The Review of Economics and Statistics Fellowship for NBER Summer Institute

Name: Sungwoo Cho
Email: sungwoo.cho@westpoint.edu
Affiliation: United States Military Academy at West Point
Date of PhD: June 2024
Website: sungwoocho.com

Current Research Activity:

My research focuses on the intersection of technology, AI, and human capital, and the economics of crime. In a recent working paper titled **"The Effect of Robot-Assistance on Skills"**, I examine how technology assistance influences human skills and investigate what happens when that technology is withdrawn. This work aims to better understand the durability of human skills when augmented by AI or technology tools and finds that individuals who were previously aided by technology experience a decline in their task performance when they resume the task without assistance. In another project titled **"Are Public Sector Performance Incentives Distortionary? Evidence from Police CompStat"**, jointly with Felipe Gonçalves, Andrew Hess, Bernardo S. Silveira and Emily Weisburst, I study the impact of CompStat on police productivity and find that police officers raise arrest activity in the week preceding CompStat meetings, and that higher arrest counts during periods used for CompStat reporting are linked to greater chances of promotion.

Building on my existing research agenda of studying human skill and decision-making, I plan to integrate these two strands of research by studying the impact of AI programs such as JusticeText and Draft One on public defenders, police officers, and other key actors in the criminal justice system. By analyzing how AI tools influence worker productivity and decision-making, I aim to shed light on the broader implications of technology adoption for human capital development in high-stakes environments.

At the Summer Institute:

As a new assistant professor working at the intersection of Artificial Intelligence and the Economics of Crime, I bring a distinctive perspective to the Summer Institute's research community. I focus on how technological innovations reshape human capital development and skill formation, with a growing interest in the effects of AI tools such as JusticeText and Draft One. Building on my broader research into expert decision-making, I aim to examine how these technologies influence productivity and outcomes within the criminal justice system. I look forward to contributing empirical insights into how digital technologies are reshaping skill development and the nature of expertise in the modern economy.

I hope to use the Summer Institute to engage with the latest questions in the economics of technology and crime. I am eager to build lasting connections with fellow scholars who are exploring similar themes and to broaden my thinking through exposure to new ideas and expert feedback. This experience will be invaluable in enhancing my research and generating new ideas that will drive my work in the future.

Sessions:

I am most interested in attending the **Labor Studies**, **Law and Economics**, and **Economics of Crime** sessions.

Sungwoo Cho

(Last Updated: April 2025)

Contact Information

United States Military Academy at West Point
Department of Social Sciences
130 Cullum Road, West Point, NY 10996

Phone: +1 (734) 545-4662
Email: sungwoo.cho@westpoint.edu
Website: sungwoocho.com

Employment

Assistant Professor of Economics, United States Military Academy at West Point 2024–

Education

Ph.D. Economics, University of California, Los Angeles 2024
M.A. Economics, Duke University 2018
B.S. Economics *with Distinction and High Honors*, University of Michigan, Ann Arbor 2016

Research

Publications

1. "The Impact of Fear on Police Behavior and Public Safety" (with Felipe Gonçalves and Emily Weisburst), Accepted, *The Review of Economics and Statistics*, 2024
2. "The Impact of Cash Transfers to Poor Mothers on Family Structure and Maternal Well-Being" (with Anna Aizer, Shari Eli and Adriana Lleras-Muney), *American Economic Journal: Applied Economics*, 16 (2): 492–529, April 2024

Working Papers

3. "The Effect of Robot Assistance on Skills"
4. "Racial Bias and Decision Fatigue"

Selected Works in Progress

5. "The Long Run Impact of Cash Transfers to Poor Families: New Estimates using Larger Samples and New Methods" (with Anna Aizer, Shari Eli, Joseph Ferrie and Adriana Lleras-Muney)
6. "Are Public Sector Performance Incentives Distortionary? Evidence from Police CompStat" (with Felipe Gonçalves, Andrew Hess, Bernardo S. Silveira and Emily Weisburst)

Media Coverage of Research

Charter

Honors, Fellowships & Grants

Best Paper Award, Albert Family Proseminar in Applied Microeconomics, UCLA 2023
Graduate Representative, Southern Economic Association Annual Meeting 2023
Dissertation Year Fellowship, UCLA 2023–2024
Summit Fellowship in Applied Economics, UCLA 2023
Lewis L. Clarke Graduate Fellowship, UCLA 2022
Graduate Research Mentorship, UCLA 2020–2021
Graduate Summer Research Mentorship, UCLA Summer 2020
Outstanding Teaching in Remote Learning, UCLA Winter 2020
Departmental Fellowship, UCLA 2018–2023
M.A. Merit Scholar Award, Duke University 2017
Duke Economics Master's Scholar Award, Duke University 2016–2018
Phi Beta Kappa, University of Michigan 2016

Professional Activities

Presentations (including scheduled)
* indicates coauthor presentation, † indicates discussant, ‡ indicates invited attendee

Lee Bun Song Population & Aging Workshop 2026

American Economic Association† 2025
Arnold Ventures (Building Research Ideas and Generating Evidence – Retail Theft; Jail Churn and Disorder)‡
Lee Bun Song Population & Aging Workshop†

Institut Mines-Télécom Business School 2024
United States Military Academy at West Point
Hong Kong University of Science and Technology
Monash University
Data-Driven and Experiment-Based Evidence for Policy
University of Pennsylvania Wharton School (AI and the Future of Work; AI Horizons)
Western Economic Association International
Lee Bun Song Population & Aging Workshop†
Yale Law School (Convening on Police Officer Standards and Training Commissions)‡

Korea-America Economic Association (Job Market Conference) 2023
Southern Economic Association

American Economic Association* 2022
Association for Public Policy Analysis and Management
National Bureau of Economics Research Summer Institute*

Society of Labor Economists* 2021
American Law and Economics Association
Southern Economic Association

Referee

American Economic Journal: Economic Policy
American Economic Review
Journal of Policy Analysis and Management
Population Research and Policy Review
Quarterly Journal of Economics

Grant Reviewer
National Science Foundation

Other
Association for Public Policy Analysis and Management Program Reviewer 2025

Teaching Experience

Instructor, USMA 2024–
Principles of Economics (SS201)

Instructor, UCLA 2022–2023
Microeconomic Theory (Econ 101)

Teaching Assistant, UCLA 2019–2024
Microeconomic Theory (Econ 11)
Microeconomic Theory (Econ 101)
Probability and Statistics for Economists (Econ 41)
Intergenerational Poverty in America (Econ 133)

Thesis Advisor, USMA
Luke Hoelzel (2025)

Research Experience	Research Assistant to Professor Adriana Lleras-Muney, UCLA	2020–2021, 2022
	Research Assistant to Professor James W. Roberts, Duke University	2016–2017
Other Experience	Legislative Assistant, National Assembly of the Republic of Korea	Summer 2014
	Volunteer Translator, Compassion Korea	2009–
Miscellaneous	**Languages:** English (native), Korean (native)	
	Citizenships: United States, Republic of Korea	

The Effect of Robot Assistance on Skills

Sungwoo Cho*

April 2025

Click for the Most Recent Version

Abstract

How does work with robots change human capital? I study a unique setting in which professional baseball leagues provided – and subsequently removed – access to robot assistance for umpires. Umpires demonstrated improved precision and accuracy in ball-strike decisions while using robot assistance, but their performance declined substantially below pre-assistance levels after it was removed. Both highly skilled and inexperienced umpires exhibited large declines in performance after the removal of robot assistance. Umpires who used robot assistance for longer periods of time faced a steeper decline in accuracy than those who used it for shorter periods. Umpires who worked a full season with robot assistance did not fully return to their initial skill level by the end of the following season. In contrast, umpires who missed an entire canceled season during the COVID-19 pandemic suffered much smaller skill depreciation effects, indicating that robot-exposure effects are not driven solely by umpires not using their skills. Finally, I show that robot-exposed umpires experience skill deterioration in determining whether a baserunner is safe, showing that skill depreciation extends beyond robot-assisted tasks.

JEL Classification: J24, J64, O33
Keywords: Robot, Labor Productivity, Skill.

*Cho: United States Military Academy at West Point, sungwoo.cho@westpoint.edu. I am grateful to Adriana Lleras-Muney, Felipe Gonçalves, Emily Weisburst and Martha Bailey for insightful comments and suggestions. I also thank Anna Aizer, Charlie Becker, A. Craig Burnside, Daniel Haanwinckel, Kareem Haggag, Byung Soo Lee, Juliana Londoño-Vélez, Yotam Shem-Tov as well as seminar and conference participants at the Albert Family Fund Proseminar in Applied Microeconomics at UCLA, Korea-America Economic Association, Southern Economic Association, Institut Mines-Télécom Business School, United States Military Academy at West Point, Hong Kong University of Science and Technology, Monash University, Data-Driven and Experiment-Based Evidence for Policy, AI and the Future of Work at the University of Pennsylvania Wharton Business School and Western Economic Association International for helpful comments. I benefited from facilities and resources provided by the California Center for Population Research at UCLA (CCPR), which receives core support (P2C-HD041022) from the Eunice Kennedy Shriver National Institute of Child Health and Human Development (NICHD). All opinions expressed in this manuscript are those of the author and do not represent the opinions of the United States Military Academy (USMA), the United States Army, or the Department of Defense. All remaining errors are my own.

"If men learn this, it will implant forgetfulness in their souls; they will cease to exercise memory because they rely on that which is written..."

- Socrates on writing, from Plato's Phaedrus

1 Introduction

As early as the time of Socrates, humans have harbored a fear that new technologies may deteriorate human knowledge and skills. This fear has animated recent popular discussions over the impacts of industrial robots and artificial intelligence (AI) technologies. Instead of fully displacing humans, however, numerous robots and technologies may collaborate with humans in practice, enhancing and complementing their work. Combining human expertise with robot assistance has the potential to yield gains that cannot be fully achieved by exclusively depending on either of them alone. However, heavy reliance on technology could lead to decreases in workers' skills and increases in workers' technological dependence.[1] The deterioration of human capital caused by dependence on technology is especially important in the labor market and can carry significant consequences. For example, while automated systems improved safety records in the airline industry, pilots now experience difficulties manually operating aircraft, contributing to some recent disasters (Nicas and Wichter, 2019).[2] Similarly, doctors' physical diagnosis skills have declined with increasing reliance on technologies, shifting focus away from patients (Aw, 2014).[3]

Across a variety of settings, expert ability is one of the key determinants of service outcomes (e.g. Chan et al., 2022; Currie and MacLeod, 2020). Expert's decision-making skills are becoming increasingly important as technologies substitute for humans in routine tasks (Deming, 2021). Yet, the effects of robot on decision-makers is ex-ante ambiguous. Experts may shirk and not fully utilize technology (Agarwal et al., 2023), receive external feedback and learn (Bhuller and Sigstad, 2024), or even unlearn by not performing the task.

[1] Cymek et al. (2023) find evidence of "social loafing" - decreased personal effort on a task completed by a group - when working with a robot.

[2] Over 60% of recent accidents can be attributed to pilots who have lost their practical skills. For example, an investigation by the National Transportation Safety Board identified pilot error as the primary contributing factor to the Asiana Airlines Flight 214 crash that occurred on July 6, 2013. The pilots did not properly monitor airspeed and altitude, and they allowed the aircraft to sink dangerously low and stall just before landing. One contributing factor was the overreliance on automation in the cockpit.

[3] Gong et al. (2019) argue that medical students now avoid specializing in radiology due to fear of job loss caused by AI assistance. However, Agarwal et al. (2023) find that radiologists-AI collaboration is still suboptimal due to radiologists' biases against AI.

Understanding how experts' decision-making evolves with technology is crucial for preparing for future labor market dynamics. Therefore, I seek to identify the impact of robot on expert decision-makers.

In this paper, I ask three questions: First, what is the impact of robot assistance on individual workers' skill and productivity? Second, how do robot exposure effects vary across the skill distribution? Third, does longer work duration with robot assistance increase skill depreciation, and do workers bounce back and regain skills upon removal of the robots?

Examining the impact of technology assistance on human capital is empirically challenging. First, direct measures of individuals' skill and productivity are often unavailable and typical measures of human capital, such as wages, can be influenced by robot assistance through various channels of complementarity. Even when data are available, it is difficult to clearly identify the effect of robots on individual skills because individuals do not randomly adopt or drop robot assistance. Firms and humans choose robot assistance if they believe that it improves their performance and stop using it when they do not. In addition, the durations of robot assistance are usually not random, making it difficult to study the effect of not using a skill over a certain amount of time on skill depreciation as those with lower skills may choose to be assisted for longer time.

I address these challenges by studying the implementation of the automated ball-strike system (ABS, or "robot umpire") in professional baseball leagues, designed to assist umpires in their game decisions. Robots are not replacing the umpires but rather providing additional data about the location of the ball, while the umpire continues to make the final decision. Professional baseball leagues provide several advantages when studying the impact of robot assistance. First, umpires are professional decision-makers who make frequent similar decisions, allowing for precise individual measures of productivity.[4] I observe the decisions solely made by the umpire as well as whether the decision is correct or not as determined using post-processed data. Second, there is substantial variation in the implementation of robots. There are 11 different leagues in the Minor League Baseball (MiLB) system, with each league providing a stage for experimental rule changes. The implementation of robot assistance only occurred in the Single-A Florida State League starting in 2021 and the Triple-A Pacific Coast League in 2022. With umpires moving across leagues, I can observe worker-specific usage of

[4]In baseball, umpires make decisions about whether a thrown pitch is a strike or a ball. A "strike" is a pitch within the defined strike zone, and a "ball" is a pitch outside of this zone.

technology and their behaviors before and after robot adoption, as well as *with* and *without* robot assistance. The *removal* of robot assistance enables a unique opportunity to evaluate human capital in isolation, disentangling it from the influence of human-robot collaboration. Third, a change to intermittent robot utilization in 2023 provides an ideal setting to study how skills depreciate as a function of how long individuals have been assisted by robots. In 2023, Triple-A leagues imposed a rotation system in which half of the season utilized robot assistance, and the other half did not. Umpires are quasi-randomly assigned to games and thus, the duration of robot assistance throughout a season is as-good-as random. Finally, the skill sets and cognitive processes involved in umpires' decisions extend beyond the realm of baseball. As the skill sets are commonly shared by professionals in various domains of expert deicsion-making, the findings can provide broader lessons that are applicable to a wide range of contexts.

I examine over 62,000 professional baseball games played between 2017 and 2023 in Minor and Major League Baseball (MLB) and employ a difference-in-differences design exploiting the staggered adoption of robots across leagues and the movement of umpires across leagues. This approach compares the change in the performance of umpires before and after being assisted by robots to the change in performance of umpires that are never assisted by robots.

The majority of studies in the literature rely on industry and firm-level data, finding uneven effects of robots (Acemoglu et al., 2023, 2020; Acemoglu and Restrepo, 2020; Aghion et al., 2020; Barth et al., 2020; Bonfiglioli et al., 2020; Dauth et al., 2021; Hirvonen et al., 2022; Humlum, 2022). Robots impact wages through multiple channels beyond just human capital, complicating the use of wage changes as a measure of changes in human capital. Using a direct measure of individual-level productivity, I find that umpires perform with higher precision and accuracy when assisted by a robot. In the season with robot assistance, umpires' decision accuracy increases by 6.4 percentage points, on average. In particular, both Type I errors (e.g., incorrectly calling a strike) and Type II errors (e.g., incorrectly calling a ball) decrease significantly. Using an event-study specification, I confirm that there are no pretrends suggesting that the timing of robot implementation is as-good-as random and that umpire moves are not correlated with the adoption of robot assistance. While the umpires of different prior skill levels all benefit from robot assistance, the skills distribution, with robots, compresses by 36%, broadly consistent with patterns in other settings (Brynjolfsson

et al., 2023; Kanazawa et al., 2022; Noy and Zhang, 2023; Tafti, 2022). When restricting attention to more ambiguous decisions, robot assistance improves the umpire's accuracy by 12.3 percentage points. Further, umpires suffer less from decision biases with robot assistance: negative autocorrelation in decisions, or the "gambler's fallacy" (Chen et al., 2016), is reduced by almost 100% and omission bias, a tendency to call more pitches strikes when the next called ball would end an at-bat (i.e., to avoid issuing a walk), by 89%.

Next, I investigate the consequences when humans lose access to technology. My finding that technology enhances the productivity of human workers is broadly consistent with a growing literature focused on AI technology, yet much less is known about productivity of these workers without technology (Almog et al., 2024; Brynjolfsson et al., 2023; Kanazawa et al., 2022; Kreitmeir and Raschky, 2024; Noy and Zhang, 2023). When making calls without robot assistance, umpires suffer a decline of 2.0 percentage points in accuracy relative to preassistance levels. The size of the decline is roughly equal to the gap between the median and the bottom 5th percentile umpires. For pitches that are harder to judge, accuracy declines by 3.7 percentage points. Umpires also suffer more from decision biases, showing increasing signs of omission bias by about 33%. Further, I find that robot-induced deterioration in skills differs across umpires, with the largest skill decline occurring for highly skilled umpires. When the robot assistance is removed, the lower-skilled umpires' skills decline by a statistically insignificant 0.9 percentage points, whereas the high-skilled umpires lose skill by 4.7 percentage points. The distribution of skills across the umpires is compressed by 64.6% following the work with robot assistance at the expense of high-skilled umpires relative to the preassistance period.

Further, umpires subject to a full season of work with the robot do not bounce back quickly when recalled to the task without a robot. In the first game back, the umpires sustained a large skill decline of 3.6 percentage points. The decline becomes smaller as the season progresses: it is 1.7 by the middle of the season and 1.0, and still marginally significant, by the end of the season. These results demonstrate that umpires require nearly an entire season to approach their initial skill level after a loss following a season of robot assistance and never fully recover.

Understanding how skill declines with the intensity and the length of the robot assistance provides important insight into the trade-offs of assigning robot assistance to workers and the transitions of individual workers after assistance stops. Using the 2023 Triple-A

implementation of robot assistance, I find that umpires who used robot assistance for longer periods experience more significant skill depreciation when making decisions without robot assistance. Umpires assisted by robots for just one game suffer a marginally significant skill decline of 0.3 percentage points while those assisted for two or three consecutive games experience a larger magnitude decline of 1.1 and 1.0 percentage points, respectively. The baseline results suggest that a full season of work with robot assistance induced a decline of 3.6 percentage points in skill in the first game back. This implies that as the duration of robot assistance increases, the decline in skill becomes more pronounced.

Umpires could lose skills for various reasons, including reliance on technology, reduced practice, and a shift in focus. To better understand the causes of skill loss, I compare to the job loss experiment in my data. If the decline in skill was caused by a lack of practice alone, one would expect at least some decline in skill based on results demonstrated in past literature on job loss and human capital depreciation (Benhenda, 2022; Dinerstein et al., 2022; Edin and Gustavsson, 2008; Jarosch, 2023). However, estimating human capital depreciation is difficult due to the correlation between the duration of unemployment and the productivity of workers (i.e., those with lower skills may take longer time to be reemployed).[5] To examine whether robot assistance has a distinct impact from time off work, I examine a different treatment: some umpires abruptly experienced a year-long interruption in their careers due to the COVID-19 pandemic. In 2020, the Minor League canceled the season whereas the Major League had a delayed and shortened season. Minor League umpires experienced a marginally significant skill decline of 0.8 percentage points in the first month of the return relative to the Major League umpires. Taken together, these results imply that the effect of time away from the task is smaller than the effect of robot assistance, perhaps due to reduced incentives to practice skills from the introduction of advanced technology.

While robots assisted in calling pitches, umpires remained responsible for a range of other decisions as they rotate through different positions such as those of the first-base umpire and home-plate umpire. The sharp decline in home-plate umpires' ability to accurately call pitches after working with robot assistance has raised concerns about its impact on their other skills. Increased reliance on robots for pitch calling may lead umpires to depend less on their own judgment and decision-making abilities, potentially diminishing their proficiency in other areas. In contrast, with robots handling routine pitch calls, umpires may have more

[5]See for discussion Machin and Manning (1999).

mental energy and focus to dedicate to other aspects of their role, such as making decisions on base calls. I examine whether the skill as the first-base umpire in deciding whether a baserunner is safe or out when attempting to reach first base was affected following the work with the robots. Umpires who have previously worked alongside robots experience an additional 0.098 replay review request per game (44.3%) when they officiate at first base, which is entirely due to an increase in overturned challenges (an increase of 0.094). Collectively, these findings strongly suggest that umpires with prior experience using robot assistance exhibit higher rates of inaccuracies and a decline in their skills. This implies that robot assistance impacts a broader range of abilities than it directly targets, resulting in a general deterioration.

The rest of the paper is organized as follows. Section 2 describes the umpire's decision and explains the implementation of robots and Section 3 introduces the data. Section 4 presents the empirical strategy. Section 5 shows the main results and Section 6 discusses the potential mechanisms of the findings. Section 7 shows additional results and discusses external validity of the results. Section 8 concludes.

2 Background

2.1 Umpire Decisions

Baseball umpires are professional decision makers applying official rules to the game as it is played.[6] Becoming a professional baseball umpire requires knowledge of the intricacies of the game rules, experience, and fitness. Formal training programs and umpire schools offer instructions on every area of the game.[7] Many umpires start at amateur levels and work

[6]Umpires' decisions have been used to test for the gambler's fallacy (Chen et al., 2016), racial discrimination (Parsons et al., 2011), rational inattention (Bhattacharya and Howard, 2022), attention scarcity (Archsmith et al., 2021) and status bias (Kim and King, 2014). These decisions have also been used to find that increased monitoring improves productivity (Mills, 2017), and that hotter temperatures decrease skills (Fesselmeyer, 2021). The availability of data and identification advantages sports settings offer contributes to the production of credible evidence for addressing numerous challenging research questions (See Kahn, 2000). Umpires' decisions are particularly useful tools as they provide key advantages over other settings, namely, observing precise individual-level decisions and whether they are correct or not.

[7]Attending an umpire school is a necessary step for those aspiring to umpire in the Major League and Minor League. It is extremely difficult for an umpire to reach the Major League. For example, out of about 150 students in a class, only about 20 will receive recommendations to advance to the umpire evaluation course conducted by the Professional Baseball Umpires Corporation (PBUC). Additionally, about 45 other students will be designated for placement in independent leagues. One umpire from each class will become a Major League umpire, on average.

their way up into professional baseball leagues and ultimately to Major League Baseball.

A baseball season starts in mid-February with spring training, and the regular season starts around the last week of March. The regular season consists of 162 games for the Major League and between 132 and 150 games for the Minor League. The season concludes with the postseason of up to 22 games. On average, a Major League Baseball umpire works in about 112 regular season games, of which 28 are behind the home plate, calling balls and strikes. Umpires work as a crew of usually four umpires, each working behind home plate and behind first, second, and third bases. The umpire crew rotates positions throughout the series of games, resulting in a quasi-random assignment of umpires to games.

The primary task of the umpires is rule enforcement.[8] Home-plate umpires are tasked with determining whether thrown pitches are balls or strikes.[9] These split-second decisions could sometimes be subject to human error and subjective judgment, undermining accuracy, fairness, and consistency in determining the outcome of each pitch.

In particular, when a pitch is thrown and the batter does not swing, the umpire must determine if the ball crosses home plate through the strike zone, resulting in a strike, or if it crosses home plate outside the zone, resulting in a ball being called (See Figure A1).[10] The strike zone is defined as an imaginary rectangular region over the home plate that extends roughly between the batter's shoulders and kneecaps, of roughly 20 by 25 inches in dimension. An umpire's ball-strike decisions are critical aspects of the game, as they directly affect the count of balls or strikes on the batter.[11] The count influences the strategies of both the pitcher and the batter, making it a pivotal element in the game's dynamics.

With umpires' decisions being such vital parts of the game, players, coaches, and fans closely monitor and scrutinize them for consistency and accuracy. Often, disagreements between players or coaches and umpires arise. Fans also criticize umpires for poor calls on even a single pitch. The Major League employs umpire supervisors and observers, and also uses the Supervisor Umpire Review and Evaluation (SURE) system to evaluate umpires. Umpires receive "report cards" following the game and also receive mid-year and postseason

[8] See Section A2 for a short description of a baseball game.

[9] A pitcher throws a ball from the mound to the catcher, who sits 60 feet and 6 inches away. A fastball thrown at 100 miles per hour takes about 0.4 seconds to reach the catcher.

[10] All pitches are subject to the umpire's judgment. However, the umpire makes a call only when the batter does not swing.

[11] The count refers to the number of balls and strikes on the batter. For example, if the umpire calls two strikes on the batter, the count is "0-2," meaning there are no balls and two strikes.

evaluations.[12] Therefore, umpires must maintain consistent and accurate strike zones within and across games.

2.2 Robot Assistance

Despite their efforts to strive for consistency and high accuracy, umpires often make controversial calls.[13] Because umpires' decisions are subject to human error and judgment, Major League Baseball started an experiment with an automated ball-strike system (ABS, or robot umpires) that could provide reliable assistance, ensuring accuracy, fairness, and consistency.

The robot utilizes sophisticated hardware and advanced software algorithms to track pitches and determine whether a pitch passes through the strike zone. As part of the "Hawk-Eye" tracking system, multiple cameras placed around the stadium capture the ball's flight path from different angles after a pitch is thrown. The system then identifies the pitch location and uses a calibrated strike zone, personalized for each batter, to make a ball-strike decision. The call is then communicated through an earpiece to an umpire who is standing behind home plate to physically make the decision.[14]

In 2019, Major League partnered with the independent Atlantic League of Professional Baseball (ALPB) to begin experimenting with robot assistance. Positive results in consistency and reliability led to an expansion of the experiment to Minor League; the Single-A Florida State League adopted it first in 2021, followed by the Triple-A Pacific Coast League in 2022.[15,16] In 2023, both of the Triple-A Leagues (i.e., the International and Pacific Coast

[12]See https://umpscorecards.com/home/ for examples regarding how umpire performance is measured by fans.

[13]A controversial called third strike can directly change the result of the game. For example, a game between the San Francisco Giants and the San Diego Padres on September 28th, 2020 had 27 incorrectly called pitches that benefited the Padres by 1.85 expected runs in a game they won by 1.

[14]The home-plate umpire still stands behind the plate and makes decisions on other aspects of the game. The home-plate umpire can overrule the decision communicated by the robot assistance for an obvious mistake.

[15]Minor League is a professional league below the Major League that is divided into four classes: Single-A, High-A, Double-A and Triple-A. Each class includes multiple leagues. Major League Baseball teams utilize the minor leagues to develop young players; players move through the ranks to eventually play in the Major League. Umpires also can be promoted and demoted through the major-minor league system. There are about 230 professional umpires working in both major and minor leagues. Compensation also varies significantly across levels: an average umpire in the lowest league earns about $3,000 per month, while a Major League umpire receives over $10,000 on average.

[16]On July 20, 2021, the league adjusted the size of the strike zone following inputs from players and umpires. The adjusted strike zone is only applied to the games utilizing robot assistance and not to other games. During those games, the strike zone is wider and shorter than the original strike zone. See Figure A2.

Leagues) implemented robot assistance in games held between Tuesday and Thursday of the week.[17,18]

3 Data

This study uses the universe of Major League and Minor League pitch-level data from the 2017 season to the 2023 season. The data are web-scraped from the MLB Stats API which provides a wide range of statistics and data related to all games and individuals involved in the league.[19]

First, I collected data from about 73,000 games, including the dates, weather conditions, time durations, attendances, teams and leagues, venues, and importantly, the umpires overseeing the games. I also collated information on personnel associated with each game that includes the names, ages, experiences, height, and handedness of the umpires and individual players.

I then gathered play-by-play data for over 22.5 million pitches that include detailed pitch characteristics.[20] Specifically, X and Y pitch coordinates of every pitch as it crosses the plate and the top and bottom of the strike zone are available.[21] Further, detailed descriptions of a pitch outcome (e.g., called strike, called ball, pitch resulted in hit or out, etc.) are collected. The main outcome is whether an umpire correctly called a pitch or not. To determine this, I utilize the exact pitch location, the dimension of the strike zone, and the umpire's call decision to generate the decision accuracy measure.[22] Various game-situational characteristics like ball counts and out counts, and whether there are runners on base at the moments of these counts, are also collected.

To generate the estimation sample, I drop non-regular season games that include spring

[17]Games held between Tuesday and Thursday of the week used robot assistance, while games between Friday and Sunday were called by umpires without assistance. The system was adopted to compare two formats and to prepare Triple-A umpires for when they appear in Major League games as replacement umpires.

[18]Robot implementation is described in more detail in Section A3. Also, see Figure A3 for chronology of these events.

[19]The data are provided by the Statcast system.

[20]In a typical 9-inning game, about 300 total pitches are thrown, of which about 150 are called.

[21]In stadiums with advanced cameras equipped, the Statcast system tracks the pitches with an accuracy of better than one inch. See Section A4 for details on how pitches are tracked in other stadiums.

[22]The X-and Y-coordinates utilized to generate the decision accuracy measure come from the MLB Stats API and are different from the coordinates plotted by the robot. These are also post-processed to correct for potential errors.

training, All-Star Games, other exhibition games, and postseason games. I also drop games with missing personnel information, games in which I cannot identify whether a robot is used, and games with too few pitches.[23] Finally, I also drop outlier pitches that are inaccurately recorded due to calibration issues. The final dataset has over 62,000 games and 18.5 million total pitches, of which 8.8 million are called with about 680 umpires and 7,500 pitchers and batters.

3.1 Summary Statistics and Preliminary Evidence

Table 1 summarizes the pitch characteristics of the sample, separately by whether a game used robot assistance or not. Out of 62,678 games in the sample, 2,611 games are called with robot assistance. In these games, umpires are harsher relative to when the assistance is not used and call fewer strikes (31.2% vs. 33.6%). These differences in the calls may be due to the differential responses of pitchers: when pitchers know the robot is assisting to make a call, they tend to pitch further away from the strike zone (0.78 vs. 0.75 feet horizontally and 0.87 vs. 0.85 feet vertically). In other words, pitchers are more confident about receiving a strike call at the edge of the strike zone with robot assistance. On average, umpires with assistance are no better than without in accuracy (93.0% vs. 93.0%), but after accounting for the pitch location and the stadium, they perform better (0.015 vs. -0.0007).[24]

Figure 1 plots the average accuracy of umpires with and without assistance as a function of the pitch's distance from the nearest border of the strike zone. Both with and without robot assistance, umpires perform extremely well when the pitches are obviously inside or obviously outside the strike zone. However, accuracy declines as the pitch gets closer to the border. Umpires with assistance are more accurate on these pitches; while not perfect, for outside pitches that are 0.2 inches away from the nearest border, they have an average accuracy of 77.4% relative to accuracy of 64.5% without robot assistance. Figure 2 reveals distinctive discrepancies between umpires with and without assistance. The "enforced" strike zone without a robot is oval-shaped despite the actual strike zone being

[23] In 2022, the Major League announced that the "select" games in Single-A Florida State League will use the "challenge" system detailed in Section A3. However, these games are not specified in the data.

[24] Robot assistance is not perfect for several reasons. Any technology has a margin of error. The system relies on a radar system that tracks the pitches, and it sometimes performs suboptimally in a crowded environment. Errors could also be due to calibration issues in these environments in locating pitches, and measuring the size of the strike zone depends on the batter's height and stance.These errors likely happen at the stadium-level, so I employ a team-by-year fixed effect to address this issue.

rectangular. Umpires without assistance incorrectly call pitches that fall within the corners. Robot assistance, on the other hand, helps the umpires match the rule-defined strike zone well.[25]

This difference in the "enforced" strike zone leads the players to strategically adjust their behaviors. In games assisted by robots, pitchers pitch further away from the center of the strike zone (1.07 vs. 0.99 feet) on average, and pitches are also more likely to fall on the "edge" of the strike zone (61.8% vs. 59.5%) (Table A1). The adjustment also increased "base on balls" in the games in these games, as shown by pitcher's walks allowed (2.20 vs. 1.94 per game) and batter walks (0.54 vs. 0.42 per game).[26]

The estimation focuses on umpires. Figure A4 plots the raw accuracy for umpires who are assisted by robots and those who are not in the year with robot assistance and years prior and after by pitch location. Panel A shows that the Major League umpires improve their accuracy slightly over the years. In contrast, Panel B displays a large increase in accuracy across pitch location when robot assists calling the game, and a drop in the year following relative to pre-robot year. This provides suggestive evidence that umpires' skills deteriorated following the implementation of the robot below preassistance levels.

4 Empirical Strategy

Main Model I exploit the staggered adoption of robot assistance over time and umpires moving across leagues in a difference-in-differences framework.[27] I observe umpires *before and after* the implementation of robots and compare them to those who were not assisted by the robot. Figure A5 describes the potential umpire moves across leagues and robot implementation.[28] A total of 17 umpires called games in the Single-A Florida State League in 2021, 11 with more than 10 games with a robot, and 66 umpires called games in the

[25]The robot-assisted umpire fails the "Turing test," which is a test to assess a machine's ability to exhibit human-like intelligence or behavior.

[26]A batter's at-bat ends in a walk (or "base on balls") when the count reaches 4 balls. The batter is awarded first base by the umpire.

[27]Table A2 shows the summary statistics separately for the Minor League umpires who are never-assisted and assisted by the robots. I also show the outcomes for those who are assisted and subsequently move to leagues without . In the year prior to robot implementation, these umpires are not observably different, on average (residualized accuracy, 0.0 vs. 0.0 and 0.0001).

[28]For example, Dane Poncsak called games in the Double-A Northeast in 2021 prior to the implementation. He moved to the Triple-A Pacific Coast League in 2022, a league that implemented, and called games with the robot. In 2023, I observe him in Triple-A in games without a robot.

Triple-A Pacific Coast League in 2022, 41 with more than 10 games with a robot. I compare these umpires to those who have never worked in these two leagues.

My primary specification estimates the effects for periods with a robot and without a robot after the implementation:

$$Y_{it} = \delta_0 D_{it}^{Robot} + \delta_1 D_{it}^{PostRobot} + \beta X_{i,p} + \pi_{j,y(t)} + \theta_t + \gamma_i + \epsilon_{it} \tag{1}$$

The main outcome Y_{it} is the decision accuracy of the umpire measured as whether a pitch is correctly called by umpire i in time t.[29] The indicator variables D_{it}^{Robot} specifies that a robot is helping to call the game and $D_{it}^{PostRobot}$ corresponds to the cases where umpires are calling the games without assistance and after having been assisted by a robot. The coefficient δ_0, therefore, measures the relative performance with a robot compared to without a robot. The coefficient δ_1 is the main object of interest and it captures the effect of robot assistance on individual skill after having worked with a robot. I include the year-by-month fixed effects, θ_t, and umpire fixed effects, γ_i, which account for variations in the outcome over time and over umpires. Standard errors are clustered at the umpire level.

The identification assumption of the empirical design is that the the adoption was not anticipated and that the decision to move or to stay in a given league was uncorrelated with the adoption of the robot. In particular, to identify δ_0, the umpires who are assisted by a robot are not on a differential trend compared to the umpires who are not assisted by a robot, conditional on controls and fixed effects. To identify δ_1, conditional on controls, fixed effects, and the indicator D_{it}^{Robot}, the performance of umpires who are no longer assisted by a robot, after having been assisted, is parallel to that of umpires who are not assisted. In other words, the decision to move or stay for the assisted umpires does not depend on robot assistance. To test for these assumptions, I estimate an event-study model to test for pretrends:

$$Y_{it} = \sum_{\substack{k \in \{-\underline{T}, \overline{T}\} \\ k \neq -1}} \delta_k D_{it}^k + \beta X_{i,p} + \pi_{j,y(t)} + \theta_t + \gamma_i + \epsilon_{it} \tag{2}$$

where the indicator variables D_{it}^k specify that the umpire is k months away from robot implementation. I check that the coefficients, δ_k, prior to the implementation are not

[29] See Footnote 22 for how the decision accuracy is measured. I use the post-processed data to determine the accuracy.

statistically significant.

Addressing Threats to Identification A potential threat to identification is that the composition of pitches changed at the time of the implementation of robot assistance. For instance, players, especially the pitchers, might strategically respond to the implementation of the system changing the nature of the game. These changes in strategies can persist further into the future when players move across leagues. Pitchers could aim at different areas of the strike zone and batters can also be more or less aggressive in the game. These further affect situational characteristics of the game: for example, if the batters are more likely to swing, then the scores of the games can increase. Changes in accuracy of umpires as a result of robot implementation could then be partially due to the changes in the compositions of pitches and game situations.

To partially address these concerns, I include a rich set of control variables and fixed effects ($X_{i,p}$) to flexibly account for pitch location and potential umpire and situational biases. Most importantly, I control for the exact pitch location, which is the only decision guideline that governs whether a pitch should be called a strike or not. Since umpires may want to avoid calling strikes in certain situations (Moskowitz and Wertheim, 2011), I also control for pitch counts and other situational characteristics, including game score, outs, and runners on base.[30] Additionally, I control for the top-of-the-inning indicator to account for potential home-team bias. I employ pitcher and batter fixed effects to consider any player-specific effects and league level-by-year fixed effects to account for differences in league quality.[31] Further, I conduct a robustness check, reweighing the data to match the distribution of pitches across the leagues and years to address the concern of potential endogenous player responses.

Another potential threat to identification is that the umpires who move across leagues can be systematically different from umpires who do not.[32] For example, if the best umpires

[30]For example, in the most extreme counts (3-0 and 0-2), umpires are known to adjust the size of the strike zone. Umpires are more likely to call a pitch a strike on a 3-0 count, because the next called ball ends an at-bat. In a lopsided game where the score difference is large, umpires become more lenient in calling pitches to increase the pace of the game.

[31]In principle, the implementation of robot assistance is random, so these are not needed except to increase precision. I test the validity of this assumption by estimating models with and without pitch-level controls and I find similar results (See Section 5.1.1).

[32]Section 5.1.1 shows that the umpires who moved to different leagues or who are promoted to upper-level leagues perform neither better nor worse, on average.

in the league are promoted, then comparing the umpires who move following the robot implementation to those who do not can partially capture the differences in these umpires that are not related to the robot and result in an upward bias. On the other hand, if the poor-performing umpires are moving across leagues, then the estimation will result in a downward bias. To alleviate this concern, I also estimate models with varying control groups such as promoted umpires, only the Major League umpires, and only the Single-A umpires.

A third concern is that there may be measurement error in how the pitch coordinates are recorded. As detailed in Section A4, the data include the X and Y coordinates for every thrown pitch, and there are two different coordinate systems: manually-plotted pitch coordinates and camera-tracked pitch coordinates. In the Major League and the Minor Leagues that adopted the robot, pitch coordinates are available from these cameras. However, in other stadiums without such technology, pitch coordinates are manually-plotted by stringers hired by the league. If the stringers plot the pitches to match the umpire's calls (i.e., plot an ambiguous pitch inside if the umpire calls strike and vice versa), then when the umpire moves to the league with advanced technology, his accuracy might drop as a result of better measurement.

To address this concern, first, I use the dosage model described in the next section that only utilizes games with "Hawk-Eye" technology plotting the pitch coordinates. I also conduct a robustness check using just the manually-plotted pitch coordinates which are available in all stadiums. Second, I employ home team(stadium)-by-year fixed effects, $\pi_{j,y(t)}$ to address how the pitch coordinates are recorded manually. I transform the manually-plotted coordinates to have the same units as the camera-tracked coordinates at the home team-by-year level (Section A4.3). While I cannot employ stringer-FEs as the data do not contain this information, home team-by-year-FEs partially address the concerns associated with the data quality of pitch coordinates. For example, unique bias in recording coordinates that could depend on the stringers' vantage point can be relieved.

Dosage Model With the 2023 implementation of robots in Triple-A leagues which assigned robots to games held on Tuesdays, Wednesdays, and Thursdays, and the random assignment of umpires to games, the treatment dosage, or the share of games with the robot assisting calling the game is as-good-as random. I therefore also consider the alternative dosage model using just the data from the 2023 Triple-A leagues:

$$Y_{it} = \beta_0 D_{it}^{Robot} + \beta_1 RobotExposure_{it} + \beta X_{i,p} + \pi_{j,y(t)} + \theta_t + \gamma_i + \epsilon_{it} \quad (3)$$

where D_{it}^{Robot} is an indicator variable that the game used the robot and $RobotExposure_{it}$ is the share of games with robot assistance.[33,34] The coefficient β_0 estimates how well the umpire performs with robot assistance relative to without. The coefficient β_1 measures the effect of treatment dosage on individual skill. A set of control variables and fixed effects is identical to those employed in Equation 1 except that θ_t is now at the week-level.

5 Results

5.1 Do Humans Lose Skills Following Robot Adoption?

Table 2 displays the main results. First, umpires using robot assistance make calls remarkably well. In months when a robot helps call games, pitches are 6.4 percentage points (about 6.9%) more likely to be correctly called. This finding confirms the general belief that robots perform higher quality work than humans in repetitive and monotonous tasks. Robots can help humans achieve a high level of precision and consistency.

Following the implementation of robot assistance, umpires can move to another league where they are asked to perform the task of calling pitches again without assistance. In this post-robot period but *without* robots, umpires are 2.0 percentage points (2.1%) less accurate relative to preassistance levels. In other words, the umpires' skill deteriorates following robot implementation below the preassistance levels. The magnitude of the decrease roughly matches the difference between umpires at the median and those in the bottom 5th percentile. The event-study version in Figure 3 also suggests the same pattern. Panel A shows that the umpires using robot assistance execute tasks with high precision, but when the robot is removed, they experience a significant skill decline. The figure also confirms that there is no pretrend. The finding suggests that overreliance on robots for tasks that were once performed solely by humans can lead to a loss of skills.

[33]The umpire crew rotates positions throughout the series of games, resulting in a random assignment of umpires to robot assistance. Throughout the season, umpires will work statistically identical number of games with and without robot assistance.

[34]$RobotExposure_{it} = \frac{\text{\# of games with robot}}{\text{\# of total games}}$. For example, $RobotExposure_{it}$ can only take the value 0 or 1 for the umpire's second game and 0, 0.5 or 1 for the umpire's third game, etc.

The skills gap between umpires with and without assistance is most pronounced for pitches that fall near the border of the strike zone (See Figure 1). Table A3 restricts the sample to the pitches that are within 0.5 feet from the nearest border of the strike zone from the outside, and within 0.2 feet from the nearest border of the strike zone from the inside. The results are more striking, with a decline in accuracy of 3.7 percentage points (4.1%) following work with robot assistance for these pitches. After receiving assistance from robots, umpires also becomes less consistent. Table A4 shows that the standard deviation of accuracy increases in the post-robot period.

With a unique setting that allows me to observe decision accuracy, I can also classify decisions into four different categories: true positive (correctly called strike), true negative (correctly called ball), false positive (Type I error, incorrectly called strike) and false negative (Type II error, incorrectly called ball). As robots improve precision, umpires assisted by a robot have higher true positive and true negative decisions (2.2 and 4.1 percentage points). On the other hand, following the work with robot assistance but without robots, umpires have lower rates of true decisions without robots (1.0 and 1.0 percentage points). Umpires, however, have many more incorrectly called strikes (false positive, 2.5 percentage points) than incorrectly called balls (false negative, -0.5 percentage points) in these months.[35] These collectively result in higher rates of pitches being called strikes; umpires are therefore more lenient to pitchers. Since Figure A6 shows that umpires tend to call more pitches as strikes for pitches that are harder to judge, I take this as suggestive evidence that, when umpires feel more uncertain about their decisions, they lean towards calling strikes.

Humans suffer from decision biases. Table A5 displays whether incorrect decisions due to biases increase following robot implementation. First, Moskowitz and Wertheim (2011) suggests that umpires suffer from omission bias: they are less likely to call a pitch strike when the count is 0-2 when the next strike would end an at-bat, and more likely to call a pitch strike when the count is 3-0 when the next ball would end an at-bat.[36] I find that the share of wrong calls when the count is 3-0 decreases in the months when the robot helps, calling the game more accurately by 7.7 percentage points relative to the base mean of 8.7%.[37] After robots are removed, the rate of omission biases when the count is 3-0 increases

[35]See Figure A7 for event-study versions of these outcomes.
[36]The size of the "enforced" strike zone changes based on game situations (Figure A8).
[37]In a 3-0 count, an incorrectly called strike due to bias is considered an inaccurate call.

by 2.9 percentage points (33.3%) relative to preassistance levels. Second, Chen et al. (2016) show that umpires suffer from the gambler's fallacy: they underestimate the probability of consecutive streaks happening by random chance. In particular, they are less likely to call a pitch a strike if the previous pitch was called a strike. The robot helps to reduce the share of incorrectly called balls following a previous pitch that was called a strike by 1.2 percentage points. Following the work with robot assistance, but without the robot, umpires have a marginally lower share of wrong calls when the count is 0-2 and when the previous pitch was called a strike. These results, however, should be interpreted with caution, as I also find that umpires lean towards calling strikes on more ambiguous pitches; these results can therefore be an artifact of umpires calling more strikes overall.

5.1.1 Robustness Checks

In this section, I conduct a number of robustness checks to verify the results.

Control Groups My estimation relies on umpires who move across leagues, particularly for the estimation of the skill decline that occurs after robot assistance is no longer available. These umpires might be systematically different in their ability if the leagues promote umpires who are better at the task. However, Table A6 presents that the umpires who moved across leagues or who are promoted to upper-tier leagues have neither higher nor lower accuracy relative to all other umpires who stayed, on average (residualized accuracy, -0.0 and 0.0 vs. -0.0002). In addition, the identification strategy depends on the presence of parallel trends between the groups, making the differences in levels acceptable.

Table A7 shows the results with varying control groups and replicates the baseline results. The results are robust to comparing treated umpires to the untreated umpires who are promoted (a decline of 2.7 percentage points) and to untreated umpires who are in the bottom and top leagues (declines of 3.1 and 2.7 percentage points, respectively).

Is it the Skills or the Effort? Umpires may exert less effort if they perceive their skills are no longer relevant due to the introduction of robots. Consequently, the decreased accuracy may be attributed to umpires putting in less effort during a game rather than an actual decline in their skills.

In Table A8, I focus on high-leverage game situations where umpires are likely to exert

a high level of effort. First, in these high-leverage circumstances (e.g., small score difference, later in the game and/or having runners on base), umpires exhibit higher baseline accuracy than in the full sample, suggesting that umpires may indeed exert high effort. Even in these context, umpires still perform with higher precision when assisted by a robot but suffer a decline in accuracy after having been assisted by a robot. Compared to the baseline results, the magnitude of increased accuracy when assisted by a robot is lower (0.064 vs. 0.038 in the most restrictive situation after the 7th inning, when the score difference is less than or equal to 1 and there are runners on base). Importantly, the decline in accuracy is also smaller in the post-robot period. In the most restrictive situation, the effect size indicates a decline of 1.5 percentage points compared to the baseline result of a 2.0 percentage point decline following the use of robot assistance, suggesting that the findings can be largely attributed to an actual decline in skill rather than lower effort.

What if the Players Respond Strategically? If a robot increases accuracy and consistency, players may need to adjust their strategies to optimize their productivity. In particular, pitchers can adjust their behavior to focus on hitting specific regions of the strike zone where the umpires assisted by a robot are more likely to call strikes.[38] Knowing that the umpire with robot assistance is more precise in its strike zone, pitchers might aim to hit those areas consistently to increase their chances of getting called strikes.

Table A9 estimates the same specification, but for players who are exposed to a robot and move to another league without a robot. Pitchers respond strategically to the use of robot assistance. In the months with robots assisting calls of the game, pitchers pitch closer to the center of the strike zone (0.016 feet), and also closer to the border of the strike zone (0.017 feet), on average. As Figure 2 showed, the umpire calls pitches that fall in the corner strikes more often with a robot, making players strategically respond by pitching closer to the border. Figure A10 shows the event-study versions of these outcomes at the game level. Panel B shows that the pitchers gradually learn and respond to the robot-assisted strike zone by pitching more frequently inside of the strike zone.[39]

The effects persist after these players move to another league without robot assistance. Pitchers maintain the adjusted pitching behavior by pitching closer to the center of the strike

[38] Pitchers are more than capable of adjusting strategies. See Figure A9 for how pitchers pitch at different levels.

[39] See Section A5 for a discussion on how strategic adjustment affects players' productivity.

zone and to the border (0.008 and 0.003 feet), but Panel B of Figure A11 show that pitchers readjust by the end of the season and revert back to the original behavior.

The endogenous player response might raise a concern that the baseline results are an artifact of the distribution of pitches changing and persisting. To address this issue, Table A10 presents the results reweighing the data to have the same distribution of pitches as the prerobot period. The results are largely similar and even more pronounced, with effect sizes of 8.0 percentage points in increased accuracy of robot assistance, and a 2.4 percentage-point decline below preassistance levels following the work with robot assistance.

The main specification includes a rich set of pitch-level controls to account for different game situational characteristics and player behaviors. These are not needed in essence as the implementation of robot assistance is as-good-as random, but used to increase precision. I confirm and find similar results with and without pitch-level controls in Table A11.

Data Quality If the stringers hired by the league align their pitch plotting with the umpire's calls (i.e., they record a pitch as inside the strike zone when the umpire calls a strike for ambiguous pitches), the umpires' accuracy could potentially decline once they transition to a league with robot assistance.

Table A12 shows the results using only the manually-plotted pitch coordinates. The results are much like the baseline results, with slightly attenuated 3.3 percentage-point increase of accuracy with the robot relative to without the robot, and a decline of skill of 1.5 percentage points following the removal of the robot.

Two-way Fixed Effects Finally, I address the concerns raised by recent literature regarding difference-in-differences models (e.g. Goodman-Bacon, 2021). In the presence of heterogeneous treatment effects, a two-way fixed effects model can provide biased results. Figure A12 presents four different estimators addressing this issue and finds a generally similar pattern of results (Borusyak et al., 2021; Callaway and Sant'Anna, 2021; De Chaisemartin and d'Haultfoeuille, 2020; Sun and Abraham, 2021).[40]

[40]In these models, I estimate at the month-level instead of the pitch-level.

5.2 Heterogeneity

Do High-Skilled Umpires Lose More Skill? A particularly important question is who loses more with robot implementation.[41] Table A14 shows the results separately by accuracy level of umpires in the year prior to the implementation. First, in the months when a robot helps calling the game, accuracy increases by 10.5 percentage points for the bottom-quartile umpires and 8.4 percentage points for the top-quartile umpires. In months without robot assistance, the umpires with varying prior skills experience skill declines of varying degrees relative to the preassistance levels. The umpires who had the least skill lose the least, showing a statistically insignificant decline of 0.9 percentage points, while the better umpires lose more, with the middle-half and top-quartile umpires suffering drops of 4.6 and 4.7 percentage points below the preassistance levels, respectively.

Brynjolfsson et al. (2023), Kanazawa et al. (2022), Tafti (2022) and Noy and Zhang (2023) also find that the productivity distribution compresses among taxi drivers, surgeons, and workers after AI, driver assistance, surgical robots and ChatGPT are used, respectively. In all four settings, advanced technology benefited the low-skilled workers. In the setting I examine, I find that the skill decline is the smallest for low-skilled umpires and larger for better-skilled umpires after robot removal. Panel A of Figure A13 shows that the skills gap among umpires of different skill levels compresses. Prior to robot implementation, the gap between the best and worst-skilled umpires is 6.0 percentage points. When they return to making calls without robot assistance following work with it, the gap shrinks to 2.1 percentage points or by about 64.6%. Collectively, high-skilled workers not only benefit the least *with* robots, but also lose the most when the robot is removed.

Do Less Experienced Umpires Lose More Skill? Table A15 shows that umpires get neither better nor worse with additional years of experience, on average. Returns to experience are positive and significant for relatively young and new umpires, but they eventually become negative: umpires start losing skills with additional years of experience as

[41] For all analyses in Section 5.2, I focus on the umpires who have worked in Triple-A or the Major League. Personnel information is more complete and better recorded for this group of umpires. Further, implementation of robot assistance in the Single-A Florida State League happened in 2021, so the preperiod statistics comes from 2019 instead of 2020 due to the COVID-19 pandemic. Table A13 shows the main results for this sample. For this sample, pitches are 8.9 percentage points more likely to be correctly called with robot assistance and umpires face a 3.6 percentage-point decline in skills relative to preassistance levels.

they age.[42] Therefore, more experienced umpires are often the less-skilled umpires.

Table A16 displays the results separately by umpires' years of experience. A robot increases accuracy for umpires of all experience levels in the months when the robot is used. Following the work with robot assistance, umpires with the most experience (i.e., more than 9 years of experience in professional leagues) lose the least, showing a 1.8 percentage points decline relative to the preassistance levels in accuracy. Umpires who are relatively inexperienced (i.e., less than 6 years) experience a significant decline of skill of 5.3 percentage points.[43,44]

Do Umpires with Prior Major League Experience Lose More Skill? Among Triple-A umpires, those with Major League experience as replacement umpires are more likely to be next in line for promotion.[45] These umpires are also less likely to be replaced sooner than those without Major League experience as the Major League has yet to plan to adopt robot assistance system completely.

Table A17 shows the results separately by whether the umpire had prior experience in the Major League or not. In the months with robot assistance, decision accuracy increases for both types of umpires (8.2 and 9.3 percentage points, respectively for those without and with Major League experience). However, the magnitude of skill decline relative to the preassistance levels is much larger for those without Major League experience (4.9 percentage points) than those with Major League experience (1.5 percentage points). These findings suggest that those with Major League experience, and potentially who are more likely to be promoted in the near future, try to retain skills more.

5.3 Does Longer Work With Robot Increase Skill Depreciation?

The duration of nonemployment spells affects future prospects of wages and reemployment opportunities (e.g. Kroft et al., 2013). While potential explanations include skill depreciation while unemployed (e.g. Benhenda, 2022; Dinerstein et al., 2022; Edin and Gustavsson,

[42]Williams (2019) finds that the best-performing umpires, on average, have fewer years of experience. Figure A14 also shows the same patterns.

[43]A large share of umpires who are assisted by robots come from Triple-A minor leagues. For umpires to reach this level, they need several years of training through Single-A and Double-A leagues.

[44]A potential explanation for this result is that the task requires physical fitness (e.g., dynamic visual acuity) to make a decision immediately after observing a fast pitch.

[45]Ahead of the 2023 season, the Major League promoted 10 umpires from the minor leagues. All of these umpires had worked in the Major League as replacement umpires.

2008; Jarosch, 2023), the length of the nonemployment spell is often correlated with skill, making it difficult to study the impact of duration on skill depreciation. The 2023 Triple-A implementation of robot assistance, paired with the umpire rotation system, makes a good setting to study this issue as the length of duration with robots is as-good-as random.

Table A18 reports the pitch characteristics for Triple-A games in 2023. A similar pattern emerges in these games as the full sample: the robot increases accuracy, on average (94.0% vs. 90.1%). A total of 71 umpires called at least one game in Triple-A leagues, with 64 umpires working with the robot for at least a game. The average umpire's share of games with the robot is 42.0%. Figure A15 shows the distribution of the dosage. Panel A presents that the share is approximately normal over the season. After the first game of the season, the umpire's treatment dosage is either 0 or 1, and as the season progresses, the dosage becomes more concentrated around 50% (Panel B, Figure A15). The dosage is as-good-as random, and more or less experienced umpires are neither more nor less likely to have more games with the robot (Figure A16).

In 2023, umpires working in Triple-A minor leagues had between 1 and 3 consecutive games with robot assistance before having to call a game solely by themselves.[46] To compare with the baseline results of full-season work with robot assistance, I consider the following model:

$$\begin{aligned} Y_{it} = \eta_0 D_{it}^{Robot} + \eta_1 RobotExposure_{it} &+ \eta_2 \mathbb{1}(\text{Previous Game Used Robot})_{it} \\ &+ \eta_3 \mathbb{1}(\text{Previous 2 Games Used Robot})_{it} \\ &+ \eta_4 \mathbb{1}(\text{Previous 3 Games Used Robot})_{it} \\ &+ \beta X_{i,p} + \pi_{j,y(t)} + \theta_t + \gamma_i + \epsilon_{it} \end{aligned} \tag{4}$$

where the omitted group is that the previous game did not use the robot.[47] I employ the same set of control variables and the fixed effects as previously explained.

Table 3 shows the results separately by the length of work with robot assistance relative

[46] On average, the number of days since the last time umpires had to call a game without robot assistance is 8.5, 15.9 and 23.7 days for 1, 2 and 3 consecutive games of robot assistance, respectively.

[47] In particular, $\mathbb{1}(\text{Previous Game Used Robot})_{it}$ is an indicator variable that specifies that the robot was used in the previous game in a previous three-game span. It therefore includes cases where the robot was used in $t-3$ and $t-1$, but not cases where the robot was used in $t-2$ and $t-1$, as this is denoted by $\mathbb{1}(\text{Previous 2 Games Used Robot})_{it}$.

to umpires who are not assisted. When an umpire returns to the task after a game with robot assistance, their skill declines by a 0.3 percentage points. The magnitude of the decline of skill increases when an umpire is assisted in 2 or 3 consecutive games by a robot to 1.1 percentage points and a statistically insignificant 1.0 percentage point, respectively.

An average umpire works as a home-plate umpire for 28 games in a season. Unfortunately, the umpires working in Triple-A minor leagues in 2023 had between 1 and 3 consecutive games of work with the robot, so I cannot fully estimate the effect of longer durations of robot assistance. The baseline results suggest that umpires with a full season of work with the robot suffered a 2.0 percentage-point decline in skills relative to the preassistance levels, on average, in the following season; the findings therefore suggest that increases in work duration with robots increase skill depreciation.

Further, Table 4 presents the results using an alternative specification (Equation 3) that exploits the variation in treatment intensity to robots. I also plot the second- and third-degree polynomial estimates of dosage effects (Figure 4). A game with robot assistance corresponds to a dosage of about 3.5% in a full season. At low to medium dosages, skill does not seem to deteriorate. However, at a dosage of 100%, the decline in skill is 1.1 percentage points, which is statistically significant at the 10% level, confirming the findings that umpires' skills worsen following the implementation of robot assistance. The overall pattern of the decline suggests that the effects take place after a certain intensity of exposure.

5.4 How Long Does It Take To Regain Skill?

Related questions include whether umpires can regain their skills after loss of robot assistance and, if so, how long it would take to return to their original skill levels. Learning by doing is an important source of productivity growth and returns to experience for workers (e.g. Arrow, 1962; Lucas Jr, 1988; Thompson, 2010; Yang and Borland, 1991). Whether workers return to a prior level of skill with recall after use of robot assistance, and who therefore "unlearned" by doing less, is relatively unknown.[48]

I investigate this question in examining the umpires who are assisted by robots in 2022 and subsequently moved to another league without robots. In particular, I employ the following event-study specification at the game level:

[48]Surprisingly, over 40% of workers who are unemployed return to the original employer (Fujita and Moscarini, 2017).

$$Y_{it} = \sum_{\substack{g \in \{-\underline{G}, \overline{G}\} \\ g \neq -1}} \xi_l D_{it}^g + \beta X_{i,p} + \pi_{j,y(t)} + \theta_t + \gamma_i + \epsilon_{it} \tag{5}$$

where D_{it}^g indicates that the umpire is g games away from the last game with the robot.

Figure 5 shows that these umpires do not recover within a season from the skill decline following the robot assistance. In the first game back after use of robot assistance, the umpires suffer a decline of skill of 3.6 percentage points. While the decline does become smaller in magnitude over the season, it is still statistically significant. By the 16th game, or about halfway through the season, skill deteriorates by 1.7 percentage points. Closer to the end of the season (after 21 games since the return to calling without robot assistance), the skill decline is 1.0 percentage points.

Table 5 presents the same results that the decline in skill is larger early in the season and gradually decreases as the season continues. In the first five games returning to the task without robot assistance, umpires face a skill decline of 2.9 percentage points relative to preassistance levels. However, in the 16-20 games back and 21 or more games since the return to calling without robot assistance, the decline is 1.7 percentage points and 1.0 percentage points, respectively. These findings imply that umpires require a considerable amount of time to regain their skills following the work with robot assistance. While they do make significant progress regaining their skill levels over the course of the season, they do not fully return to their initial skill levels by the end of the following season.

6 Mechanisms

6.1 Are Robot Effects the Same as the Time-Away Effects?

Is skill depreciation a result of time away from the task? Past literature on whether skill depreciates during unemployment spells remains inconclusive as Cohen et al. (2023) find that cognitive and non-cognitive skills do not depreciate for German workers, but Dinerstein et al. (2022) show that Greek teachers become less skilled while unemployed.

In 2020, the COVID-19 pandemic delayed and shortened the 2020 Major League Baseball season, but canceled the entire Minor League Baseball season. Therefore, the Minor League umpires experienced a season-long "unemployment" in which they did not engage in

the task, unlike Major League umpires.

I compare the Minor League and Major League umpires in a difference-in-differences framework following the COVID-19-canceled 2020 season.[49,50] In particular, I employ the following specification to study whether the COVID-19 pause resulted in skill depreciation:

$$Y_{it} = \rho_0 D_{it}^{FirstSeason} + \rho_1 D_{it}^{SeasonsAfter} + \beta X_{i,p} + \pi_{j,y(t)} + \theta_t + \gamma_i + \epsilon_{it} \tag{6}$$

where the $D_{it}^{FirstSeason}$ and $D_{it}^{SeasonsAfter}$ indicate the 2021 season when the umpires returned from the pause and all following seasons, respectively.[51]

Table 6 reports that in 2021, the first season back from the canceled season, the Minor League umpires do not suffer from any skill depreciation, showing a statistically insignificant decline of 0.4 percentage points, on average.

However, the event-study version reveals that the umpires did marginally lose skill following the pause. Figure 6 shows that in the first month since the leagues restarted, the umpires faced a skill decline of 0.8 percentage points, statistically significant at the 10% level. They quickly bounced back, however, with a statistically insignificant decline of 0.5 percentage points starting from the second month of return.

Collectively, these results suggest that the skill depreciation stemming from robot implementation is larger than the time-away effects. A first potential explanation is a reduced incentive to practice skills: umpires may perceive that their skills are no longer relevant, leading to a faster skill depreciation compared to individuals who are temporarily off work. Another potential reason is loss of confidence: when umpires are continually assisted by the robot, they may lose confidence in their abilities to perform tasks manually. This lack of confidence could further erode their skills because they become less willing to engage in tasks that require their expertise.

[49]Table A19 compares the Minor League umpires who did and did not return from the COVID-19-canceled season. Out of 167 Minor League umpires who worked in 2019 and were never robot assisted, 84 returned in 2021 while 83 did not. In 2019, the year prior to the pause, these umpires are not observably different, on average.

[50]This analysis limits the sample to the never-assisted group.

[51]While the umpires can "practice" before the start of the season in the preseason (i.e., spring training), the same is true for the umpires assisted by a robot for the entire season.

6.2 Complementarity or Substitutability of Skills

Home-plate umpires experience a noticeable drop in their ability to accurately call pitches once they have worked with robot assistance. This raises the question of whether the introduction of robot assistance impacts other skills as well. If two skills are substitutable, a decline in skill in one area can lead to an increase in competency for the other skill. Conversely, when two skills are complementary, a decrease in proficiency in one task can also lead to a decline in skills for the other task.

The first-base umpire in baseball is responsible for several tasks during a game and his primary role is to make calls related to plays that occur at first base. The most crucial task of the first-base umpire is to make decisions on whether a baserunner is safe or out when attempting to reach first base. They must closely observe the timing of the runner's arrival at first base and determine whether the ball reached the base before the runner.

A team can request a review of a play that occurred at first base to determine if the call made by the umpire was correct. The replay officials review the available camera angles and video footage of the play and make a determination on whether the call on the field was correct or if it should be overturned. As replay reviews provide a mechanism to correct any errors made by the first base umpire during the game, the numbers of the requested reviews and overturned calls are good proxies for the first-base umpire's skills.

I compare the outcomes of umpires who are assisted by the robots in 2022 to those of who are not.[52] The outcomes are only available for the Major League games, so I restrict the set to all Major League games held in 2023.[53] In particular, I employ the following specification:

$$Y_{it} = \phi_0 D_{it}^{Treated2022} + \beta X_i + \theta_t + \epsilon_{it} \tag{7}$$

where the $D_{it}^{Treated2022}$ indicate that the umpire i is treated in 2022. I include team fixed effects to account for variations in the outcome across teams and control for umpire's years of experience. θ_t are the month fixed effects.[54]

[52]The umpire crew rotates positions throughout the series of games. For example, an umpire will work behind the home plate in one game and will work behind first base in the next.

[53]Replay reviews are only available in stadiums with cameras installed. Many minor league stadiums still lack this technology.

[54]The identification assumption is that umpires who have not previously received robot assistance serve as good counterfactuals for those who have been assisted. The two groups exhibit some observable differences,

Table 7 first establishes that there is no significant difference in the number of ground outs, or calls for which the first-base umpire is responsible, between the two groups. However, the umpires who are previously assisted by the robots face about 0.1 more challenges (i.e., replay reviews) per game (44.3%), indicating that umpires' decisions are under increased scrutiny. Considering that the number of challenges being overturned has also increased by 0.09 per game (68.6%), these findings collectively imply that umpires who have received previous robot assistance exhibit higher rates of inaccuracies and diminished skills in the task.

As robots and automation systems are integrated into various industries and tasks, the demand for specific skills may shift to adapt to the new working environment. In particular, workers may need to adapt and acquire new skills to perform other tasks. The two tasks, calling pitches and determining whether a baserunner is safe or out, are similar and exhibit complementarities, suggesting that the skill shift following robot implementation may occur at a greater distance along the spectrum.

6.3 Are Umpires Learning from the Robot?

When automated systems are used to track pitches and make calls, umpires can receive immediate feedback on the accuracy of their calls compared to the technology. Over time, umpires may adapt their calling style to align more closely with the robot assistance, especially if the technology is proven to be highly accurate.

On July 20, 2021, following inputs from players and umpires, the league widened the strike zone by 2 inches on each side of plate and lowered the top of the strike zone by 3.5 inches (Figure A2). If the umpires who are assisted by the robot adjust their calling style to match it, then their skills may have seemed to drop as they call with the adjusted rule instead of the old rule. Similarly, if the umpires did match the robot well, using the alternative definition of accuracy with the adjusted strike zone will show that the skills would have increased.

Table A20 shows that even with the decision accuracy measure using the adjusted

as the umpires who were assisted by robots in 2022 are, on average, younger and less experienced. However, in previous sections, I demonstrated that these umpires also perform better on average. As part of a placebo test, I found that the number of ejections does not differ between these two groups which show that the assisted umpires do not make more controversial calls. While the results should be interpreted with caution, I argue that mean comparisons between these groups provide insights into the impact of robot assistance.

strike zone, the umpires experience a decline in accuracy following robot implementation (0.9 percentage points). Therefore, these estimates rule out umpires learning and adjusting to a new standard.

7 Discussion

7.1 Does the League Gain from Robot Adoption?

While the workers in firms that have adopted robots typically suffer wage and employment losses, robot-adopting firms enjoy increases in output gains and productivity (e.g. Acemoglu et al., 2022). In this section, I explore whether the adopting leagues benefited from implementing the robot.

I compare the Single-A Florida State League that adopted the robot in 2021, and the Single-A Carolina League that did not, in the following difference-in-differences framework:

$$Y_{it} = \tau_0 D_{it}^{FirstSeason} + \tau_1 D_{it}^{SeasonsAfter} + \beta X_i + \theta_t + \omega_i + \epsilon_{it} \tag{8}$$

where the $D_{it}^{FirstSeason}$ and $D_{it}^{SeasonsAfter}$ indicate the months in the 2022 and 2023 seasons, respectively.[55] θ_t and ω_i are the year-by-month and the league fixed effects, respectively. I also include team fixed effects to account for variations in the outcome across teams.

Table A21 shows the results comparing the outcomes in those two leagues. First, league attendance increased by 265.2 per game (12.4%) in the first year of implementation of the robot. A back-of-the-envelope calculation suggests that this translates to an increase of $1.25 million in just ticket revenue and $2.5 million considering parking, food and beverages.[56] Figure A17 shows the event-study version and confirms the finding. Panel A also suggests that the increase in attendance is concentrated at the beginning and end of the season, which insinuates that increased publicity about the robot might have prompted the increase. In comparison, the cost of "Hawk-Eye" installation is about $300,000 and the robot operator's

[55] The Single-A Florida State League adopted the robot in 2021, and the "First Season" refers to the 2021 season. Other Minor League leagues do not provide good comparisons for several reasons. First, there are large differences in league quality across the league classes. Second, the Minor League has several experimental rule changes each year. For example, in 2021, the Single-A California League adopted onfield timers to reduce game length.

[56] A total of 589 games is played in the Single-A Florida State League in 2021 and an average ticket price in Single-A is about $8.00. The average cost for a family of four to attend a game is about $65.00 in total.

hourly wage is about $25.00.[57] The league can more than bear the cost of the robots in its first season of implementation just from increased revenue generated from ticket sales.

Upon the implementation of robot assistance, fans are more likely to enjoy games with fewer controversial calls and disputes, leading to increased viewership and attendance, which can positively impact revenue through ticket sales, merchandise purchases, and advertising revenue. In addition, the use of robot assistance could appeal to tech-savvy audiences and younger generations of viewers, who may be drawn to the game due to its integration with cutting-edge technology. Over time, the adoption of the robots could also lead to cost savings by reducing the number of onfield umpires needed for games.[58]

Second, the duration of the game increased by 21.1 minutes (12.7%) in the first season of implementation and the total score increased by 1.1 runs (12.9%). While the Major League wants to shorten game length, it also wants to generate more in-the-field-of-play offense.[59] Increased duration is likely an artifact of increased scoring and more activity in the game. Indeed, the number of total pitches thrown in a game also increased by 11.2 (4.1%) in the first season. Viewers generally prefer watching high-scoring to low-scoring games and thus, an increased total score in the first season is also likely to enlarge the fan base.

7.2 External Validity

The impact of robot assistance on umpires' skills in making ball-strike decisions has significant relevance and transferability to a wide range of other occupational settings. Beyond the specific context of baseball, the skill set and cognitive processes involved in these decisions are shared by professionals across various domains.[60] These skills encompass a combination of vision, physical conditioning, and mental sharpness.

Sharp visual perception and mental acuity required for umpires to make accurate ball-strike decisions are widespread attributes that apply across various professions. They are especially important in any occupation that relies on precise observation and decision making. Healthcare professionals, such as surgeons and radiologists, require sharpness of vision to

[57] In 2013, the Premier League, an English football league, installed a "Hawk-Eye" system with 14 cameras for £250,000.

[58] Average umpires in Single-A and in the Major League earn $3,000 and $10,000 per month, respectively.

[59] The Major League experimented with various rule changes in the Minor Leagues, including limiting defensive shifts, limiting pickoff throws and moving the mound back to generate more offense.

[60] For the first-base umpire, the skill set closely resembles that of the home-plate umpire, as he must make decisions based on the timing of the ball and the runner's arrival at first base.

accurately diagnose medical conditions and perform surgeries. In the aviation industry, pilots and air traffic controllers must possess sharp visual perception to navigate aircraft. Similarly, workers in manufacturing and quality control settings use visual perception to inspect products and ensure quality standards are met.

The cognitive demands on an umpire mirror those in professions that require quick and accurate judgments, especially in high-stakes settings. The cognitive processes involved in decision making, such as attention, are therefore fundamental beyond the setting of baseball. For instance, surgeons and air traffic controllers must maintain high level of focus throughout procedures and navigating air traffic as lapses in attention can lead to potentially catastrophic consequences.

The application of sharp visual perception and acute decision making extends across diverse occupational fields, emphasizing the broader implications and relevance of the findings of this paper. Hence, the deterioration of these skills could have significant consequences for accuracy and safety in numerous occupational domains.

8 Conclusion

How does working alongside robots influence human skills, and do individuals perform worse in tasks when robot assistance is removed? Even as robots become more accessible, the value of human skills endures, especially in domains that require adaptability and complex decision making.[61]

Importantly, I examine the impact of working with robots on human capital, focusing on the decision accuracy of umpires in professional baseball settings. By analyzing data from over 62,000 games played between 2017 and 2023 in both Major and Minor Baseball Leagues, I find that umpires perform more effectively with robot assistance, affirming the role of robots in enhancing efficiency and productivity across various contexts. However, umpires suffer a large skill decline below preassistance levels when they are required to perform tasks independently once again, suggesting that relying too heavily on robots can have detrimental effects on skill retention.

Second, I investigate the impact of robot implementation on different groups. The

[61] In 2018, the automobile manufacturing company, Tesla, failed to meet production targets after relying too much on robots, and replaced the automation system with humans. The company's CEO, Elon Musk, stated that "humans are underrated."

umpires with the highest skills see the smallest improvements with robot assistance and experience the most pronounced decline once the assistance is removed. Consequently, the skills gap diminishes with the introduction of robot assistance and persists even after its removal.

Third, I study the impact of the duration of working with the robot on the decline of skill and the time umpires take to recover their skills after the recall following the work with robot assistance. I find that as the duration of robot assistance lengthens, the decline in skill becomes increasingly pronounced. On the contrary, although umpires make substantial progress, they do not completely return to their original skill level by the conclusion of the season.

Fourth, by studying a canceled season during the COVID-19 pandemic, I reject that skill depreciation can be attributed solely to umpires not utilizing their skills, as skill depreciation resulting from robot implementation is more substantial than the effects of time away. Further, after working with robots, umpires also exhibit a decline in their ability to determine whether a baserunner is safe, suggesting applicability to a broad spectrum of other occupational settings that demand similar skill sets.

Finally, I find that robot implementation affects other workers who respond strategically and benefits the implementing firm. With robots increasing the accuracy and consistency of calls, pitchers adjusted their strategies to optimize their chances of getting favorable calls. In addition, game attendance and total scores in games increased following robot implementation suggesting an increase in revenue.

These findings raise important questions for future research. Robots and automation can often perform tasks more quickly, accurately, and consistently than humans. This increased efficiency can lead to higher productivity and output gains for businesses (Acemoglu et al., 2022, 2020; Dixon et al., 2021; Humlum, 2022; Koch et al., 2021). However, one of the most significant concerns with automation is the potential displacement of human workers. Jobs that involve repetitive or rule-based tasks are more susceptible to automation and workers in heavily impacted industries can suffer unemployment or wage decline (Acemoglu et al., 2023, 2020; Acemoglu and Restrepo, 2020; Barth et al., 2020; Bonfiglioli et al., 2020; Dauth et al., 2021; Humlum, 2022).

Understanding the impact of automation on human workers' skills when they are assisted by robots is crucial but understudied. First, identifying the skills affected by au-

tomation allows policymakers to identify potential skill gaps and develop strategies to help workers adapt to the changing job market. Second, analyzing the skills that are becoming less relevant due to automation can inform targeted reskilling programs to help affected workers transition into new roles. Studying the impact of automation on human workers' skills is essential for proactively addressing the challenges and opportunities brought by technological advancements. More research, therefore, is needed to provide evidence on which skills are affected and how to plan for the future labor market. The finding that robots decrease human skills in tasks they assist is important for future policies. As human capital without robot assistance is critical even when robots are readily accessible, effective labor policies need to address skill decline for a smooth but inevitable transition to the future labor market with robots.

References

Aaronson, D., B. Mazumder, and S. Schechter (2010): "What is Behind the Rise in Long-Term Unemployment?" *Economic Perspectives*, 34.

Acemoglu, D., G. W. Anderson, D. N. Beede, C. Buffington, E. E. Childress, E. Dinlersoz, L. S. Foster, N. Goldschlag, J. C. Haltiwanger, Z. Kroff, et al. (2022): "Automation and the Workforce: A Firm-Level View from the 2019 Annual Business Survey," Tech. rep., National Bureau of Economic Research.

Acemoglu, D., H. R. Koster, and C. Ozgen (2023): "Robots and Workers: Evidence from the Netherlands," Tech. rep., National Bureau of Economic Research.

Acemoglu, D., C. Lelarge, and P. Restrepo (2020): "Competing with Robots: Firm-Level Evidence from France," in *AEA Papers and Proceedings*, American Economic Association 2014 Broadway, Suite 305, Nashville, TN 37203, vol. 110, 383–388.

Acemoglu, D. and P. Restrepo (2020): "Robots and Jobs: Evidence from US labor markets," *Journal of Political Economy*, 128, 2188–2244.

——— (2022): "Tasks, Automation, and the Rise in US Wage Inequality," *Econometrica*, 90, 1973–2016.

Agarwal, N., A. Moehring, P. Rajpurkar, and T. Salz (2023): "Combining Human Expertise with Artificial Intelligence: Experimental Evidence from Radiology," Tech. rep., National Bureau of Economic Research.

Aghion, P., C. Antonin, S. Bunel, and X. Jaravel (2020): "What are the Labor and Product Market Effects of Automation? New Evidence from France," .

Agrawal, A., J. S. Gans, and A. Goldfarb (2019): "Artificial Intelligence: the Ambiguous Labor Market Impact of Automating Prediction," *Journal of Economic Perspectives*, 33, 31–50.

Almog, D., R. Gauriot, L. Page, and D. Martin (2024): "AI Oversight and Human Mistakes: Evidence from Centre Court," *arXiv preprint arXiv:2401.16754*.

Archsmith, J. E., A. Heyes, M. J. Neidell, and B. N. Sampat (2021): "The Dynamics of Inattention in the (Baseball) Field," Tech. rep., National Bureau of Economic Research.

Arellano-Bover, J. (2022): "The Effect of Labor Market Conditions at Entry on Workers' Long-Term Skills," *Review of Economics and Statistics*, 104, 1028–1045.

Arrow, K. J. (1962): "The Economic Implications of Learning by Doing," *The Review of Economic Studies*, 29, 155–173.

Autor, D. H. (2015): "Why are There Still So Many Jobs? The History and Future of Workplace Automation," *Journal of Economic Perspectives*, 29, 3–30.

Autor, D. H. and D. Dorn (2013): "The Growth of Low-Skill Service Jobs and the Polarization of the US Labor Market," *American Economic Review*, 103, 1553–1597.

Aw, J. (2014): "Digitally Numb: Doctors are Losing Hands-On Diagnosis Skills by Relying Too Much on Technology," *National Post*.

Bandiera, O., I. Barankay, and I. Rasul (2005): "Social Preferences and the Response to Incentives: Evidence from Personnel Data," *The Quarterly Journal of Economics*, 120, 917–962.

Barth, E., M. Roed, P. Schøne, and J. Umblijs (2020): "How Robots Change Within-Firm Wage Inequality," .

Benhenda, A. (2022): "Absence, Substitutability and Productivity: Evidence from Teachers," *Labour Economics*, 76, 102167.

Bhattacharya, V. and G. Howard (2022): "Rational Inattention in the Infield," *American Economic Journal: Microeconomics*, 14, 348–393.

Bhuller, M. and H. Sigstad (2024): "Feedback and learning: The causal effects of reversals on judicial decision-making," *Review of Economic Studies*, rdae073.

Bonfiglioli, A., R. Crino, H. Fadinger, and G. Gancia (2020): "Robot Imports and Firm-Level Outcomes," .

Borusyak, K., X. Jaravel, and J. Spiess (2021): "Revisiting Event Study Designs: Robust and Efficient Estimation," *arXiv preprint arXiv:2108.12419*.

Boustan, L. P., J. Choi, and D. Clingingsmith (2022): "Automation after the Assembly Line: Computerized Machine Tools, Employment and Productivity in the United States," Tech. rep., National Bureau of Economic Research.

Brynjolfsson, E., D. Li, and L. R. Raymond (2023): "Generative AI at Work," Tech. rep., National Bureau of Economic Research.

Burke, M., V. Tanutama, S. Heft-Neal, M. Hino, and D. Lobell (2023): "Game, Sweat, Match: Temperature and Elite Worker Productivity," Tech. rep., National Bureau of Economic Research.

Burke, M. A. and T. R. Sass (2008): "Classroom Peer Effects and Student Achievement," *Federal Reserve Bank of Boston*.

Callaway, B. and P. H. Sant'Anna (2021): "Difference-in-Differences with Multiple Time Periods," *Journal of Econometrics*, 225, 200–230.

Cardoso, A. R., P. Guimaraes, P. Portugal, and H. Reis (2018): "The Returns to Schooling Unveiled," .

Chan, D. C., M. Gentzkow, and C. Yu (2022): "Selection with variation in diagnostic skill: Evidence from radiologists," *The Quarterly Journal of Economics*, 137, 729–783.

Chen, D. L., T. J. Moskowitz, and K. Shue (2016): "Decision Making under the Gambler's Fallacy: Evidence from Asylum Judges, Loan Officers, and Baseball Umpires," *The Quarterly Journal of Economics*, 131, 1181–1242.

Chiappori, P.-A., S. Levitt, and T. Groseclose (2002): "Testing Mixed-Strategy Equilibria when Players are Heterogeneous: The Case of Penalty Kicks in Soccer," *American Economic Review*, 92, 1138–1151.

Cohen, J. P., A. C. Johnston, and A. S. Lindner (2023): "Skill Depreciation during Unemployment: Evidence from Panel Data," Tech. rep., National Bureau of Economic Research.

Cooley, J. (2009): "Can Achievement Peer Effect Estimates Inform Policy? A View from Inside the Black Box," *Review of Economics and Statistics*, 93.

Cornelissen, T., C. Dustmann, and U. Schönberg (2017): "Peer Effects in the Workplace," *American Economic Review*, 107, 425–456.

Currie, J. M. and W. B. MacLeod (2020): "Understanding doctor decision making: The case of depression treatment," *Econometrica*, 88, 847–878.

Cymek, D. H., A. Truckenbrodt, and L. Onnasch (2023): "Lean back or lean in? Exploring social loafing in human–robot teams," *Frontiers in Robotics and AI*, 10, 1249252.

Dauth, W., S. Findeisen, J. Suedekum, and N. Woessner (2021): "The Adjustment of Labor Markets to Robots," *Journal of the European Economic Association*, 19, 3104–3153.

De Chaisemartin, C. and X. d'Haultfoeuille (2020): "Two-Way Fixed Effects Estimators with Heterogeneous Treatment Effects," *American Economic Review*, 110, 2964–2996.

Deming, D. J. (2021): "The growing importance of decision-making on the job," Tech. rep., National Bureau of Economic Research.

Dinerstein, M., R. Megalokonomou, and C. Yannelis (2022): "Human Capital Depreciation and Returns to Experience," *American Economic Review*, 112, 3725–62.

Dixon, J., B. Hong, and L. Wu (2021): "The Robot Revolution: Managerial and Employment Consequences for Firms," *Management Science*, 67, 5586–5605.

Duggan, M. and S. D. Levitt (2002): "Winning isn't Everything: Corruption in Sumo Wrestling," *American Economic Review*, 92, 1594–1605.

Edin, P.-A. and M. Gustavsson (2008): "Time Out of Work and Skill Depreciation," *ILR Review*, 61, 163–180.

Falk, A. and A. Ichino (2006): "Clean Evidence on Peer Effects," *Journal of Labor Economics*, 24, 39–57.

Farber, H. S. (2005): "Is Tomorrow Another Day? The Labor Supply of New York City Cabdrivers," *Journal of Political Economy*, 113, 46–82.

Fesselmeyer, E. (2021): "The Impact of Temperature on Labor Quality: Umpire Accuracy in Major League Baseball," *Southern Economic Journal*, 88, 545–567.

Fujita, S. and G. Moscarini (2017): "Recall and Unemployment," *American Economic Review*, 107, 3875–3916.

Glennon, B., F. J. Morales, S. Carnahan, and E. Hernandez (2022): "Does Employing Skilled Immigrants Enhance Competitive Performance? Evidence from Football Clubs," in *Academy of Management Proceedings*, Academy of Management Briarcliff Manor, NY 10510, vol. 2022, 16452.

Goldin, C. and C. Rouse (2000): "Orchestrating Impartiality: The Impact of "Blind" Auditions on Female Musicians," *American Economic Review*, 90, 715–741.

Gong, B., J. P. Nugent, W. Guest, W. Parker, P. J. Chang, F. Khosa, and S. Nicolaou (2019): "Influence of Artificial Intelligence on Canadian Medical Students' Preference for Radiology Specialty: A National Survey Study," *Academic radiology*, 26, 566–577.

Goodman-Bacon, A. (2021): "Difference-in-Differences with Variation in Treatment Timing," *Journal of Econometrics*, 225, 254–277.

Graetz, G. and G. Michaels (2018): "Robots at Work," *Review of Economics and Statistics*, 100, 753–768.

Grennan, J. and R. Michaely (2020): "Artificial Intelligence and High-Skilled Work: Evidence from Analysts," *Swiss Finance Institute Research Paper*.

Guryan, J., K. Kroft, and M. J. Notowidigdo (2009): "Peer Effects in the Workplace: Evidence from Random Groupings in Professional Golf Tournaments," *American Economic Journal: Applied Economics*, 1, 34–68.

Hirvonen, J., A. Stenhammar, and J. Tuhkuri (2022): "New Evidence on the Effect of Technology on Employment and Skill Demand," *Available at SSRN 4081625*.

Hockenberry, J. M. and L. A. Helmchen (2014): "The Nature of Surgeon Human Capital Depreciation," *Journal of Health Economics*, 37, 70–80.

Hoffman, M., L. B. Kahn, and D. Li (2018): "Discretion in Hiring," *The Quarterly Journal of Economics*, 133, 765–800.

Hong, L. and S. Lattanzio (2022): "The Peer Effect on Future Wages in the Workplace," *Available at SSRN 4052587*.

Hortaçsu, A., F. Luco, S. L. Puller, and D. Zhu (2019): "Does strategic ability affect efficiency? Evidence from electricity markets," *American Economic Review*, 109, 4302–4342.

Hoxby, C. M. and G. Weingarth (2005): "Taking Race Out of the Equation: School Reassignment and the Structure of Peer Effects," Tech. rep., Citeseer.

Humlum, A. (2022): *Robot Adoption and Labor Market Dynamics*, Rockwool Foundation Research Unit.

Imberman, S. A., A. D. Kugler, and B. I. Sacerdote (2012): "Katrina's Children: Evidence on the Structure of Peer Effects from Hurricane Evacuees," *American Economic Review*, 102, 2048–2082.

Jacobson, L. S., R. J. LaLonde, and D. G. Sullivan (1993): "Earnings Losses of Displaced Workers," *The American Economic Review*, 685–709.

Jarosch, G. (2023): "Searching for Job Security and the Consequences of Job Loss," *Econometrica*, 91, 903–942.

Kahn, L. M. (2000): "The Sports Business as a Labor Market Laboratory," *Journal of Economic Perspectives*, 14, 75–94.

Kanazawa, K., D. Kawaguchi, H. Shigeoka, and Y. Watanabe (2022): "AI, Skill, and Productivity: The Case of Taxi Drivers," Tech. rep., National Bureau of Economic Research.

Kim, J. W. and B. G. King (2014): "Seeing Stars: Matthew Effects and Status Bias in Major League Baseball Umpiring," *Management Science*, 60, 2619–2644.

Kleinberg, J., H. Lakkaraju, J. Leskovec, J. Ludwig, and S. Mullainathan (2018): "Human Decisions and Machine Predictions," *The Quarterly Journal of Economics*, 133, 237–293.

Koch, M., I. Manuylov, and M. Smolka (2021): "Robots and Firms," *The Economic Journal*, 131, 2553–2584.

Kovash, K. and S. D. Levitt (2009): "Professionals Do Not Play Minimax: Evidence from Major League Baseball and the National Football League," Tech. rep., National Bureau of Economic Research.

Kreitmeir, D. and P. A. Raschky (2024): "The Heterogeneous Productivity Effects of Generative AI," *arXiv preprint arXiv:2403.01964*.

Kroft, K., F. Lange, and M. J. Notowidigdo (2013): "Duration Dependence and Labor Market Conditions: Evidence from a Field Experiment," *The Quarterly Journal of Economics*, 128, 1123–1167.

Lavy, V., M. D. Paserman, and A. Schlosser (2012): "Inside the Black Box of Ability Peer Effects: Evidence from Variation in the Proportion of Low Achievers in the Classroom," *The Economic Journal*, 122, 208–237.

Levitt, S. D., J. A. List, and C. Syverson (2013): "Toward an Understanding of Learning by Doing: Evidence from an Automobile Assembly Plant," *Journal of Political Economy*, 121, 643–681.

Lucas Jr, R. E. (1988): "On the Mechanics of Economic Development," *Journal of Monetary Economics*, 22, 3–42.

Machin, S. and A. Manning (1999): "The Causes and Consequences of Long-Term Unemployment in Europe," *Handbook of Labor Economics*, 3, 3085–3139.

Maestas, N., K. J. Mullen, A. Strand, et al. (2015): "Does Delay Cause Decay? The Effect of Administrative Decision Time on the Labor Force Participation and Earnings of Disability Applicants," Tech. rep., National Bureau of Economic Research.

Maguire, E. A., D. G. Gadian, I. S. Johnsrude, C. D. Good, J. Ashburner, R. S. Frackowiak, and C. D. Frith (2000): "Navigation-Related Structural Change in the Hippocampi of Taxi Drivers," *Proceedings of the National Academy of Sciences*, 97, 4398–4403.

Malueg, D. A. and A. J. Yates (2010): "Testing Contest Theory: Evidence from Best-of-Three Tennis Matches," *The Review of Economics and Statistics*, 92, 689–692.

Mas, A. and E. Moretti (2009): "Peers at Work," *American Economic Review*, 99, 112–145.

Miller, J. B. and A. Sanjurjo (2021): "Is it a Fallacy to Believe in the Hot Hand in the NBA Three-Point Contest?" *European Economic Review*, 138, 103771.

Mills, B. M. (2017): "Technological Innovations in Monitoring and Evaluation: Evidence of Performance Impacts among Major League Baseball Umpires," *Labour Economics*, 46, 189–199.

Moskowitz, T. and L. J. Wertheim (2011): *Scorecasting: The Hidden Influences Behind How Sports are Played and Games are Won*, Crown Archetype.

Nicas, J. and Z. Wichter (2019): "A Worry for Some Pilots: Their Hands-On Flying Skills are Lacking," *New York Times. https://www. nytimes. com/2019/03/14/business/automated-planes. html*.

Noy, S. and W. Zhang (2023): "Experimental Evidence on the Productivity Effects of Generative Artificial Intelligence," *Available at SSRN 4375283*.

Parsons, C. A., J. Sulaeman, M. C. Yates, and D. S. Hamermesh (2011): "Strike Three: Discrimination, Incentives, and Evaluation," *American Economic Review*, 101, 1410–1435.

Peng, S., E. Kalliamvakou, P. Cihon, and M. Demirer (2023): "The Impact of AI on Developer Productivity: Evidence from Github Copilot," *arXiv preprint arXiv:2302.06590*.

Pope, D. G., J. Price, and J. Wolfers (2018): "Awareness Reduces Racial Bias," *Management Science*, 64, 4988–4995.

Price, J., L. Lefgren, and H. Tappen (2013): "Interracial Workplace Cooperation: Evidence from the NBA," *Economic Inquiry*, 51, 1026–1034.

Price, J. and J. Wolfers (2010): "Racial Discrimination Among NBA Referees," *The Quarterly Journal of Economics*, 125, 1859–1887.

Schmieder, J. F., T. von Wachter, and S. Bender (2016): "The Effect of Unemployment Benefits and Nonemployment Durations on Wages," *American Economic Review*, 106, 739–777.

Schwabe, H. and F. Castellacci (2020): "Automation, Workers? Skills and Job Satisfaction," *PLOS One*, 15, e0242929.

Smith, A. and M. Anderson (2017): "Automation in Everyday Life," .

Sun, L. and S. Abraham (2021): "Estimating Dynamic Treatment Effects in Event Studies with Heterogeneous Treatment Effects," *Journal of Econometrics*, 225, 175–199.

Tafti, E. A. (2022): *Technology, Skills, and Performance: The Case of Robots in Surgery*, Institute for Fiscal Studies.

Thompson, P. (2010): "Learning by Doing," *Handbook of the Economics of Innovation*, 1, 429–476.

Williams, M. T. (2019): "MLB Umpires Missed 34,294 Ball-Strike Calls in 2018," *Bring on Robo-Umps*.

Wiswall, M. (2013): "The Dynamics of Teacher Quality," *Journal of Public Economics*, 100, 61–78.

Yang, X. and J. Borland (1991): "A Microeconomic Mechanism for Economic Growth," *Journal of Political Economy*, 99, 460–482.

Tables & Figures

Table 1: Summary Statistics

	Full Sample		With Robot		Without Robot	
	Mean	SD	Mean	SD	Mean	SD
Pitch Characteristics						
Called Strike	0.335	(0.472)	0.312	(0.463)	0.336	(0.472)
Called Correctly	0.930	(0.255)	0.930	(0.255)	0.930	(0.255)
Residualized Accuracy	-0.0000	(0.2392)	0.0152	(0.2461)	-0.0007	(0.2388)
Horizontal distance	0.753	(0.453)	0.782	(0.495)	0.752	(0.451)
Vertical distance	0.850	(0.567)	0.869	(0.584)	0.850	(0.566)
Number of Games	62,678		2,611		60,067	
Number of Pitches	8,864,801		390,823		8,473,978	
Number of Umpires	678		121		666	

Notes: A pitch is correctly called if it crosses the strike zone and called strike or missed the strike zone and called ball. "Residualized accuracy" residualizes whether a decision was correct for pitch location and team-by-year fixed effect to account for pitch coordinates that may depend on calibration for each stadium and stringer plotting coordinates. Distances are in feet. Robot is used in games in Single-A Florida from 2021, in Triple-A Pacific Coast League in 2022 and in select games in both Triple-A leagues in 2023.

Table 2: Do Umpires Lose Skills Following Robot Implementation?

	Treated Umpires					
	Robot Mo.	S.E.	Post-Robot Mo.	S.E.	Baseline Mean	N
Pitch-level Outcomes						
Correctly Called	0.064***	(0.009)	-0.020***	(0.003)	0.932	8,864,667
Correctly Called Strike	0.022***	(0.005)	-0.010***	(0.002)	0.281	8,864,667
Correctly Called Ball	0.041***	(0.006)	-0.010***	(0.002)	0.651	8,864,667
Incorrectly Called Strike	-0.058***	(0.009)	0.025***	(0.003)	0.059	8,864,667
Incorrectly Called Ball	-0.006***	(0.001)	-0.005***	(0.001)	0.008	8,864,667
Called Strike	-0.036***	(0.006)	0.015***	(0.003)	0.340	8,864,667

Notes: All regressions include a vector of covariates at the pitch-level, year-by-month, umpire, league-by-year and home team-by-year fixed effects. Standard errors are clustered at the umpire-level. "Robot Mo." indicate that robot is assisting umpires calling the game and "Post-Robot Mo." indicate that the umpire returned following robot-assistance. A pitch is correctly called if it crosses the strike zone and called strike or missed the strike zone and called ball. A pitch is incorrectly called if it misses the strike zone but called strike or crosses the strike zone but called ball. Baseline mean is calculated using the data before 2021, the year of first robot implementation. * p<0.1,** p<0.05, *** p<0.01.

Table 3: Do Skill Depreciation Vary with Length of Work Duration With Robot?

	Treated Umpires									
			Num. of Consecutive Games w. Robot							
	With Robot	S.E.	1 Game	S.E.	2 Games	S.E.	3 Games	S.E.	Baseline Mean	N
Pitch-level Outcomes										
Correctly Called	0.049***	(0.001)	-0.003*	(0.001)	-0.011***	(0.003)	-0.010	(0.008)	0.895	300,427
Correctly Called Strike	0.031***	(0.002)	-0.002	(0.002)	-0.009**	(0.004)	-0.006	(0.008)	0.251	300,427
Correctly Called Ball	0.018***	(0.002)	-0.000	(0.002)	-0.002	(0.004)	-0.004	(0.008)	0.644	300,427
Incorrectly Called Strike	-0.062***	(0.001)	0.002	(0.001)	0.008**	(0.003)	0.012	(0.008)	0.089	300,427
Incorrectly Called Ball	0.012***	(0.001)	0.001	(0.001)	0.003***	(0.001)	-0.002	(0.002)	0.016	300,427
Called Strike	-0.030***	(0.002)	-0.000	(0.002)	-0.001	(0.004)	0.006	(0.008)	0.340	300,427

Notes: All regressions include a vector of covariates at the pitch-level, year-by-month, umpire, league-by-year and team-by-year fixed effects. Standard errors are clustered at the umpire-level. "With Robot" indicate that the robot is assisting umpires calling the game. "Num. of Consecutive Games w. Robot" indicate that the umpire was assisted by the robot for the last specified number of games. Omitted group is the previous game that did not have the robot. A pitch is correctly called if it crosses the strike zone and called strike or missed the strike zone and called ball. A pitch is incorrectly called if it misses the strike zone but called strike or crosses the strike zone but called ball. Baseline mean is calculated using the first game of the year without robot-assistance. * p<0.1,** p<0.05, *** p<0.01.

Table 4: Do Skill Depreciation Vary with Treatment Intensity?

	Treated Umpires					
	With Robot	S.E.	Dosage	S.E.	Baseline Mean	N
Pitch-level Outcomes						
Correctly Called	0.050***	(0.001)	-0.009	(0.007)	0.903	330,005
Correctly Called Strike	0.033***	(0.001)	-0.006	(0.007)	0.244	330,005
Correctly Called Ball	0.017***	(0.001)	-0.002	(0.007)	0.659	330,005
Incorrectly Called Strike	-0.062***	(0.001)	0.010	(0.006)	0.076	330,005
Incorrectly Called Ball	0.012***	(0.000)	-0.001	(0.002)	0.021	330,005
Called Strike	-0.029***	(0.001)	0.004	(0.007)	0.320	330,005

Notes: All regressions include a vector of covariates at the pitch-level, week, umpire, league-by-year and team-by-year fixed effects. Standard errors are clustered at the umpire-level. "With Robot" indicate that robot is assisting umpires calling the game and "Dosage" is $\frac{\text{\# of games with robot}}{\text{\# of total games}}$. A pitch is correctly called if it crosses the strike zone and called strike or missed the strike zone and called ball. A pitch is incorrectly called if it misses the strike zone but called strike or crosses the strike zone but called ball. Baseline mean is calculated using the first game of the year without robot-assistance. * p<0.1,** p<0.05, *** p<0.01.

Table 5: Do Umpires Regain Skill?

	Num. of Games Since Return											
	1-5 Games	S.E.	6-10 Games	S.E.	11-15 Games	S.E.	16-20 Games	S.E.	21+ Games	S.E.	Baseline Mean	N
Pitch-level Outcomes												
Correctly Called	-0.029***	(0.004)	-0.027***	(0.005)	-0.029**	(0.012)	-0.017**	(0.007)	-0.010**	(0.004)	0.929	4,800,807
Correctly Called Strike	-0.013***	(0.004)	-0.007*	(0.004)	-0.006	(0.007)	-0.011	(0.008)	-0.005*	(0.003)	0.275	4,800,807
Correctly Called Ball	-0.016***	(0.006)	-0.019***	(0.005)	-0.023**	(0.010)	-0.006	(0.005)	-0.004	(0.004)	0.654	4,800,807
Incorrectly Called Strike	0.031***	(0.004)	0.026***	(0.004)	0.006	(0.005)	0.017***	(0.007)	0.010***	(0.004)	0.063	4,800,807
Incorrectly Called Ball	-0.001	(0.002)	0.001	(0.001)	0.023*	(0.013)	-0.000	(0.003)	-0.001	(0.001)	0.008	4,800,807
Called Strike	0.018***	(0.005)	0.019***	(0.005)	0.000	(0.008)	0.007	(0.005)	0.005	(0.003)	0.338	4,800,807

Notes: All regressions for pitch-level outcomes include year-by-month, umpire, league-by-year and team-by-year fixed effects. Standard errors are clustered at the umpire-level. "Num. of Games Since Return' indicate that the umpire is calling the game in the specified numbers of game since the return. A pitch is correctly called if it crosses the strike zone and called strike or missed the strike zone and called ball. A pitch is incorrectly called if it misses the strike zone but called strike or crosses the strike zone but called ball. Baseline mean is calculated using the data before 2021, the year of first robot implementation. * p<0.1,** p<0.05, *** p<0.01.

Table 6: Do Umpires Lose Skill? - From COVID-19 Pause

	Treated Umpires			
	First Season	S.E.	Baseline Mean	N
Pitch-level Outcomes				
Correctly Called	-0.004	(0.003)	0.907	2,817,021
Correctly Called Strike	0.001	(0.003)	0.261	2,817,021
Correctly Called Ball	-0.005**	(0.002)	0.647	2,817,021
Incorrectly Called Strike	0.003	(0.003)	0.082	2,817,021
Incorrectly Called Ball	0.000	(0.001)	0.010	2,817,021
Called Strike	0.005**	(0.002)	0.343	2,817,021

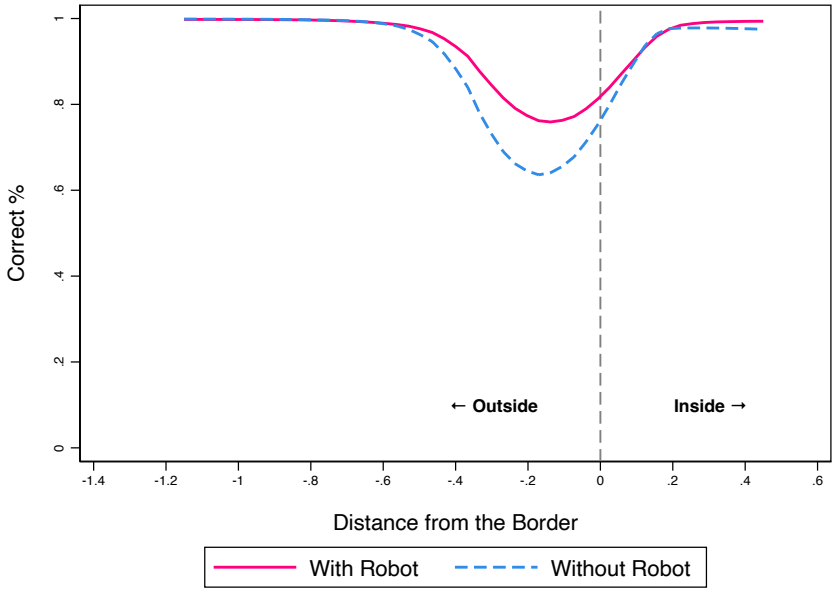
Notes: All regressions for pitch-level outcomes include a vector of covariates at the pitch-level, year-by-month, umpire, league-by-year and team-by-year fixed effects. Standard errors are clustered at the umpire-level. "First Season" indicates that umpires returned from a year-long COVID-19 interruption relative to Major League umpires who didn't have such canceled season. A pitch is correctly called if it crosses the strike zone and called strike or missed the strike zone and called ball. A pitch is incorrectly called if it misses the strike zone but called strike or crosses the strike zone but called ball. * p<0.1,** p<0.05, *** p<0.01.

Table 7: Do Umpires Lose Other Skill? - First Base Umpire and Replay Review

	Post-Robot Umpire	S.E.	Baseline Mean	N
First Base Umpires				
Num. of Ground Outs to 1B	-0.601	(0.399)	19.769	2,403
Num. of Challenges for Calls on the 1B	0.098**	(0.041)	0.221	2,403
Num. of Challenges Overturned	0.094***	(0.028)	0.137	2,403
Num. of Challenges Upheld	0.005	(0.029)	0.084	2,403
Num. of Ejections	-0.013	(0.013)	0.060	2,403

Notes: All regressions include a control for umpire experience and month and team fixed effects. Standard errors are clustered at the umpire-level. "Post-Robot Umpire" indicate that the umpires are assisted by the robots in 2022. The outcomes are from the Major League in 2023 season. "Num. of Ground Outs to 1B" include all ground outs with a play at the first base. "Num. of Challenges for Calls on the 1B" are the number of replay reviews requested by the teams for plays happening at the first base. "Num. of Challenges Overturned" and "Num. of Challenges Upheld" are the number of replay reviews that are overturned and upheld for plays happening at the first base, respectively. "Num. of Ejections" are the number of ejections in a game. * $p<0.1$,** $p<0.05$, *** $p<0.01$.

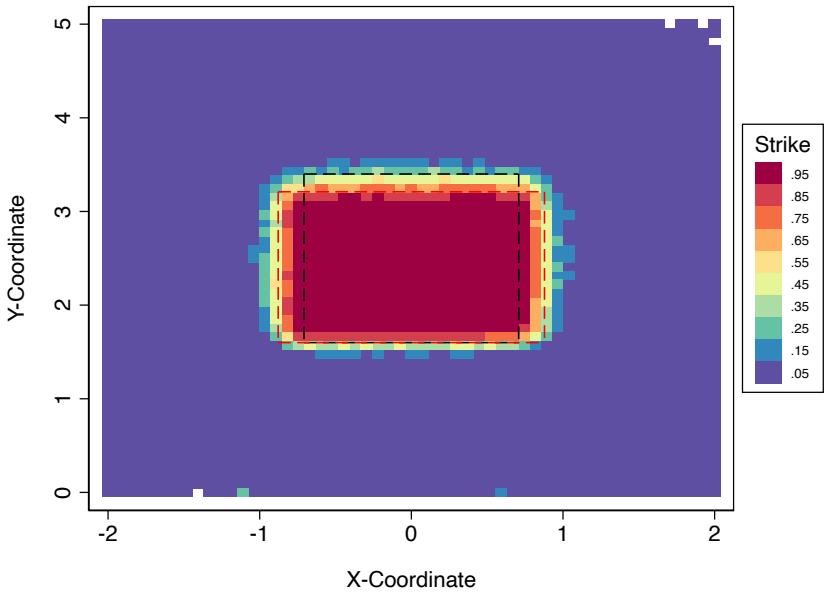
Figure 1: Pitch Distribution



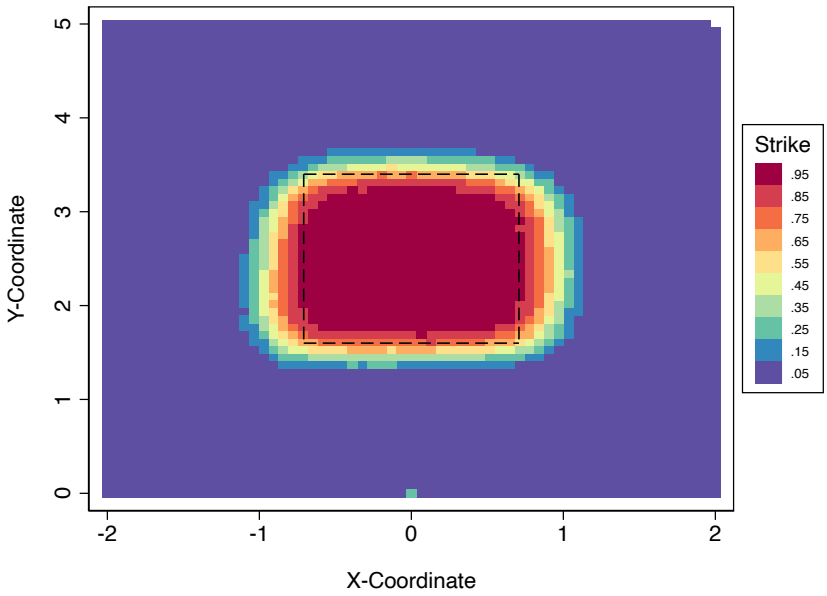
Notes: X-axis is the pitch distance from the nearest border of the strike zone in feet and y-axis shows the average accuracy rate. To the left of origin are pitches falling outside of the strike zone, and to the right are pitches falling inside the strike zone.

Figure 2: Called Strike Heatmaps

A. With Robot



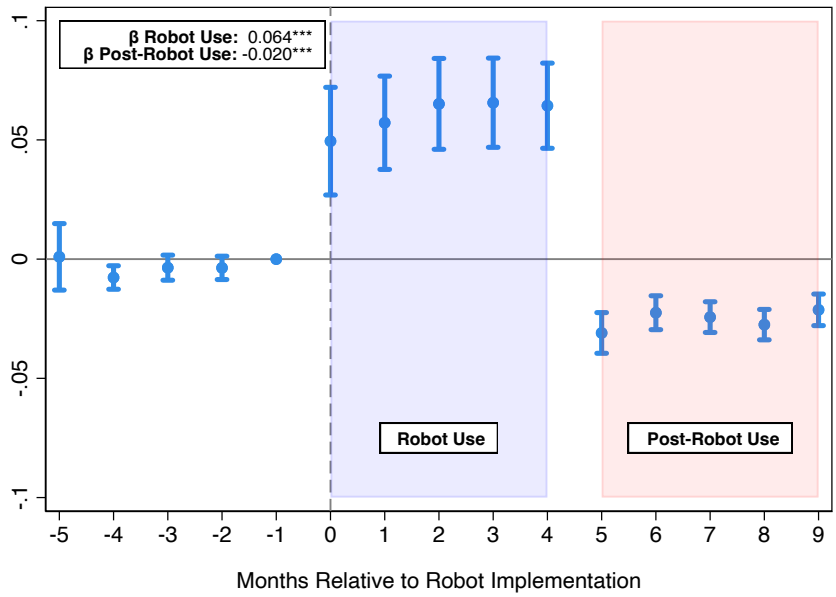
B. Without Robot



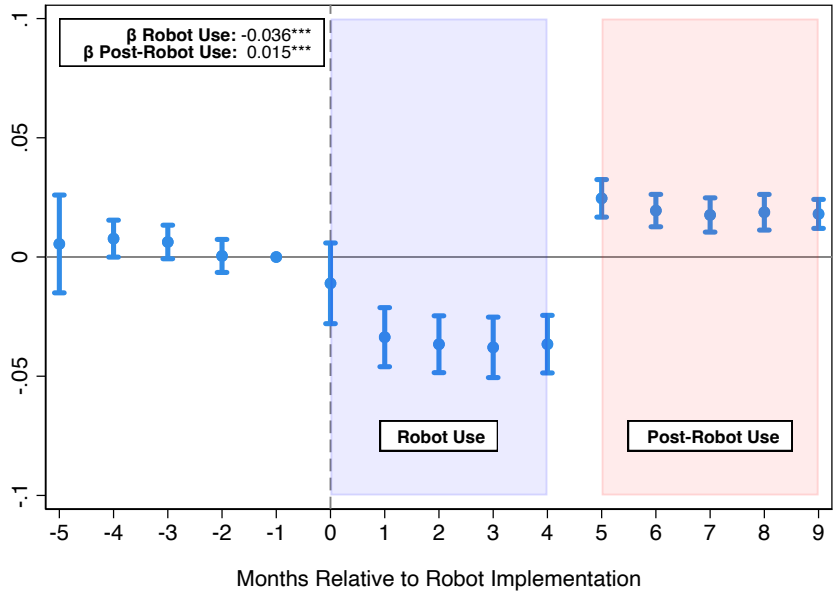
Notes: The figures plot the share of pitches that are called strike by pitch location. The black dotted line shows the strike zone for an average batter and the red dotted line in Panel A shows the adjusted strike zone that was implemented on July 20, 2021 for games with robot.

Figure 3: Do Umpires Lose Skill? - Event-Study

A. Accuracy



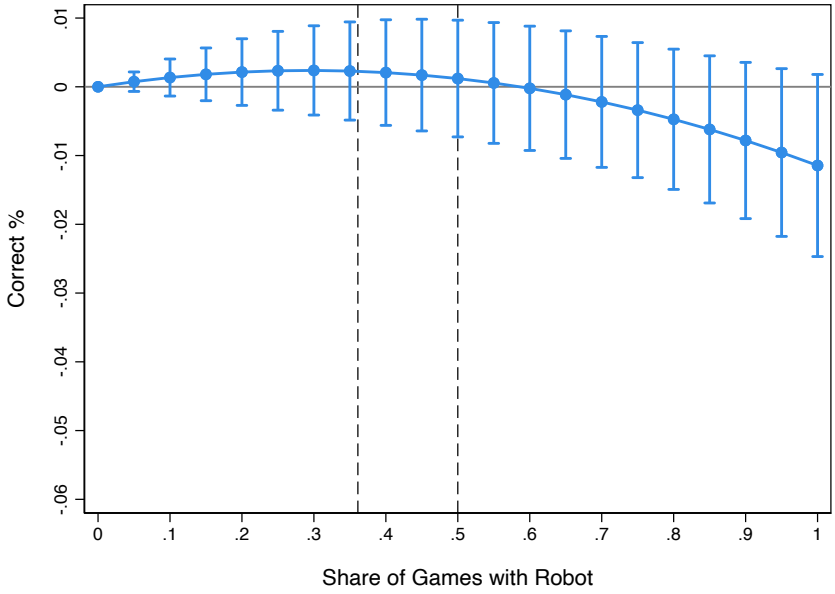
B. Called Strike



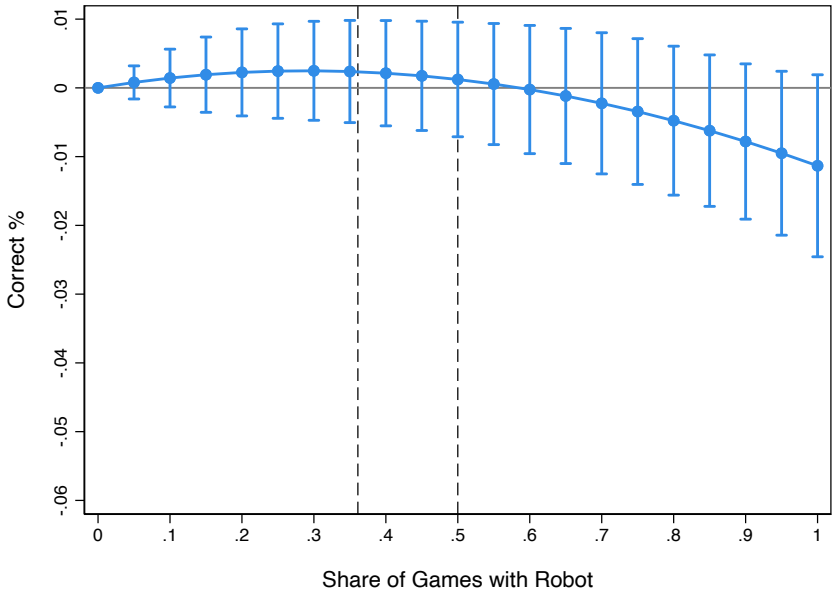
Notes: All regressions include a vector of covariates at the pitch-level, year-by-month, umpire, league-by-year and team-by-year fixed effects. Standard errors are clustered at the umpire-level. "Robot Use" indicate that robot is assisting umpires calling the game and "Post-Robot Use" indicate that the umpire returned following robot-assistance. X-axis is months relative to the first month of robot implementation. A pitch is correctly called if it crosses the strike zone and called strike or missed the strike zone and called ball. * p<0.1,** p<0.05, *** p<0.01.

Figure 4: Do Skill Depreciation Vary with Dosage?

A. Second-Degree Polynomial

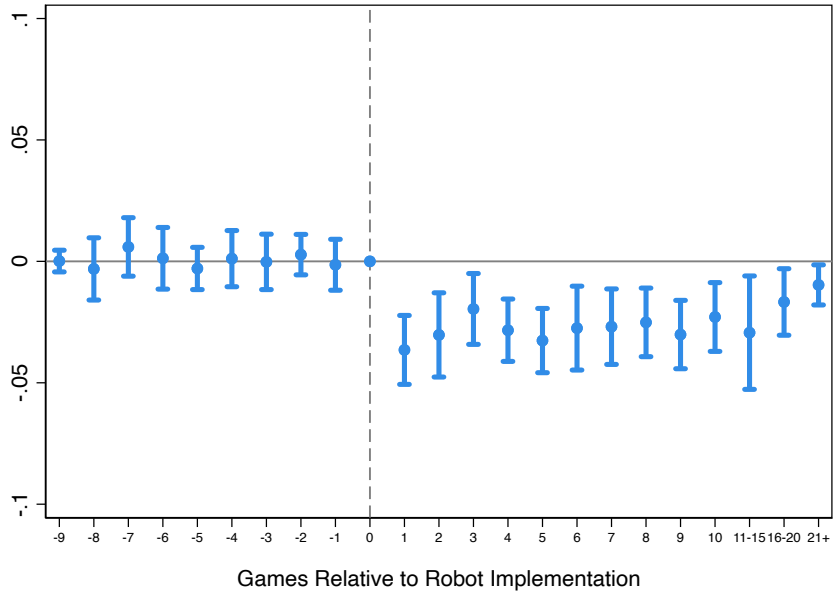


B. Third-Degree Polynomial



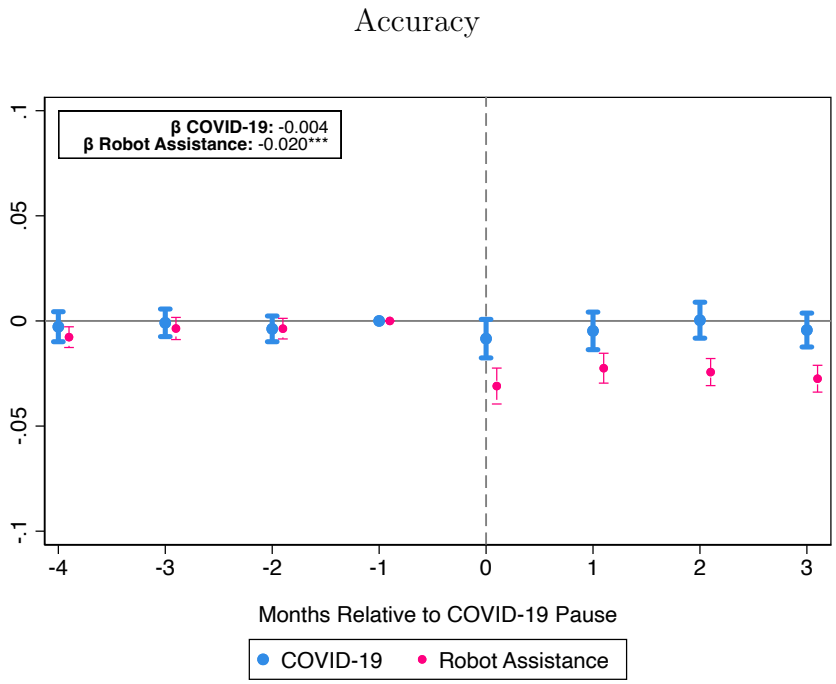
Notes: X-axis plots the dosage where it is $\frac{\#\text{ of games with robot}}{\#\text{ of total games}}$. Y-axis plots the treatment effect on decision accuracy. The figure plots the second and third-degree polynomial estimates of dosage effect. The regression also includes a vector of covariates at the pitch-level, week, umpire, league-by-year and team-by-year fixed effects. Standard errors are clustered at the umpire-level. The gray dotted lines represent the 25th and 75th percentile of dosage. A pitch is correctly called if it crosses the strike zone and called strike or missed the strike zone and called ball.

Figure 5: Do Umpires Regain Skill?



Notes: All regressions include a vector of covariates at the pitch-level, year-by-month, umpire, league-by-year and team-by-year fixed effects. Standard errors are clustered at the umpire-level. X-axis is the number of games relative to the last game with the robot. A pitch is correctly called if it crosses the strike zone and called strike or missed the strike zone and called ball.

Figure 6: Event-Study - Umpire Skills Following COVID-19



Notes: All regressions include a vector of covariates at the pitch-level, year-by-month, umpire, league-by-year and team-by-year fixed effects. Standard errors are clustered at the umpire-level. X-axis is months relative to the first month of 2021 season following the COVID-19 pause. Blue lines show the effect of returning from COVID-19 pause and red line shows the effect of robot implementation in comparison. A pitch is correctly called if it crosses the strike zone and called strike or missed the strike zone and called ball. A pitch is incorrectly called if it misses the strike zone but called strike or crosses the strike zone but called ball. * p<0.1,** p<0.05, *** p<0.01.

A1 Appendix Tables & Figures

Table A1: Player Outcomes Summary Statistics

	Full Sample		With Robot		Without Robot	
	Mean	SD	Mean	SD	Mean	SD
Pitch Characteristics						
Pitch Distance from Center	0.991	(0.544)	1.072	(0.559)	0.988	(0.543)
Pitch Distance from Border	0.403	(0.282)	0.426	(0.344)	0.402	(0.278)
Pitched Inside	0.511	(0.500)	0.426	(0.494)	0.515	(0.500)
Pitched Edge	0.596	(0.491)	0.618	(0.486)	0.595	(0.491)
Batter Swung	0.489	(0.500)	0.478	(0.500)	0.489	(0.500)
Game Characteristics - Pitchers						
Pitcher Strike Out	4.718	(2.351)	4.584	(2.234)	4.724	(2.356)
Pitcher Hit Allowed	4.946	(2.254)	4.606	(2.228)	4.959	(2.253)
Pitcher Home Run Allowed	0.589	(0.802)	0.615	(0.817)	0.588	(0.801)
Pitcher Walk Allowed	1.952	(1.394)	2.196	(1.471)	1.942	(1.390)
Game Characteristics - Batters						
Batter Struck Out	0.959	(0.899)	1.025	(0.920)	0.956	(0.898)
Batter Hit	0.926	(0.892)	0.928	(0.895)	0.926	(0.891)
Batter Home Run	0.108	(0.330)	0.119	(0.346)	0.107	(0.329)
Batter Walked	0.428	(0.652)	0.543	(0.732)	0.423	(0.648)
Number of Games	66,297		2,731		63,566	
Number of Pitchers	7,821		1,948		7,657	
Number of Batters	7,496		1,537		7,383	

Notes: Distances are measured in feet. "Pitched Inside" indicates that a pitch falls inside the strike zone and "Pitched Edge" indicates that a pitch falls between 0.5 and 1.5 feet from the center of the strike zone. Outcomes in "Game Characteristics" panel are collapsed at the game-level. I restrict the sample to pitchers who faced at least 15 batters in an appearance and batters with at least 3 at-bats. A batter's at-bat ends in a walk (or base on balls) when the count reaches 4 balls. Robot is used in games in Single-A Florida from 2021, in Triple-A Pacific Coast League in 2022 and in select games in both Triple-A leagues in 2023.

Table A2: Summary Statistics - Are Robot-Assisted Umpires Different?

	Never-Treated Umpires		Robot-Assisted Umpires		Post-Robot Umpires	
	Mean	SD	Mean	SD	Mean	SD
Pitch Characteristics						
Called Strike	0.338	(0.473)	0.337	(0.473)	0.337	(0.473)
Called Correctly	0.953	(0.212)	0.953	(0.211)	0.953	(0.211)
Residualized Accuracy	0.0000	(0.2016)	0.0000	(0.2013)	0.0001	(0.2006)
Horizontal distance	0.740	(0.434)	0.737	(0.435)	0.739	(0.436)
Vertical distance	0.844	(0.567)	0.840	(0.557)	0.838	(0.557)
Number of Games	15,067		10,750		9,630	
Number of Pitches	2,089,927		1,493,225		1,337,175	
Number of Umpires	313		70		62	

Notes: "Never-Treated Umpires" are those who have never been assisted by the robots, "Robot-Assisted Umpires" include those who have been assisted by the robots and "Post-Robot Umpires" include those who have been assisted and subsequently moved to another league without the robots. The statistics are from before any robot implementation and from the Minor Leagues. A pitch is correctly called if it crosses the strike zone and called strike or missed the strike zone and called ball. "Residualized accuracy" residualizes whether a decision was correct for pitch location and team-by-year fixed effect to account for pitch coordinates that may depend on calibration for each stadium and stringer plotting coordinates. Distances are in feet.

Table A3: Do Umpires Lose Skill? - Restricting to More Ambiguous Decisions

	Treated Umpires					
	Robot Mo.	S.E.	Post-Robot Mo.	S.E.	Baseline Mean	N
Pitch-level Outcomes						
Correctly Called	0.123***	(0.018)	-0.037***	(0.005)	0.864	4,166,996
Correctly Called Strike	0.015***	(0.005)	0.004**	(0.002)	0.267	4,166,996
Correctly Called Ball	0.108***	(0.017)	-0.041***	(0.006)	0.597	4,166,996
Incorrectly Called Strike	-0.111***	(0.018)	0.047***	(0.007)	0.126	4,166,996
Incorrectly Called Ball	-0.012***	(0.002)	-0.011***	(0.001)	0.010	4,166,996
Called Strike	-0.097***	(0.017)	0.052***	(0.007)	0.393	4,166,996

Notes: All regressions for pitch-level outcomes include a vector of covariates at the pitch-level, year-by-month, umpire, league-by-year and team-by-year fixed effects. Standard errors are clustered at the umpire-level. "Robot Mo." indicate that robot is assisting umpires calling the game and "Post-Robot Mo." indicate that the umpire returned following robot-assistance. A pitch is correctly called if it crosses the strike zone and called strike or missed the strike zone and called ball. A pitch is incorrectly called if it misses the strike zone but called strike or crosses the strike zone but called ball. Sample is restricted to pitches that are within 0.5 feet from the nearest border of the strike zone from the outside and within 0.2 feet from the nearest border of the strike zone from the inside. Baseline mean is calculated using the data before 2021, the year of first robot implementation. * p<0.1,** p<0.05, *** p<0.01.

Table A4: Do Umpires Become Less Consistent?

	Treated Umpires					
	Robot Mo.	S.E.	Post-Robot Mo.	S.E.	Baseline Mean	N
Game-level Outcomes						
Correctly Called	0.123***	(0.019)	-0.036***	(0.005)	0.853	47,960
Correctly Called (SD)	-0.098***	(0.016)	0.032***	(0.005)	0.313	47,960

Notes: All regressions include year-by-month, umpire, league-by-year and team-by-year fixed effects. Standard errors are clustered at the umpire-level. "Robot Mo." indicate that robot is assisting umpires calling the game and "Post-Robot Mo." indicate that the umpire returned following robot-assistance. A pitch is correctly called if it crosses the strike zone and called strike or missed the strike zone and called ball. A pitch is incorrectly called if it misses the strike zone but called strike or crosses the strike zone but called ball. "Correctly Called (SD)" is the standard deviation of accuracy at the game-level. Sample is restricted to pitches that are within 0.5 feet from the nearest border of the strike zone from the outside and within 0.2 feet from the nearest border of the strike zone from the inside. Baseline mean is calculated using the data before 2021, the year of first robot implementation. * p<0.1,** p<0.05, *** p<0.01.

Table A5: Do Umpires Suffer More From Decision Biases?

	Treated Umpires					
	Robot Mo.	S.E.	Post-Robot Mo.	S.E.	Baseline Mean	N
Game-level Outcomes						
Omission Bias (3-0)	-0.077***	(0.023)	0.029***	(0.008)	0.088	58,580
Omission Bias (0-2)	-0.002	(0.003)	-0.003**	(0.001)	0.010	62,490
Gambler's Fallacy	-0.012*	(0.006)	-0.003*	(0.002)	0.011	61,839

Notes: All regressions include year-by-month, umpire, league-by-year and team-by-year fixed effects. Standard errors are clustered at the umpire-level. "Robot Mo." indicate that robot is assisting umpires calling the game and "Post-Robot Mo." indicate that the umpire returned following robot-assistance. Outcomes are indicators that a call is incorrectly called given a situation. The outcome for "Omission Bias (3-0)" is that a pitch was incorrectly called strike when the count is 3-0. The outcome for "Omission Bias (0-2)" is that a pitch was incorrectly called ball when the count is 3-0. The outcome for "Gambler's Fallacy" is that a pitch was incorrectly called ball when the previous pitch was called strike. Baseline mean is calculated using the data before 2021, the year of first robot implementation. * $p<0.1$,** $p<0.05$, *** $p<0.01$.

Table A6: Summary Statistics - Are Umpires who Move Different?

	Umpires who Stayed		Moved Umpire		Promoted Umpire	
	Mean	SD	Mean	SD	Mean	SD
Pitch Characteristics						
Called Strike	0.330	(0.470)	0.328	(0.470)	0.325	(0.469)
Called Correctly	0.950	(0.219)	0.944	(0.230)	0.936	(0.244)
Residualized Accuracy	-0.0000	(0.2070)	0.0000	(0.2151)	-0.0002	(0.2253)
Horizontal distance	0.738	(0.432)	0.723	(0.424)	0.722	(0.424)
Vertical distance	0.837	(0.553)	0.846	(0.550)	0.862	(0.553)
Number of Games	915		5,444		2,850	
Number of Pitches	132,937		782,866		407,161	
Number of Umpires	86		161		80	

Notes: "Umpires who Stayed" are umpires who have not moved between the leagues in 2022, "Moved Umpire" represents the umpires who primarily called games in different league than in 2022 and "Promoted Umpire" represents the umpires who moved up the league class from 2022. The statistics are from 2021 and from the Minor Leagues as 2020 season was canceled due to the COVID-19 pandemic. A pitch is correctly called if it crosses the strike zone and called strike or missed the strike zone and called ball. "Residualized accuracy" residualizes whether a decision was correct for pitch location and team-by-year fixed effect to account for pitch coordinates that may depend on calibration for each stadium and stringer plotting coordinates. Distances are in feet.

Table A7: Are Results Sensitive to Different Control Groups?

	Treated Umpires					
	Robot Mo.	S.E.	Post-Robot Mo.	S.E.	Baseline Mean	N
Pitch-level Outcomes						
Control Group: Umpires Switched League						
Correctly Called	0.062***	(0.009)	-0.025***	(0.004)	0.951	4,814,684
Correctly Called Strike	0.017***	(0.004)	-0.011***	(0.002)	0.294	4,814,684
Correctly Called Ball	0.045***	(0.007)	-0.014***	(0.003)	0.657	4,814,684
Incorrectly Called Strike	-0.056***	(0.009)	0.031***	(0.004)	0.042	4,814,684
Incorrectly Called Ball	-0.006***	(0.001)	-0.007***	(0.001)	0.007	4,814,684
Called Strike	-0.039***	(0.007)	0.021***	(0.003)	0.336	4,814,684
Control Group: Promoted Umpires						
Correctly Called	0.061***	(0.009)	-0.027***	(0.004)	0.951	3,879,268
Correctly Called Strike	0.019***	(0.004)	-0.011***	(0.002)	0.296	3,879,268
Correctly Called Ball	0.042***	(0.007)	-0.015***	(0.003)	0.656	3,879,268
Incorrectly Called Strike	-0.055***	(0.009)	0.034***	(0.004)	0.042	3,879,268
Incorrectly Called Ball	-0.006***	(0.001)	-0.007***	(0.001)	0.007	3,879,268
Called Strike	-0.036***	(0.007)	0.023***	(0.003)	0.338	3,879,268
Control Group: Single-A Umpires						
Correctly Called	0.061***	(0.009)	-0.031***	(0.004)	0.951	3,191,497
Correctly Called Strike	0.019***	(0.004)	-0.013***	(0.002)	0.297	3,191,497
Correctly Called Ball	0.042***	(0.007)	-0.018***	(0.003)	0.655	3,191,497
Incorrectly Called Strike	-0.054***	(0.009)	0.039***	(0.004)	0.042	3,191,497
Incorrectly Called Ball	-0.006***	(0.002)	-0.008***	(0.001)	0.007	3,191,497
Called Strike	-0.035***	(0.007)	0.026***	(0.003)	0.338	3,191,497
Control Group: Major League Umpires						
Correctly Called	0.065***	(0.008)	-0.027***	(0.003)	0.913	4,723,320
Correctly Called Strike	0.029***	(0.005)	-0.014***	(0.002)	0.266	4,723,320
Correctly Called Ball	0.036***	(0.005)	-0.013***	(0.002)	0.647	4,723,320
Incorrectly Called Strike	-0.057***	(0.008)	0.034***	(0.004)	0.077	4,723,320
Incorrectly Called Ball	-0.008***	(0.001)	-0.007***	(0.001)	0.010	4,723,320
Called Strike	-0.028***	(0.005)	0.020***	(0.002)	0.343	4,723,320

Notes: All regressions for pitch-level outcomes include a vector of covariates at the pitch-level, year-by-month, umpire, league-by-year and team-by-year fixed effects. Standard errors are clustered at the umpire-level. "Robot Mo." indicate that robot is assisting umpires calling the game and "Post-Robot Mo." indicate that the umpire returned following robot-assistance. A pitch is correctly called if it crosses the strike zone and called strike or missed the strike zone and called ball. A pitch is incorrectly called if it misses the strike zone but called strike or crosses the strike zone but called ball. Each panel uses different control groups: "Umpires Switched League" refers to any umpires who switched league between seasons; "Promoted Umpires" refers to umpires who moved up in the league system; "Single-A Umpires" and "Major League Umpires" are the umpires working in the lowest and highest system, respectively. Baseline mean is calculated using the data before 2021, the year of first robot implementation. * p<0.1,** p<0.05, *** p<0.01.

Table A8: Is it the Skills or the Effort?

	Treated Umpires					
	Robot Mo.	S.E.	Post-Robot Mo.	S.E.	Baseline Mean	N
Full Sample						
Correctly Called	0.064***	(0.009)	-0.020***	(0.003)	0.932	8,864,667
High Leverage Situations						
Score Difference $\leq$ 1						
Correctly Called	0.055***	(0.010)	-0.017***	(0.003)	0.932	4,256,396
After 7th Inning						
Correctly Called	0.066***	(0.009)	-0.020***	(0.003)	0.933	2,845,494
Runners on Base						
Correctly Called	0.053***	(0.008)	-0.017***	(0.003)	0.942	4,422,108
Score Difference $\leq$ 1 & After 7th Inning						
Correctly Called	0.058***	(0.010)	-0.016***	(0.004)	0.933	930,918
Runners on Base & After 7th Inning						
Correctly Called	0.054***	(0.008)	-0.016***	(0.003)	0.943	1,454,314
Score Difference $\leq$ 1 & Runners on Base						
Correctly Called	0.045***	(0.008)	-0.015***	(0.003)	0.942	2,088,734
Score Difference $\leq$ 1 & Runners on Base & After 7th Inning						
Correctly Called	0.038***	(0.008)	-0.015***	(0.004)	0.944	475,044

Notes: All regressions for pitch-level outcomes include a vector of covariates at the pitch-level, year-by-month, umpire, league-by-year and team-by-year fixed effects. Standard errors are clustered at the umpire-level. "Robot Mo." indicate that robot is assisting umpires calling the game and "Post-Robot Mo." indicate that the umpire returned following robot-assistance. A pitch is correctly called if it crosses the strike zone and called strike or missed the strike zone and called ball. A pitch is incorrectly called if it misses the strike zone but called strike or crosses the strike zone but called ball. Each panel restricts to respective game situations. Baseline mean is calculated using the data before 2021, the year of first robot implementation. * $p<0.1$,** $p<0.05$, *** $p<0.01$.

Table A9: Do Players Alter Their Behaviors?

	Treated Players					
	Robot Mo.	S.E.	Post-Robot Mo.	S.E.	Baseline Mean	N
Pitch-level Outcomes						
Pitch Distance from Center	-0.016***	(0.003)	-0.008***	(0.003)	0.980	17,327,268
Pitch Distance from Border	-0.017***	(0.002)	-0.003**	(0.001)	0.404	17,327,268
Pitched Inside	0.070***	(0.002)	0.005**	(0.002)	0.527	17,327,268
Pitched Edge	0.001	(0.002)	0.002	(0.002)	0.588	17,327,268
Batter Swung	-0.014***	(0.003)	0.000	(0.003)	0.491	17,325,584

Notes: All regressions include a vector of covariates at the pitch-level, year-by-month, umpire and player, league-by-year and team-by-year fixed effects, otherwise specified. For distance outcomes, pitch location controls are omitted. Standard errors are clustered at the player-level. "Robot Mo." indicate that robot is assisting umpires calling the game and "Post-Robot Mo." indicate that the umpire returned following robot-assistance. Distances are measured in feet. "Pitched Inside" indicates that a pitch falls inside the strike zone and "Pitched Edge" indicates that a pitch falls between 0.5 and 1.5 feet from the center of the strike zone. Baseline mean is calculated using the data before 2021, the year of first robot implementation. * p<0.1,** p<0.05, *** p<0.01.

Table A10: Do Umpires Lose Skill? - Reweighing for Player Responses

	Treated Umpires					
	Robot Mo.	S.E.	Post-Robot Mo.	S.E.	Baseline Mean	N
Pitch-level Outcomes						
Correctly Called	0.080***	(0.012)	-0.024***	(0.004)	0.931	8,685,514
Correctly Called Strike	0.013***	(0.004)	-0.005***	(0.001)	0.286	8,685,514
Correctly Called Ball	0.067***	(0.010)	-0.020***	(0.004)	0.645	8,685,514
Incorrectly Called Strike	-0.074***	(0.012)	0.030***	(0.004)	0.060	8,685,514
Incorrectly Called Ball	-0.006***	(0.001)	-0.006***	(0.001)	0.009	8,685,514
Called Strike	-0.061***	(0.010)	0.025***	(0.004)	0.346	8,685,514

Notes: All regressions for pitch-level outcomes include a vector of covariates at the pitch-level, year-by-month, umpire, league-by-year and team-by-year fixed effects. Regressions reweigh the data to have same distribution of pitches as the pitches in the Major League in 2021. Standard errors are clustered at the umpire-level. "Robot Mo." indicate that robot is assisting umpires calling the game and "Post-Robot Mo." indicate that the umpire returned following robot-assistance. A pitch is correctly called if it crosses the strike zone and called strike or missed the strike zone and called ball. A pitch is incorrectly called if it misses the strike zone but called strike or crosses the strike zone but called ball. Baseline mean is calculated using the data before 2021, the year of first robot implementation. * $p<0.1$,** $p<0.05$, *** $p<0.01$.

Table A11: Do Pitch-Level Controls Matter?

	Treated Umpires					
Main Specification	Robot Mo.	S.E.	Post-Robot Mo.	S.E.	Baseline Mean	N
without Pitch-level Controls						
Correctly Called	0.051***	(0.008)	-0.015***	(0.002)	0.932	8,864,801
with Pitch-level Controls						
Correctly Called	0.064***	(0.009)	-0.020***	(0.003)	0.932	8,864,667
Dosage Specification	With Robot	S.E.	Dosage	S.E.	Baseline Mean	N
without Pitch-level Controls						
Correctly Called	0.039***	(0.001)	-0.006	(0.007)	0.903	330,008
with Pitch-level Controls						
Correctly Called	0.050***	(0.001)	-0.008	(0.007)	0.903	330,005

Notes: All regressions for pitch-level outcomes include year-by-month, umpire, league-by-year and team-by-year fixed effects. Standard errors are clustered at the umpire-level. "Treated Games" indicate that robot is assisting umpires calling the game and "Dosage" is $\frac{\#\text{ of games with robot}}{\#\text{ of total games}}$. A pitch is correctly called if it crosses the strike zone and called strike or missed the strike zone and called ball. A pitch is incorrectly called if it misses the strike zone but called strike or crosses the strike zone but called ball. Baseline mean is calculated using the data before 2021, the year of first robot implementation for the Main Specification panel. Baseline mean is calculated using the first game of the year without robot-assistance for the Dosage Specification panel. * p<0.1,** p<0.05, *** p<0.01.

Table A12: Does Data Quality Matter?

	Treated Umpires					
	Robot Mo.	S.E.	Post-Robot Mo.	S.E.	Baseline Mean	N
Pitch-level Outcomes						
Correctly Called	0.033***	(0.006)	-0.015***	(0.002)	0.919	8,815,175
Correctly Called Strike	0.038***	(0.007)	-0.020***	(0.003)	0.267	8,815,175
Correctly Called Ball	-0.005	(0.004)	0.005***	(0.001)	0.652	8,815,175
Incorrectly Called Strike	-0.035***	(0.005)	0.017***	(0.003)	0.075	8,815,175
Incorrectly Called Ball	0.002	(0.002)	-0.002***	(0.000)	0.006	8,815,175
Called Strike	0.003	(0.004)	-0.003**	(0.001)	0.342	8,815,175

Notes: All regressions for pitch-level outcomes include a vector of covariates at the pitch-level, year-by-month, umpire, league-by-year and team-by-year fixed effects. I use the manually-plotted coordinates for all pitches. Standard errors are clustered at the umpire-level. "Robot Mo." indicate that robot is assisting umpires calling the game and "Post-Robot Mo." indicate that the umpire returned following robot-assistance. A pitch is correctly called if it crosses the strike zone and called strike or missed the strike zone and called ball. A pitch is incorrectly called if it misses the strike zone but called strike or crosses the strike zone but called ball. Baseline mean is calculated using the data before 2021, the year of first robot implementation. * p<0.1,** p<0.05, *** p<0.01.

Table A13: Do Umpires Lose Skill? - Restricting to Triple-A and Major League Umpires

	Treated Umpires					
	Robot Mo.	S.E.	Post-Robot Mo.	S.E.	Baseline Mean	N
Pitch-level Outcomes						
Correctly Called	0.089***	(0.003)	-0.036***	(0.003)	0.913	1,920,905
Correctly Called Strike	0.046***	(0.003)	-0.021***	(0.002)	0.261	1,920,905
Correctly Called Ball	0.043***	(0.003)	-0.015***	(0.002)	0.652	1,920,905
Incorrectly Called Strike	-0.081***	(0.003)	0.044***	(0.003)	0.079	1,920,905
Incorrectly Called Ball	-0.008***	(0.001)	-0.008***	(0.001)	0.008	1,920,905
Called Strike	-0.035***	(0.004)	0.024***	(0.002)	0.340	1,920,905

Notes: All regressions for pitch-level outcomes include a vector of covariates at the pitch-level, year-by-month, umpire, league-by-year and team-by-year fixed effects. Standard errors are clustered at the umpire-level. "Robot Mo." indicate that robot is assisting umpires calling the game and "Post-Robot Mo." indicate that the umpire returned following robot-assistance. A pitch is correctly called if it crosses the strike zone and called strike or missed the strike zone and called ball. A pitch is incorrectly called if it misses the strike zone but called strike or crosses the strike zone but called ball. I restrict the sample to the Major League umpires and umpires treated in Triple-A implementation in 2022. Baseline mean is calculated using the data before 2022, the year of robot implementation for the Triple-A umpires. * p<0.1,** p<0.05, *** p<0.01.

Table A14: Do High-Skilled Umpires Lose More Skill?

	Treated Umpires					
	Robot Mo.	S.E.	Post-Robot Mo.	S.E.	Baseline Mean	N
Bottom Quartile						
Correctly Called	0.105***	(0.009)	-0.009	(0.006)	0.901	459,518
Correctly Called Strike	0.058***	(0.011)	-0.005	(0.004)	0.250	459,518
Correctly Called Ball	0.047***	(0.009)	-0.004	(0.004)	0.651	459,518
Incorrectly Called Strike	-0.096***	(0.005)	0.011	(0.007)	0.090	459,518
Incorrectly Called Ball	-0.010*	(0.005)	-0.003*	(0.001)	0.009	459,518
Called Strike	-0.038***	(0.011)	0.007	(0.005)	0.340	459,518
IQR						
Correctly Called	0.091***	(0.007)	-0.046***	(0.005)	0.948	353,885
Correctly Called Strike	0.036***	(0.004)	-0.022***	(0.003)	0.287	353,885
Correctly Called Ball	0.054***	(0.007)	-0.024***	(0.003)	0.661	353,885
Incorrectly Called Strike	-0.082***	(0.005)	0.056***	(0.006)	0.048	353,885
Incorrectly Called Ball	-0.009**	(0.003)	-0.010***	(0.001)	0.004	353,885
Called Strike	-0.046***	(0.006)	0.034***	(0.004)	0.335	353,885
Top Quartile						
Correctly Called	0.084***	(0.008)	-0.047***	(0.002)	0.961	232,069
Correctly Called Strike	0.043***	(0.008)	-0.032***	(0.003)	0.291	232,069
Correctly Called Ball	0.041***	(0.010)	-0.015***	(0.003)	0.670	232,069
Incorrectly Called Strike	-0.063***	(0.006)	0.061***	(0.003)	0.034	232,069
Incorrectly Called Ball	-0.021***	(0.007)	-0.014***	(0.001)	0.004	232,069
Called Strike	-0.019*	(0.009)	0.029***	(0.003)	0.325	232,069

Notes: All regressions for pitch-level outcomes include a vector of covariates at the pitch-level, year-by-month, umpire, league-by-year and team-by-year fixed effects. Standard errors are clustered at the umpire-level. "Robot Mo." indicate that robot is assisting umpires calling the game and "Post-Robot Mo." indicate that the umpire returned following robot-assistance. A pitch is correctly called if it crosses the strike zone and called strike or missed the strike zone and called ball. A pitch is incorrectly called if it misses the strike zone but called strike or crosses the strike zone but called ball. I restrict the sample to never-treated umpires and umpires treated in Triple-A implementation in 2022. Umpires' skill are measured in 2021 and divided into quartiles. Baseline mean is calculated using the data before 2022, the year of robot implementation for the Triple-A umpires. * p<0.1,** p<0.05, *** p<0.01.

Table A15: Do Umpires Improve with Experience?

	Experience	S.E.	Outcome Mean	N
All Umpires				
Accuracy	0.0007	(0.0005)	0.9162	893
Umpires Experience $\leq$ 5 yrs.				
Accuracy	0.0020***	(0.0006)	0.9467	317
Umpires Experience $\in$ (5,10] yrs.				
Accuracy	-0.0052***	(0.0017)	0.9216	208

Notes: All regressions include umpire fixed effects. Standard errors are clustered at the umpire-level. Experience is measured in years. A pitch is correctly called if it crosses the strike zone and called strike or missed the strike zone and called ball. A pitch is incorrectly called if it misses the strike zone but called strike or crosses the strike zone but called ball. Outcomes are collapsed at the umpire-by-year level. * p<0.1,** p<0.05, *** p<0.01.

Table A16: Do More-Experienced Umpires Lose More Skill?

	Treated Umpires					
	Robot Mo.	S.E.	Post-Robot Mo.	S.E.	Baseline Mean	N
Umpire Experience ≤ 6 years						
Correctly Called	0.094***	(0.008)	-0.053***	(0.003)	0.953	157,135
Correctly Called Strike	0.014*	(0.008)	-0.031***	(0.002)	0.288	157,135
Correctly Called Ball	0.080***	(0.011)	-0.022***	(0.002)	0.666	157,135
Incorrectly Called Strike	-0.084***	(0.008)	0.064***	(0.002)	0.042	157,135
Incorrectly Called Ball	-0.010***	(0.003)	-0.011***	(0.001)	0.004	157,135
Called Strike	-0.069***	(0.011)	0.033***	(0.002)	0.330	157,135
Umpire Experience ∈ (6, 9] years						
Correctly Called	0.079***	(0.004)	-0.042***	(0.003)	0.948	460,100
Correctly Called Strike	0.039***	(0.004)	-0.022***	(0.002)	0.284	460,100
Correctly Called Ball	0.040***	(0.005)	-0.020***	(0.003)	0.664	460,100
Incorrectly Called Strike	-0.074***	(0.004)	0.052***	(0.004)	0.046	460,100
Incorrectly Called Ball	-0.005**	(0.002)	-0.011***	(0.001)	0.005	460,100
Called Strike	-0.035***	(0.005)	0.031***	(0.003)	0.331	460,100
Umpire Experience > 9 years						
Correctly Called	0.092***	(0.008)	-0.018***	(0.005)	0.895	1,100,884
Correctly Called Strike	0.058***	(0.006)	-0.009***	(0.003)	0.249	1,100,884
Correctly Called Ball	0.035***	(0.008)	-0.009***	(0.003)	0.646	1,100,884
Incorrectly Called Strike	-0.084***	(0.008)	0.021***	(0.006)	0.096	1,100,884
Incorrectly Called Ball	-0.008***	(0.002)	-0.003***	(0.001)	0.009	1,100,884
Called Strike	-0.026***	(0.008)	0.012***	(0.004)	0.345	1,100,884

Notes: All regressions for pitch-level outcomes include a vector of covariates at the pitch-level, year-by-month, umpire, league-by-year and team-by-year fixed effects. Standard errors are clustered at the umpire-level. "Robot Mo." indicate that robot is assisting umpires calling the game and "Post-Robot Mo." indicate that the umpire returned following robot-assistance. A pitch is correctly called if it crosses the strike zone and called strike or missed the strike zone and called ball. A pitch is incorrectly called if it misses the strike zone but called strike or crosses the strike zone but called ball. I restrict the sample to the Major League umpires and umpires treated in Triple-A implementation in 2022. Umpires' years of experience are measured in 2021. Baseline mean is calculated using the data before 2022, the year of robot implementation for the Triple-A umpires. * p<0.1,** p<0.05, *** p<0.01.

Table A17: Do Umpires Lose More Skill if They Are Further Away from Promotion?

	Treated Umpires					
	Robot Mo.	S.E.	Post-Robot Mo.	S.E.	Baseline Mean	N
Umpire without MLB Experience						
Correctly Called	0.082***	(0.005)	-0.049***	(0.002)	0.953	512,674
Correctly Called Strike	0.028***	(0.004)	-0.026***	(0.002)	0.290	512,674
Correctly Called Ball	0.054***	(0.006)	-0.023***	(0.002)	0.664	512,674
Incorrectly Called Strike	-0.073***	(0.004)	0.060***	(0.002)	0.042	512,674
Incorrectly Called Ball	-0.009***	(0.003)	-0.011***	(0.001)	0.004	512,674
Called Strike	-0.045***	(0.006)	0.035***	(0.002)	0.332	512,674
Umpire with MLB Experience						
Correctly Called	0.093***	(0.005)	-0.015***	(0.004)	0.897	1,205,489
Correctly Called Strike	0.053***	(0.006)	-0.009***	(0.003)	0.249	1,205,489
Correctly Called Ball	0.040***	(0.005)	-0.007***	(0.003)	0.647	1,205,489
Incorrectly Called Strike	-0.085***	(0.005)	0.020***	(0.005)	0.094	1,205,489
Incorrectly Called Ball	-0.007***	(0.002)	-0.004***	(0.001)	0.009	1,205,489
Called Strike	-0.032***	(0.006)	0.011***	(0.003)	0.343	1,205,489

Notes: All regressions for pitch-level outcomes include a vector of covariates at the pitch-level, year-by-month, umpire, league-by-year and team-by-year fixed effects. Standard errors are clustered at the umpire-level. "Robot Mo." indicate that robot is assisting umpires calling the game and "Post-Robot Mo." indicate that the umpire returned following robot-assistance. A pitch is correctly called if it crosses the strike zone and called strike or missed the strike zone and called ball. A pitch is incorrectly called if it misses the strike zone but called strike or crosses the strike zone but called ball. I restrict the sample to the Major League umpires and umpires treated in Triple-A implementation in 2022. Umpires' prior experience in the Major League is measured in 2021. Baseline mean is calculated using the data before 2022, the year of robot implementation for the Triple-A umpires. * $p<0.1$,** $p<0.05$, *** $p<0.01$.

Table A18: 2023 Triple-A Implementation Summary Statistics

	Full Sample		With Robot		Without Robot	
	Mean	SD	Mean	SD	Mean	SD
Pitch Characteristics						
Called Strike	0.310	(0.462)	0.308	(0.462)	0.311	(0.463)
Called Correctly	0.918	(0.275)	0.940	(0.237)	0.901	(0.299)
Residualized Accuracy	-0.0000	(0.2630)	0.0297	(0.2316)	-0.0220	(0.2820)
Horizontal distance	0.763	(0.490)	0.757	(0.488)	0.768	(0.492)
Vertical distance	0.839	(0.575)	0.836	(0.573)	0.842	(0.576)
Number of Games	2,126		888		1,238	
Number of Pitches	330,188		140,588		189,600	
Number of Umpires	71		64		70	
Dosage	0.420	(0.137)				

Notes: A pitch is correctly called if it crosses the strike zone and called strike or missed the strike zone and called ball. "Residualized accuracy" residualizes whether a decision was correct for pitch location and team-by-year fixed effect to account for pitch coordinates that may depend on calibration for each stadium and stringer plotting coordinates. Distances are in feet. Robot is used in games played between Tuesday and Thursday in both Triple-A leagues in 2023. "Dosage" is $\frac{\text{\# of games with robot}}{\text{\# of total games}}$.

Table A19: 2020 COVID Pause Summary Statistics

	Never-Returned Umpires		Returned Umpires	
	Mean	SD	Mean	SD
Pitch Characteristics				
Called Strike	0.334	(0.472)	0.333	(0.471)
Called Correctly	0.957	(0.203)	0.952	(0.213)
Residualized Accuracy	-0.0001	(0.1945)	0.0003	(0.2034)
Horizontal distance	0.745	(0.437)	0.738	(0.431)
Vertical distance	0.849	(0.575)	0.857	(0.575)
Number of Games	1,382		3,358	
Number of Pitches	193,254		465,809	
Number of Umpires	83		84	

Notes: A pitch is correctly called if it crosses the strike zone and called strike or missed the strike zone and called ball. Distances are in feet. "Never-Returned Umpires" are those who have not returned to call games since the COVID-19 pandemic and "Returned Umpires" are those who have returned from the COVID-19 canceled season. The statistics are from 2019 before the COVID-19 canceled 2020 season and from the Minor Leagues.

Table A20: Are Umpires Following the Other Guideline?

	Treated Umpires					
	Robot Mo.	S.E.	Post-Robot Mo.	S.E.	Baseline Mean	N
Pitch-level Outcomes						
Correctly Called	0.038***	(0.006)	-0.009***	(0.002)	0.932	8,864,667
Correctly Called Strike	0.013***	(0.004)	-0.006***	(0.001)	0.281	8,864,667
Correctly Called Ball	0.025***	(0.003)	-0.004**	(0.002)	0.651	8,864,667
Incorrectly Called Strike	-0.024***	(0.004)	0.007***	(0.001)	0.059	8,864,667
Incorrectly Called Ball	-0.014***	(0.002)	0.002***	(0.001)	0.008	8,864,667
Called Strike	-0.011***	(0.003)	0.002	(0.002)	0.340	8,864,667

Notes: All regressions for pitch-level outcomes include a vector of covariates at the pitch-level, year-by-month, umpire, league-by-year and team-by-year fixed effects. Standard errors are clustered at the umpire-level. "Robot Mo." indicate that robot is assisting umpires calling the game and "Post-Robot Mo." indicate that the umpire returned following robot-assistance. A pitch is correctly called if it crosses the strike zone and called strike or missed the strike zone and called ball. A pitch is incorrectly called if it misses the strike zone but called strike or crosses the strike zone but called ball. Accuracy is measured using the alternate strike zone implemented for the robot-assisted games on July 20, 2021. Baseline mean is calculated using the data before 2021, the year of first robot implementation. * $p<0.1$,** $p<0.05$, *** $p<0.01$.

Table A21: Do the Leagues Benefit from Robot Implementation?

	Treated Team					
	First Season	S.E.	Second Season	S.E.	Baseline Mean	N
Game-level Outcomes						
Attendance	265.2**	(123.1)	3.3	(121.2)	2135.2	6,826
Duration (in min.)	21.1***	(2.9)	15.5***	(2.9)	166.2	7,219
Num. of Pitches	11.2***	(2.7)	0.5	(3.6)	272.4	7,220
Total Score	1.1*	(0.5)	-0.1	(0.5)	8.5	7,220
Delay	-17.9	(24.3)	-16.8	(20.3)	64.2	845

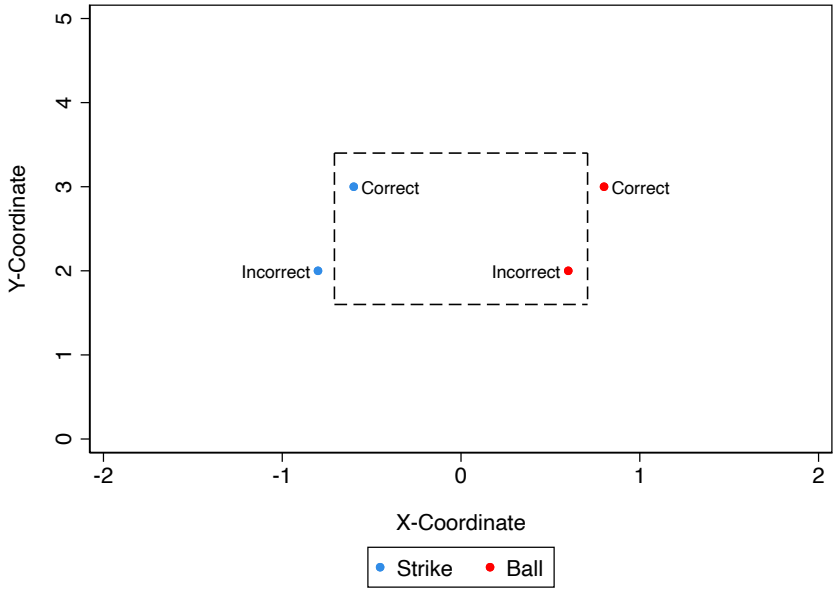
Notes: All regressions include year-by-month and league and team fixed effects and league specific time trend. Standard errors are clustered at the team-level. "First Season' indicate that it is the first season of robot implementation and "Second Season' indicate that it is the second season of robot implementation. The sample includes games from Single-A Florida and Carolina leagues. Baseline mean is calculated using the data from 2019, the season before robot implementation. * p<0.1,** p<0.05, *** p<0.01.

Table A22: Do Players Perform Better Following Robot Implementation?

	Treated Players					
	Robot Mo.	S.E.	Post-Robot Mo.	S.E.	Baseline Mean	N
Pitcher Outcomes						
Pitcher Strike Out	-0.057	(0.111)	-0.145*	(0.082)	4.650	118,876
Pitcher Walk Allowed	0.081	(0.072)	-0.026	(0.047)	1.908	118,876
Pitcher Hit Allowed	-0.111	(0.112)	-0.026	(0.060)	5.170	118,876
Pitcher Home Run Allowed	0.009	(0.040)	0.019	(0.018)	0.559	118,876
Batter Outcomes						
Batter Struck Out	0.031	(0.020)	0.040***	(0.014)	0.914	1,089,716
Batter Walked	0.032**	(0.014)	-0.011	(0.009)	0.393	1,089,716
Batter Hit	-0.011	(0.016)	-0.018**	(0.009)	0.938	1,089,716
Batter Home Run	0.009	(0.007)	0.003	(0.003)	0.099	1,089,716

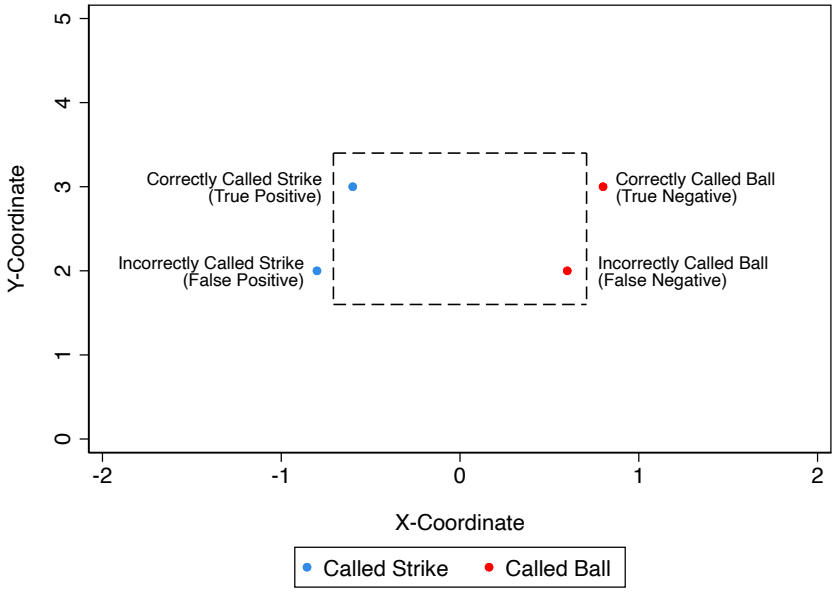
Notes: All regressions include year-by-month, umpire and player, league-by-year and team-by-year fixed effects. Standard errors are clustered at the player-level. "Robot Mo." indicate that robot is assisting umpires calling the game and "Post-Robot Mo." indicate that the players are playing in games called by umpires following robot-assistance. The outcomes are collapsed at the player-by-game level. I restrict the sample to pitchers who faced at least 15 batters in an appearance and batters with at least 3 at-bats. A batter's at-bat ends in a walk (or base on balls) when the count reaches 4 balls. Baseline mean is calculated using the data before 2021, the year of first robot implementation. * p<0.1,** p<0.05, *** p<0.01.

Figure A1: Strike Zone



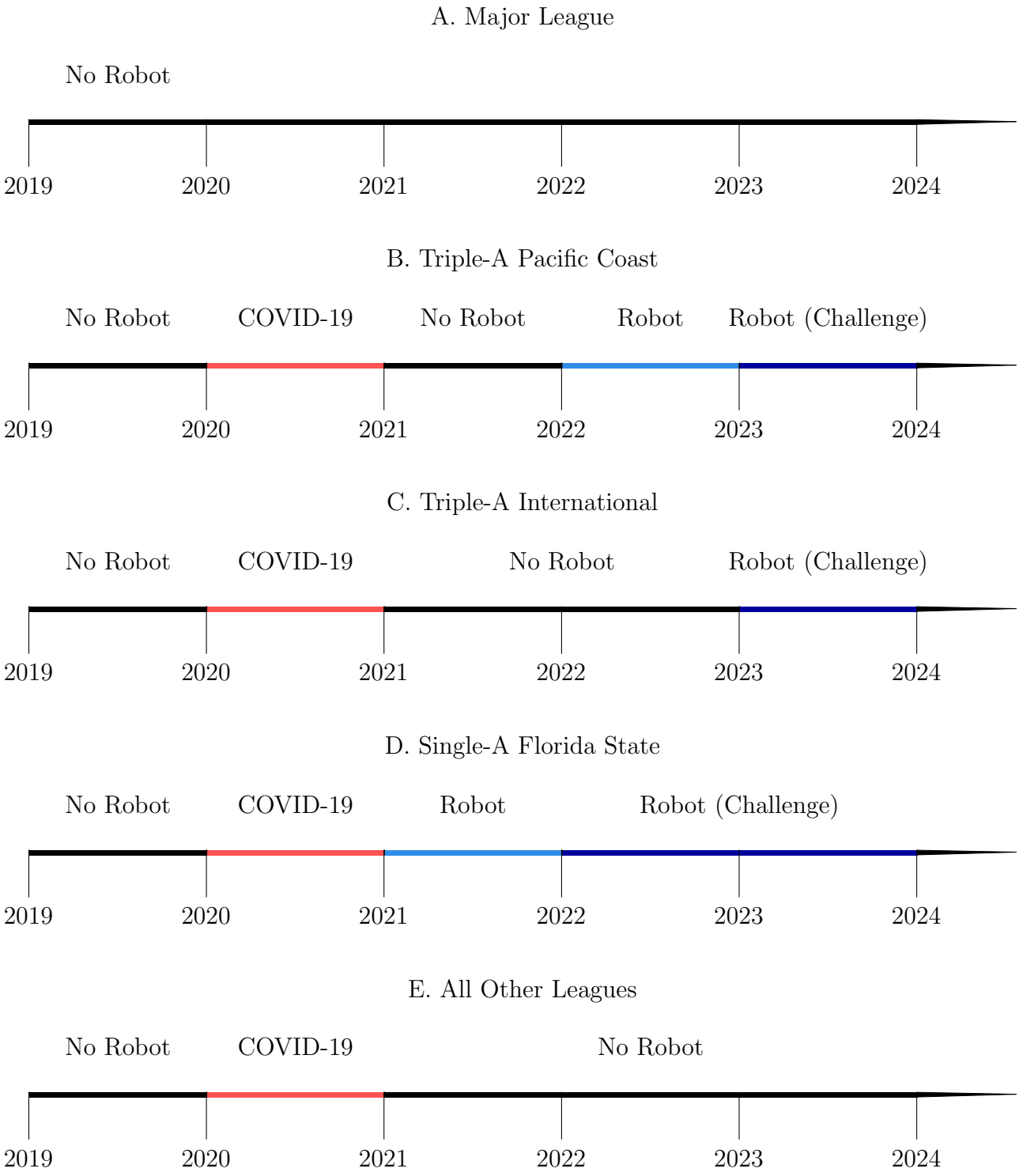
Notes: The figure describes the strike zone and the umpire's decision classification. Pitches that are called strikes and ball are in blue and red, respectively. A pitch is correctly called if it crosses the strike zone and called strike or missed the strike zone and called ball. A pitch is incorrectly called if it misses the strike zone but called strike or crosses the strike zone but called ball. The strike zone is defined as an imaginary rectangular region over the home plate that extends roughly between the batter's shoulders and kneecap and the dimension is about 20 by 25 inches. Axes are in feet.

Figure A2: Strike Zone - Robot Implementation Adjustment



Notes: The figure describes the strike zones. The red-dashed lines represent the adjusted strike zone implemented for the robot since July 20, 2021 and the black-solid line represents the original strike zone. The strike zone is defined as an imaginary rectangular region over the home plate that extends roughly between the batter's shoulders and kneecap and the dimension is about 20 by 25 inches. Axes are in feet.

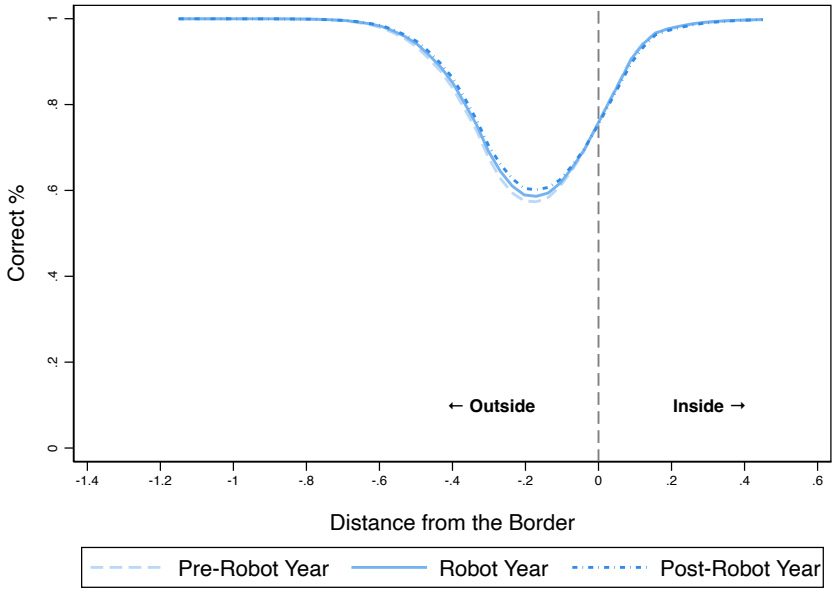
Figure A3: Chronology of Robot Implementation



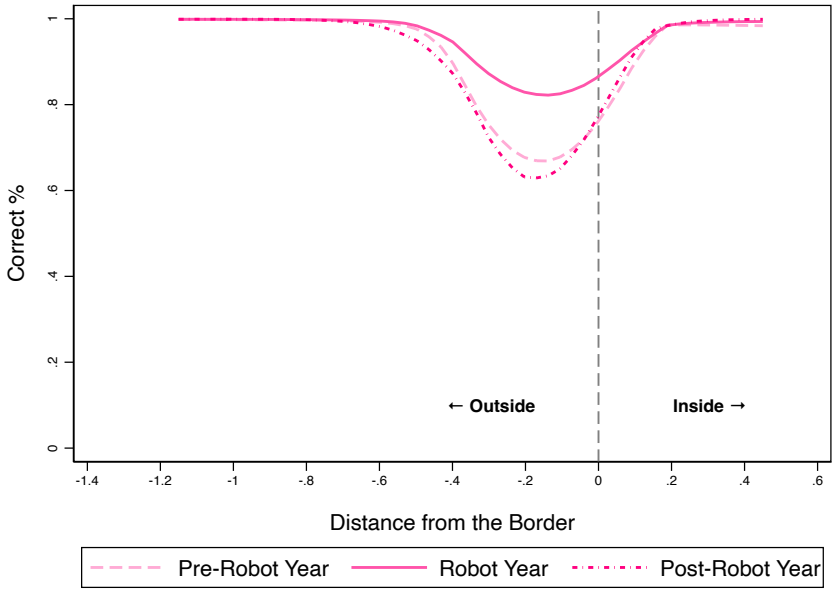
Notes: The figure shows timeline of the implementation of the robot in different leagues. COVID-19 pause is shown in red. "Robot" indicate that the league implemented the robot and "Robot (Challenge)" indicate that the challenge system described in Section A3 is used.

Figure A4: Pitch Distribution and Accuracy

A. Major League Umpires

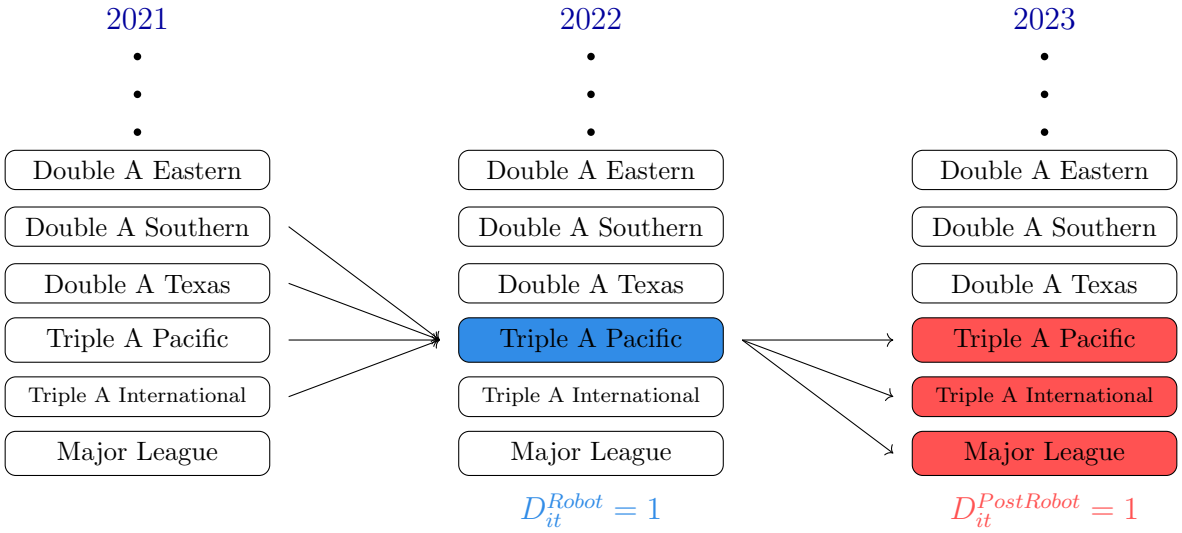


B. Minor League Treated Umpires



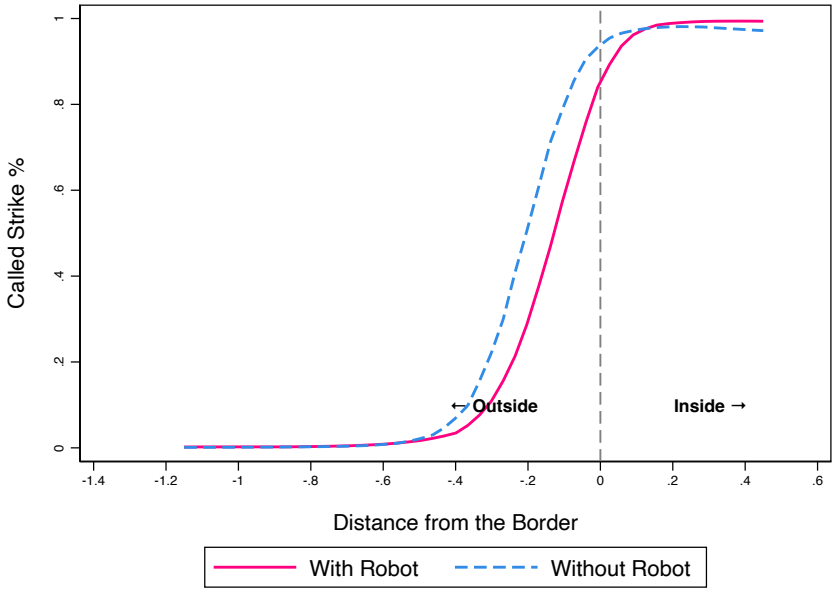
Notes: X-axis is the pitch distance from the nearest border of the strike zone in feet and y-axis shows the average accuracy rate. To the left of origin are pitches falling outside of the strike zone, and to the right are pitches falling inside the strike zone. Panel A shows the distribution of pitches for the umpires who worked in the Major League from 2021 to 2023 and Panel B shows for the umpires who worked in Triple-A Pacific in 2022 and games without robot in 2021 and 2023.

Figure A5: Robot Implementation and Umpire Moves



Notes: The figure describes potential lateral and vertical movements of umpires across leagues and robot implementation. In 2022, Triple-A Pacific Coast League implemented robot and 66 umpires (41 with more than 10 games with robot) are assisted by the robot. These umpires, in 2021, worked in Double-A leagues and Triple-A leagues. In the following year, in 2023, these umpires stayed in Triple-A and/or moved up to the Major League. In 2023, games held on Tuesdays to Thursdays used robot and games on Fridays to Sundays are called by umpires without robot. $D_{it}^{Robot} = 1$ indicates the games with robot and $D_{it}^{PostRobot} = 1$ indicates the games without robot in 2023.

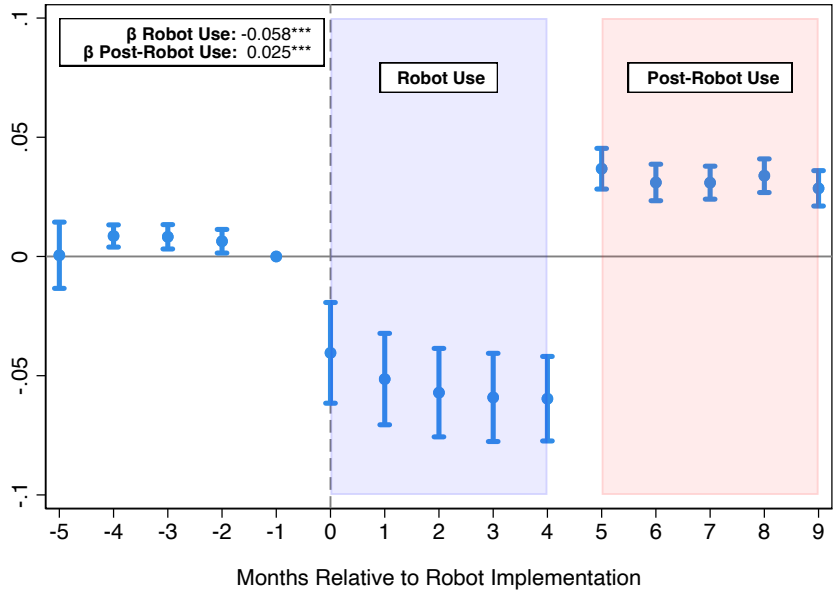
Figure A6: Pitch Distribution and Called Strike



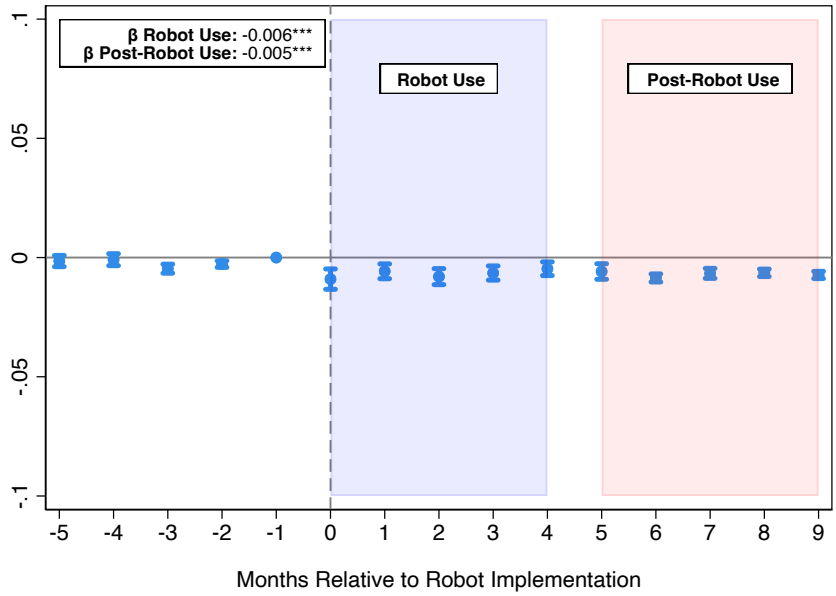
Notes: X-axis is the pitch distance from the nearest border of the strike zone in feet and y-axis shows the average called strike rate. To the left of origin are pitches falling outside of the strike zone, and to the right are pitches falling inside the strike zone.

Figure A7: Event-Study - Umpire Decisions

A. Incorrectly Called Strikes



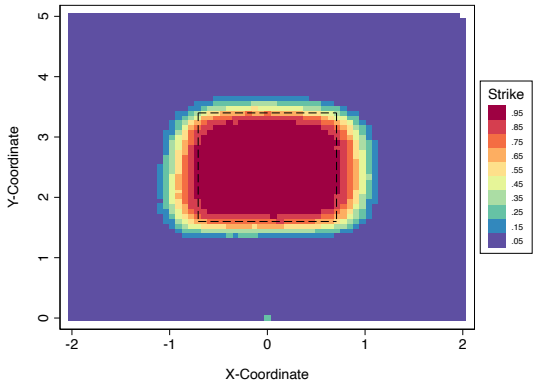
B. Incorrectly Called Balls



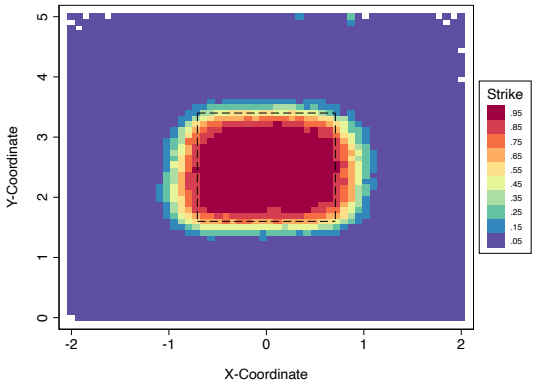
Notes: All regressions include year-by-month, umpire, league-by-year and team-by-year fixed effects. Standard errors are clustered at the umpire-level. "Robot Use" indicate that robot is assisting umpires calling the game and "Post-Robot Use" indicate that the umpire returned following robot-assistance. X-axis is months relative to the first month of robot implementation. A pitch is correctly called if it crosses the strike zone and called strike or missed the strike zone and called ball. A pitch is incorrectly called if it misses the strike zone but called strike or crosses the strike zone but called ball. * p<0.1,** p<0.05, *** p<0.01.

Figure A8: Called Strike Heatmaps - By Situations without Robot

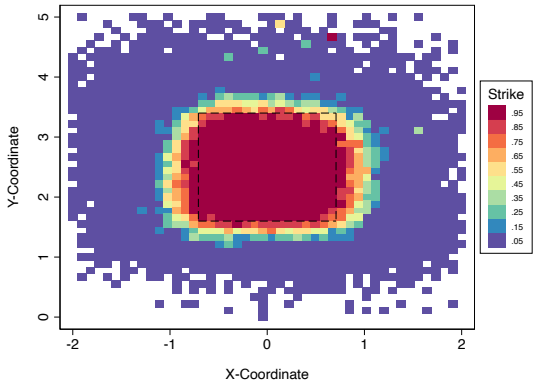
A. All Games



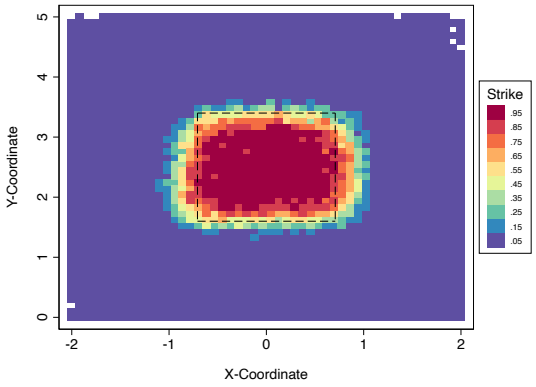
B. Previous Pitch was Called Strike



C. Count is "3-0"



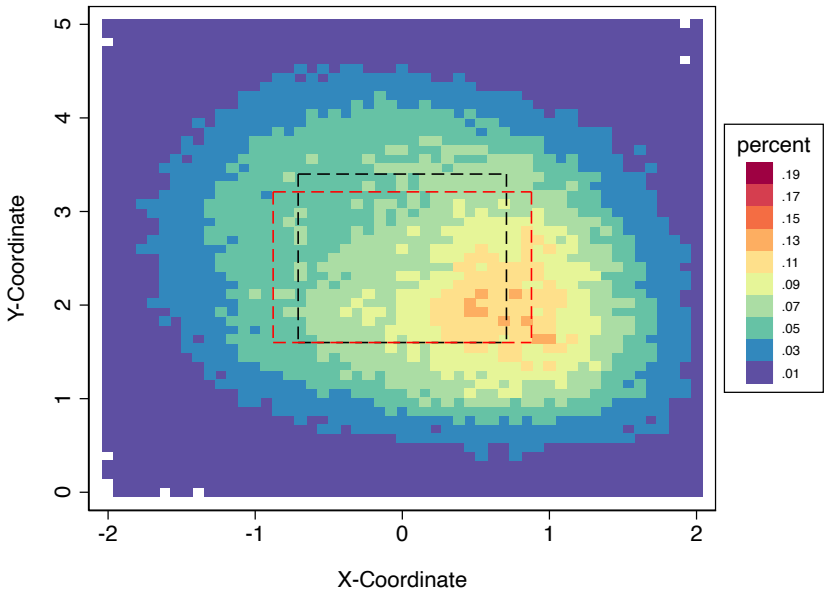
D. Count is "0-2"



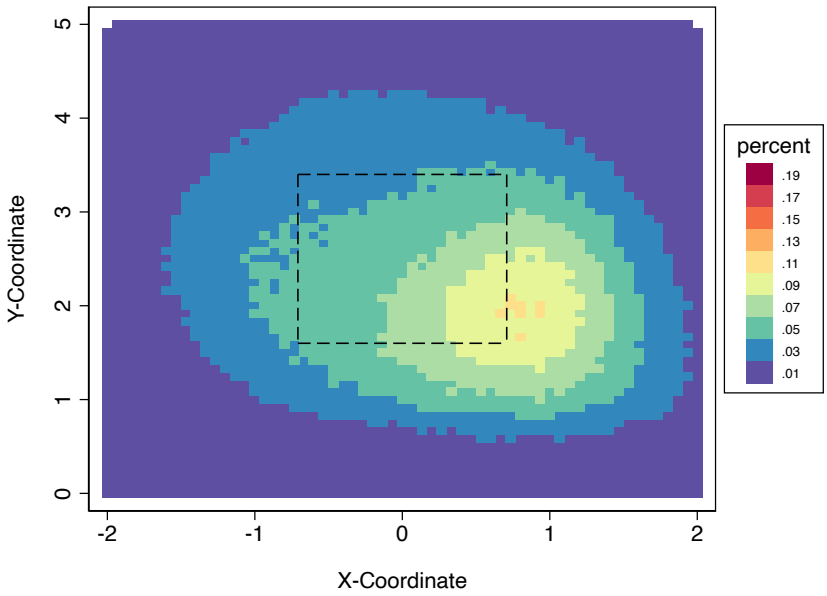
Notes: The figures plot the share of pitches that are called strike by pitch location and by game situations. The black dotted line shows the strike zone for an average batter.

Figure A9: Player Pitch Heatmaps

A. Pitchers Facing Umpire with Robot-Assistance



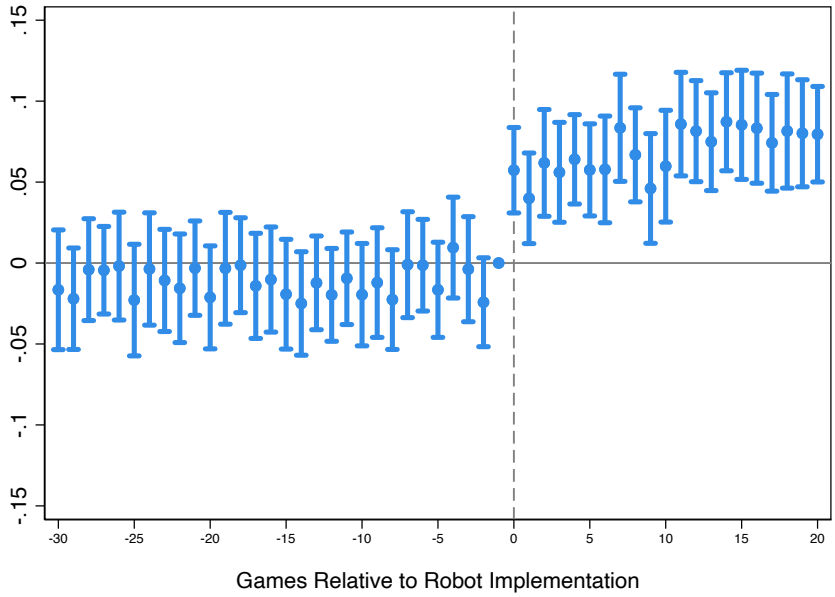
B. Major League Pitchers



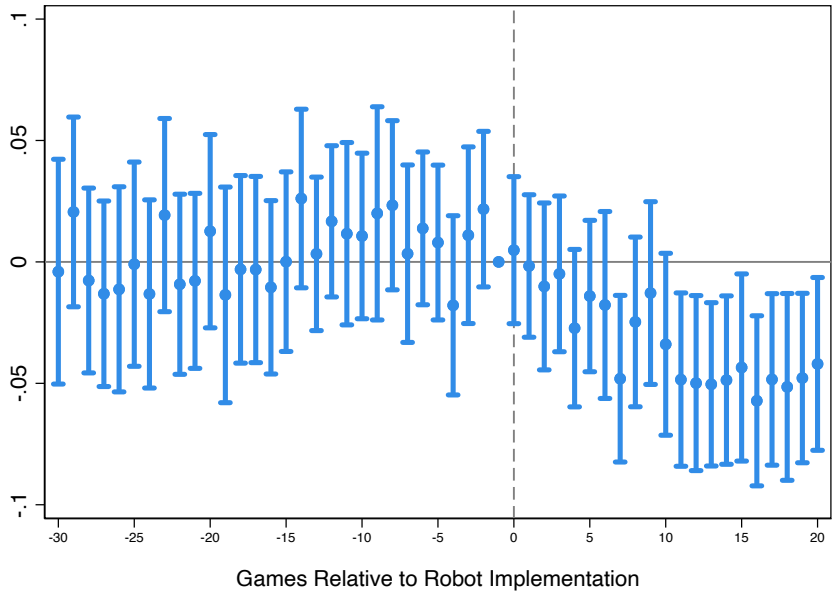
Notes: The figures plot the share of pitches by pitch location of pitchers facing right-handed batters. The black dotted line shows the strike zone for an average batter and the red dotted line in Panel A shows the adjusted strike zone that was implemented on July 20, 2021 for games with robot.

Figure A10: Event-Study - Pitchers Game-Level

A. Pitched Inside



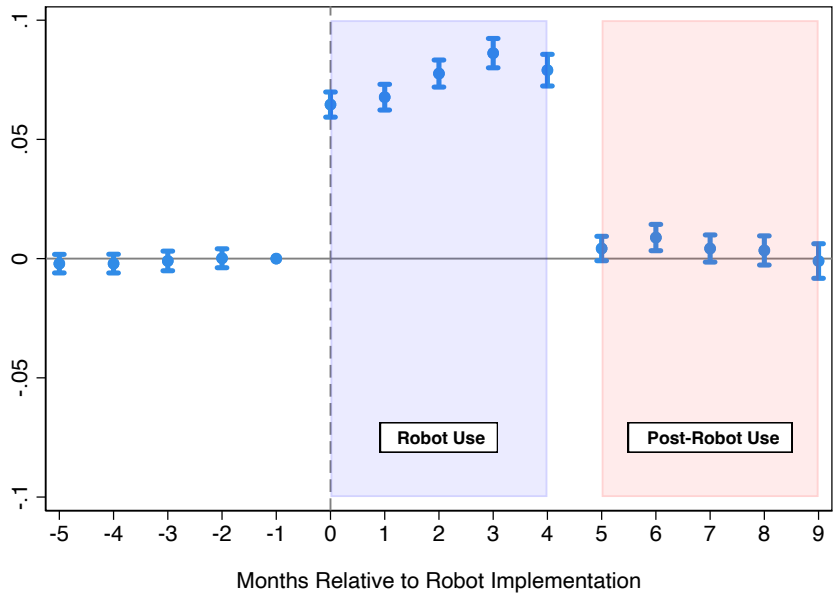
B. Distance from Center



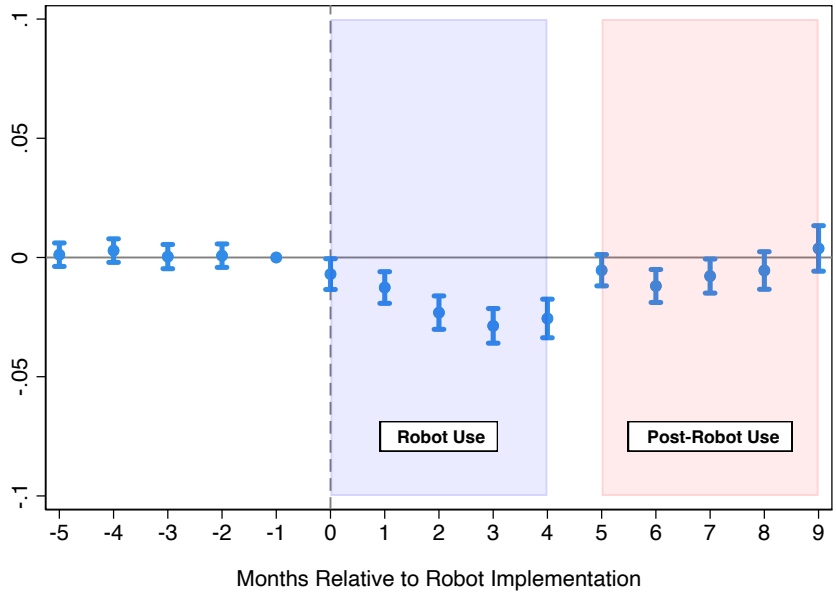
Notes: All regressions include a vector of covariates at the pitch-level (excluding pitch location controls), year-by-month, umpire and player, league-by-year and team-by-year fixed effects. Standard errors are clustered at the player-level. "Pitched Inside" indicates that a pitch falls inside the strike zone. Distances are measured in feet. X-axis is number of games relative to the last game with robot.

Figure A11: Event-Study - Pitchers Month-Level

A. Pitched Inside



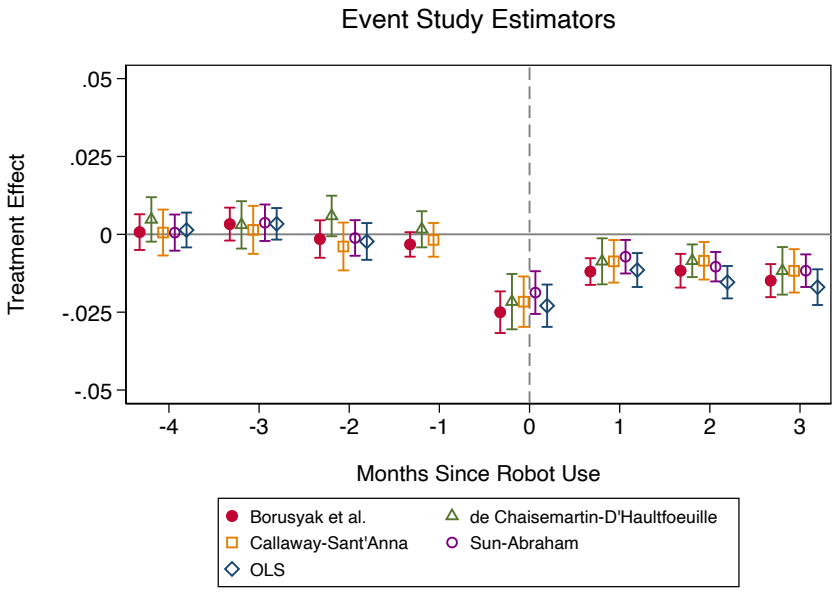
B. Distance from Center



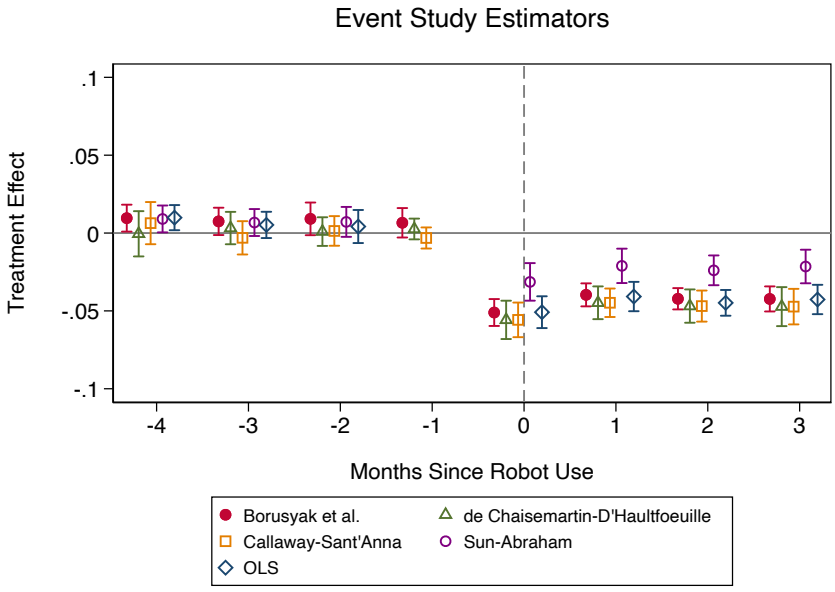
Notes: All regressions include a vector of covariates at the pitch-level (excluding pitch location controls), year-by-month, umpire and player, league-by-year and team-by-year fixed effects. Standard errors are clustered at the player-level. "Pitched Inside" indicates that a pitch falls inside the strike zone. Distances are measured in feet. X-axis is months relative to the first month of robot implementation.

Figure A12: Event-Study - Different Estimators

A. Residualized Accuracy



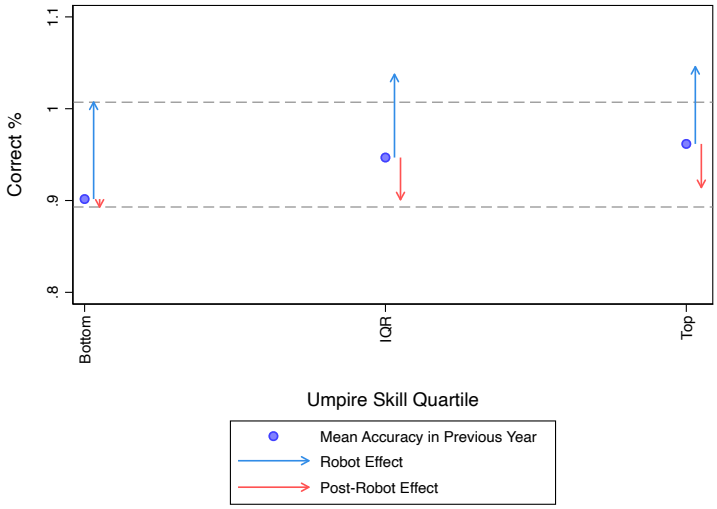
B. Accuracy



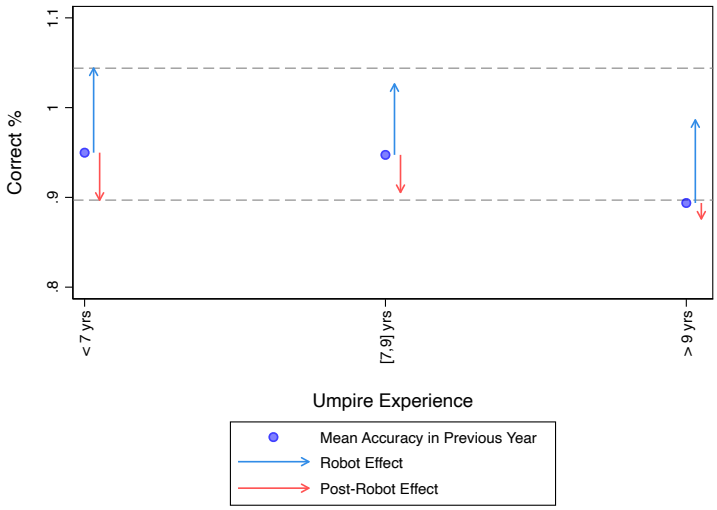
Notes: The figure plots coefficients from four different event study estimators and two-way fixed effects model (Borusyak et al., 2021; Callaway and Sant'Anna, 2021; De Chaisemartin and d'Haultfoeuille, 2020; Sun and Abraham, 2021). X-axis is months relative to the last month of robot implementation. A pitch is correctly called if it crosses the strike zone and called strike or missed the strike zone and called ball. The outcomes in Panel A and B are decision accuracy residualized for pitch location and fixed-effects and raw decision accuracy. The outcomes are collapsed at the month-level.

Figure A13: Does Skill Distribution Compress?

A. Skill



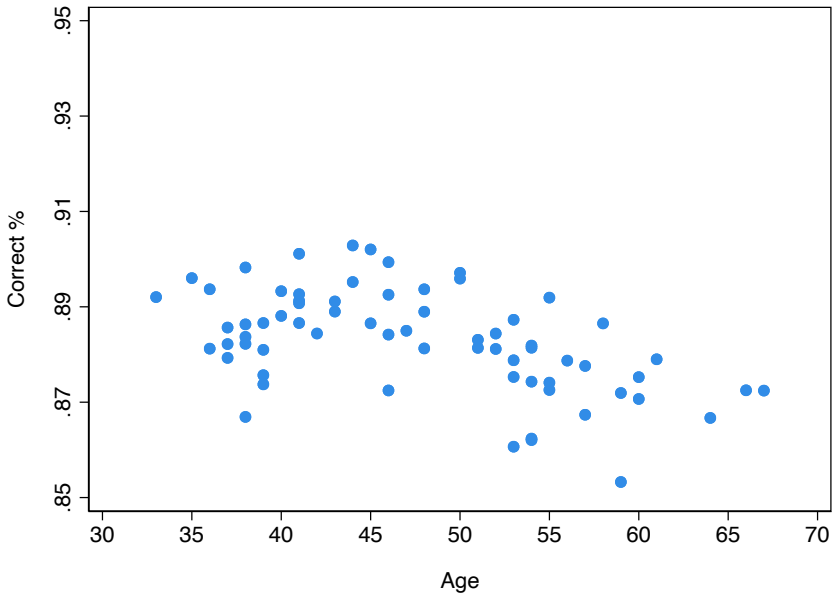
B. Experience



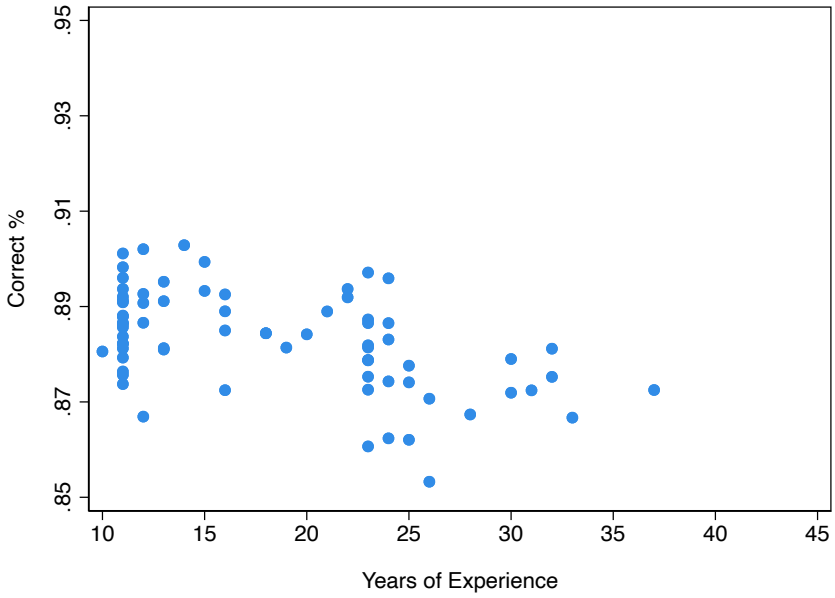
Notes: The figure plots the results in Table A14 and Table A16. All regressions include a vector of covariates at the pitch-level, year-by-month, umpire, league-by-year and team-by-year fixed effects. "Mean Accuracy" is measured in 2021. "Robot Effect" indicate the effect of robot assisting the umpire calling the game and "Post-Robot Effect." indicate that the effect of human umpire returning following robot-assistance. A pitch is correctly called if it crosses the strike zone and called strike or missed the strike zone and called ball. A pitch is incorrectly called if it misses the strike zone but called strike or crosses the strike zone but called ball. I restrict the sample to never-treated umpires and umpires treated in Triple-A implementation in 2022. Umpires' skill are measured in 2021 and divided into quartiles. Umpires' years of experience are measured in 2021.

Figure A14: Umpire Skill by Age and Experience

A. Age



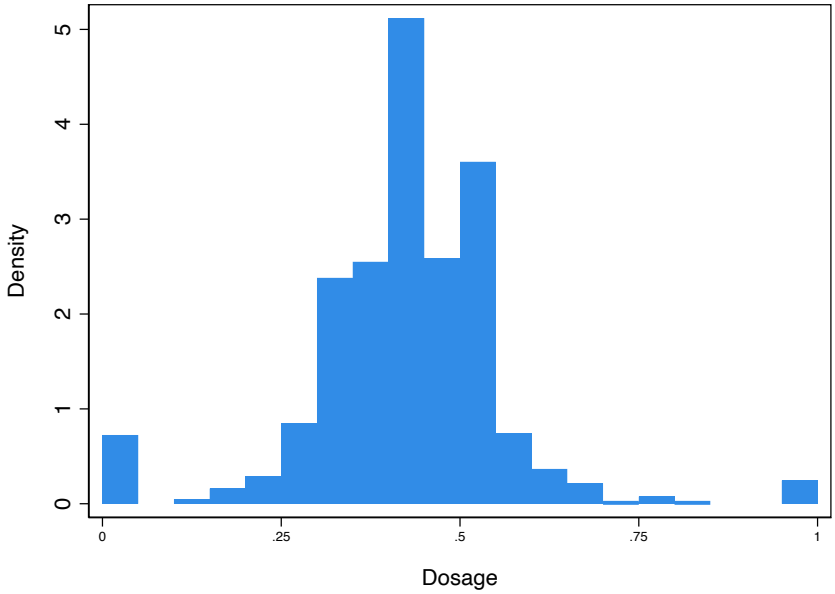
B. Experience



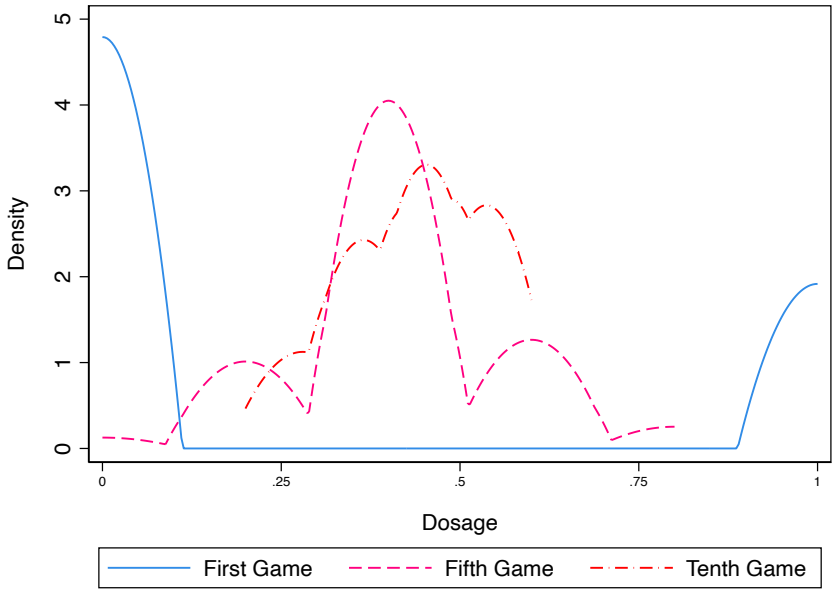
Notes: The figure plots the average accuracy of the Major League umpires in 2022 by age and years of experience.

Figure A15: Treatment Intensity Distribution

A. Full Sample

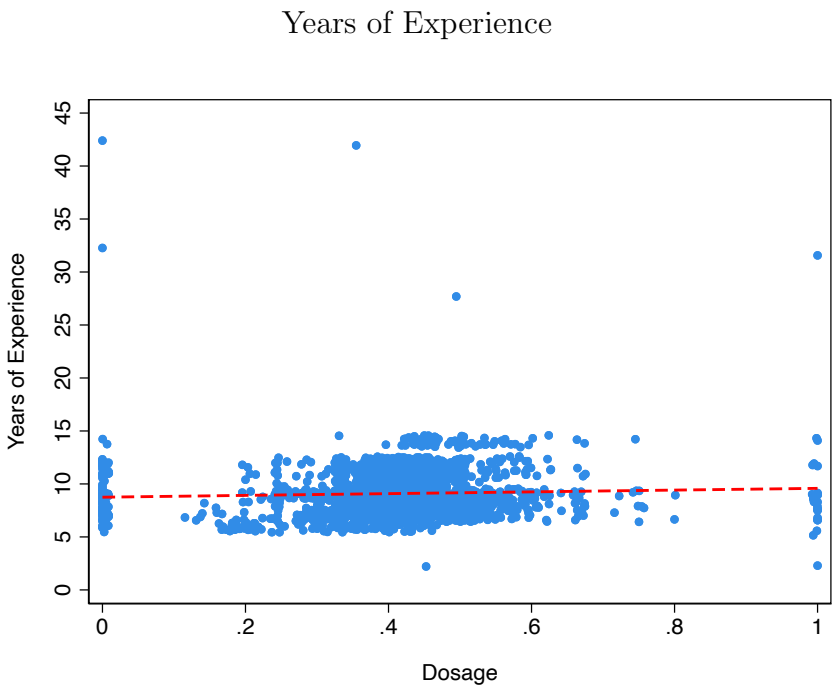


B. Distributions by the Number of Games Called



Notes: In 2023, games held on Tuesdays to Thursdays used robot and games on Fridays to Sundays are called without robot. "Dosage" is $\frac{\#\text{ of games with robot}}{\#\text{ of total games}}$. Panel A plots the dosage for full sample of Triple-A league games in 2023 at the umpire-game level. Panel B plots the dosage after the umpire's first, fifth and tenth game of the season.

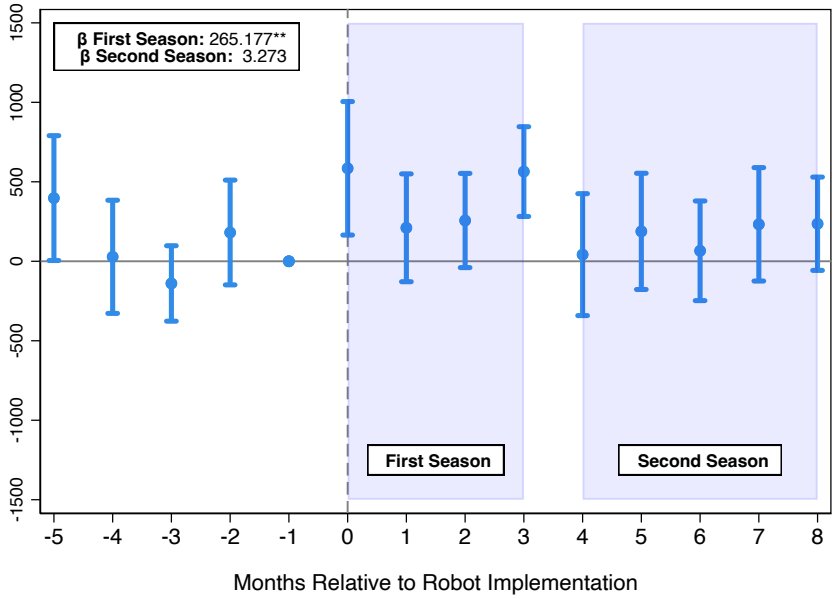
Figure A16: Are Dosages Quasi-Random?



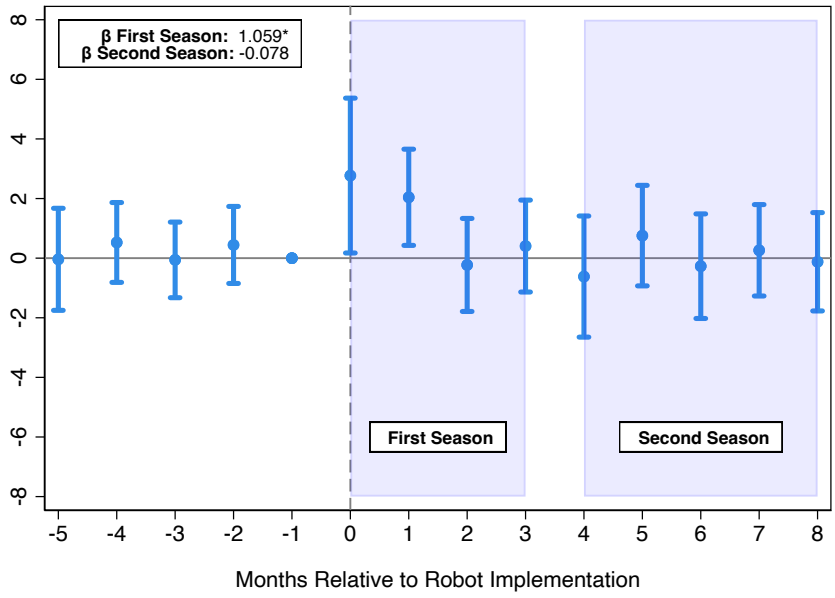
Notes: In 2023, games held on Tuesdays to Thursdays used robot and games on Fridays to Sundays are called without robot. "Dosage" is $\frac{\#\text{ of games with robot}}{\#\text{ of total games}}$. The figure plots the umpire's years of experience on the Y-axis and the dosage on the X-axis.

Figure A17: Event-Study - Game Outcomes

A. Attendance



B. Total Score



Notes: All regressions include year-by-month and league and team fixed effects and league specific time trend. Standard errors are clustered at the team-level. "First Season' indicate that it is the first season of robot implementation and "Second Season' indicate that it is the second season of robot implementation. The sample includes games from Single-A Florida State and Carolina leagues. * p<0.1,** p<0.05, *** p<0.01.

A2 Baseball Game Description

Baseball is often referred to as America's pastime and has captivated fans for over a century. In this section, I briefly describe a typical baseball game.

Setup A baseball game is played between two teams, each consisting of nine players. One team takes on the role of the batting team, while the other becomes the fielding team. The objective of the batting team is to score runs, while the fielding team aims to prevent runs. The field is divided into four bases arranged in a diamond pattern: first base, second base, third base, and home plate.

Game The game begins with a pitcher, standing on the pitcher's mound, throwing a pitch. The pitcher's objective is to throw the baseball to the batter while trying to prevent them from hitting it. The batter stands at home plate, ready to swing at any pitch that comes his way. The batter's goal is to hit the pitched ball and safely reach a base. To score a run, a batter must make a complete circuit around all four bases, returning to home plate.

A game consists of nine innings, with each team having the opportunity to bat and field. Each inning is divided into two halves: the top and bottom. The team with the most runs at the end of the nine innings wins.

Inning An inning refers to the period of play. A standard baseball game comprises nine innings, although extra innings can occur if the game is tied after the regulation nine innings. In the top half of the inning, the team (usually an away team) is designated as the "batting team" and gets its chance to score runs. Each player gets a chance to bat in a specific order ("batting lineup") until three outs are accumulated. At the end of the top half of an inning, there is a brief break as the teams switch roles. At the end of the bottom half of an inning, another short break occurs as the teams switch roles again. This pattern continues until nine innings are completed.

At-Bat An "at-bat" refers to the specific plate appearance of a batter during a game. Each at-bat involves a sequence of pitches thrown by the pitcher, and the count represents the number of balls and strikes during this plate appearance. The count consists of two numbers. The first number represents the number of balls, and the second number represents

the number of strikes. The count starts with 0 balls and 0 strikes. A ball is called when the pitcher throws a pitch outside the strike zone, and the batter does not swing. A strike is called when the pitcher throws a pitch inside the strike zone and the batter does not swing, or the batter swings and misses. If a batter accumulates four balls, he is awarded first base, a situation known as a "walk." If a batter accumulates three strikes, he is called "out."

A3 More Details about Robot Implementation

In 2021, the Major League successfully implemented an automated ball-and-strike system at Single-A Florida State League. The system expanded to the Triple-A Pacific Coast League in 2022 and the entire Triple-A in 2023. The technology is used in two different systems: full ABS in which the game is called with the robot and a "Challenge" system.

Challenge System In the Single-A Florida State League, in 2022, the Major League explored an alternative system called a "Challenge" system. In select games with the challenge system, umpires call pitches as before without the robot, but the players (pitcher, catcher and batter) can appeal these calls and check with the robot. Each team were given three challenges in a game and if an appeal is successful, then the team retains the challenge.[62,63] The challenge system expanded to Triple-A leagues in 2023 and was also used in select games.[64]

Data In the data, I observe the number of successful and unsuccessful challenges in these select games, but I do not know the exact pitch that resulted in a challenge. I observe the ultimate call on the pitch.

In an average game, the umpire calls about 150 pitches. The average number of challenges is about 4.4 per game and the average success rate is about 33%. Therefore, about 1.4 pitches are overturned in a game with the challenge system (or about 0.9%). The average accuracy in these games is 90.1%, but with 1.4 overturned pitches, the accuracy will be about 89.2%. The main result of the paper is that the umpires' accuracy declines following the implementation of robots. If anything, the challenge system biases me away from finding the negative results.

Sample Restriction To account for the alternating implementation of the challenge system and full robot-assisted games, I impose the following sample restrictions.

In the main difference-in-differences specification (Equation 1), I drop the games with the full robot system in the second year of the implementation. For example, if an umpire

[62] A challenge must be made within 2 seconds of the call.
[63] The average number of challenges is about 4.4 per game and the average success rate is about 33%.
[64] Games held on Fridays to Sundays used the challenge system.

worked in the Triple-A Pacific in 2022 and continued to work in the Triple-A Leagues in 2023, I drop the full robot-assisted games in 2023. This restriction is made to estimate the effect of robot implementation on human skills.

In the dosage model (Equation 3), I exploit the alternating system and keep all these games.

A4 Pitch Coordinates

The Major League provides detailed pitch-level data including the X and Y coordinates for every thrown pitch. There are two different coordinate systems: manually-plotted and pitch-tracking technology-plotted coordinates.

A4.1 Manually-Plotted Pitch Coordinates

The league hires "stringers" who digitally score games to provide data that are available on the MLB Stats API. Most stadiums hire about three stringers and each work about 25-30 games in a season. New stringers go through an 8-10 weeks training program and score practice games with the more experienced stringers before working alone. The main task of these stringers is to enter the results of every pitch and play. They work at the press box in each stadium with the Major League representative who oversees the job. In particular, to record pitch coordinates, the stringer "clicks" on the software and records where he sees the pitch go through the plate. These are, therefore, recorded in the pixel coordinate system.

A4.2 Technology-Plotted Pitch Coordinates

Starting in 2003, the Major League started installing technology that tracks pitches in Major League stadiums. "Questec" was used to measure umpires' performances, yet it was criticized for reliability.

In 2005, the league started installing cameras in stadiums to track pitches and completed in 2007. "PITCHf/x" used three cameras, strategically located in the stadium, to track pitches as it goes through the strike zone. Sportvision, the company that developed PITCHf/x, claims that the system tracks pitches with an accuracy of better than one inch.

In 2017, the league replaced "PITCHf/x" with "TrackMan" which uses cameras and Doppler radar to track pitches. And in 2020, "TrackMan" was replaced with "Hawk-Eye" which uses the optical-tracking and vision-processing system that enables tracking pitches with an accuracy of better than 0.1 inch.

The pitch coordinates from these systems are recorded as X and Y coordinates in feet.

A4.3 Transforming Coordinates

To create a dataset with a uniform coordinate system, I convert the manually-plotted pitch coordinates in pixels into feet. First, each stringer might have a unique bias that could depend on the vantage point. To account for this, I transform the coordinates at the team(stadium)-by-year level.[65] At the team-by-year level, I generate a heatmap of umpires' ball-strike decisions and find the corners of the strike zone. In particular, I find the region where the pitches are called strike more than 95% of the time and take the top, bottom, left and right points. Using these four points, I linearly transform the pitch coordinates into feet-unit as I know the four corners of the strike zone in feet.

All of the analyses include team-by-year fixed effects to account for these transformations unless otherwise mentioned.

[65] I do not observe who the stringer is for each game, so I cannot employ stringer-FEs.

A5 Players' Productivity

Table A22 displays per-game outcomes for players.[66] In the season with robot assistance, as they adjusted their behaviors, the number of walks increased for both pitchers and batters. This is likely due to the pitcher pitching closer to the edge and the batter swinging less (See Table A9). Batters, following robot exposure, become less offensively productive in the following season. They strike out 0.03 more and generate 0.02 hits less per game. Collectively, these findings suggest the players' productivity is negatively impacted by the robot adoption despite small change in their strategic behaviors. A potential explanation includes an adjustment in the prior season that hindered the growth of players relative to players who didn't have to adjust. Readjusting to inconsistent calls by readjusting strategies might have affected performances.

[66] I restrict to pitchers who faced at least 15 batters in an appearance and batters with at least 3 at-bats.