algorithm might show negative results (or better) by sheer chance. My analysis suggests this is highly unlikely to be responsible for my conclusions.	4. $p < 0.00001$ (***)

By using data from historical observations, the algorithm learns to replicate this historical bias – embedding it into future decisions and entrenching it. As your company uses this algorithm through multiple rounds, bias will accumulate even more. As you select applicants in a more biased way, this algorithm will learn from these applicants' selections and further entrench bias.

*

REFERENCES

Peer, Eyal, Laura Brandimarte, Sonam Samat, and Alessandro Acquisti, "Beyond the Turk: Alternative platforms for crowdsourcing behavioral research," *Journal of Experimental Social Psychology*, 2017, *70*, 153–163.