

Blaming Humans vs. Algorithms

Cjache Kang, Ajan Adriano, Pranavi Javangula, William Ryan
JDM Lab, Haas School of Business, University of California, Berkeley

1 INTRODUCTION

Background: In 2018, the Harvard Business Review published an article titled “Do People Trust Algorithms More Than Companies Realize?” The article explores previously studied research in “algorithm aversion” and “algorithm appreciation”, which illustrates people’s tendencies to either distrust or heavily rely on decisions made by algorithms. It was found that people often trust human advisors more for subjective decisions and algorithms for objective decisions.

Question: How do people assign blame to companies when either an algorithm or humans make erroneous or controversial decisions?

Approach: We created novel scenarios so people wouldn’t have pre-existing biases about whether a human or algorithm should receive more blame for a faulty decision.

Why is this important? Algorithms make impactful decisions and how people hold companies accountable could have important repercussions for these domains.

2 EXPERIMENT

Study 1: Exploratory study with all scenarios (N=266)

Study 2: Hiring decisions only (N=618)

Study 3: Weather channel (N= 360)

Experiment design:

- S1-3: Asked how much people blamed the specific team or algorithm responsible for decision making, how much they blame the company as a whole, what action should be taken, and how bad the actions were
- Across multiple scenarios (S1): discrimination in hiring, showing inappropriate ads, and biased college admissions or only one (S2/3)
- S1-3: Every person saw each scenario, but for half an algorithm was responsible and for half a human was responsible

Scenario Example

In 2017, **Evers Inc.**, a tech company, employs a hiring team to take full responsibility of the hiring process. The hiring team rates candidates based on a list of characteristics from the resumes of current employees. Candidates that were ranked highest were hired.

However, a report evaluating employee diversity revealed a year later that **65% of Evers Inc. employees were white** and 80% of executives were white, meaning diversity in the company had decreased. Upon further inspection, it was found that **the hiring team had been ranking minorities lower when screening candidates’ resumes**, because their chosen characteristics were biased toward the qualities of Evers Inc. employees, who were primarily white.

Survey Items: Algorithm Example

To what degree do you believe Evers Inc.’s algorithm should be blamed for their hiring practices?

To what degree do you believe Evers Inc. as a whole should be blamed for their hiring practices?

To what degree do you believe policy makers should take action against Evers. Inc and regulate its hiring practices to prevent similar practices from occurring in the future?

How wrong do you think the result of Evers. Inc’s actions were?

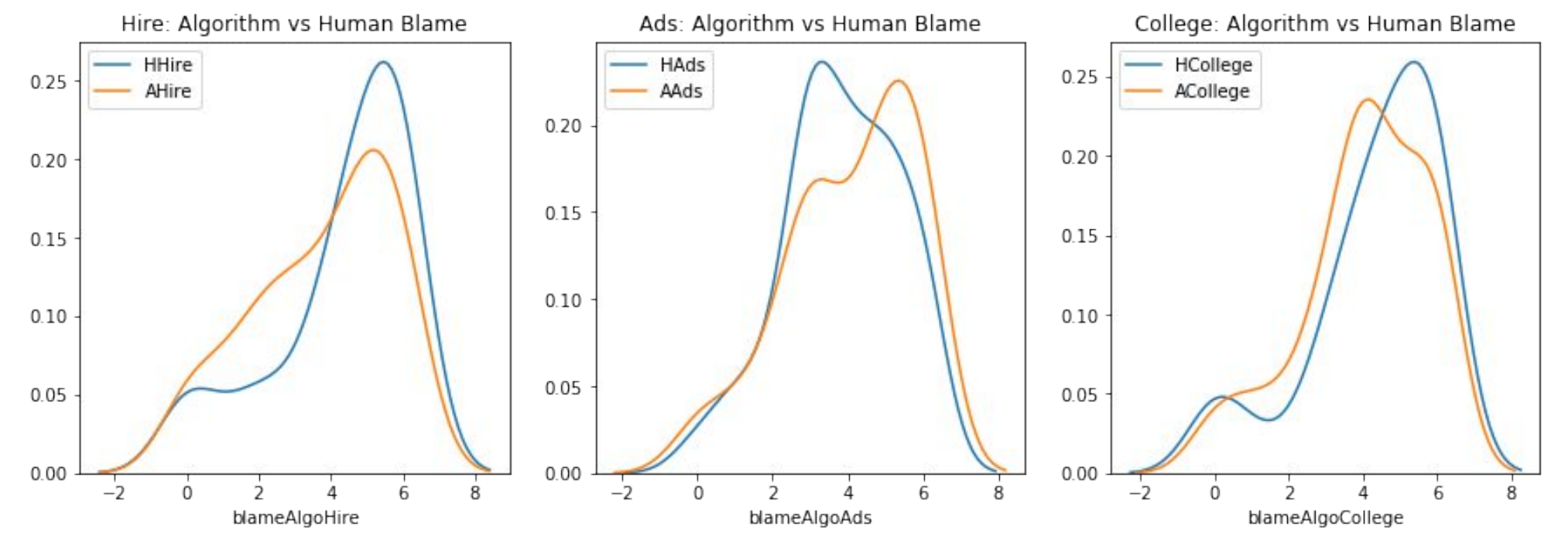
Hires a bunch of white people

Companies:

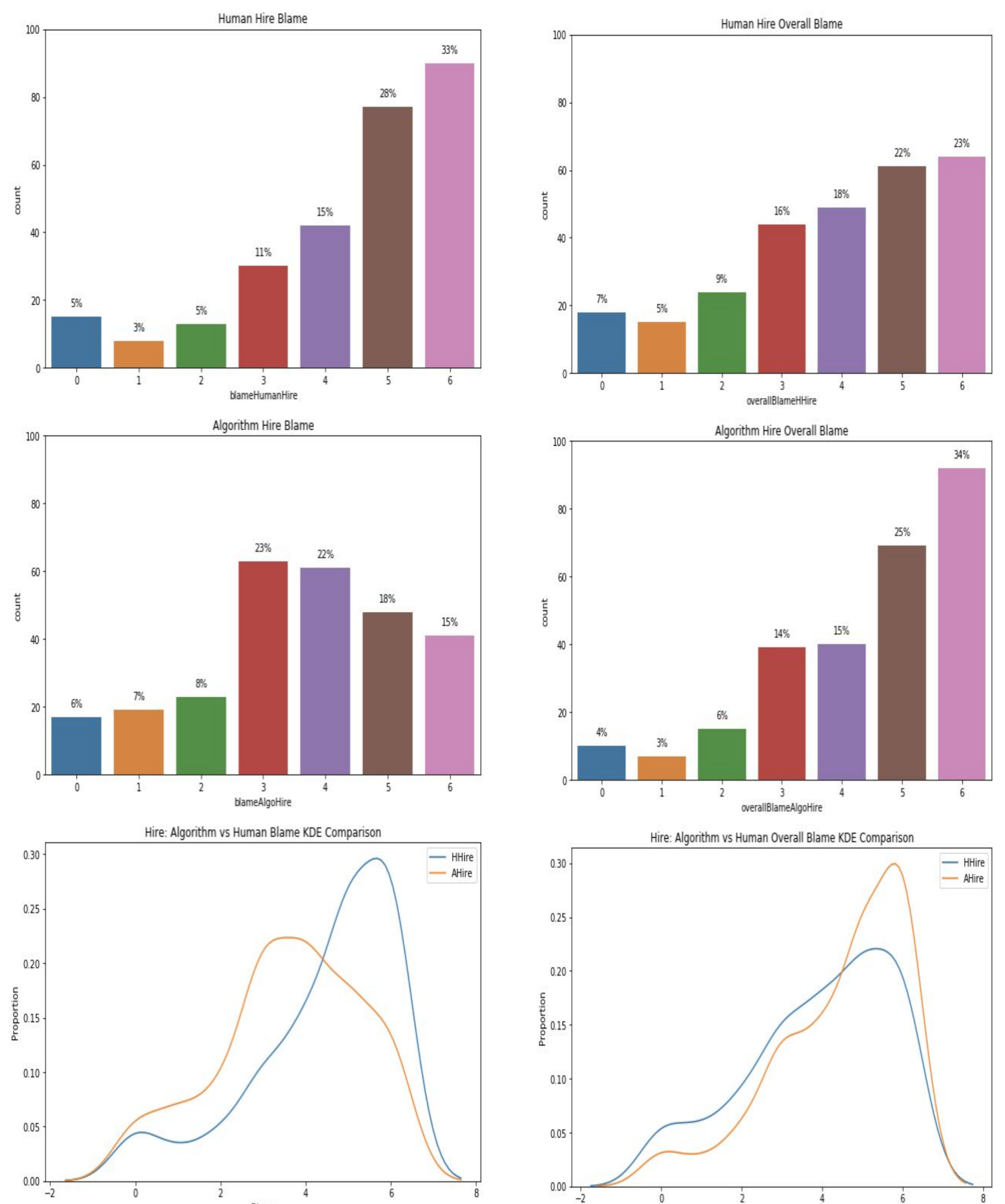


3 OUR FINDINGS

Study 1 (N=266)



Study 2 (N=618)



Figures 1. Comparing the means of the human scenario and the algorithm scenario.

	Human Mean (n=304)	Algorithm Mean (n=314)	P-value
Blame	4.43	3.62	3.43E-08
Overall Blame	3.93	4.45	3.43E-04
Government Action	3.69	3.6	0.59
Wrongness	4.1	3.89	0.16

Figure 2. The p-value of comparing the means between the same question in each scenario. (0-6)

4 CONCLUSION: What did we find?

- Study 1 (N=226) did not find significant results, but pattern mirrored S2
- Study 2 on hiring decisions (N=618) found human teams were blamed more, but the company overall was blamed more when algorithms are used.
- Study 3 (N=360) on weather channels found no significant differences

5 NEXT STEPS: What’s next?

- We believe there are definitely confounding variables. For example, there might exist established perceptions on, for example, algorithms in self-driving cars, that bias a certain individuals’ blame in certain scenarios.
- Our initial sample sizes could have been too small.
- Moving forward, we want to test blame on algorithms versus humans in scenarios where individuals do not have preconceptions about algorithm use.