Correspondence:
semir_tatlidil@brown.edu
https://brown.zoom.us/j/99171808730

Algorithms as advisors in the future of the workplace

Semir Tatlidil[1], Erin Bugbee[2], Madison Dick[1], Babak Hemmatian[1], and Steven Sloman[1]

[1]Department of Cognitive, Linguistic, & Psychological Sciences, Brown University
[2]Department of Social and Decision Sciences, Carnegie Mellon University

Abstract

Artificial Intelligence has an increasing role in the workplace. In three experiments, we look at how likely people are to trust and to take advice from an algorithm compared to a human coworker, for decisions such as hiring a new employee. We find that participants' agreement with the advisor is a good predictor of their trust of the advisor, both for the algorithm and the human. The results also suggest that although people trust the human more than the algorithm on average, they do not always follow the advice of the human.

Introduction

Previous studies suggest people are averse to taking advice from algorithms, an effect called **"algorithm aversion"**. We investigate if this result is dependent on:

- The trustworthiness of advisors
- The potential impact of the decision on other people and its moral weight: **hiring** a new employee vs **assigning** an employee to a team

Methods (contd.)

Example stimulus for hiring judgment

You are tasked with increasing the productivity of a team at your company. An applicant has applied to join the team. You have to decide whether to hire this applicant to increase the team's productivity.

This team has three members. One member is a skilled engineer and enjoys writing poetry. Another member is skilled at mathematics and enjoys playing the piano. The third team member has strong communication skills and enjoys building cars. The applicant is skilled at critical reading and enjoys programming video games.

After analyzing the group and portfolio of the applicant, a human resources employee recommends that you **hire *(do not hire)*** the applicant to improve the team's productivity and an algorithm designed for this purpose recommends that you **do not hire *(hire)*** the applicant to improve the team's productivity.

Note. Text in bold shows the advisors' advice in human "recommends/algorithm does not" condition, text in parentheses shows advice in "algorithm recommends/human does not condition".

DVs (1-7 Likert Scale):

Hiring: How likely are you to hire this applicant?
Assigning: How beneficial for maximizing productivity do you predict the employee will be on this team?
Trust: How much do you trust the recommendation from

Results (contd.)

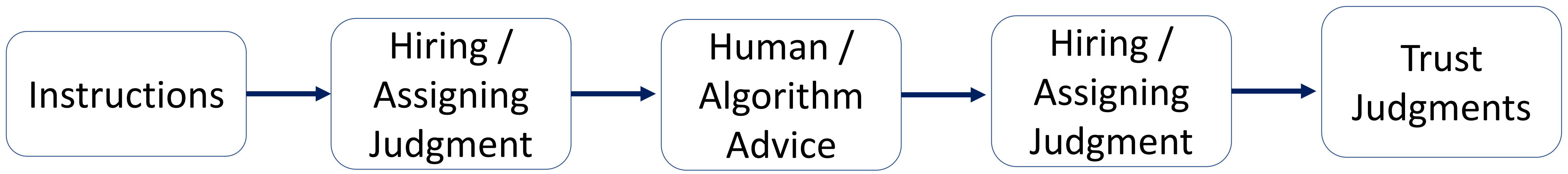
Table 1. Pre-advice judgments predicting trust in advisors, when the advisor recommends hiring/assigning

	Trust Algorithm	Trust Human
Experiment 1	0.43 (.13)**	0.18 (.03)
Experiment 2	0.22 (.03)	0.28 (.06)*
Experiment 3	0.39 (.14)**	0.41 (.14)**

Note. Values represent beta weights, values in parentheses are R^2. * p < .05, ** p < .001

Methods

The general procedure of the experiments was the following:


Instructions → Hiring / Assigning Judgment → Human / Algorithm Advice → Hiring / Assigning Judgment → Trust Judgments

Participants were asked to make a judgment about hiring an employee in Experiment 1 (N = 320, Prolific) and about assigning an employee to a team in Experiment 2 (N = 316). For half of the participants, the human recommended to hire, and the algorithm recommended not to hire. For the other half, the opposite pattern was presented. Only half of the participants provided pre-advice judgments.

In Experiment 3 (N=400) only one advisor was presented in a 2 (Advice) x 2 (Advisor) x 2 (Judgment Type) design with all factors being between-subjects. All subjects provided pre- and post-advice judgments.

Results

Figure 1. Hiring/Assigning judgments (Experiments 1 and 2)

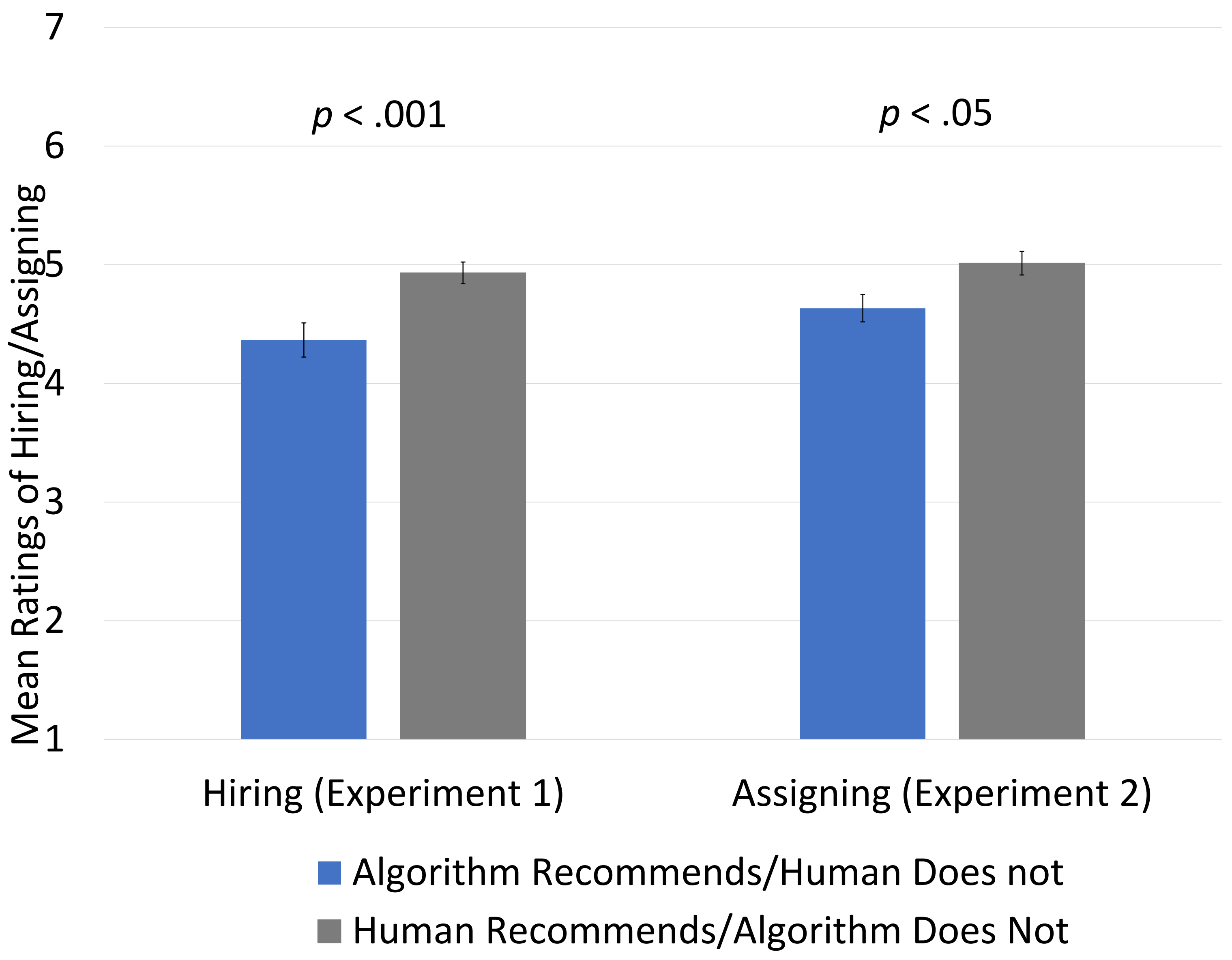


Figure 2. Hiring/Assigning judgments (Experiment 3)

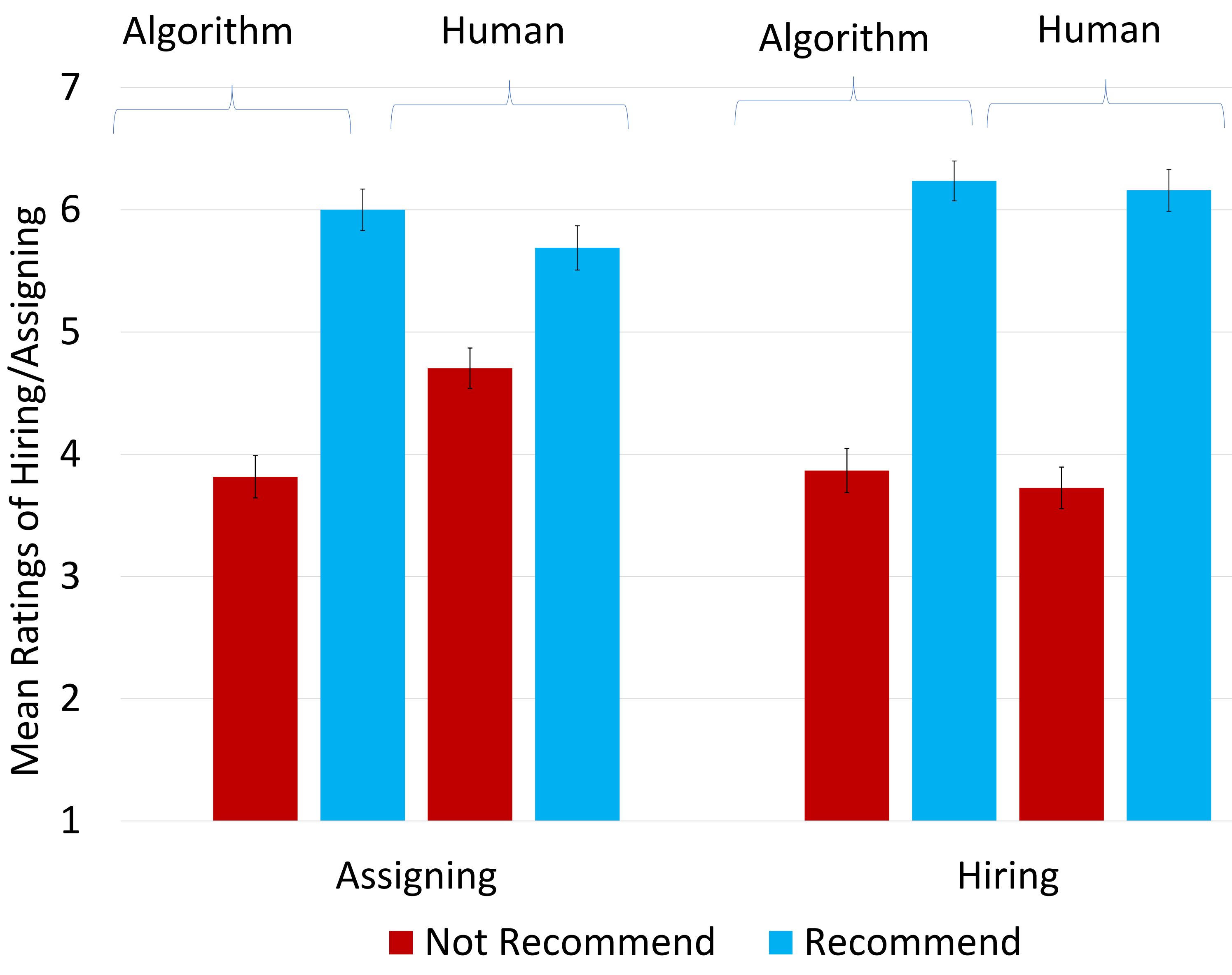
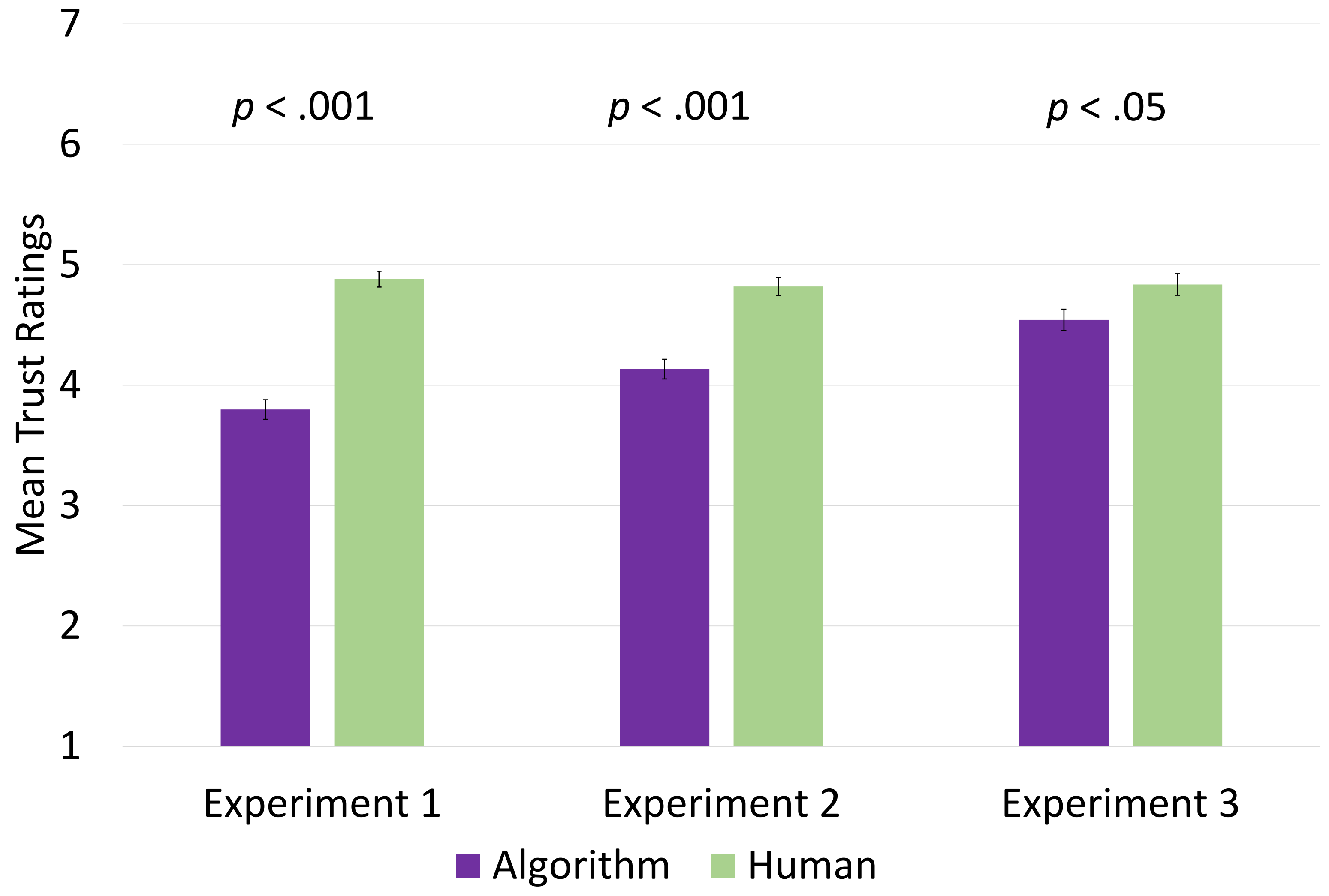


Figure 3. Trust judgments across experiments.



Discussion

- Across three experiments, people consider the **human advisor to be more trustworthy than the algorithm** (Figure 3).
- People are more likely to take the advice of the human in Experiments 1 and 2 (Figure 1), but not in Experiment 3 (Figure 2). This implies that **algorithm aversion may disappear** when people are not explicitly comparing an algorithm to a human advisor.
- **Agreement** between the opinion of the participant and the advisor is associated with **higher trust in the advisor**, both for the human and the algorithm (Table 1). Consistent with this conclusion, pre-advice judgments did not predict trust in advisors when advisors recommended not to hire/assign, as less than 6% of the participants shared this view before advice (ratings < 4).