# Attention and Perception

## Lecture 7: AI Oversight

Daniel Martin
danielmartin@ucsb.edu

# Questions

- Does AI oversight change human behavior?
  - Does $P_{\text{No AI}}(a, \omega) = P_{\text{AI}}(a, \omega)$?
- Could this be due to utility changes?
  - Does $U_{\text{No AI}}(a, \omega) = U_{\text{AI}}(a, \omega)$?
  - Maintained assumption: the attention technology $K(\mu, \pi)$ is stable
  - Idea: utility changes can change attention, revealed posteriors, and choices
- What is the utility impact?
  - How does identification improve as we add assumptions on $K(\mu, \pi)$?
  - BEU/NIAS $\rightarrow$ general RI/NIAC $\rightarrow$ state-weighted Shannon costs

# Motivation

- Many firms will soon have the option of using AI to correct worker mistakes
  - Better prediction and lower data-processing costs expand feasible monitoring
  - Algorithms are increasingly strong in prediction problems: bail, medicine, pricing, hiring, and auditing
- But oversight is not only a technological intervention
  - It changes the environment in which humans choose effort, attention, and actions
  - Shame, pride, embarrassment, stress, and relief can all enter the perceived payoff matrix

# Example: Zalando

- At a leading e-commerce company, category managers suggest discounts based on private information about fashion trends
- Human interventions can be valuable, but poor interventions can offset the benefits of good interventions
- Huelden et al. (2024) show that AI tools can predict and block undesirable human interventions
- This is the organizational logic of oversight:
  - keep human judgment where it is useful,
  - use AI to detect and correct the mistakes

# Objections

- Why not give workers AI recommendations instead of allowing AI to correct mistakes?
  - Recommendations do not always work
  - Final decision rights can matter for accountability and trust
- If AI predictions are good, why not replace the human decision-maker?
  - Humans often have private information, context, and edge-case knowledge
  - The AI may only need to be good at identifying mistakes, not doing the whole job
- That makes hybrid human-AI systems an important empirical object

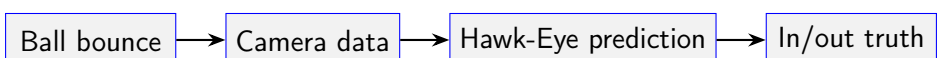
# Our Questions

- Using AI to correct human mistakes seems like a clear win
  - Potential gains in high-stakes settings: law, medicine, hiring, pricing, and management
- But adoption requires knowing whether oversight changes human decision-making
- We study one high-visibility setting: AI review of professional tennis line calls
- **Questions:** How did umpires respond to the introduction of AI review? Could psychological factors have played a role?

**This provides field evidence that AI oversight can affect human decision-making.**

# AI in Tennis

- Hawk-Eye predicts a ball's landing spot using input from 6 to 10 cameras
- The task is difficult because many line calls are very close
- Average Hawk-Eye error is about 3.6 mm; a tennis ball has diameter about 67 mm
- Like the ATP, we treat Hawk-Eye as the ground truth in the empirical exercise



# What is Hawk-Eye Review?

- March 2006: Hawk-Eye challenges were officially used for the first time in a tournament
- If a call is challenged, the AI system can overrule the on-court call
- If the ball was out but called in, the score is corrected directly
- If the ball was in but called out, the umpire must decide whether the player could have made a play
  - If so, the point is replayed
  - This is a more awkward and public correction

# Our Empirical Setting

- Rare field setting where AI is plausibly more accurate than elite humans
- This helps isolate short-run behavioral responses to oversight
- The setting avoids many standard difficulties:
  - Outcomes are observed
  - Ground truth is measured
  - Private information about the state is limited
- Data granularity and rollout timing generate control data
  - Pre-Hawk-Eye review: 7 tournaments and 109 matches
  - Post-Hawk-Eye review: 28 tournaments and 589 matches
- There were few changes in the umpiring pool and no contemporaneous changes in standard training

# What Makes Hawk-Eye Interesting?

Hawk-Eye was not used by the ATP for formal umpire assessment, but it introduced new forces that could change calls

1. Do we see an attentional response from umpires?
   - They might relax because mistakes can be corrected
   - They might become more attentive because some mistakes are corrected publicly
2. Do we see an asymmetric response based on the relative cost of each error?
   - Type I: ball is out, call is in. If overturned, the point ends
   - Type II: ball is in, call is out. If overturned, the umpire may need to order a replay

# Data Construction

1. **Hawk-Eye base data**: precise point information and the location of every ball bounce
2. **Challenge data**: outcomes of player challenges during early Hawk-Eye-review tournaments
3. **Video auditing**: validation of the merge and mistake-identification rules

| | Pre-review | Post-review | Total |
|---|---|---|---|
| Tournaments | 7 | 28 | 35 |
| Matches | 109 | 589 | 698 |
| Points | 15,439 | 83,898 | 99,337 |
| Bounces | 71,638 | 402,446 | 474,102 |

# Data Updates in the Current Paper

- Consolidated sample: 698 matches across 35 tournaments
- 2,038 of 2,108 challenges in the 28 matched post-review tournaments were merged: **97% merge rate**
- Video audit: 43 matches and 144 challenges
  - Merging algorithm matched 143 of 144 audited challenges
  - 136 were matched to the correct point: 94.4% audited accuracy
- Pre- and post-review samples are similar on close-call shares and shot-type composition

# Reduced-Form Patterns

- Within 100 mm of the line, Hawk-Eye review reduces the average incorrect-call rate by about 1.1 p.p.
- The average masks an important change in the composition of mistakes
- Close to the line, umpires shift toward calling balls **in**
  - More Type I errors: ball is out, call is in
  - Fewer Type II errors: ball is in, call is out
- This is exactly where the institutional costs of being overturned are asymmetric

# The Threshold Shift

- For balls within 20 mm of the line, the rate at which umpires called the ball **in** rose from **42.8%** to **49.0%**
- In 22 of the 28 post-review tournaments, umpires called the ball in more often than the pre-review mean
- The increase appears gradual: the tournament-level trend is 0.45 p.p. per month
- Psychometric curves have similar middle-region slopes, consistent with a threshold shift rather than a wholesale change in perceptual precision

**AI oversight lowers average error while changing which errors humans make**

# Structural Model of Rational Inattention

- The reduced-form estimates suggest that psychological costs of being overruled by AI changed umpire behavior
- The structural exercise asks how large those costs must be
- Umpires choose attention before making a call
  - More attention is cognitively costly
  - More attention improves perception and prediction
- The umpire trades off cognitive costs, ordinary mistake costs, and the cost of being publicly overruled

# Problem Setup

- Actions: $a_I$ = call in and $a_O$ = call out
- States: $\omega_I$ = ball in and $\omega_O$ = ball out
- Challenge rates conditional on incorrect calls:
  - $\eta_I$: ball is in, call is out
  - $\eta_O$: ball is out, call is in
- AI oversight penalties:
  - $c_I$: penalty when the ball is in and the call is out
  - $c_O$: penalty when the ball is out and the call is in

# Utility Under AI Oversight

Correct calls are normalized to utility 0. An unchallenged incorrect call has utility $-1$

$$U_{\mathrm{AI}}(a,\omega) = \begin{array}{c} \\ a_I \\ a_O \end{array} \begin{array}{c} \begin{array}{cc} \omega_I & \omega_O \end{array} \\ \begin{pmatrix} 0 & -1+\eta_O(1+c_O) \\ -1+\eta_I(1+c_I) & 0 \end{pmatrix} \end{array}$$

- If $c = 0$, AI correction has no direct psychological downside
- If $c = -1$, correction exactly balances the ordinary benefit of fixing the mistake
- If $c < -1$, being overruled is worse than simply being wrong

# Revealed Posteriors

- The data are joint probabilities $P_t(a, \omega)$, where $t \in \{\text{No AI}, \text{AI}\}$
- Unconditional action probabilities are

$$P_t(a) = \sum_{\omega \in \{\omega_I, \omega_O\}} P_t(a, \omega)$$

- The revealed posterior after call $a$ is

$$\gamma_{a,t}(\omega) = \frac{P_t(a, \omega)}{P_t(a)}$$

- The prior in the attention problem is $\mu_t(\omega) = \sum_a P_t(a, \omega)$
- Distance from the line defines the empirical sample; it is not a state in the attention problem

# Subjective Perception: NIAS

- NIAS requires each chosen action to be optimal at its revealed posterior

$$\sum_{\omega} \gamma_{a_I,\mathsf{AI}}(\omega)\left[U_{\mathsf{AI}}(a_I,\omega) - U_{\mathsf{AI}}(a_O,\omega)\right] \geq 0$$

$$\sum_{\omega} \gamma_{a_O,\mathsf{AI}}(\omega)\left[U_{\mathsf{AI}}(a_O,\omega) - U_{\mathsf{AI}}(a_I,\omega)\right] \geq 0$$

- This gives bounds on one unobserved penalty in terms of the other

$$c_I \leq \frac{P_{\mathsf{AI}}(a_I,\omega_I)(1-\eta_I) - P_{\mathsf{AI}}(a_I,\omega_O)(1-\eta_O)}{P_{\mathsf{AI}}(a_I,\omega_I)\eta_I} + \frac{P_{\mathsf{AI}}(a_I,\omega_O)\eta_O}{P_{\mathsf{AI}}(a_I,\omega_I)\eta_I} c_O,$$

$$c_I \geq \frac{P_{\mathsf{AI}}(a_O,\omega_I)(1-\eta_I) - P_{\mathsf{AI}}(a_O,\omega_O)(1-\eta_O)}{P_{\mathsf{AI}}(a_O,\omega_I)\eta_I} + \frac{P_{\mathsf{AI}}(a_O,\omega_O)\eta_O}{P_{\mathsf{AI}}(a_O,\omega_I)\eta_I} c_O$$

# Subjective Perception: Bounds

- Use observed $P_{\text{AI}}(a, \omega)$ as the empirical joint distribution of actions and objective states
- Use observed challenge rates as estimates of $\eta_I$ and $\eta_O$
- NIAS becomes

$$0.6207 + 0.5012 c_O \geq c_I \geq -1.2115 + 1.7744 c_O$$

- If $c_O = -1$, so the benefit of correction balances the oversight penalty for Type I errors,

$$0.1195 \geq c_I \geq -2.9859$$

- NIAS alone is informative, but the interval is wide

# General Rational Inattention: NIAC

- Now assume the umpire chooses an information structure $\pi$ with additively separable cost $K(\mu, \pi)$
- The same cost technology is available before and after Hawk-Eye review
- NIAC says the chosen information structure in each environment must beat swapping in the other environment's information structure
- This sharpens the upper bound on the Type II oversight penalty:

$$c_I \leq \frac{(P_{\text{AI}}(a_O, \omega_I) - P_{\text{No AI}}(a_O, \omega_I))\eta_I + (P_{\text{AI}}(a_I, \omega_O) - P_{\text{No AI}}(a_I, \omega_O))\eta_O}{(P_{\text{No AI}}(a_O, \omega_I) - P_{\text{AI}}(a_O, \omega_I))\eta_I}$$

$$+ \frac{(P_{\text{AI}}(a_I, \omega_O) - P_{\text{No AI}}(a_I, \omega_O))\eta_O}{(P_{\text{No AI}}(a_O, \omega_I) - P_{\text{AI}}(a_O, \omega_I))\eta_I} c_O$$

# General Rational Inattention: Bounds

- For calls within 20 mm of the line, the NIAC bound is

$$c_I \leq 1.0105 + 2.0105 c_O$$

- If $c_O = -1$, NIAC gives

$$c_I \leq -1$$

- Combined with the NIAS restrictions,

$$-1 \geq c_I \geq -2.9859$$

- General RI already implies that the Type II oversight penalty must be meaningfully negative

# State-Weighted Shannon Costs

- To recover point estimates, specialize to a Shannon-style cost
- The model allows different marginal attention costs across states:

$$K(\mu,\pi) = \kappa_I \left( \sum_{\gamma \in \Gamma(\pi)} \pi(\gamma)\gamma(\omega_I) \ln \gamma(\omega_I) - \mu(\omega_I) \ln \mu(\omega_I) \right) + \kappa_O \left( \sum_{\gamma \in \Gamma(\pi)} \pi(\gamma)\gamma(\omega_O) \ln \gamma(\omega_O) - \mu(\omega_O) \ln \mu(\omega_O) \right),$$

- where $\kappa_I, \kappa_O > 0$
- This is the Shannon-cost problem with state-specific marginal costs

# Two-Step Parameter Recovery

By the invariant likelihood ratio logic, optimal revealed posteriors obey

$$\frac{\gamma_{a_I,t}(\omega_I)}{\gamma_{a_O,t}(\omega_I)} = \exp\left(\frac{U_t(a_I,\omega_I) - U_t(a_O,\omega_I)}{\kappa_I}\right)$$

$$\frac{\gamma_{a_O,t}(\omega_O)}{\gamma_{a_I,t}(\omega_O)} = \exp\left(\frac{U_t(a_O,\omega_O) - U_t(a_I,\omega_O)}{\kappa_O}\right)$$

- Step 1: use matches without Hawk-Eye review to recover $\kappa_I$ and $\kappa_O$

$$\kappa_I = \frac{1}{\ln \gamma_{a_I,\text{No AI}}(\omega_I) - \ln \gamma_{a_O,\text{No AI}}(\omega_I)}$$

$$\kappa_O = \frac{1}{\ln \gamma_{a_O,\text{No AI}}(\omega_O) - \ln \gamma_{a_I,\text{No AI}}(\omega_O)}$$

# Two-Step Parameter Recovery

- Step 1: use matches without Hawk-Eye review to recover marginal attention costs

$$\kappa_I = \frac{1}{\ln \gamma_{a_I,\text{No AI}}(\omega_I) - \ln \gamma_{a_O,\text{No AI}}(\omega_I)}$$

$$\kappa_O = \frac{1}{\ln \gamma_{a_O,\text{No AI}}(\omega_O) - \ln \gamma_{a_I,\text{No AI}}(\omega_O)}$$

- Step 2: use matches with Hawk-Eye review and observed challenge rates to recover the oversight penalties

$$c_I = \frac{1 - \kappa_I \left(\ln \gamma_{a_I,\text{AI}}(\omega_I) - \ln \gamma_{a_O,\text{AI}}(\omega_I)\right)}{\eta_I} - 1$$

$$c_O = \frac{1 - \kappa_O \left(\ln \gamma_{a_O,\text{AI}}(\omega_O) - \ln \gamma_{a_I,\text{AI}}(\omega_O)\right)}{\eta_O} - 1$$

# Parameters Recovered

| Parameter | Estimate | Interpretation |
|---|---|---|
| $\gamma_{a_I,\text{No AI}}(\omega_I)$ | 0.849 | posterior accuracy when calling in |
| $\gamma_{a_O,\text{No AI}}(\omega_O)$ | 0.876 | posterior accuracy when calling out |
| $\kappa_I$ | 0.580 | marginal cost, ball-in state |
| $\kappa_O$ | 0.510 | marginal cost, ball-out state |
| $\eta_I$ | 0.449 | challenge rate, Type II errors |
| $\eta_O$ | 0.415 | challenge rate, Type I errors |
| $c_I$ | -1.374 | Type II oversight penalty |
| $c_O$ | -0.903 | Type I oversight penalty |

$c < -1$ means being overruled is worse than being wrong without correction

# Interpretation

- Type I overturned calls: $c_O = -0.903$, not meaningfully worse than the ordinary incorrect-call cost
- Type II overturned calls: $c_I = -1.374$
- Umpires appear to care about Type II errors about 37% more when those errors are publicly overturned
- This matches the institutional asymmetry:
  - calling a ball in when it was out is corrected cleanly,
  - calling a ball out when it was in can require an awkward replay decision

# Conclusion

- AI oversight improved average accuracy for close line calls
- It also changed behavior: umpires shifted toward calling more balls in near the line
- A rational-inattention model translates this shift into an implied utility cost of being overturned
- NIAS gives wide bounds, NIAC tightens them, and Shannon costs recover point estimates
- The broader lesson: even accurate oversight technologies can change the human decision problem they are meant to improve

*Thank You!*

# Backup: How Much Does AI Matter?

- A live implementation question is whether humans respond differently when AI rather than another human performs oversight
- Existing evidence suggests people care who evaluates them, and AI can make evaluation feel different
- But the main reason AI matters is scalability: it expands where frequent, low-cost oversight can occur

Back

# Backup: Tournaments and Matches

| | Before Hawk-Eye review | After Hawk-Eye review |
|---|---|---|
| Tournaments | 7 | 28 |
| Matches | 109 | 589 |
| Points | 15,439 | 83,898 |
| Bounces | 71,638 | 402,446 |

- The study uses early rollout variation across tournaments
- Umpiring pool and standard instructions were stable over the relevant period

# Backup: Where the Mistakes Are

| Sample before Hawk-Eye review | Share of bounces | Mistake rate |
|---|---|---|
| All bounces | 100% | 0.61% |
| Within 100 mm of the line | 3.6% | 13.89% |
| Within 20 mm of the line | 0.78% | 32.91% |

- More than 93% of challenges involve calls within 100 mm of the line
- Structural estimates focus on the close-call regions where mistakes are meaningful and observable

# Backup: Alternative Mechanisms

- Explicit instruction?
  - A leading official reported no instruction to call more balls in on close calls
- Different challenge exposure?
  - Type I errors: 41.5% challenged
  - Type II errors: 44.9% challenged
- More performance information?
  - Before stadium review, many incorrect calls were already visible through broadcast coverage
- The main change was public, on-the-spot overturning of on-court decisions