Man, Umpire, or Machine?

Challenges of the Challenge System in Tennis

Melbourne Data Science Week

June 2, 2017

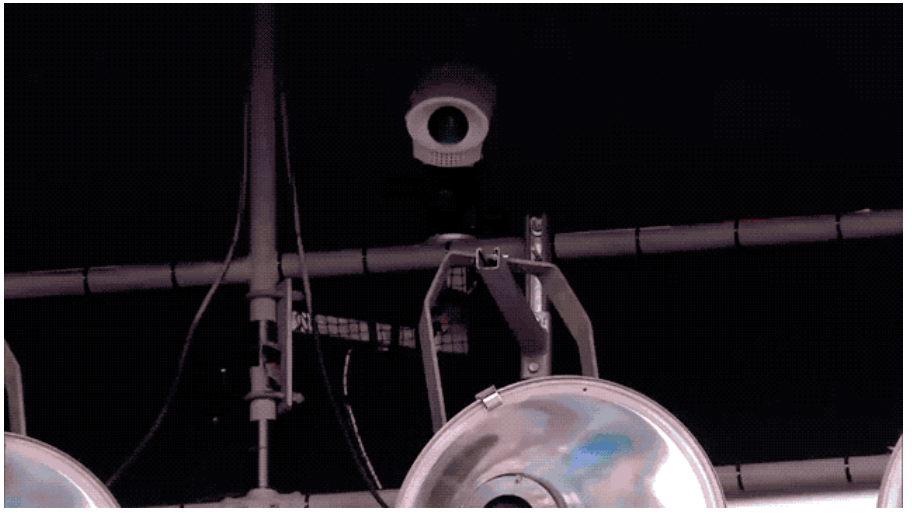
About Me

- ISEAL Research Fellow at Victoria University
- Principal Data Scientist for Tennis Australia's Game Insight Group
- Tennis Blog: on-the-t.com
- @StatsOnTheT

Research Collaborators

- Machar Reid, GIG Director
- Jeff Sackmann, *Tennis Abstract*

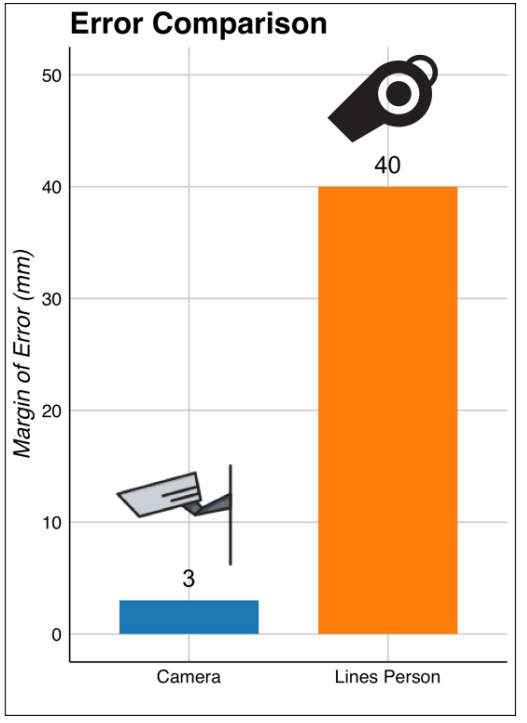
History of Tennis' Challenge System



- Introduced in 2006
- Rolled out gradually
- Now available at most hard and grass court ATP and WTA matches, but *not* all

How the System Works

- Multiple cameras
- Estimation of 3-D trajectories
- 3 mm region of uncertainty



Current Rules

1. 3 challenges per set
2. One additional challenge for tiebreaks
3. When a player is wrong, the player loses a challenge

Previous Theoretical Work

Nearly all prior work on the challenge system has been theoretical. The aim has been to define *optimal* strategies for how challenges are used.

Previous Theoretical Work

Nearly all prior work on the challenge system has been theoretical. The aim has been to define *optimal* strategies for how challenges are used.

In an optimal world, players would:

Previous Theoretical Work

Nearly all prior work on the challenge system has been theoretical. The aim has been to define *optimal* strategies for how challenges are used.

In an optimal world, players would:

- Reserve challenges for most important points

Previous Theoretical Work

Nearly all prior work on the challenge system has been theoretical. The aim has been to define *optimal* strategies for how challenges are used.

In an optimal world, players would:

- Reserve challenges for most important points
- Use more challenges later in a set

Previous Theoretical Work

Nearly all prior work on the challenge system has been theoretical. The aim has been to define *optimal* strategies for how challenges are used.

In an optimal world, players would:

- Reserve challenges for most important points
- Use more challenges later in a set
- Use more challenges in tiebreaks

Previous Theoretical Work

Nearly all prior work on the challenge system has been theoretical. The aim has been to define *optimal* strategies for how challenges are used.

In an optimal world, players would:

- Reserve challenges for most important points
- Use more challenges later in a set
- Use more challenges in tiebreaks
- Be less selective in close matches

Prior Empirical Work

Abramitzky et al. performed one of the few empirical studies of challenges looking at 2,000 men's challenges.

Prior Empirical Work

Abramitzky et al. performed one of the few empirical studies of challenges looking at 2,000 men's challenges.

They conclude players are basically optimal but the study was limited by:

- Lack of data on points not challenged
- No information on ball or player position
- No data for women players
- A number of questions about the function of the challenge system were not examined

But Do Players *Really* Use Challenges Optimally?

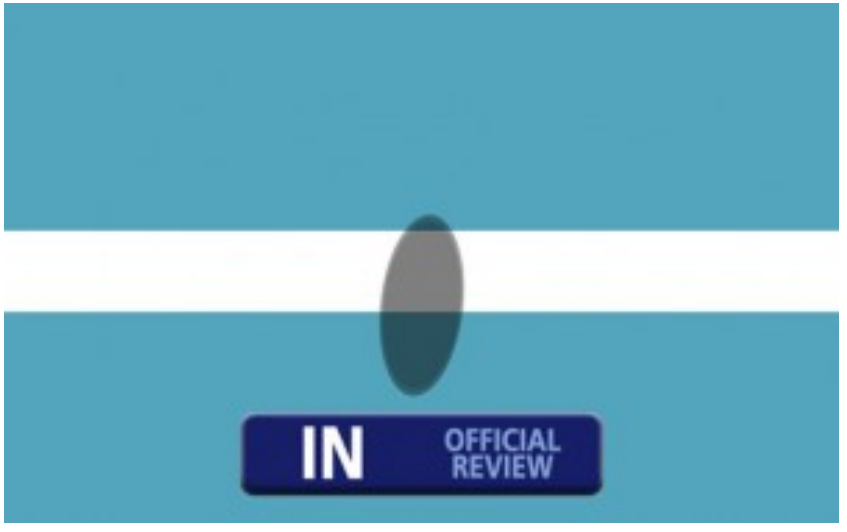
Research Objectives

Our research addressed questions on four different topic concerning the challenge system.

Topic	Questions
Player Decision-Making	What influences the use and success of challenges?
Gender	How does decision-making with challenges differ between male and female players?
Misuses of Challenges	Are there unintended uses?
Official Bias	Do tennis chairs show impact bias?

Capturing Challenges

We developed an challenge capture program in *OpenCV* to flag the occurrence of challenges in single-camera match broadcasts.

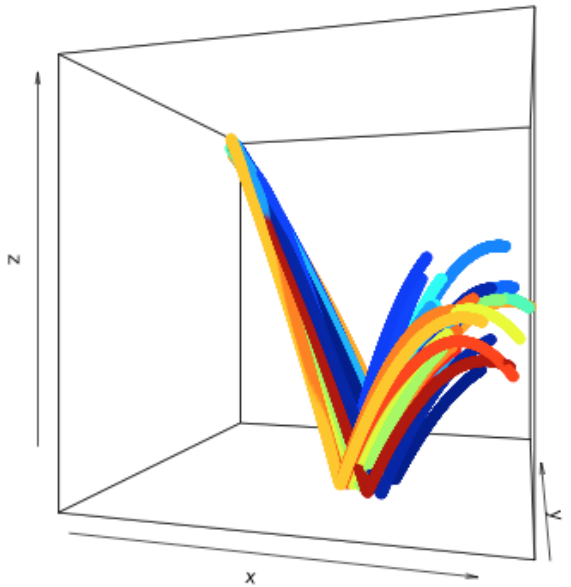


Coding Challenges

Three independent coders reviewed each challenge and recorded information about the match, players, score, shot, call, and outcome of the challenge.

Data

Coded challenges from 86 men's and 82 women's matches at the 2016 Australian Open were combined with shot-level tracking data.

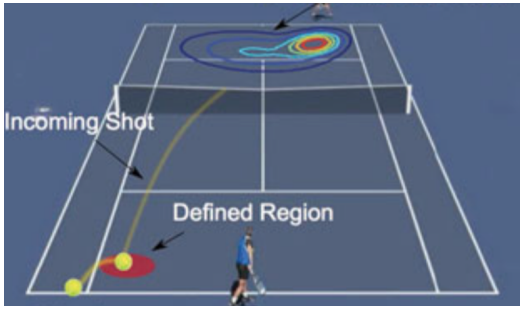


Attribute	Men	Women
Matches	86	82
Challenged Shots	697	340
Unchallenged Shots	110923	80196

Player Decision-Making

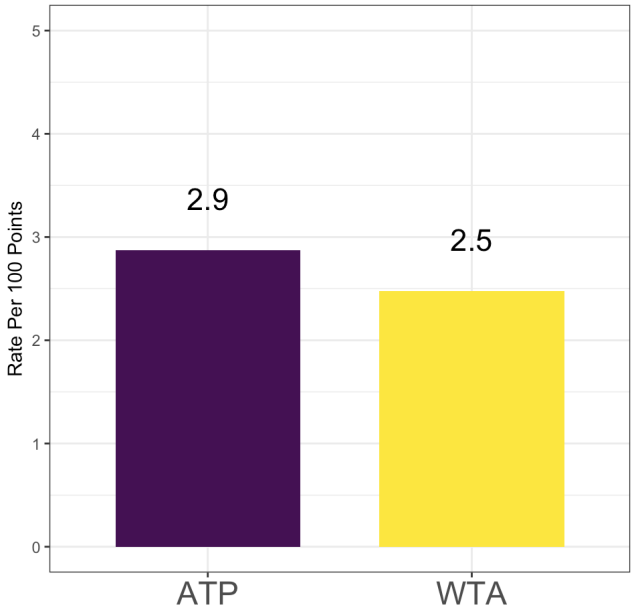
Player Decision-Making: Methods and Variables

We used logistic regression at the point- and shot-level to understand when a challenge is used and when a challenge is successful.



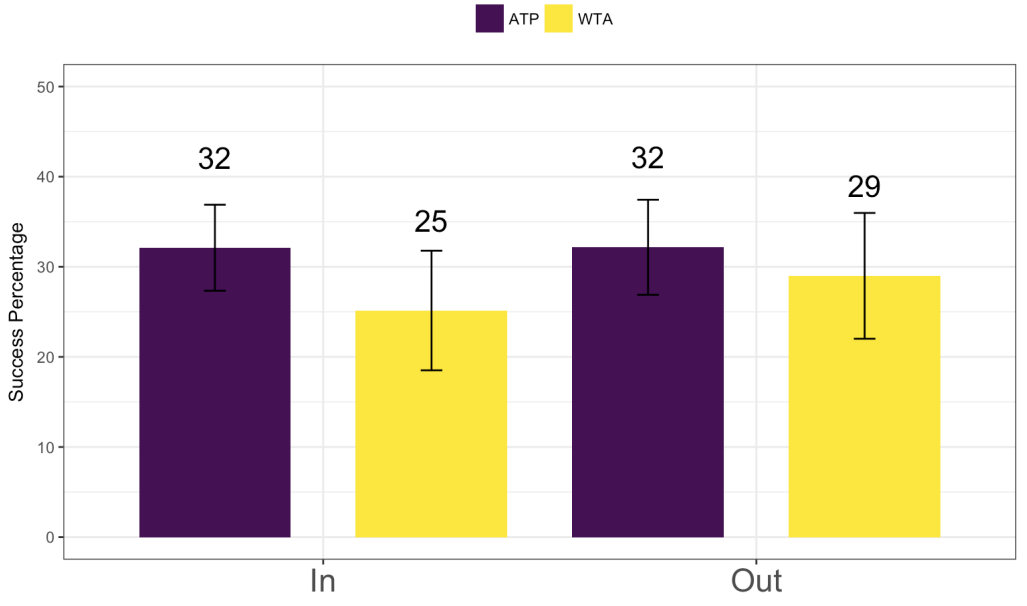
Type	Variable
Score	Set
	Serve
	Tibreak
	Importance
History	Points remaining
	Challenges remaining
	Challenge outcomes
Shot attributes	Shot speed
	Shot location
Player position	Distance from shot
	Orientation

Frequency Of Challenges



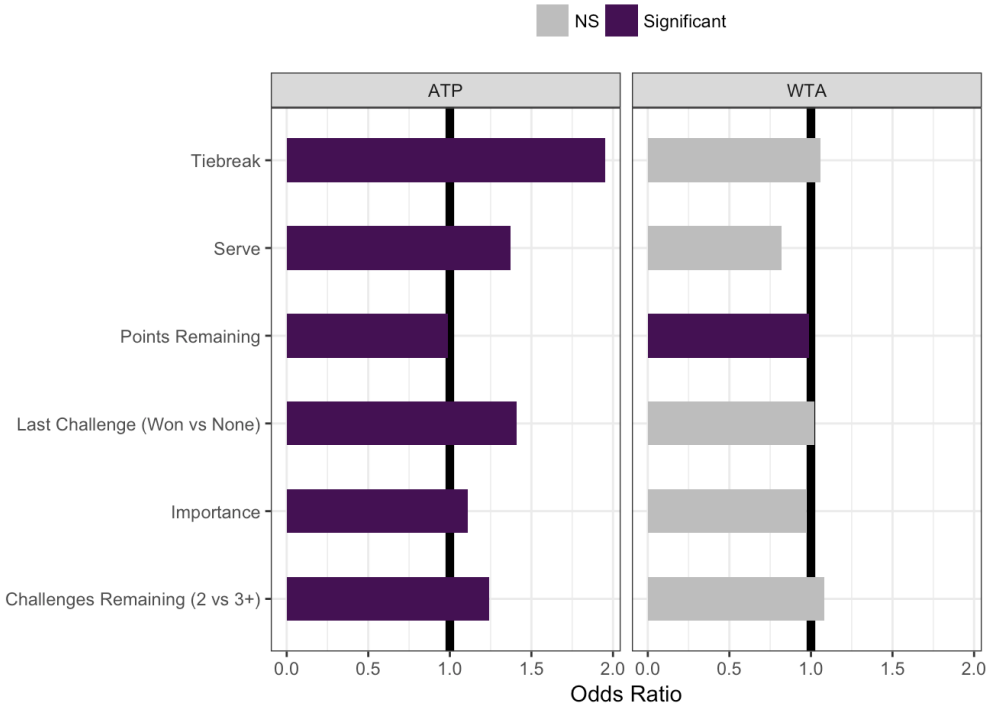
- On average, 3 challenges occur for every 100 points played in a professional men's match
- In a professional women's match, the number is 2.5 challenges

Success of Challenges



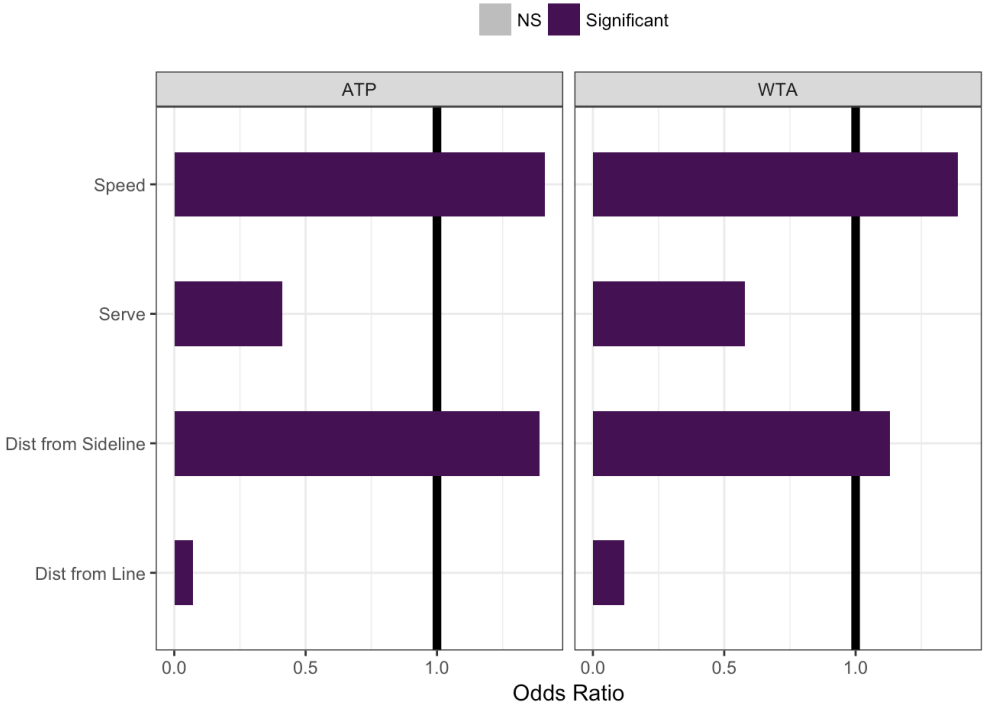
- Both male and female players are successful with approximately 1 of every 3 challenges
- There is no evidence that the marginal rate of success depends on the lines person's call

Point-Level Regression: Use of Challenge



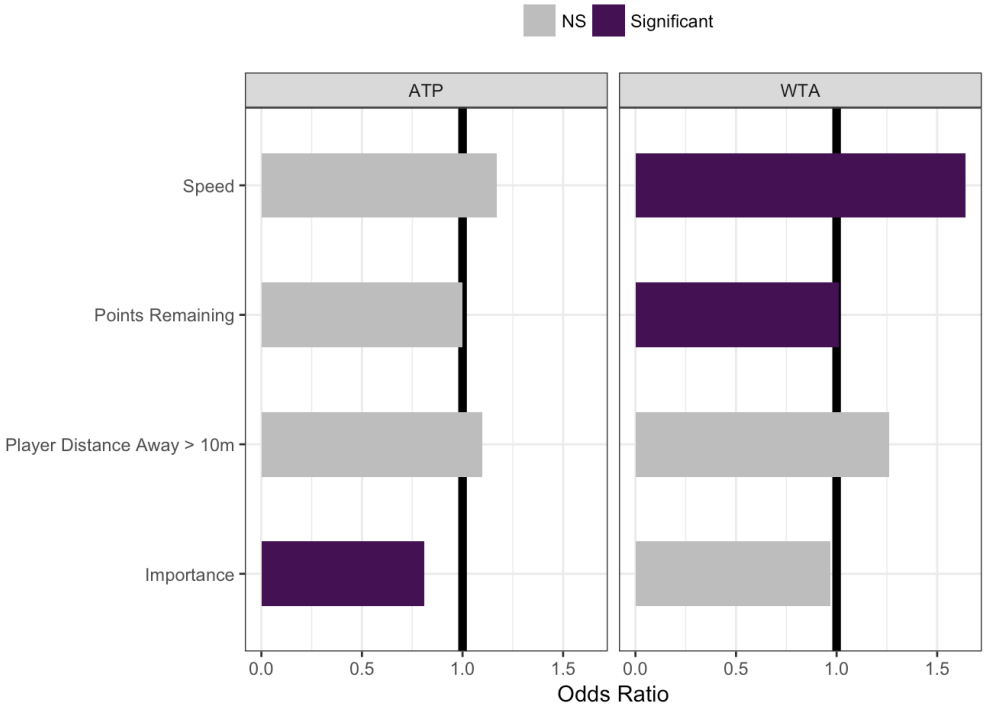
[1] Adjusted for set and previous rally length, which were not statistically significant factors

Shot-Level Regression: Use of Challenge



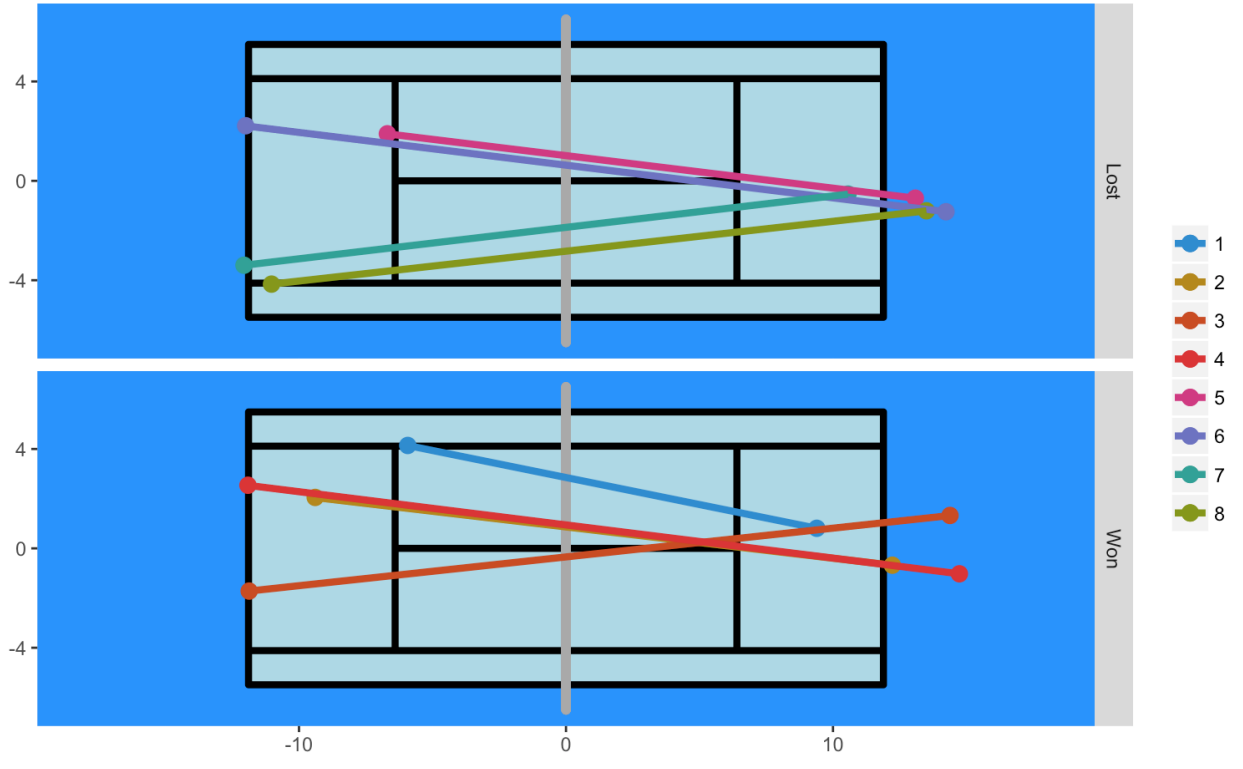
[1] Adjusted for all point-level factors and only those with changed effects (compared to the point-level analysis) are shown

Shot-Level Regression: Success



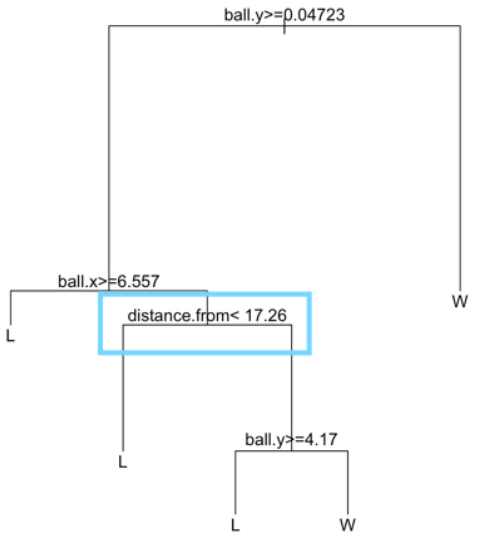
[1] Adjusted for all point-level factors, ball, and player location

Role of Player Position

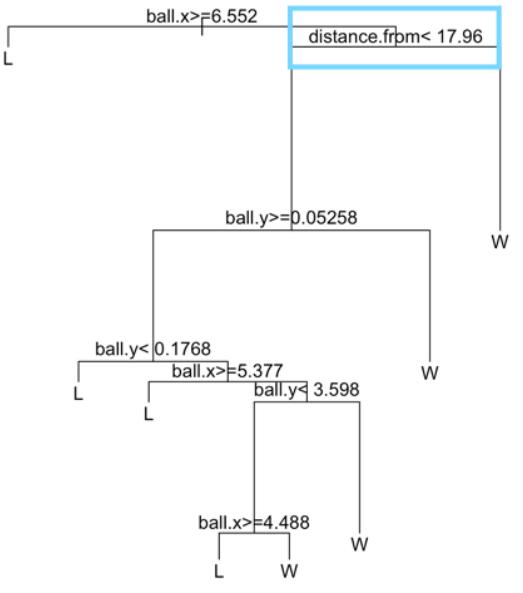


Decision Tree Analysis

When Server Challenges



Men



Women

Gender Differences

Gender Differences

- Female players tend to use challenges less frequently but are as successful with challenges as male players

Gender Differences

- Female players tend to use challenges less frequently but are as successful with challenges as male players
- Points remaining in a set was the *only* strong factor found to influence the use of challenges for female players

Gender Differences

- Female players tend to use challenges less frequently but are as successful with challenges as male players
- Points remaining in a set was the *only* strong factor found to influence the use of challenges for female players
- Importance was strongly associated with the use of challenges for male players but *not* female players

Gender Differences

- Female players tend to use challenges less frequently but are as successful with challenges as male players
- Points remaining in a set was the *only* strong factor found to influence the use of challenges for female players
- Importance was strongly associated with the use of challenges for male players but *not* female players
- Distance from shot was a strong influence on success of servers challenging their own for both men and women

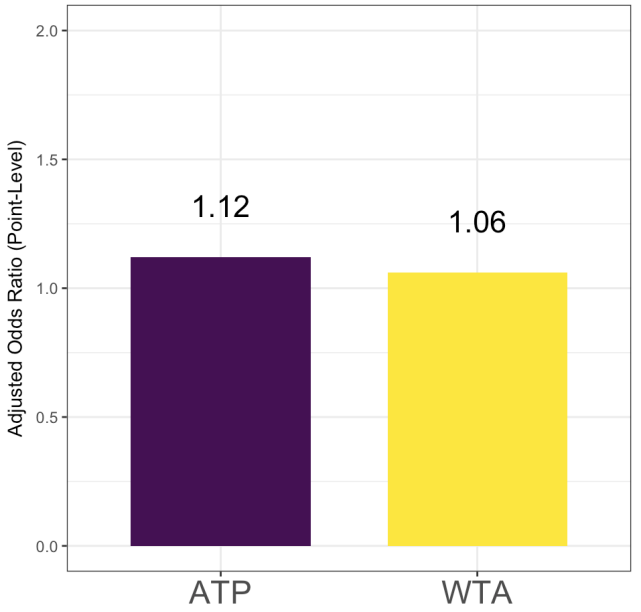
Misuses of Challenges

Misuses: Questions and Methods

Question	Approach
Do players use challenges to gain time?	Adjust regression models for shot length
Do players 'delay' challenges in an effort to gain information?	Calculate the frequency of delays

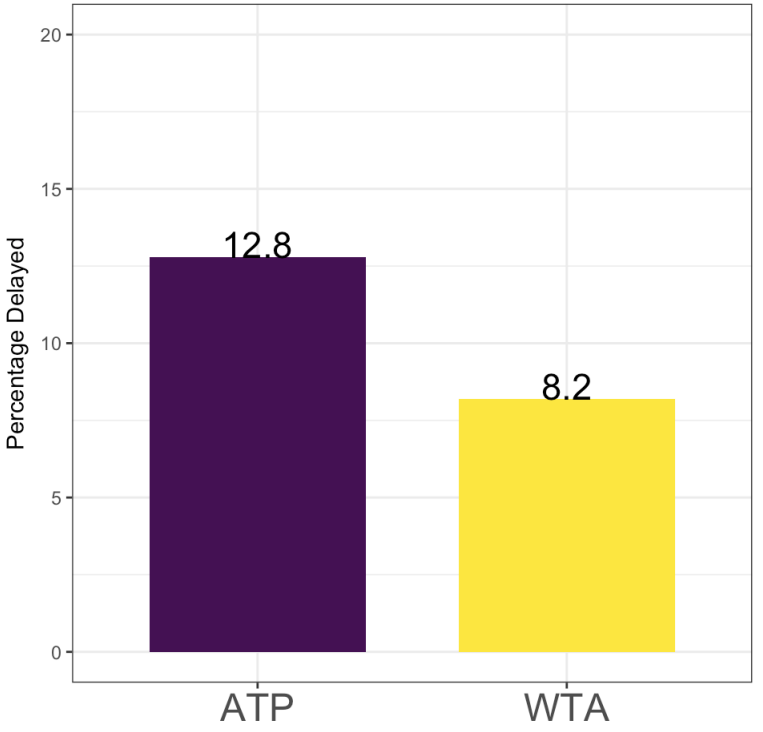
[1] We consider a 'delay' to have occurred if an additional shot was played after the shot that was eventually challenged.

Challenges & Recovery



We found a positive but non-significant relationship between extended rallies (>10 shots) and the use of a challenge.

Delayed Challenges



Official Bias

Potential for Bias

- On an overturned 'Out' call, an official decides whether to award the point or replay the point.
- Rewarding the point is equivalent to concluding that the challenged shot *was not* returnable.

Evaluating Bias: Methods

- We had 3 coaches independently review the points on every overturned out call and decide whether the shot was *returnable* or not

Evaluating Bias: Methods

- We had 3 coaches independently review the points on every overturned out call and decide whether the shot was *returnable* or not
- We treated the consensus classification as the 'ground truth'

Evaluating Bias: Methods

- We had 3 coaches independently review the points on every overturned out call and decide whether the shot was *returnable* or not
- We treated the consensus classification as the 'ground truth'
- The agreement, false positive, and false negative rates of official decisions to reward points were calculated by comparison to the ground truth labels

Official Impact Aversion

Attribute	Men	Women
Overturned Out Calls	101	43
Overall Agreement	74.5%	79.2%
False Positive Rate	21%	14%
False Negative Rate	28%	26%

- *False negatives* here are cases where chairs replayed the point when experts judged the challenged shot to be a winner
- *False negatives* indicate the presence of impact aversion

Practical Implications

Practical Implications

- Professional tennis players should use *more* of their challenges!

Practical Implications

- Professional tennis players should use *more* of their challenges!
- Success rates aren't strongly influenced by context so players should focus on concentrating challenges on important situations

Practical Implications

- Professional tennis players should use *more* of their challenges!
- Success rates aren't strongly influenced by context so players should focus on concentrating challenges on important situations
- There is a non-negligible percentage of challenge requests that are made with a delay, which tournaments should take steps to reduce

Practical Implications

- Professional tennis players should use *more* of their challenges!
- Success rates aren't strongly influenced by context so players should focus on concentrating challenges on important situations
- There is a non-negligible percentage of challenge requests that are made with a delay, which tournaments should take steps to reduce
- Officials tend to be conservative when making a decision to replay or award points after a successful challenge, which currently weakens the integrity of the challenge system

References

- Kovalchik, S, Reid, M. "Player, Official, or Machine?: The Uses and Misuses of the Challenge System in Professional Tennis". Under Review.
- Abramitzky, Ran, et al. "On the optimality of line call challenges in professional tennis." International Economic Review 53.3 (2012): 939-964.

skovalchik@tennis.com.au

@StatsOnTheT