Design Topic:

Decision Making

Jin L.C. Guo
SOCS McGill University

ML-Assisted Decision Making

- Why might the ML-assisted decision making be preferred?
- What are the criteria to accept the ML-assisted decision-making process?

ML-Assisted Decision Making

- Accuracy

  People using the algorithm should make more accurate predictions than they could without the algorithm.

- Reliability

  People should accurately evaluate their own and the algorithm's performance and should calibrate their use of the algorithm to account for its accuracy and errors.

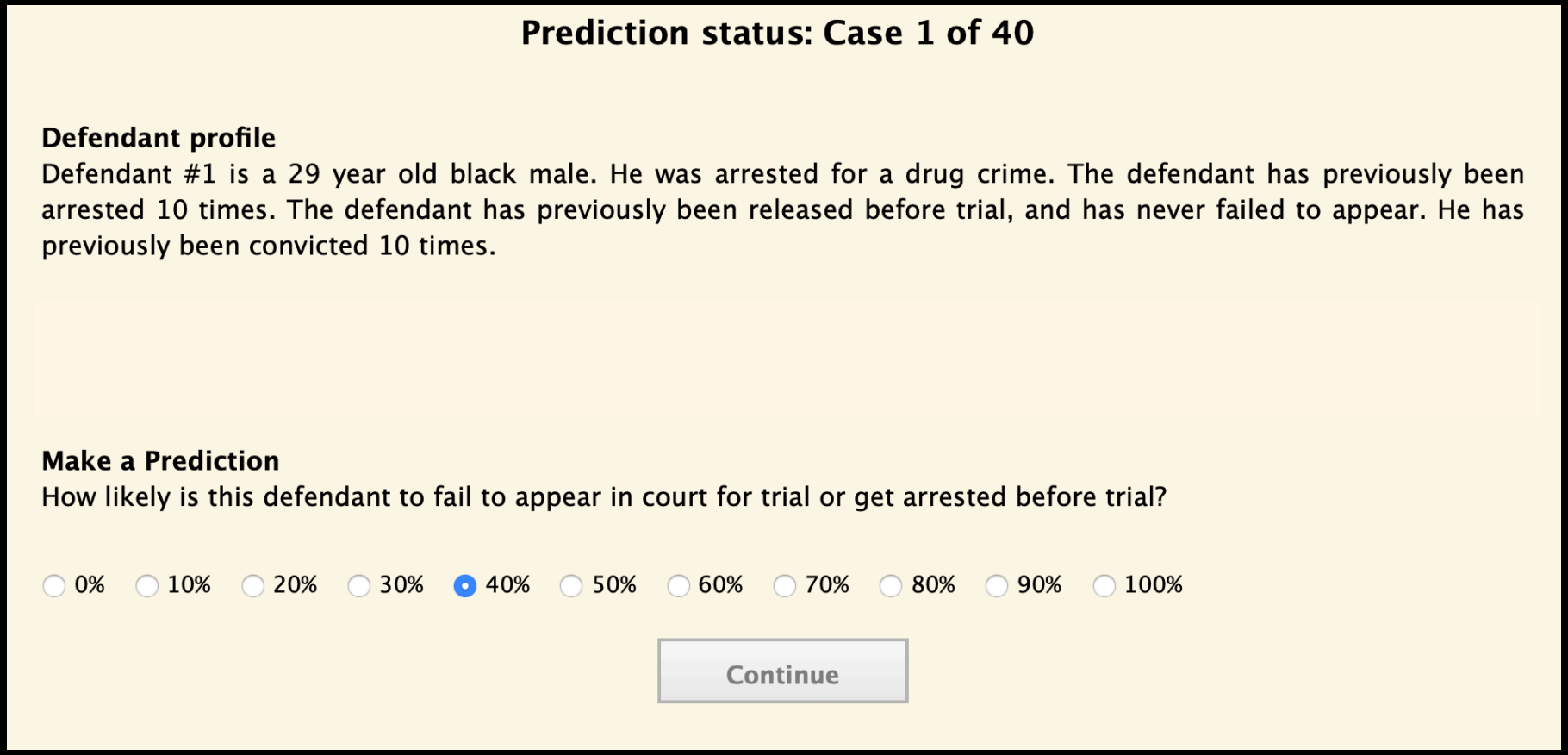
- Fairness

  People should interact with the algorithm in ways that are unbiased with regard to race, gender, and other sensitive attributes.

- ……

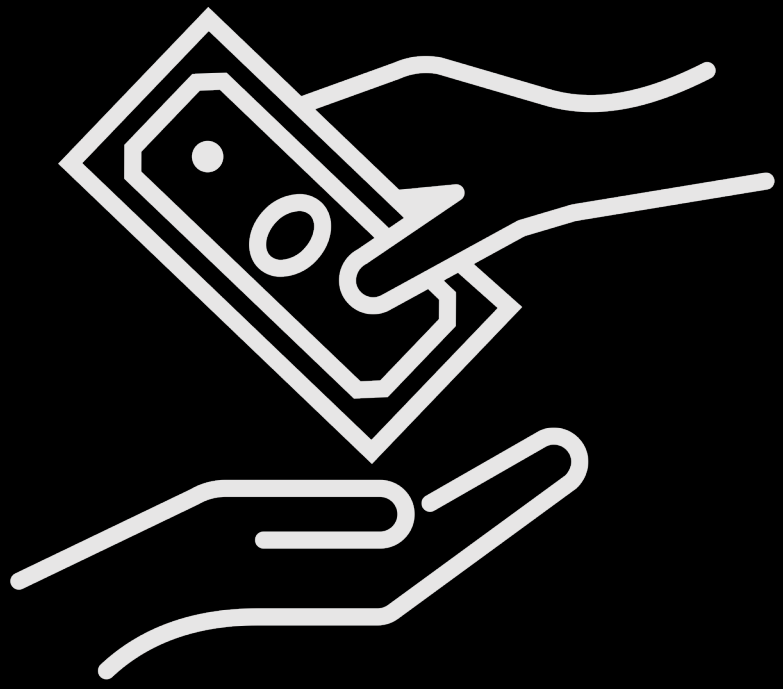
Green, Ben, and Yiling Chen. "The principles and limits of algorithm-in-the-loop decision making." *Proceedings of the ACM on Human-Computer Interaction* 3, no. CSCW (2019): 1-24.

Estimate the likelihood that criminal defendants will be arrested before trial or fail to appear in court for trial.



Green, Ben, and Yiling Chen. "The principles and limits of algorithm-in-the-loop decision making." *Proceedings of the ACM on Human-Computer Interaction* 3, no. CSCW (2019): 1-24.

Estimate the likelihood that a loan applicant will fail to pay back the money (“defaulting” on the loan).



Baseline

Prediction status: Case 1 of 40

Applicant profile

Loan applicant #1 has applied for a loan of $30,375, with an interest rate of 19.52%. The loan will be paid in 36 monthly installments of $1,121.43. The applicant has an annual income of $80,000 and a "Good" credit score. The applicant has a mortgage out on their home.

Make a Prediction

How likely is this applicant to default on their loan?

○ 0% ○ 10% ○ 20% ○ 30% ○ 40% ○ 50% ○ 60% ○ 70% ○ 80% ○ 90% ○ 100%

Continue

Green, Ben, and Yiling Chen. "The principles and limits of algorithm-in-the-loop decision making." *Proceedings of the ACM on Human-Computer Interaction* 3, no. CSCW (2019): 1-24.

RA Prediction

Prediction status: Case 1 of 40

Applicant profile

Loan applicant #1 has applied for a loan of $30,375, with an interest rate of 19.52%. The loan will be paid in 36 monthly installments of $1,121.43. The applicant has an annual income of $80,000 and a "Good" credit score. The applicant has a mortgage out on their home.

Risk assessment

The risk score algorithm predicts that this person is 40% likely to default on their loan.

Make a Prediction

How likely is this applicant to default on their loan?

○ 0% ○ 10% ○ 20% ○ 30% ○ 40% ○ 50% ○ 60% ○ 70% ○ 80% ○ 90% ○ 100%

Continue

Green, Ben, and Yiling Chen. "The principles and limits of algorithm-in-the-loop decision making." *Proceedings of the ACM on Human-Computer Interaction* 3, no. CSCW (2019): 1-24.

Explanation

Prediction status: Case 1 of 40

Applicant profile

Loan applicant #1 has applied for a loan of $30,375, with an interest rate of 19.52%. The loan will be paid in 36 monthly installments of $1,121.43. The applicant has an annual income of $80,000 and a "Good" credit score. The applicant has a mortgage out on their home.

Risk assessment

The risk score algorithm predicts that this person is 40% likely to default on their loan. Compared to the average applicant, the following attributes make this applicant notably

- Higher risk: Interest rate.
- Lower risk: Home ownership.

Make a Prediction

How likely is this applicant to default on their loan?

○ 0% ○ 10% ○ 20% ○ 30% ○ 40% ○ 50% ○ 60% ○ 70% ○ 80% ○ 90% ○ 100%

Continue

Green, Ben, and Yiling Chen. "The principles and limits of algorithm-in-the-loop decision making." *Proceedings of the ACM on Human-Computer Interaction* 3, no. CSCW (2019): 1-24.

Default

Prediction status: Case 1 of 40

Defendant profile

Defendant #1 is a 29 year old black male. He was arrested for a drug crime. The defendant has previously been arrested 10 times. The defendant has previously been released before trial, and has never failed to appear. He has previously been convicted 10 times.

Risk assessment

The risk score algorithm predicts that this person is 40% likely to fail to appear in court for trial or get arrested before trial. **The prediction has been set to this value, but you are free to predict another value.**

Make a Prediction

How likely is this defendant to fail to appear in court for trial or get arrested before trial?

○ 0% ○ 10% ○ 20% ○ 30% ◉ 40% ○ 50% ○ 60% ○ 70% ○ 80% ○ 90% ○ 100%

Continue

Green, Ben, and Yiling Chen. "The principles and limits of algorithm-in-the-loop decision making." *Proceedings of the ACM on Human-Computer Interaction* 3, no. CSCW (2019): 1-24.

Update

Participants were first presented with the Baseline condition; after making a prediction, participants were presented with the RA Prediction condition (for the same case) and asked to make the prediction again.

Feedback

Participants were presented with the RA Prediction condition; after submitting each prediction, participants were presented with an alert indicating the outcome of that case (e.g., whether the loan applicant actually defaulted on their loan).

Green, Ben, and Yiling Chen. "The principles and limits of algorithm-in-the-loop decision making." *Proceedings of the ACM on Human-Computer Interaction* 3, no. CSCW (2019): 1-24.

Accuracy

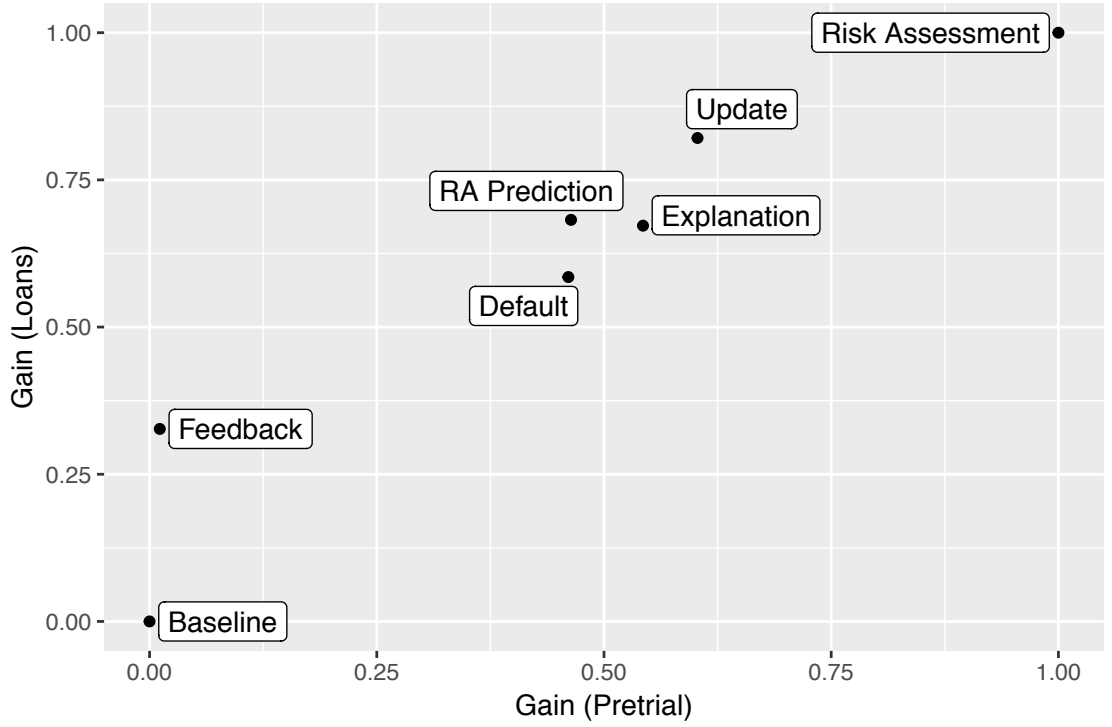


Fig. 2. The relative performance gain (Equation 1) achieved by each experimental condition across the pretrial and loans settings. In both settings, the Update treatment performed statistically significantly better than RA Prediction and the Feedback treatment performed statistically significantly worse. Across the two settings, the gain of the conditions was highly correlated.

Green, Ben, and Yiling Chen. "The principles and limits of algorithm-in-the-loop decision making." *Proceedings of the ACM on Human-Computer Interaction* 3, no. CSCW (2019): 1-24.

Reliability

Table 4. Summary of participant abilities to evaluate performance (first two columns) and to calibrate their predictions (third column). The columns measure the relationships between between participant confidence and actual performance (Confidence), participant estimates of the algorithm's performance and its actual performance (RA Accuracy), and participant reliance on the risk assessment and the risk assessment's performance (Calibration). + signifies a positive and statistically significant relationship, - signifies a negative and statistically significant relationship, and 0 signifies no statistically significant relationship. In all cases, + means that the desired behavior was observed.

	Confidence		**RA Accuracy**		**Calibration**	
	Pretrial	Loans	Pretrial	Loans	Pretrial	Loans
RA Prediction	0	0	0	0	-	0
Default	0	-	0	-	0	0
Update	0	-	-	-	0	0
Explanation	0	0	0	0	-	+
Feedback	0	0	0	0	-	0

Green, Ben, and Yiling Chen. "The principles and limits of algorithm-in-the-loop decision making." *Proceedings of the ACM on Human-Computer Interaction* 3, no. CSCW (2019): 1-24.

Fairness

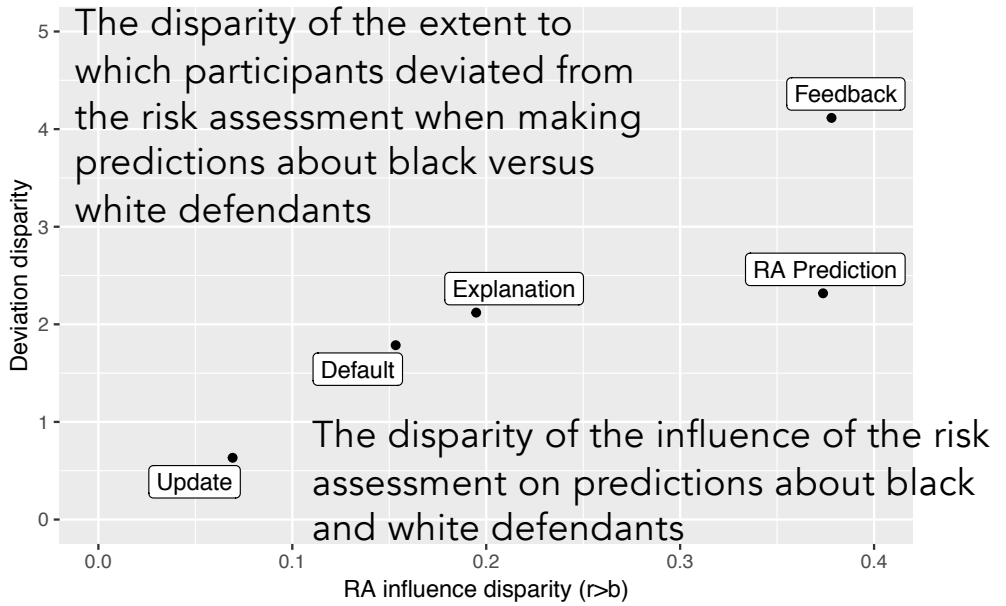
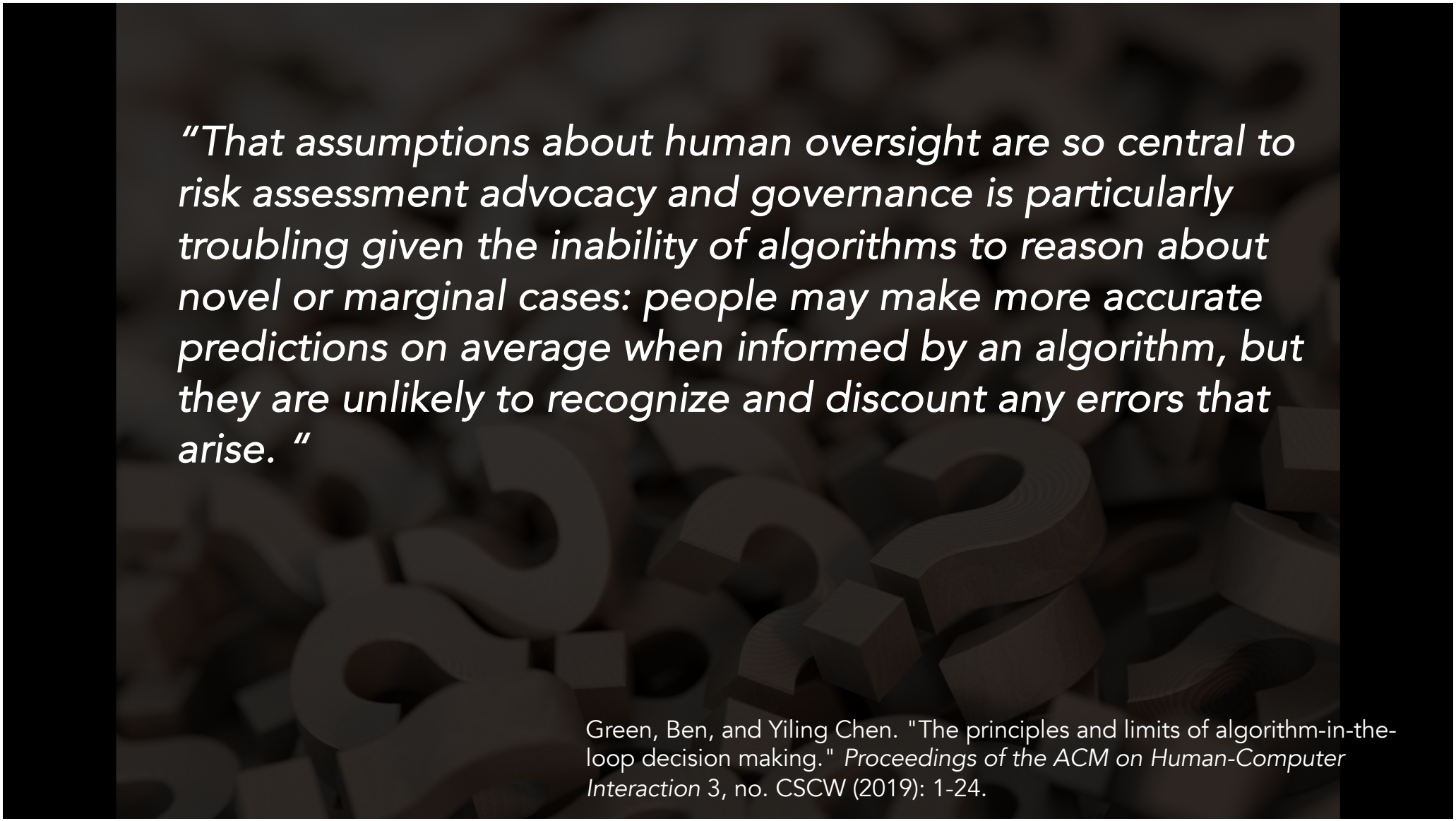
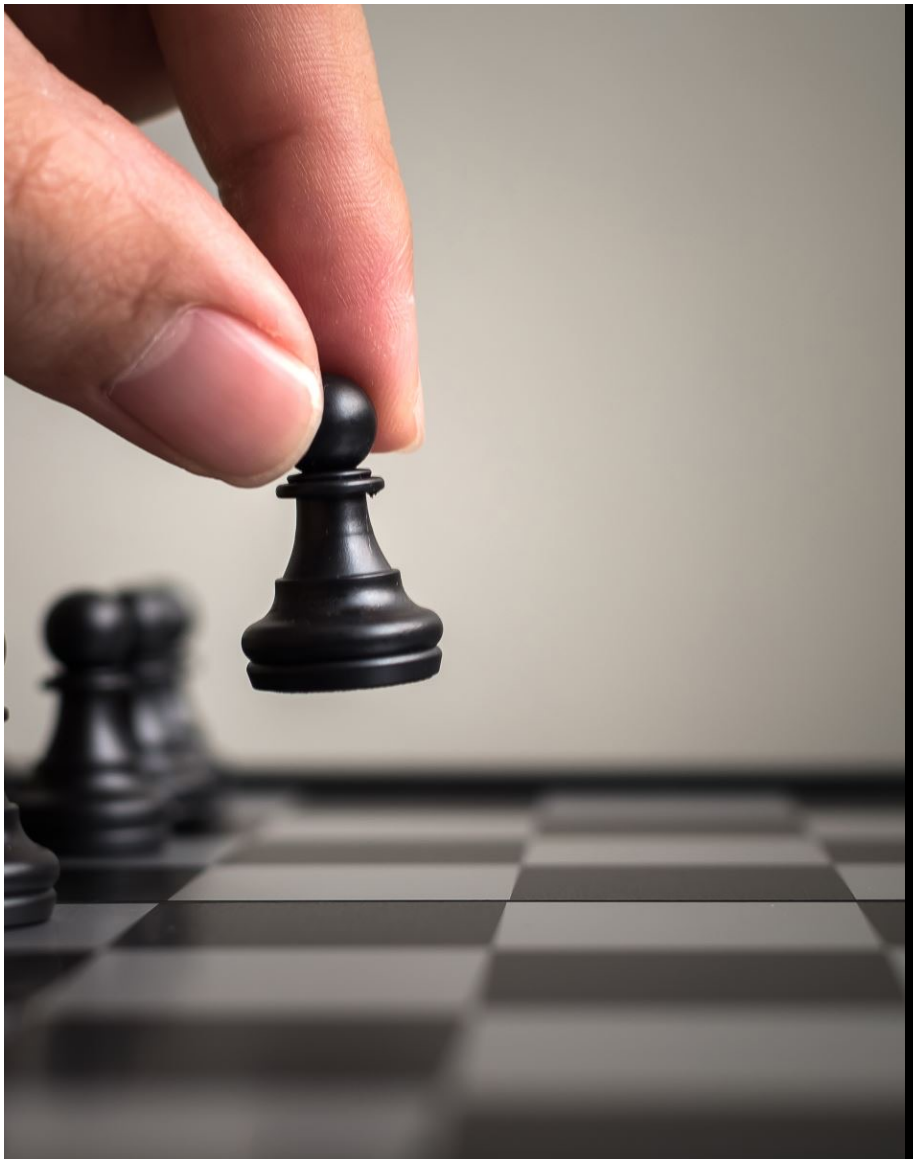


Fig. 3. The disparate interactions present in each treatment in the pretrial setting, measured by the disparities in risk assessment influence (Equation 3) and in participant deviations (Equation 4) for black versus white defendants. In both cases, values closer to 0 indicate lower levels of bias. The Update treatment yielded the smallest disparate interactions along both metrics, reducing the disparities (compared to RA Prediction) by 81.5% and 73.9%, respectively.

Green, Ben, and Yiling Chen. "The principles and limits of algorithm-in-the-loop decision making." *Proceedings of the ACM on Human-Computer Interaction* 3, no. CSCW (2019): 1-24.


“That assumptions about human oversight are so central to risk assessment advocacy and governance is particularly troubling given the inability of algorithms to reason about novel or marginal cases: people may make more accurate predictions on average when informed by an algorithm, but they are unlikely to recognize and discount any errors that arise. ”
Green, Ben, and Yiling Chen. "The principles and limits of algorithm-in-the-loop decision making." *Proceedings of the ACM on Human-Computer Interaction* 3, no. CSCW (2019): 1-24.



Human Judgement

under uncertainty

Heuristics | Biases

The following subject is randomly sampled from a group of 30 engineers and 70 lawyers:

> *Steve is a 30 year old man. He is married with no children. A man of high ability and high motivation, he promises to be quite successful in his field. He is well liked by his colleagues.*

What is the probability of Steve being an engineer rather than an lawyer?

Tversky, Amos, and Daniel Kahneman. "Judgment under uncertainty: Heuristics and biases." *science* 185, no. 4157 (1974): 1124-1131.

Human Decision

Representativeness

People respond differently when given no evidencee and when given worthless evidence.

Tversky, Amos, and Daniel Kahneman. "Judgment under uncertainty: Heuristics and biases." *science* 185, no. 4157 (1974): 1124-1131.

Human Decision

Representativeness

Insensitive to prior probability of outcomes

Illusion of validity

Tversky, Amos, and Daniel Kahneman. "Judgment under uncertainty: Heuristics and biases." *science* 185, no. 4157 (1974): 1124-1131.

Within 5 sec, estimate the product of :

8x7x6x5x4x3x2x1

versus

1x2x3x4x5x6x7x8

Tversky, Amos, and Daniel Kahneman. "Judgment under uncertainty: Heuristics and biases." *science* 185, no. 4157 (1974): 1124-1131.

Human Decision

Adjustment and Anchoring

Insufficient adjustment

Biases in evaluation of conjunctive and disjunctive events

Tversky, Amos, and Daniel Kahneman. "Judgment under uncertainty: Heuristics and biases." *science* 185, no. 4157 (1974): 1124-1131.

Human Decision

Availability

Biases due to the retrievability of instances

Biases of imaginability

Illusion of correlation

Tversky, Amos, and Daniel Kahneman. "Judgment under uncertainty: Heuristics and biases." *science* 185, no. 4157 (1974): 1124-1131.

Summary

- Decision Making today has become a socio technical affair.
  - ML-Assisted decision making is more than model's statistical properties.
  - No matter how the risk assessment was presented, participants
    - could not determine their own or the model's accuracy
    - failed to calibrate their use of the model to the quality of its predictions,
    - and exhibited disparate interactions when making predictions.
- Human decision heavily replies on heuristics and therefore can introduce biases
  - Representativeness
  - Adjustment and Anchoring
  - Availability

Next Lecture

Explainable AI