How Does AI Improve Human Decision-Making? Evidence from the AI-Powered Go Program*

Sukwoong Choi[a] Namil Kim[b,*] Junsik Kim[c] Hyo Kang[d]

July 2021

Abstract

How does AI improve human decision-making? Answering this question is challenging because it is difficult to assess the quality of each decision and to disentangle AI's influence on decisions. We study professional Go games, which provide a unique opportunity to overcome such challenges. In 2016 an AI-powered Go program (APG) unexpectedly beat the best human player, surpassing the best human knowledge and skills accumulated over thousands of years. To investigate the impact of APGs, we compare human moves to AI's superior solutions, before and after the initial public release of an APG. Our analysis of 750,990 moves in 25,033 games by 1,242 professional players reveals that APGs significantly improved the quality of the players' moves as measured by the changes in winning probability with each move. We also show that the key mechanisms are reductions in the number of human errors and in the magnitude of the most critical mistake during the game. Interestingly, the improvement is most prominent in the early stage of a game when uncertainty is higher. Further, young players—who are more open to and better able to utilize APG—benefit more than senior players, suggesting generational inequality in AI adoption and utilization.

Keywords: Artificial Intelligence, Technology adoption, Decision-making, Human capital, Professional Go players, AI adoption inequality.

---

* We gratefully acknowledge valuable comments from Neil Thompson, Abhishek Nagaraj, and Nur Ahmed. We also thank organizers of the 2021 NBER Economics of AI Conference, the 2021 Academy of Management Meeting, and the 2021 Strategic Management Society Conference for inviting this paper for presentation (all forthcoming). This paper was nominated by the SMS Strategic Human Capital Interest Group as the finalist for the Best Interdisciplinary Paper Award (winner decision pending). This paper was previously circulated under the title, "Strategic Choices with Artificial Intelligence." All errors are our own.
[a] MIT Sloan School of Management, MIT Initiative on the Digital Economy. sukwoong@mit.edu.
[b,*] Corresponding author. School of Management, Harbin Institute of Technology. namil.kim@hit.edu.cn.
[c] School of Electrical Engineering, KAIST. mibastro@gmail.com.
[d] Marshall School of Business, University of Southern California. hyokang@marshall.usc.edu.

1. Introduction

Artificial Intelligence (AI) is the newest general-purpose technology (GPT) in the foreseeable future (Goldfarb et al. 2020, Trajtenberg 2018). It has the potential to affect all aspects of the economy—including the ways we innovate, do business, and organize (Cockburn et al. 2019). One remarkable characteristic of AI is its ability to provide humans with high-quality predictions at relatively low cost and to automate a wide range of predictions (Agrawal et al. 2018). AI has already outperformed human professionals in many domains—including strategic gameplay (Silver et al. 2017), medical diagnosis (Wang et al. 2019), and new drug development (Wallach et al. 2015). Furthermore, as a GPT, the domains where AI outperforms humans are expanding at a fast pace (Agrawal et al. 2018). The rapid development and adoption of AI thus raise an interesting yet pressing question about how AI affects human tasks in these various domains.

Studies of AI and human capital, for instance, have examined the performance gap between humans and AI, highlighting AI's potential for replacing jobs (e.g., Acemoglu et al. 2019, Acemoglu and Restrepo 2020, Luo et al. 2019, Webb 2020). For example, AI's advancement is associated with changes in task description in job postings, leading to a significant decline in demand for certain skills (Acemoglu et al. 2019). Luo et al. (2019) also reveal that AI chatbots are as effective as experienced workers and four times more effective than inexperienced workers in engendering customer purchases.

The impact of AI is not limited to the substitution of jobs; rather, AI can complement and assist human skills. A number of recent studies have examined how AI complements human tasks at different skill and job levels (e.g., Cao et al. 2021, Cowgill et al. 2020, Kleinberg et al. 2017, Wang et al. 2019). For example, human analysts who use an AI-based program make more accurate stock price forecasts for financial investments than those who do not (Cao et al. 2021). AI solutions for recording patient conditions using standardized codes in medical charts also improved the productivity of medical coders who previously did the job manually (Wang et al. 2019). This evidence shows that human workers can increase their productivity and performance when they employ AI technology directly at work.

Despite abundant findings on the benefits of utilizing AI, less is known about whether AI can improve native human abilities in making decisions—even when the AI technology is not directly available at work. The improved performance resulting from AI's assistant roles does not necessarily mean that AI nurtured fundamental human capabilities. For instance, while a calculator advances the operator's calculation speed and accuracy, it may atrophy his or her arithmetic ability. Thus it is important to distinguish the instructional roles of AI from the assistant roles. We argue in this paper that the effect of AI could go beyond the assistant roles to instructional roles, training human professionals to make better decisions.

This study examines how AI enables human professionals to make better choices by improving their heuristics or everyday practices in decision-making. We ask (1) whether AI improves the quality of human decisions, (2) by what mechanisms is performance improved, and (3) how does the effect vary by differential access to and attitudes toward AI. Empirical studies in this area are challenging because of several difficulties: finding a context where AI can train human professionals (but does not perform the task directly); observing a decision (or a series of decisions) by humans and assessing the results; and disentangling AI's clout on such decisions. Furthermore, given that AI has progressed dramatically only recently, researchers have been constrained from examining the impact of AI by limited data availability (Seamans and Raj 2018).

To address these concerns, we study the impact of AI on professional players of the strategy board game, Go, which provides a unique opportunity to overcome these challenges. Over thousands of years professional Go players have accumulated knowledge, wisdom, and skills from playing Go games. Yet the introduction of the AI-powered Go program (henceforth, APG), which is far superior to the best professional player, suddenly changed how Go players learn and play the game. The historic Go match (AlphaGo vs. Sedol Lee) was held in 2016; in this game, AI beat the best human professional player for the first time and by a large margin. Shortly after this event, the first open APG, Leela, became available to players in February 2017. Our quantitative and qualitative investigation indicates that professional Go players have used APGs heavily in their training since its release.

The great advantage of this context is that it allows us to observe every single decision of professional Go players before and after the public release of APGs; a game's entire move history is well archived and maintained for all major games. Furthermore, using the APG's best solution as a benchmark, we can calculate the probability of winning for every move (i.e., 750,990 decisions) by 1,242 professional Go players in 25,033 major games held from 2015 through 2019; note that this can be done even for the games played before APG's release. We then compare the move-level probability of winning to that of APG's best solution.

The results show that the quality of moves by professional Go players improved substantially following the release of APG. Before the release, the winning probability of each move by professional Go players averaged 2.47 percentage points lower than the moves of APG. This gap decreased by about 0.756 percentage points (or 30.5 percent) after the release of APG. Additional analyses indicate that the improvement in move quality eventually leads to the final win of the game. Interestingly, this effect is most prominent in the early stage of a game where higher uncertainty is exhibited and there is more opportunity for players to learn from AI. Furthermore, quality improvement is more prominent among young players who are open to and capable of utilizing APGs; this has important implications for digital literacy and inequality in AI utilization.

We also explore the mechanisms through which professional players achieve a higher probability of winning. Our mediation analysis reveals that improvements in the quality of moves are driven mainly by reducing the number of *errors* (moves where the winning probability drops by 10 or more percentage points compared to the immediately preceding move by a focal player) and by reducing the magnitude of the most *critical mistake* (the biggest drop in winning probability during the game). Specifically, the number of errors per game decreased by 0.15–0.50 and the magnitude of the most *critical mistake* decreased by 4–7 percentage points.

To the best of our knowledge, this study is one of the initial studies to provide micro-level evidence of the instructional role of AI in human decisions and performance. Our empirical analysis of 750,990 moves in Go games show that AI-trained professionals substantially improve the quality of their moves and increase their probability of winning through reducing errors and the magnitude of mistakes.

Importantly, the decisions made in games are similar to those made by managers and policymakers in that both use similar intuitive techniques under uncertainty and time constraints, as highlighted in the literature (e.g., Bechara et al. 1997, Mintzberg 1987, 1994, Miric et al. 2020, Simon 1987, Simon and Chase 1988). Our findings therefore have meaningful implications for the complementary and instructional role of AI, notably how AI could nurture decision-making by business managers and policymakers. The improvements brought by APG show that even the best human professionals have biases or heuristics in decision-making and that AI can train and educate human decision-making in a fast-paced, uncertain environment. The power of this improvement is significant to the extent that AI overturned an area of human knowledge and wisdom accumulated over thousands of years. Further, the fact that the young benefited more from the AI-powered program has important implications for potential inequality in accessing, adopting, and utilizing AI.

The paper proceeds as follows. Section 2 discusses the related literature on human decision-making and the roles of AI. Section 3 explains the data, research design, and identification strategy used for empirical analysis. Section 4 presents the results and robustness checks. Section 5 provides further analyses on the mechanisms and heterogeneity. The discussion and conclusion are provided in Section 6.

2. AI and Decision-Making

2.1. The Impact of AI on Human Decision-Making

When making decisions, humans tend to draw on their conceptualization of the future as input into the decision-making process (Lindebaum et al. 2020; Mintzberg 1987, 1994). Humans also depend on knowledge of causality, which they actively develop to understand how past actions impact future outcomes. Through these processes, humans can judge and learn from situations—even unexpected situations—to improve their decision-making processes and outcomes (Lindebaum et al. 2020, Mintzberg 1994). However, individuals are limited in their ability to process information, which slows down learning and limits its scope (Cyert and March 1963; Galbraith 1974; Simon 1955, 1958), and which in turns leads to failure to optimize decision-making (Kalberg 1980). For instance, managers' choices are often affected by rigidity to change and other routines, which lead to learning myopia (Levinthal and March 1993).

Acknowledging these limitations, researchers have studied incompleteness in managerial decision-making (Dane and Pratt 2007, Eisenhardt 1990, Shepherd et al. 2015). Managers typically predict possible options by collecting and evaluating all relevant information and making decisions they perceive will best maximize their economic returns. Even when managers must make very significant decisions for firms, they often fail to follow the procedures for rational choices (Mousavi and Gigerenzer 2014, Simon 1955). Rather, they rely on their heuristics, a simple decision-making process that utilizes only a fraction of the available information (Bingham et al. 2007, Bingham and Haleblian 2012). Research on managerial decision-making has also shown that making consistently optimal decisions is difficult due to bounded rationality (Simon 1991), cognitive biases (Thaler 1993), or perceptions deviating from economic optimality (Kahneman 2003). To mitigate these biases and errors, researchers propose to set goals and aspirations to guide decision-making and to use backward- and forward-looking decision models (Chen 2008, Cyert and March 1963, Gavetti and Levinthal 2000). However, benefits of these choice models are marginal in alleviating the aforementioned limitations to optimal decision-making.

Information technology (IT) literature provides yet another set of solutions and argues that the adoption and utilization of new technologies compensate for these shortcomings. Information theory (e.g., Blackwell 1953) and the information-processing view of the organization (Galbraith 1974) propose that the more accurate and precise the information used in decision-making, the higher the firm performance. This is primarily because IT improves a firm's ability to collect, analyze, and process information for internal operational decisions. Specifically, IT complements organizational practices, which in turn leads to higher productivity (e.g., Bapna et al. 2013, Bresnahan et al. 2002, Brynjolfsson and Hitt 2000). The positive relationship between the volume and quality of information and optimal decision-making has been supported by a plethora of studies (e.g., Ayres 2007, Brynjolfsson et al. 2011, Davenport and Harris 2017, Loveman 2003).

As data availability has grown, researchers have extended these arguments to data-driven decision-making. The data about consumers, suppliers, competitors, and partners and the utilization of large-scale analytics have supported managerial decision-making (Brynjolfsson et al. 2011, Wu et al. 2019). For example, Brynjolfsson et al. (2011) find that the adoption of data-driven decision-making practices increases financial returns. Saunders and Tambe (2013) reveal that firms with data-driven decision-making at an executive level have higher productivity and market valuations. Data analytics also support decision-making for R&D search and incremental process improvements (Wu et al. 2020). Overall, the adoption of new IT plays an important role in decision-making at both organizational and individual levels.

Researchers have recently extended this discussion to the adoption and utilization of AI. The advance in AI with the development of machine learning and deep-learning algorithms contributes to the avoidance of mistakes and errors stemming from human judgments (Danziger et al. 2011). AI algorithms are fundamentally different from previous algorithmic technologies for several reasons (Agrawal et al. 2018, Smith 2019). First, AI improves performance through self-learning, making inferences on new data based on prior learning. Because of the ability to discover hidden patterns, AI can conduct insightful tasks that need human-like "intuition." Second, AI obtains predictions and judgments with high accuracy and their accuracy increases with the number of training sessions and the quantity of voluminous data. Third, models using AI algorithms could improve over time to achieve a superior performance. These distinct characteristics enable AI to outperform humans not only in repetitive work and recognition tasks but also in creative tasks in some domains (He et al. 2015, Mnih et al. 2015). Researchers find that AI performs well even in high-level cognitive tasks such as making a legal decision in the court (Kleinberg et al. 2017), discovering protein structure in biology (Senior et al. 2020), and playing strategic games (Schrittwieser et al. 2020), among others.

Organizations therefore have begun to use AI algorithms for tasks requiring accurate judgment, improving the allocation of valuable resources, designating work schedules, and analyzing employee performance (De Cremer 2020, Rock 2020). For example, AI algorithms are helpful in new drug development—in particular, at the early stage when the tasks are heavily dependent on automatic data processing and pattern recognition (Lou and Wu 2021). Medical coders in hospitals also use AI suggestions about chart coding and thereby improve their productivity (Wang et al. 2019).

Considering the assumption of bounded rationality—that decision-makers tend to balance the quality of their decisions with the cost, such as the cognitive effort and time required to reach their decisions (Kahneman 2003)—AI can contribute to lowering cost, which in turn rebalances the accuracy of decisions. That is, AI helps the process of human decision-making by evaluating a broader scope of options at a lower cost and by performing a more accurate evaluation of the options available. Since AI can provide an extensive set of precisely assessed alternatives that informs and trains human professionals, it can revisit their decision-making practices (which may have yielded inferior decisions if they were not trained with AI). Our main prediction, therefore, is that AI can train human professionals and improve the quality of their decisions—especially when the performance of AI is superior to that of humans and when tasks are complex and uncertain.

2.2. Differential Adoption and Utilization of AI by Age

Studies suggest that age is an important factor in adopting new technology (Weinberg 2004). Age could affect the adoption and utilization of AI through three major channels: risk-aversion, search behavior, and absorptive capacity. First, junior (younger) professionals are more open to new technologies and less risk-averse than senior professionals. Hambrick and Mason (1984) argue that an executive's age affects strategic decisions; for example, younger managers pursue riskier strategies in terms of the usage and development of new technology (Tyler and Steensma 1998). Most products powered with AI have come onto the market recently. Although superior performance is expected from it, the product's credibility and stability of performance have not yet been verified at the initial stage, so the public is not convinced of its usefulness. Because there is no track record, people making career-determining decisions are more likely to view utilizing the AI-powered product as riskier. In such situations, junior professionals are more inclined to utilize AI in their decision-making than are senior professionals.

Second, senior professionals tend to rely more on the knowledge and experience they have accumulated than do juniors. In other words, senior professionals conduct exploitative searches and make choices that are path-dependent on past records. In contrast, junior professionals have less experience and are more likely to explore and make path-independent decisions—that is, to adopt and utilize AI in training and decision-making.

Third, younger professionals have better *absorptive capacity* (Cohen and Levinthal 1990) for new technology; the young are better able to recognize the value of new technology, to assimilate it, and to apply it to their professional tasks. Empirical evidence indicates that younger workers are more qualified and more likely to adopt new information and communications technologies (de Koning and Gelderblom 2006, Meyer 2011, Morris and Venkatesh 2000, Schleife 2006).

All the above arguments suggest that the impact of AI on human decision-making will have differential effects on young and old professionals and that junior professionals benefit more from this new technology.

3. Empirical Strategy

"In a short space of time, AlphaGo Zero has understood all of the Go knowledge that has been accumulated by humans over thousands of years of playing … it's actually chosen to go beyond that and discovered something that the humans hadn't even discovered in this time period."

– David Silver, Google's DeepMind Lab (Kennedy 2017)

3.1. Setting

3.1.1. The Game of Go

Go (or Baduk) is a two-player strategy board game that originated in China at least 3,000 years ago. The board consists of nineteen lines by nineteen lines. Players compete to obtain more of the board's territory by alternating the placement of stones at the intersection of the lines. The professional Go industry is substantial—especially in China, Japan, South Korea, and Taiwan. Each country holds more than ten major professional tournaments, held throughout the year and sponsored by large corporations. For example, the Kisei tournament in Japan—held annually since 1977 and sponsored by the Yomiuri Shimbun newspaper—awards 4,500,000 yen (or $413,000) to the first-place winner in addition to per-game compensations.[1]

Demis Hassabis, head of the Google DeepMind team, noted that "Go is the most complex and beautiful game ever devised by humans … the richest in terms of intellectual depth." (Knight 2016) Go has about 250^{150} possible moves, and the search space is often described as "a number greater than there are atoms in the universe" (Silver et al. 2016)[2] The seemingly unlimited number of possible moves in Go cannot be exactly identified by brute force calculation (as supercomputers have done with chess); in the past two decades, several Go software programs—such as GnuGo, Pachi, and Crazy Stone—were released, but the performance of these programs was far inferior to that of professional Go players who use superlative "intuition" and evaluation skills in making certain moves (Knight 2016).

3.1.2. AI's Entrance into Go

Even if the latest supercomputers cannot calculate all possible moves in Go, the recent advancement in deep-learning algorithms in AI has made a remarkable improvement. Instead of evaluating all possible solutions, AI uses deep learning to reduce the potential moves to be considered and predicts sequential outcomes and winning probabilities. AlphaGo, the initial APG with these algorithms, was invented by Google DeepMind. After several quality tests, Google held a historic Go match in 2016 between AlphaGo and the human Go master, Sedol Lee. Prior to this match, Lee and other Go experts expected that Lee would sweep all five games. Yet AlphaGo beat Lee 4–1, "a feat previously thought to be at least a decade away" (Silver et al. 2016). This event has been described as one of the milestones in the history of AI (Press 2021).

[1] Other examples of major competitions include the Nongshim Cup, the competition between Team China, Japan, and South Korea, which awards $450,000 to the winning team. The Ing Cup (also known as Go Olympics), is held every four years and awards $400,000 to the winning player. In 2020, Jinseo Shin, a twenty-one-year-old from South Korea, earned $920,754 in award money; Imaya Yuta,a thirty-year-old from Japan, earned $1,179,456.

[2] For comparison, chess has about 35^{80} possible moves. After the first two moves, chess has 400 possible next moves, while Go has 130,000 possible next moves (Muoio 2016).

The result shocked not only Go players but also the public who believed computers to be far interior at intuitive judgments made amid enormous complexity. The match suddenly and unexpectedly demonstrated that AI-powered Go software surpassed the best human Go player. The match completely changed how professional Go players learn and practice Go; since the release in 2017 of public APGs—such as Leela Zero, KataGo, and Handol—all professional players have learned from APGs (Somers 2018). Figure A.1 in Appendix A shows a snapshot of a Go game between two professional players and how APG analyzes the game, illustrating how the winning probability changed with each move made (on the upper left corner) and the winning probability for potential next moves (on the main board).

3.1.3. How Much Better at Go Is AI Compared to Humans?

Go players are ranked and evaluated by the Elo rating system.[3] Figure 1 shows how Elo scores have evolved among Go programs. Non-AI Go software—GnuGo, Pachi, and Crazy Stone—scored under 2,000. The best human players scored around 3,800. In contrast, the scores of recent APGs based on deep reinforcement learning far exceeded 4,000. Given this gap in the Elo ratings, even the top professional Go players have no chance of winning against APGs. Ke Jie, who ranked second in the 2020 World Go Ranking, admitted that "AlphaGo is more like the god of Go" (Mozur 2017). Put differently, the moves by APG yield the highest probability of winning and even the best professional Go player could learn a lot from APGs.

"Insert Figure 1 here"

3.1.4. How Does Go Resemble Human Decision-Making?

The decisions made in each move in Go share many aspects with managerial decision-making in a complex, competitive environment. Studies have emphasized the link between decision-making in games and real businesses. For example, Simon (1987) argues that the intuitive skills required for managers are similar to the intuitive skills of chess masters. Mintzberg (1973, 1990) empirically finds that half of the activities of chief executives last less than nine minutes. Managers perform an average of 583 activities per eight-hour shift and perform one activity every 48 seconds (Guest 1956). These executives and supervisors rely on intuitions and routines when making decisions (Mintzberg, 1973). Likewise, professional Go players face a complex, competitive environment and are forced to make a series of important decisions within a strict time restriction; after using up the regular time, they must make each move within thirty seconds.

Executives and practitioners also confirm this point. John Koo, chairman of LS Future Center and a pan-LG group family member, noted that "Go is a battle that starts out from a small part of the board and later expands to the entire board. You need to make your move while seeing the bigger picture from the

[3] The Elo rating is calculated based on the relative capabilities of two players and their game outcome. The system has been widely used in other sports such as chess, football, basketball, and soccer.

very beginning. Business management is the same" (Korea Herald 2014). Also, LG Economic Research Institute (2004) published a report, *Learning Business Strategy from the Principles of Go*, highlighting that both executives and professional Go players must make decisions ceaselessly under uncertainty.

3.2. Research Design

We compare changes in the quality of moves by professional players around the first public release of APG. Although AlphaGo was the first APG to beat the best professional Go player, in 2016 only a scientific article about its algorithm—not the program itself—was available to players. The first public APG that performed at least as well as the best human player was Leela with its February 2017 update that utilized the deep-learning algorithm used in AlphaGo. A few months later, a new version, Leela Zero, was developed after AlphaGo Zero. Leela and Leela Zero gained wide attention from media as well as professional players upon their release and are viewed as the world's most successful open source Go engines based on AI (Somers 2018). The impact of Leela—which provided a set of best possible moves with the winning probability of each alternative—was substantial for Go players. We describe how Leela and Leela Zero work in Appendix A.

Importantly, the development of APGs did not come from the demand of or request by Go players; before AlphaGo, Go programs could only play at the level of human amateurs, and professional players did not believe that computer programs could ever beat professional Go players. DeepMind, the developer of AlphaGo, decided to develop the Go program solely because Go is profoundly complex (Burton-Hill 2016). The developer of Leela, Gian-Carlo Pascutto, also made it clear that he wanted to learn how deep learning works (but had no interest in playing Go). AI's entrance into Go, therefore, is not correlated with preexisting conditions in the Go industry.

We first use the event-study method to estimate the impact of APG on the quality of moves by professional Go players. The event of interest is the major update of Leela in February 2017 that adopted the AlphaGo-based deep-learning algorithm and outperformed the best human player. We conduct the analyses at the player-game level. Our sample consists of major professional Go games held from January 2015 through December 2019.

We focus primarily on early moves—the first 30 moves for each game—because, like many other games, starting with a great opening is critical to winning at Go. Chang-ho Lee, a once-in-a-century player, pointed out the importance of the opening and likened it to a blueprint for architecture; the opening strategies are general roadmaps and ideas on how players lead the game (Noh 2016). As such, professional Go players accumulate their knowledge and wisdom particularly for the first thirty moves. Several opening sequences have been long established as standard procedure (called *joseki)*; players then deviate from these standards as the game proceeds beyond the thirtieth move. The early stage of the game, therefore, is where APGs can best challenge the knowledge accumulated historically by professional players. We also analyze later stages and compare the results.

3.3 Data

3.3.1. Go Games and Players

We collect data on professional Go games held from 2015 through 2019 from the Go4Go database, which has been widely used in studies of Go (e.g., Chao et al. 2018, Ramon and Struyf 2003, Wu et al. 2018). The data contains detailed information on the game, its players, Komi (the number of bonus points given to the second mover), the sequence of all moves, and the game outcome. From Go Ratings we gather additional data on the ages, nationalities (e.g., China, Japan, South Korea, Taiwan, and others), genders, and annual rankings of professional players. We multiplied negative one by the ranking and divide it by 1,000 to ease the interpretation of the result; the higher the value, the better the player. To control for the difference in players' capabilities for each game, we create a variable, *Rank difference*, as the difference between the raw rankings of two players; we divide this difference by 1,000 such that a positive value indicates that the focal player's ranking is lower than the opponent's ranking.

3.3.2. Measuring the Quality of Moves

To evaluate the quality of moves by professional Go players, we use Leela Zero as a benchmark. Leela Zero is one of the highest performing APGs and is widely used by professional players and the public. For example, the Korea Baduk (Go) Association and the South Korean National Go Team use Leela Zero for learning and training. Because Leela Zero provides the probability of winning for any possible move made at any particular point of the game, we can compare the difference in winning probability between a move made by a professional player and Leela Zero's suggested move, which would increase the winning probability more than any other alternative moves.

Our main dependent variable is $Move\ Quality_{ig}$, representing the average difference in winning probability of the focal player i's move compared to APG's corresponding solution for the first thirty moves in a game g. This variable ranges from –100 (lowest quality) to 0 (highest quality). A smaller gap in winning probability between a player's moves and those of APG indicates higher-quality moves by the player. For instance, if a player places stones as suggested by APG for all moves, the average difference in winning probability between them is zero ($Move\ Quality_{ig} = 0$). $Move\ Quality_{ig}$ is *negative* and becomes larger in absolute value as the player's moves deviate from the best moves suggested by APG.

For each game, we separately calculate the value for the two players: black stone holder (first mover) and white stone holder (second mover). We used Leela Zero (May 23, 2020 version) along with the

GoReviewPartner program to analyze all 25,033 games played from 2015 through 2019. Using two to eight Nvidia Titan-X GPUs running in parallel, the computational analysis of games took about three months. In Appendix A, we describe how Leela Zero provides in-depth analysis of a game between two professional players. The implementation and calculation details are provided in Appendix B.

3.3.3. Summary Statistics

Table 1 provides descriptive statistics on the key variables at the player-game and player levels. Table 1(a) includes two observations for each game: one for the first mover (black stone holder) and another for the second mover (white stone holder). After omitting games that lack information on players' ages or ranks, our final sample has 46,454 observations. The mean of our main dependent variable, $Move\ Quality_{ig}$, is –2.01 over the sample period. This means that the professional player's winning probability for the first thirty moves in a game averages 2.01 percentage points less than that of APG's best move. This is a substantial difference because the two percentage point difference for each move accumulates throughout the game—from the first to the last move. The average (raw) rank of the players is 280th before transformation. The average rank difference is, by definition, zero (the positive and negative differences of the two players cancel out).

Table 1(b) shows the descriptive statistics at the player level. We identified 1,242 professional players from Go matches from 2015 through 2019. Due to the missing information on the ages and ranks of some players, our final sample contains 1,088 players. The average age of players is 32.73, and the median age is 27.32.

"Insert Table 1 here"

4. Results

4.1. Does APG Improve the Quality of Moves by Professional Players?

4.1.1. Model-Free Evidence

We first graphically present our main outcome of interest. Figure 2 shows the weekly average value of $Move\ Quality_{ig}$ from 2015 through 2019. The vertical line on February 2017 represents the public release of the first APG that surpassed human performance, Leela. This model-free illustration shows that the $Move\ Quality_{ig}$ was relatively low and stable over time before APG. The $Move\ Quality_{ig}$ increased immediately after Leela's public release.

"Insert Figure 2 here"

4.1.2. Event-Study Analysis

We then use a formal OLS regression model to estimate the $Move\ Quality_{ig}$ of professional Go players around the release of APG. The baseline event-study regression specification at the player-game level is:

$$Y_{ig} = \alpha + \beta_1 \times Post_g + \gamma_i + \delta_{-i} + \epsilon_{ig},$$

where indices i and g represent player and game respectively. γ_i represents focal-player fixed effects while δ_{-i} represents fixed effects for the opponent player. Y_{ig} is $Move\ Quality_{ig}$. $Post_g$ is equal to 1 if a Go game is played in the quarters after the first public introduction of APG in February 2017 and 0 otherwise. Standard errors are clustered at the focal-player level to address a concern that the error terms are correlated across the players. We are interested in β_1, which captures how APG improved the quality of moves played by professional players.

The results are shown in Table 2. Column 1 shows that the coefficient of $Post_g$ is positive and significant ($\beta = 0.756, p < 0.01$), indicating that the $Move\ Quality_{ig}$ increased by 0.756 percentage points (or about 30.5 percent) on average after APG's public release.

"Insert Table 2 here"

Yet, it is possible that professional players' performance had been improving over time and is driving the results, although Figure 2 does not indicate evidence of this. To control for this trend, we add a $Trend_g$ variable (i.e., the number of quarters passed since the first quarter in our sample) and an interaction term ($Post_g \times Trend_g$). The results are shown in column 2. We find a small yet positive trend ($\beta = 0.007, p < 0.05$), suggesting that the performance of professional players had been improving slowly over time. Importantly, the coefficient of the interaction term ($\beta = 0.116, p < 0.01$) shows that there were much larger (about seventeen times greater) improvements following the public release of APG, even after the performance trends are taken into account.

4.2. Are There Differential Effects of AI Adoption and Utilization by Age?

As discussed in Section 2.2, age is an important factor that could affect the adoption and utilization of new technology. We plot in Figure 3 the model-free illustration of two different age groups, young and old. The figure shows that the $Move\ Quality_{ig}$ was relatively stable and similar among two groups before APG, while the increase in the $Move\ Quality_{ig}$ is notably greater for the young group after the release of Leela.

"Insert Figure 3 here"

We thus test whether APG indeed has differential effects on the move quality of professional players of different ages. We estimate the following regression model at the player-game level:

$$Y_{ig} = \alpha + \beta_1 \times Post_g \times Young_i + X_{ig} + \gamma_i + \delta_{-i} + \theta_g + \epsilon_{ig},$$

where γ_i, δ_{-i}, and θ_g represent focal-player-, opponent-player-, and quarter-fixed effects, respectively for game g. X_{ig} includes player- or game-level control variables such as *Komi*, *White*, *Rank*, and *Rank* *differences between players*. $Young_i$ is an indicator variable equal to 1 if the age of players is below the median age of all players (i.e., less than twenty-eight years) at the date of APG's public release and 0 otherwise.

Table 3 shows the results. Column 1 includes only $Young_i$ and control variables with quarter-time-fixed effects. Column 2 then adds the interaction term, $Post_g \times Young_i$. The coefficient of the interaction term ($\beta = 0.273, p < 0.01$) is positive and significant; the quality improvement for young players is 0.273 percentage points (or 11 percent) greater than that for old players.

To check whether our results are robust to the consideration of players' inborn characteristics, column 3 adds the player-fixed effect; column 4 adds the opponent-player-fixed effect. We find that the effect of AI is consistently more prominent for the junior group, whose quality of moves improved by 0.209–0.273 percentage points (or 8.5–11 percent) over the senior group, even after including the fixed effects for the players. We conduct further robustness checks and report the results in Section 4.3.

"Insert Table 3 here"

4.3. Robustness Checks

We check the robustness of the results in five ways: 1) an estimation with distributed leads and lags, 2) a sensitivity test by age conditions, 3) a placebo permutation test using the pseudo age assignment, 4) an analysis using monthly data, and 5) the different number of moves for an opening strategy (the first 40, 50, or 60 moves).

4.3.1. Event Study with Distributed Leads and Lags

To check the pre-APG trend and the time-varying effects of the APG, we include the distributed time leads and lags in our regression and estimate the following:

$$Y_{ig} = \alpha + \Sigma_z \beta_z \times Z \times Young_i + X_{ig} + \gamma_i + \delta_{-i} + \theta_g + \epsilon_{ig},$$

where γ_i, δ_{-i}, and θ_g represent focal-player-, opponent-player-, and time-quarter-fixed effects, respectively. z represents the indicators for time leads and lags—that is, the number of quarters before or after the public release of APG.

Table 4, columns 1–4, shows the detailed regression results, and Figure 4 graphically illustrates the results. We do not find any pre-APG trend for $Move\ Quality_{ig}$; the estimates for pre-APG quarters are close to and statistically not distinguishable from zero. For quarters after the public release of APG, the estimates are large and statistically significant. The improved quality by young players is large and persistent after four quarters of the APG release.

"Insert Table 4 here"

"Insert Figure 4 here"

4.3.2. Sensitivity Test for Age Groups

We test whether the results are sensitive to our operationalization of age groups. We conduct two robustness checks using different age classifications and adding more granular age categories. First, we use the average age (instead of median age) as the cutoff for the junior versus senior group; this increases the cutoff age from twenty-eight years to thirty-three years. The results, provided in Table C.1 of Appendix C, are robust to this alternative classification ($\beta = 0.289, p < 0.01$ in column 4).

In addition, we conduct the sensitivity test with more granular age groups. We investigate the same model with three age groups: those younger than twenty ("Young"), those in their twenties ("Middle"), and those thirty or older ("Old"). The results are provided in Table C.2 of Appendix C. The estimates for $Post_g \times Young_i$ ($\beta = 0.302, p < 0.01$) and $Post_g \times Middle_i$ ($\beta = 0.198, p < 0.01$) are large and statistically significant. Importantly, the effect is largest for young players and decreases monotonically for middle and then for old players.

4.3.3. Placebo Permutation Test for Age Groups

To check whether we capture the spurious variations when testing the age effects, we conduct a placebo test. We *randomly* reassign players to age groups and estimate the models. If our suggested logics hold, we expect to find null effects and thus cannot reject the null hypothesis that the age effect is zero. As shown in Table C.3 of Appendix C, the estimate for $Post_g \times Young_i$ is close to zero and not statistically significant with the randomly assigned age group.

4.3.4. Alternative Time-Fixed Effects

To consider the time effect on a more granular level, we estimate the model with month-fixed effects instead of quarter-fixed effects. Table C.4 of Appendix C shows that the results are consistent with this alternative.

4.3.5. Opening Strategy with the Different Number of Moves

Our results could have been influenced by the choice of the number of moves. To check this possibility, we estimate our models with different definitions for early opening moves: the first 40, 50, and 60 moves. The results, shown in Table C.5 of Appendix C, are robust to these alternative definitions, confirming that the operationalization of early moves does not drive the findings.

5. Further Analyses

5.1. Difference-In-Differences Estimation Using Cross-Country Variations

One may still worry that there was a general improvement in the performance of professional Go players around the release of APG in 2017. To further address this concern, we estimate the difference-in-differences model using country-level variations in APG adoption and utilization. Among the three major countries holding the largest professional Go leagues—(mainland) China, Japan, and South Korea— Japan had relatively low awareness of or interest in APGs. For example, two historical matches between the countries' best players and APG were held in South Korea (AlphaGo vs. Sedol Lee, February 2016) and in China (AlphaGo vs. Ke Jie, May 2017), but none was held in Japan. This reflects the countries' interest levels in APG, which further affected their post-event adoption and utilization of APG. A Google Trends search also reveals that, from 2016 through 2017, the term *AlphaGo* was searched most by two countries: China and South Korea; in contrast, Japan was ranked seventh with an interest score of 5 (compared to China's score of 100 as a reference point and South Korea's score of 97).

"Insert Figure 5 here"

This motivates us to estimate the difference-in-difference model that compares players in countries that are significantly affected by APG (i.e., China and Korea) to those in a country that is less affected (i.e., Japan). Note that Japan is to some extent affected by APG. Figure 5 illustrates that the model-free average $Move\ Quality_{ig}$ was similar among the three countries before APG. However, the average $Move\ Quality_{ig}$ increases more rapidly for Chinese and South Korean players, while the improvement is rather smaller for the Japanese. The fact that our control group is also affected by the treatment in the same way as the treatment group (but to a weaker extent) will bias our estimates toward zero. That is, this works *against* our findings, and the results provide conservative estimates. We estimate the following:

$$Y_{icg} = \alpha + \beta_1 \times Post_g \times Treat_{ic} + X_g + \gamma_i + \delta_{-i} + \theta_g + \epsilon_{ig},$$

where γ_i, δ_{-i}, and θ_g denote focal-player-, opponent-player-, and time-quarter-fixed effects, respectively. c denotes the nationality of a focal player i. $Treat_{ic}$ is an indicator variable having the value of 1 if a focal player i belongs to a treated country group c and 0 otherwise. The treatment group consists of (mainland) China and Korea. The control group, Japan, takes into account any non-country-specific changes in performance of professional Go players.

Table 5 shows the results from the difference-in-differences estimation. In the main analysis, the quality increased by 0.756 percentage points (Table 1, column 1). In the difference-in-differences model the magnitude is smaller: 0.315 to 0.230 percentage points (Table 5, columns 2–4), statistically significant at the 0.01 level. The smaller estimate was expected because our empirical design uses Japan—which was also affected by the release of APG (although to a lesser extent)—as a counterfactual. In other words, if Japan were not affected by the APG at all, we would have obtained a larger estimate.

"Insert Table 5 here"

Columns 5–6 in Table 5 further show the time-varying effects of the APG on move quality and Figure 6 illustrates the graphical version of this based on the results in column 6. We find no evidence for the increase in the move quality for Chinese and South Korean players compared to the Japanese for the pre-APG period; for quarters after the public release of APG, there is a positive and significant improvement in move quality. From this stringent model, we once again confirm that AI is indeed what improves the quality of moves by professional players.

"Insert Figure 6 here"

We also conducted a placebo test to check whether our findings are driven by the spurious variations in players' nationalities. We *randomly* reassign players to one of three nationalities—China, Japan, and South Korea—and estimate the same models. Under this placebo test, we expect to find no evidence that the nationality effect is different from zero. As shown in Table C.6 of Appendix C, the estimate for $Post_g \times Treat_{ic}$ is close to zero and not statistically significant.

5.2. Mechanisms for Quality Improvement: Errors and a Critical Mistake

In this part, we extend the analysis beyond $Move\ Quality$ and delve into two important channels through which AI-based training improves the quality of moves: *errors* and *critical mistake*. This analysis is motivated by the norm that, after Go games, players spend significant time and effort analyzing and evaluating each move—especially if the move was an error or a mistake. In an interview with news media, Jin-seo Shin (who was ranked first in the world in 2020) stated:

> *"Before APG, players and their peers replayed the game and discussed which move was an error and which was a critical mistake. After the public release of APG, this replay and discussion by players became almost meaningless. APG teaches us by showing the accurate winning probability with each move. If the winning probability drops from 60 percent to 40 percent after a move, that is an error. If it drops from 80 percent to 20 percent, that is a critical mistake. … I have to admit that the APG-based training provides limitless help in developing my Go skills* (Sohn 2021)*."*

To test these mechanisms, we measure the error as the number of bad moves in the game in which the winning probability drops by 10 or more percentage points compared to the winning probability of the immediately preceding move by a focal player. The critical mistake is the magnitude of the biggest drop in winning probability among all the moves in a game. Figure 7 shows the model-free trend of errors (in Panel A) and the critical mistake (in Panel B). Both errors and the critical mistake show a substantial decrease after the release of APG.

"Insert Figure 7 here"

We then conduct regression analyses on errors and the critical mistake. Table 6, columns 1 and 3, shows that the number of errors and the magnitude of a critical mistake decreased after APG release.

Columns 2 and 4 show the results after controlling for the linear trend. The estimates for the interaction term ($Post_g \times Trend_g$) show that the (preexisting) negative trend is discontinuously accelerated after the introduction of APG. These results confirm that AI improved the quality of moves of professional players by reducing both the number of errors and the magnitude of the critical mistake.

"Insert Table 6 here"

5.3. Do AI-Driven Improvements of Move Quality Lead to Winning?

Building upon our finding that junior players improve more than senior players after APG, we further investigate whether this improvement leads to a higher probability of winning the game.

We conduct the three-step mediation analysis suggested by (Baron and Kenny 1986). As a baseline model, we run a logit regression of winning a game on an interaction between an indicator for a junior player and an indicator for a post-APG period. We find from Table 7, column 1, that the estimate for $Post_g \times Young_i$ ($\beta = 0.120, p < 0.01$) is positive and statistically significant; the improvements in move quality indeed lead to a higher chance of winning. This implies that the chances of young players winning are on average 2.96 percentage points (5.82 percent) higher after the release of the APG if other variables are set to mean values.

We then conduct the mediation analysis to test for the channels. The first step is to check whether $Post_g \times Young_i$ is statistically related to the proposed mediators: *Move quality*, *Errors*, and *Critical mistake*. Table 7, columns 2–4, shows that *Move Quality* is positively associated with the junior group after APG, while *Errors* and *Critical mistake* are negatively associated.

The second step is to check whether the move quality is positively associated with the probability of winning, while errors and the magnitude of critical mistake are negatively associated with the probability of winning, without the explanatory variable ($Post_g \times Young_i$). We confirm that this is the case from the results in Table 7, columns 5–7.

As the last step, we examine whether the magnitude of the estimated effect of the explanatory variable ($Post_g \times Young_i$) decreases with inclusion of the mediators. In Table 7, columns 8–11, the estimates for the explanatory variable ($Post_g \times Young_i$) decrease for all cases after adding the mediator variables, compared to those in the baseline model (column 1).

The mediation analysis confirms that junior players are more likely to win after APG through their improvements in three dimensions: *Move quality*, *Errors*, and *Critical mistake*.

"Insert Table 7 here"

5.4. How Does the AI Effect Vary throughout the Game?

Although we focus on the early (first to thirtieth) moves in the main analyses, the role of AI is not restricted to this particular phase. In this section, we extend the analysis to include later stages of the game. We add thirty moves incrementally to our baseline analysis—up to 180 moves—and compare the effects.

We first graphically present model-free results on $Move\ Quality_{ig}$ in Figure 8. The AI effect is most prominent at early opening moves (for moves 1–30) and gradually decreases as we include later moves in the analysis.

"Insert Figure 8 here"

Formal analyses confirm these observations. Table 8 shows the results from six different regression specifications. The estimate for $Post_g \times Young_i$ gradually shrinks from 0.209 (for moves 1–30) to 0.053 (for moves 1–180). The event-study estimates with distributed leads and lags are graphically illustrated in Figure C.1 of Appendix C. This clearly shows that junior players' improvement by APG is highest for the opening strategy and becomes weaker as moves from later stages of the game are included.

"Insert Table 8 here"

One explanation for this can be uncertainty. At the early stage of a game, when only a few stones are placed, players have the highest number of possible moves, and their ability to assess all alternatives and subsequent moves is significantly limited. In other words, prior to the APG, players relied more on heuristics or conventional opening strategies to alleviate such uncertain environments where complete evaluations are not possible. This is where players' training with AI can help most in improving the quality of moves, allowing players to consider all possibilities. As the game progresses into the mid-to-late stages, uncertainty is reduced as more stones are put on the board, and it becomes less difficult to evaluate the potential moves and make decisions.

Another explanation is unexpectedness. As stones are placed sequentially in a game, the likelihood of facing another board with the same moves decreases exponentially. This implies that, for later stages, more time is required to learn strategies from AI. For example, when professional players play 100 games with APG, they are likely to make similar moves within the first thirty moves in all games (i.e., to learn the early moves 100 times). The game then develops into very different forms in later stages—that is, players learn a specific situation only once.

In either case, the results altogether suggest that AI's help with decision-making can vary depending on the uncertainty of the environment and the opportunity to train and learn from AI.

6. Discussion and Conclusions

> *Humans have evolved through thousands of years of actual combat exercises, but the AI-powered Go program has told us that humans are all wrong. Go players will combine with computers to enter a whole new field and reach a whole new level.*
>
> —Ke Jie, top professional Go player (Ke Jie's Weibo Message, December 31, 2016)

This study examines whether and how AI improves human decision-making. We exploit a unique setting, professional Go games, where AI algorithms clearly surpass the best human players and the knowledge humans have accumulated over thousands of years. We use the APG to find the best move at each point of the game and to measure the difference between the moves by professional players and AI's solution. This is a rare opportunity to observe every single decision made by human players and to evaluate the quality of these decisions. We find that AI algorithms improve the move quality of professional players (i.e., players are able to reduce the quality gap between their moves and those of AI) by decreasing the number of errors and the magnitude of the critical mistake. The effect is more prominent for the early stages of the game where higher uncertainty is exhibited and players have had more training opportunity with AI. Young players benefited more from AI than old players, suggesting a potential inequality in AI utilization by age and generation.

The findings from AI in professional Go games provide important and timely implications for human decisions and knowledge. First, AI reveals that what humans believe to be the best solution may not be the best; AI could bring breakthroughs in human knowledge, heuristics, or routines that have been developed and improved over a long time. In this sense, AI should have broader effects (beyond a mere substitution of or assistance to human tasks) on the practices and performance of individuals and organizations; it can pave a way for new paradigms.

Importantly, there exists a widespread concern that AI will replace human jobs. Contrary to this prediction, AI can be utilized as an instructional tool to improve the skills and performance of humans in the professional Go industry. Our findings can also be generalized to some domains where AI has already outperformed or will outperform. Yet, we note that the results from the professional Go industry may not be readily applicable to other contexts. It is important to carefully assess different contexts and study how best to utilize AI to complement human jobs and help humans focus on more creative or value-added tasks.

Second, not everyone may enjoy the benefits of AI at the same level. Existing studies have explored the differential effects of AI. Ahmed and Wahed (2020) find that modern AI research and utilization is concentrated among elite universities and a small number of large corporate labs. Other studies show that the experienced tend to benefit more than rookies from utilizing AI (Choudhury et al. 2020, Miric 2020).

We add to this stream of literature by showing that openness to new technology and the ability to utilize it could also contribute to reaping gains from AI; the young benefited more from APG than did the old. Further investigation is called for on differential access to and utilization of AI and the potential inequality in outcome.

Third, the impact of AI also depends on the complexity and uncertainty of a situation. AI-driven improvement is most prominent in the early stages of the game. This boundary condition of the AI effect is consistent with the findings in drug discovery and development (Lou and Wu 2021). This suggests that a uniform application of AI would not yield the best outcome and could lead to inefficient allocation of AI and human resources. A careful consideration of where to adopt AI and to what extent is therefore required; our findings indicate that AI's complementary role is most prominent when the task is more uncertain or complex.

This paper contributes to several literature streams. First, we contribute to the information systems (IS) literature about the ability to utilize the advanced data-analytics technology in the decision support system (Atasoy et al. 2017, Goldin and Katz 1998, Polites and Karahanna 2013, Wu et al. 2019). As AI technology advances, researchers have recently expanded their interest to the role of AI in supporting human judgment (Choudhury et al. 2020, Kleinberg et al. 2017, Wang et al. 2019). Our study on the (complementary) impact of AI on human decision-making provides new insights to this area by showing that AI can help reevaluate knowledge and train professionals, reducing their errors and mistakes (that stem from heuristics and routines), and thereby improve the quality of decisions. Second, we contribute to the literature on digital economy and the economics of AI. IS researchers and economists have studied AI as a tool for automating predictions (Agrawal et al. 2019, Brynjolfsson et al. 2018) and as a GPT (Cockburn et al. 2018, Trajtenberg 2018). We show that AI can provide new knowledge that humans do not have or have not yet recognized and train them. For instance, over hundreds of years humans Go players have developed opening strategies (called *joseki*) that they believed to be optimal. AI helps humans evaluate the classic strategies and provides new opening strategies if *joseki* does not provide the highest probability of winning. Lastly, we contribute to the innovation literature on how the adoption of new technology could have differential impacts on human capital (Agarwal and Prasad 1999, Choudhury et al. 2020, Miric et al. 2020). While existing studies point out the importance of prior knowledge and experience through, for example, an absorptive capacity (e.g., Cohen and Levinthal 1990), our findings show that AI's impact is more prominent for young than old professionals and are consistent with Wang et al.'s (2019) finding in the context of coding a medical chart.

Although the domains where AI outperforms humans have broadened to include different organizations such as hospitals (Cadario et al. 2021), law firms (Kahn 2020) and sports teams (Zarley 2021), the application of our findings to different contexts requires careful consideration. Decisions in professional Go games are made under well-defined rules of the game. Managers also face a set of rules and restrictions; the codification of routines is indeed important in a firm (e.g., Bingham et al. 1997, Foss 2003). Still, managers may have much more discretion in making decisions than do Go players. We hope that future study advances this line of research in different contexts.

References

Acemoglu D, Autor D, Hazell J, Restrepo P (2019) AI and Jobs: Evidence from Online Vacancies. *NBER Work. Pap.* (October):1–59.

Acemoglu D, Restrepo P (2020) Robots and Jobs: Evidence from US Labor Markets. *J. Polit. Econ.* 128(6):2188–2244.

Agarwal R, Prasad J (1999) Are Individual Differences Germane to the Acceptance of New Information Technologies? *Decis. Sci.* 30(2):361–391.

Agrawal A, Gans J, Goldfarb A (2018) *Prediction Machines: The Simple Economics of Artificial Intelligence* (Harvard Business Press, Boston, MA).

Ahmed N, Wahed M (2020) The De-democratization of AI: Deep Learning and the Compute Divide in Artificial Intelligence Research. *arXiv*.

Atasoy H, Chen P yu, Ganju K (2017) The Spillover Effects of Health IT Investments on Regional Healthcare Costs. *Manage. Sci.* (June):mnsc.2017.2750.

Ayres I (2007) *Super crunchers: Why thinking-by-numbers is the new way to be smart* (Bantam Books).

Bapna R, Langer N, Mehra A, Gopal R, Gupta A (2013) Human Capital Investments and Employee Performance: An Analysis of IT Services Industry. *Manage. Sci.* 59(3):641–658.

Baron RM, Kenny DA (1986) The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *J. Pers. Soc. Psychol.* 51(6):1173–1182.

Bechara A, Damasio H, Tranel D, Damasio AR (1997) Deciding Advantageously Before Knowing the

Advantageous Strategy. *Science* 275(5304):1293–1295.

Bingham CB, Eisenhardt KM, Davis JP (1997) Opening the Black Box of Organizational Expertise: Understanding What Firms Learn from their Process Experience and How that learning Unfolds over Time. *Work. Pap. Univ. Maryl.*

Bingham CB, Eisenhardt KM, Furr NR (2007) What makes a process a capability? Heuristics, strategy, and effective capture of opportunities. *Strateg. Entrep. J.* 1(1–2):27–47.

Bingham CB, Haleblian JJ (2012) How firms learn heuristics: Uncovering missing components of organizational learning. *Strateg. Entrep. J.* 6(2):152–177.

Blackwell D (1953) Equivalent Comparisons of Experiments. *Ann. Math. Stat.* 24(2):265–272.

Bresnahan TF, Brynjolfsson E, Hitt LM (2002) Information Technology, Workplace Organization, and the Demand for Skilled Labor: Firm-Level Evidence. *Q. J. Econ.* 117(1):339–376.

Brynjolfsson E, Hitt LM (2000) Beyond Computation: Information Technology. *J. Econ. Perspect.* 14(4):23–48.

Brynjolfsson E, Hitt LM, Kim HH (2011) Strength in Numbers: How Does Data-Driven Decisionmaking Affect Firm Performance? *SSRN Electron. J.*

Cadario R, Longoni C, Morewedge CK (2021) Understanding, explaining, and utilizing medical artificial intelligence. *Nat. Hum. Behav.*

Cao S, Jiang W, Wang J, Yang B (2021) *From Man vs. Machine to Man + Machine: The Art and AI of Stock Analyses* (Cambridge, MA).

Chen WR (2008) Determinants of firms' backward- and forward-looking R&D search behavior. *Organ. Sci.* 19(4):609–622.

Choudhury P, Starr E, Agarwal R (2020) Machine learning and human capital complementarities: Experimental evidence on bias mitigation. *Strateg. Manag. J.* 41(8):1381–1411.

Cockburn IM, Henderson RM, Stern S (2018) The Impact of Artificial Intelligence on Innovation. *NBER Work. Pap.*

Cohen WM, Levinthal DA (1990) Absorptive Capacity: A New Perspective on Learning and Innovation. *Adm. Sci. Q.* 35(1):128.

Cowgill B, Dell'Acqua F, Matz S (2020) The Managerial Effects of Algorithmic Fairness Activism. *AEA Pap. Proc.* 110:85–90.

De Cremer D (2020) What Does Building a Fair AI Really Entail? *Harvard Bus. Rev. Digit. Artic.* (September).

Cyert RM, March JG (1963) *A behavioral theory of the firm* (Prentice- Hall, Englewood Cliffs, NJ).

Dane E, Pratt MG (2007) Exploring Intuition and its Role in Managerial Decision Making. *Acad. Manag. Rev.* 32(1):33–54.

Danziger S, Levav J, Avnaim-Pesso L (2011) Extraneous factors in judicial decisions. *Proc. Natl. Acad. Sci.* 108(17):6889–6892.

Davenport T, Harris J (2017) *Competing on analytics: Updated, with a new introduction: The new science of winning* (Harvard Business Press, Boston, MA).

Eisenhardt KM (1990) Speed and Strategic Choice: How Managers Accelerate Decision Making. *Calif. Manage. Rev.* 32(3):39–54.

Foss NJ (2003) Bounded rationality and tacit knowledge in the organizational capabilities approach: an assessment and a re-evaluation. *Ind. Corp. Chang.* 12(2):185–201.

Galbraith JR (1974) Organization Design: An Information Processing View. *Interfaces (Providence).* 4(3):28–36.

Gavetti G, Levinthal D (2000) Looking forward and looking backward: Cognitive and experiential search. *Adm. Sci. Q.* 45(1):113–137.

Goldfarb A, Taska B, Teodoridis F (2020) Could Machine Learning Be a General-Purpose Technology? Evidence from Online Job Postings. *Work. Pap.*

Goldin C, Katz LF (1998) The Origins of Technology-Skill Complementarity. *Q. J. Econ.* 113(3):693–732.

Guest RH (1956) Of time and the foreman. *Personnel* 32:478–486.

Hambrick DC, Mason PA (1984) Upper Echelons: The Organization as a Reflection of Its Top Managers. *Acad. Manag. Rev.* 9(2):193.

He K, Zhang X, Ren S, Sun J (2015) Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification. *2015 IEEE Int. Conf. Comput. Vis.* (IEEE), 1026–1034.

Kahn J (2020) DeepMind unveils new details of game-mastering A.I. it says may help in complex real-world environments. *Fortune*.

Kahneman D (2003) Maps of Bounded Rationality: Psychology for Behavioral Economics. *Am. Econ. Rev.* 93(5):1449–1475.

Kalberg S (1980) Max Weber's Types of Rationality: Cornerstones for the Analysis of Rationalization Processes in History. *Am. J. Sociol.* 85(5):1145–1179.

Kennedy M (2017) Computer Learns To Play Go At Superhuman Levels "Without Human Knowledge." *Natl. Public Radio*.

Kleinberg J, Lakkaraju H, Leskovec J, Ludwig J, Mullainathan S (2017) Human Decisions and Machine Predictions. *Q. J. Econ.* (January):237–293.

Knight W (2016) Google's AI Masters the Game of Go a Decade Earlier Than Expected. *MIT Technology Review* (January 27) https://www.technologyreview.com/s/546066/googles-ai-masters-the-game-of-go-a-decade-earlier-than-expected/.

de Koning J, Gelderblom A (2006) ICT and older workers: No unwrinkled relationship. *Int. J. Manpow.* 27(5):467–490.

Korean Herald (2014) [SUPER RICH] Game of go makes great businesspeople. *The Korea Herald* (December 8).

Levinthal DA, March JG (1993) The myopia of learning. *Strateg. Manag. J.* 14(S2):95–112.

LGERI (LG Economic Research Institute) (2004) Learning Business Strategy from the Principles of Go. *LG Economic Research Institute (LGERI)*

Lindebaum D, Vesa M, den Hond F (2020) Insights From "The Machine Stops" to Better Understand Rational Assumptions in Algorithmic Decision Making and Its Implications for Organizations. *Acad. Manag. Rev.* 45(1):247–263.

Lou B, Wu L (2021) Artificial Intelligence on Drugs: A Large-Scale Examination of AI and Drug Innovation. *MIS Q.* Forthcoming.

Loveman G (2003) Diamonds in the Data Mine. *Harvard Buiness Rev.*

Luo X, Tong S, Fang Z, Qu Z (2019) Frontiers: Machines vs. Humans: The Impact of Artificial Intelligence Chatbot Disclosure on Customer Purchases. *Mark. Sci.* (December).

Meyer J (2011) Workforce age and technology adoption in small and medium-sized service firms. *Small Bus. Econ.* 37(3):305–324.

Mintzberg H (1987) *Crafting strategy* (Harvard Business School Press, Boston, MA).

Mintzberg H (1994) Rethinking strategic planning part I: Pitfalls and fallacies. *Long Range Plann.* 27(3):12–21.

Miric M, Lu J, Teodoridis F (2020) Decision-Making Skills in an AI World: Lessons from Online Chess. *SSRN Electron. J.*:1–43.

Mnih V, Kavukcuoglu K, Silver D, Rusu A a, Veness J, Bellemare MG, Graves A, et al. (2015) Human-level control through deep reinforcement learning. *Nature* 518(7540):529–533.

Morris MG, Venkatesh V (2000) Age Differences in Technology Adoption Decisions: Implications for a Changing Work Force. *Pers. Psychol.* 53(2):375–403.

Mousavi S, Gigerenzer G (2014) Risk, uncertainty, and heuristics. *J. Bus. Res.* 67(8):1671–1678.

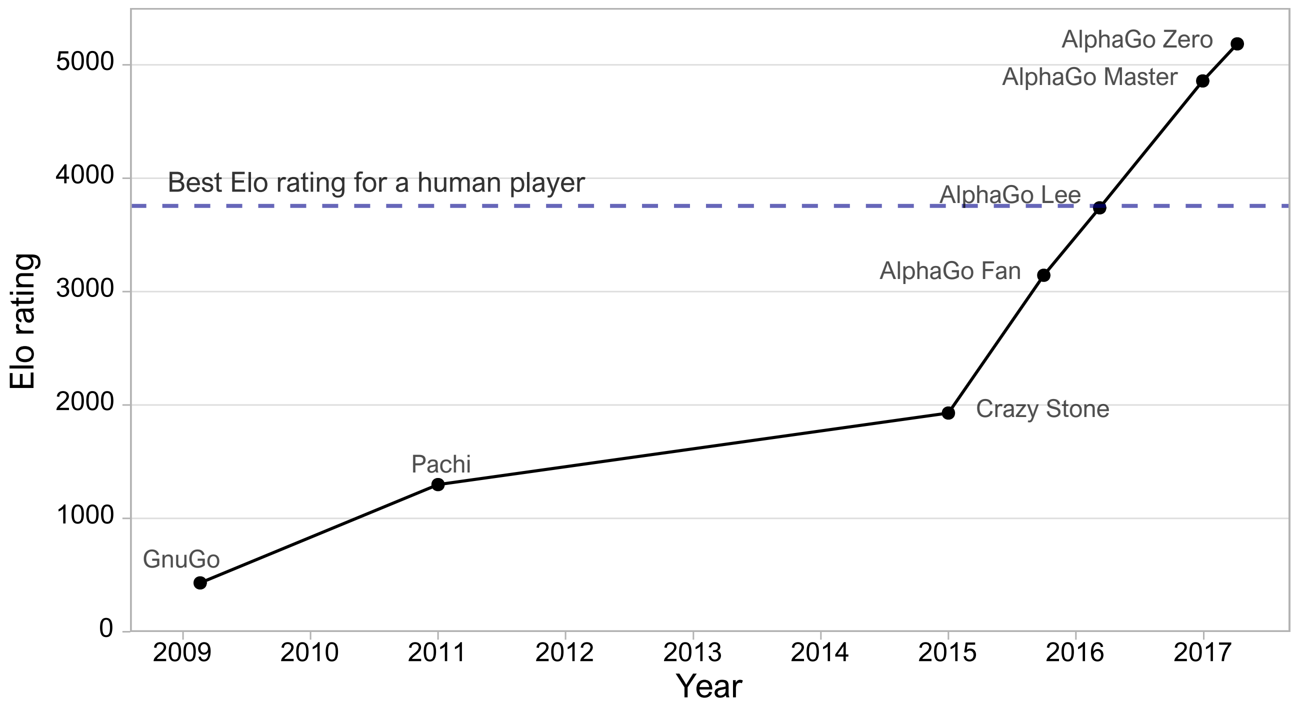
Mozur P (2017) Google's AlphaGo Defeats Chinese Go Master in Win for A.I. *The New York Times* (May 23) https://www.nytimes.com/2017/05/23/business/google-deepmind-alphago-go-champion-defeat.html.

Muoio D (2016) Why Go is so much harder for AI to beat than chess. *Business Insider* (March 10) https://www.businessinsider.com/why-google-ai-game-go-is-harder-than-chess-2016-3.

Noh S (2016) 5 Reasons AlphaGo Can't "Never" Beat Lee Sedol - Weak Paving Stones, Narrow Vision... still one number down. *Maeil Business Newspaper* (February 22)

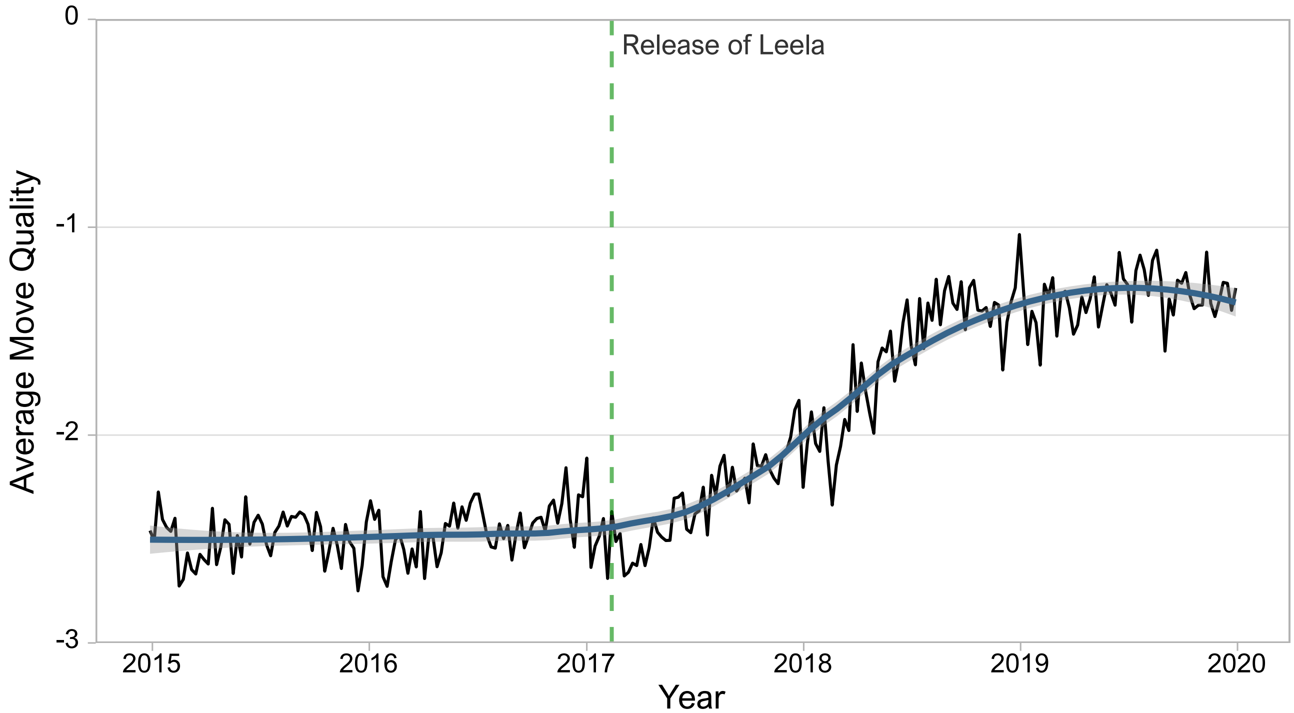
https://www.mk.co.kr/news/economy/view/2016/02/140725/ (Written in Korean).
Polites GL, Karahanna E (2013) The Embeddedness of Information Systems Habits in Organizational and Individual Level Routines: Development and Disruption. *MIS Q.* 37(1):221–246.
Press G (2021) 114 Milestones In The History Of Artificial Intelligence (AI). *Forbes* (May 19).
Ramon J, Struyf J (2003) Computer science issues in baduk. *Proc. 2nd Int. Conf. Baduk*:163–182.
Rock D (2020) How Managers Can Enable AI Talent in Organizations. *MIT Sloan Manag. Rev.*:1–11.
Saunders A, Tambe P (2013) A Measure of Firms' Information Practices Based on Textual Analysis of 10-K Filings. *Work. Pap.*
Schleife K (2006) Computer Use and Employment Status of Older Workers - An Analysis Based on Individual Data. *Labour* 20(2):325–348.
Schrittwieser J, Antonoglou I, Hubert T, Simonyan K, Sifre L, Schmitt S, Guez A, et al. (2020) Mastering Atari, Go, chess and shogi by planning with a learned model. *Nature* 588(7839):604–609.
Seamans R, Raj M (2018) AI, Labor, Productivity and The Need for Firm-Level Data. *NBER Work. Pap.*
Senior AW, Evans R, Jumper J, Kirkpatrick J, Sifre L, Green T, Qin C, et al. (2020) Improved protein structure prediction using potentials from deep learning. *Nature* 577(7792):706–710.
Shepherd DA, Williams TA, Patzelt H (2015) Thinking About Entrepreneurial Decision Making. *J. Manage.* 41(1):11–46.
Silver D, Huang A, Maddison CJ, Guez A, Sifre L, van den Driessche G, Schrittwieser J, et al. (2016) Mastering the game of Go with deep neural networks and tree search. *Nature* 529(7587):484–489.
Silver D, Schrittwieser J, Simonyan K, Antonoglou I, Huang A, Guez A, Hubert T, et al. (2017) Mastering the game of Go without human knowledge. *Nature* 550(7676):354–359.
Simon HA (1955) A Behavioral Model of Rational Choice. *Q. J. Econ.* 69(1):99.
Simon HA (1987) Making Management Decisions: the Role of Intuition and Emotion. *Acad. Manag. Perspect.* 1(1):57–64.
Simon HA (1991) Bounded Rationality and Organizational Learning. *Organ. Sci.* 2(1):125–134.
Simon HA, March JG (1958) *Organizations* (Wiley-Blackwell).
Smith BC (2019) *The promise of artificial intelligence: reckoning and judgment* (MIT Press).
Sohn H (2021) "Shin Artificial Intelligence" World Go No. 1 Shin Jin-seo 'AI is a great trick, my heart beats.' *Maeil Business Newspaper* (March 5) https://www.mk.co.kr/news/society/view/2021/03/215074/ (Written in Korean).
Somers J (2018) How the Artificial-Intelligence Program AlphaZero Mastered Its Games. *The New Yorker* (December 28).
Thaler R (1993) *Advances in Behavioral Finance Russell Sage Foundation* (New York).
Trajtenberg M (2018) AI as The Next GPT: A Political-Economy Perspective. *NBER Work. Pap.*
Tyler BB, Steensma HK (1998) The effects of executives' experiences and perceptions on their assessment of potential technological alliances. *Strateg. Manag. J.* 19(10):939–965.
Wallach I, Dzamba M, Heifets A (2015) AtomNet: A Deep Convolutional Neural Network for Bioactivity Prediction in Structure-based Drug Discovery. *arXiv*:1–11.
Wang W, Gao G (Gordon), Agarwal R (2019) Friend or Foe? The Influence of Artificial Intelligence on Human Performance in Medical Chart Coding. *SSRN Electron. J.*
Webb M (2020) The impact of artificial intelligence on behavior of people in the labor market. *Work. Pap.*
Weinberg B (2004) Experience and Technology Adoption. *IZA Discuss. Pap.* (1051).
Wu L, Hitt L, Lou B (2020) Data Analytics, Innovation, and Firm Productivity. *Manage. Sci.* 66(5):2017–2039.
Wu L, Lou B, Hitt L (2019) Data Analytics Supports Decentralized Innovation. *Manage. Sci.* 65(10):4863–4877.
Wu TR, Wu IC, Chen GW, Wei TH, Wu HC, Lai TY, Lan LC (2018) Multilabeled Value Networks for Computer Go. *IEEE Trans. Games* 10(4):378–389.
Zarley BD (2021) Liverpool and DeepMind are bringing AI to the beautiful game. *Freethink* (May 27) https://www.freethink.com/culture/ai-in-sports.

Figure 1: Elo rating comparing the best professional human player to APGs



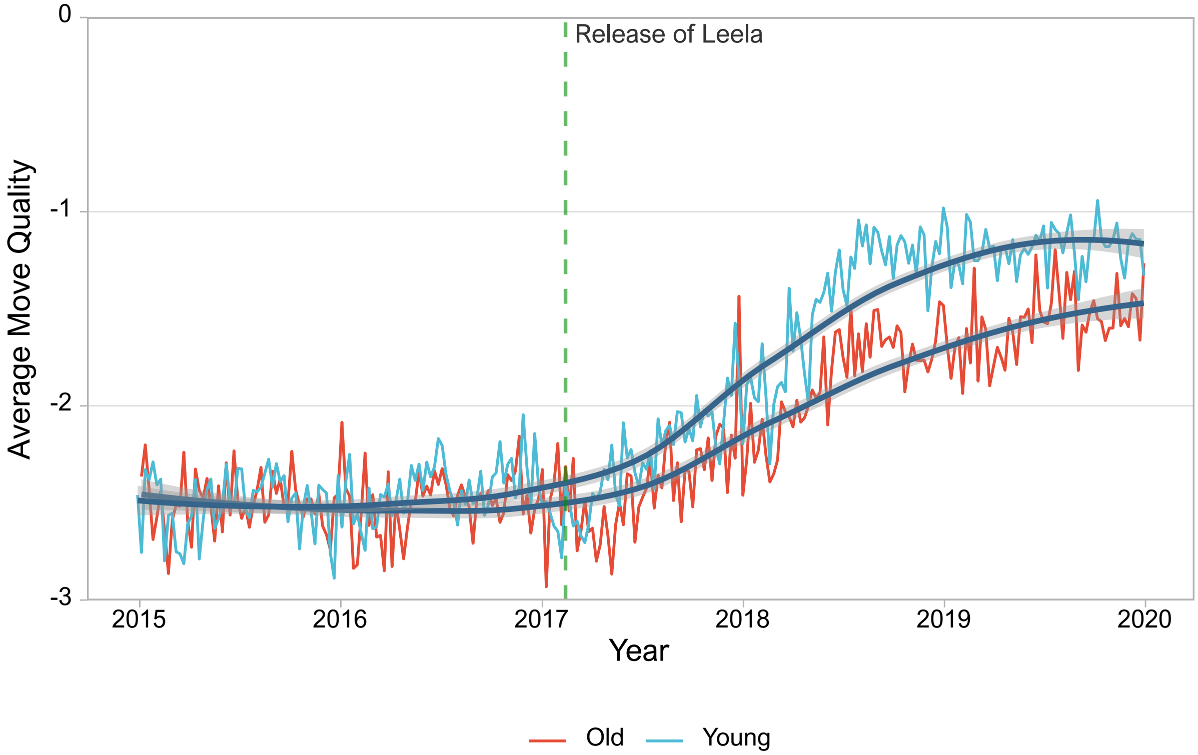
Note: This figure illustrates the advancement of Go programs from 2009 through 2017. The y-axis represents Elo ratings, which measure the performance of Go players/programs. The horizontal dashed line represents the highest score by a human, while the solid line indicates the Elo ratings of Go programs over time. Note that earlier programs including *GnuGo*, *Pachi*, and *Crazy Stone*, are not based on AI technology. *AlphaGo* and its variants are AI-powered Go Programs (APGs). The information on the Elo ratings of professional Go players comes from GoRatings (https://www.goratings.org/en/) and Go4Go (https://www.go4go.net/go/players/rank/). The performance of APGs has surpassed that of the best human player since March 2016 when *AlphaGo Lee* was introduced.

Figure 2: Effects of APG on average move quality of professional players: Model-free evidence



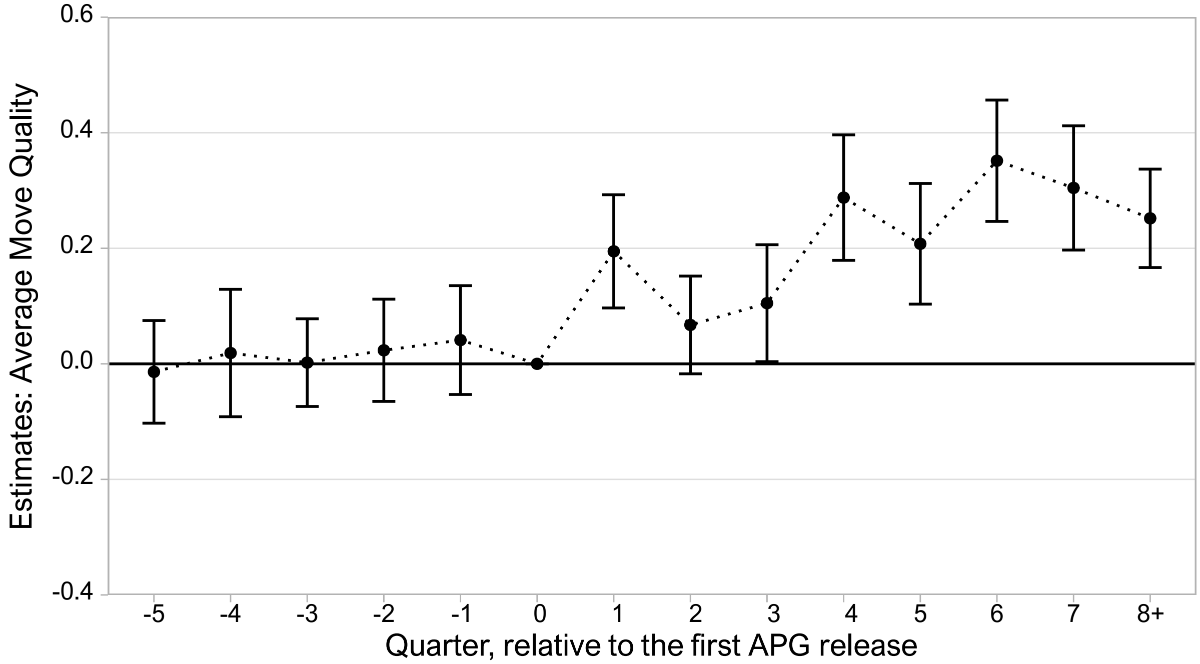
Note: This figure illustrates the weekly average $Move\,Quality$ of players from 2015 through 2019. The black solid line represents the raw (unprocessed) weekly average value. The blue solid line and the gray area around it show the local smoothed trend and the 95% confidence interval, respectively. The vertical line on February 2017 represents the first public release of an APG, *Leela*.

Figure 3: Differential effects of APGs on move quality by player age: Model-free evidence



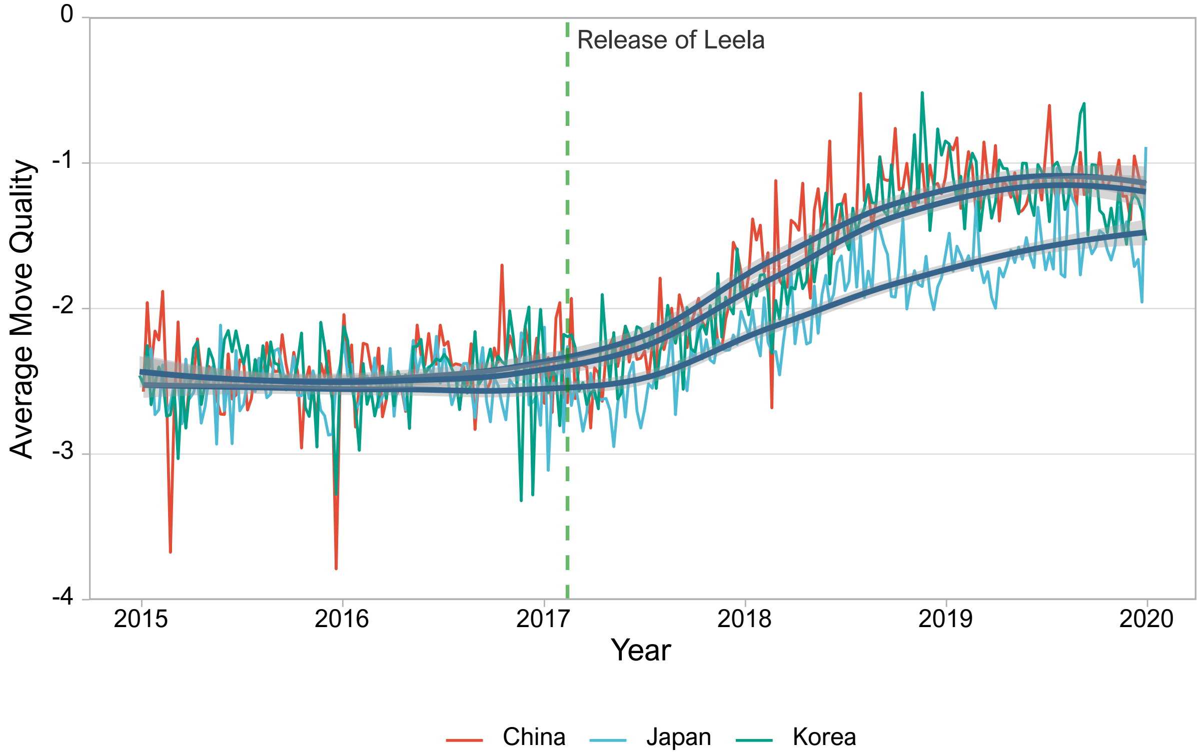
Note: This figure illustrates the average $Move\,Quality$ of professional players by their ages. The cyan and red lines show the raw (unprocessed) weekly average values for young players (below median age) and old players (above median age), respectively. The blue lines and the gray areas around them show the locally smoothed trends and the 95% confidence intervals. The vertical line on February 2017 represents the first public release of an APG, *Leela*.

Figure 4: Differential effects of APG by player age:
Estimates on move quality of young players compared to that of old players



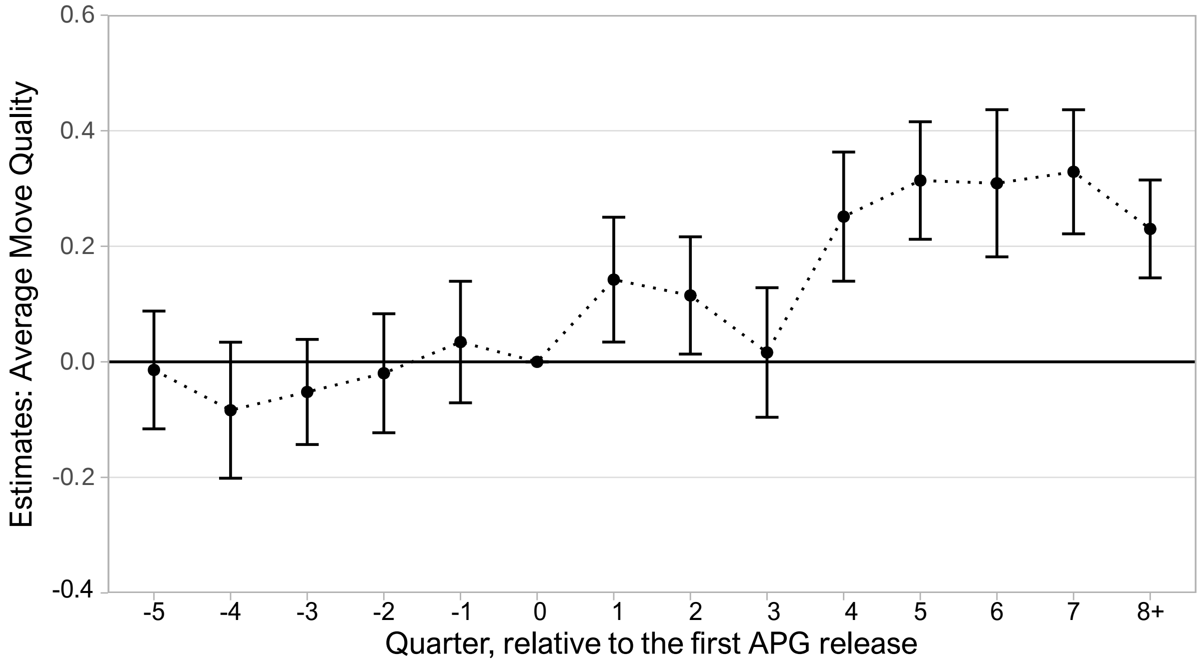
Note: This figure illustrates the differential effects of APGs on $Move\,Quality$ by player age. The points and dotted lines graphically present the $Move\,Quality$ of young players (below median age) compared to that of old players (above median age), based on the regression estimates in Table 4, column 4. The vertical error bars show the 95% confidence intervals. Before APG, we do not find a difference in $Move\,Quality$ by age. After APG, the increase in $Move\,Quality$ is greater for young players than for old players.

Figure 5: Differential effects of APG on move quality by country: Model-free evidence



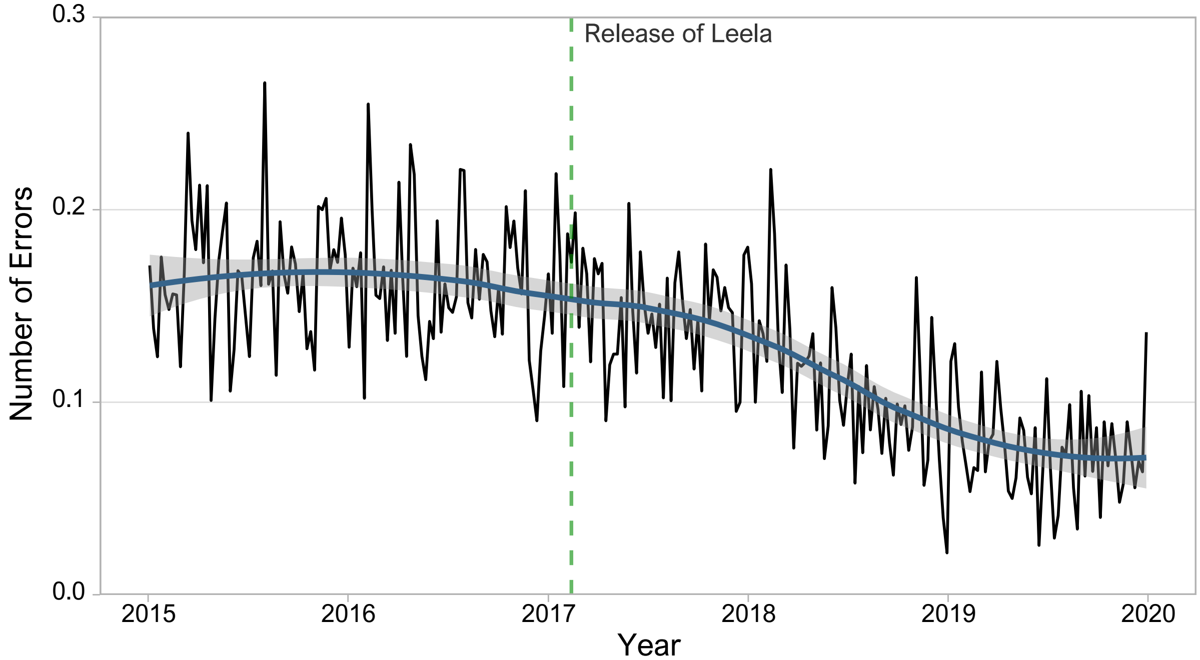
Note: This figure illustrates the average $Move\,Quality$ of professional players by their nationality. The red, green, and cyan lines show the raw (unprocessed) weekly average of $Move\,Quality$ for Chinese, Korean, and Japanese players, respectively. The blue lines and the grey areas around them show the locally smoothed trend and the 95% confidence interval. The vertical line on February 2017 represents the first public release of an APG, *Leela*.

Figure 6: Effects of APG on move quality:
Difference-in-differences estimation using cross-country variation in exposure to APG

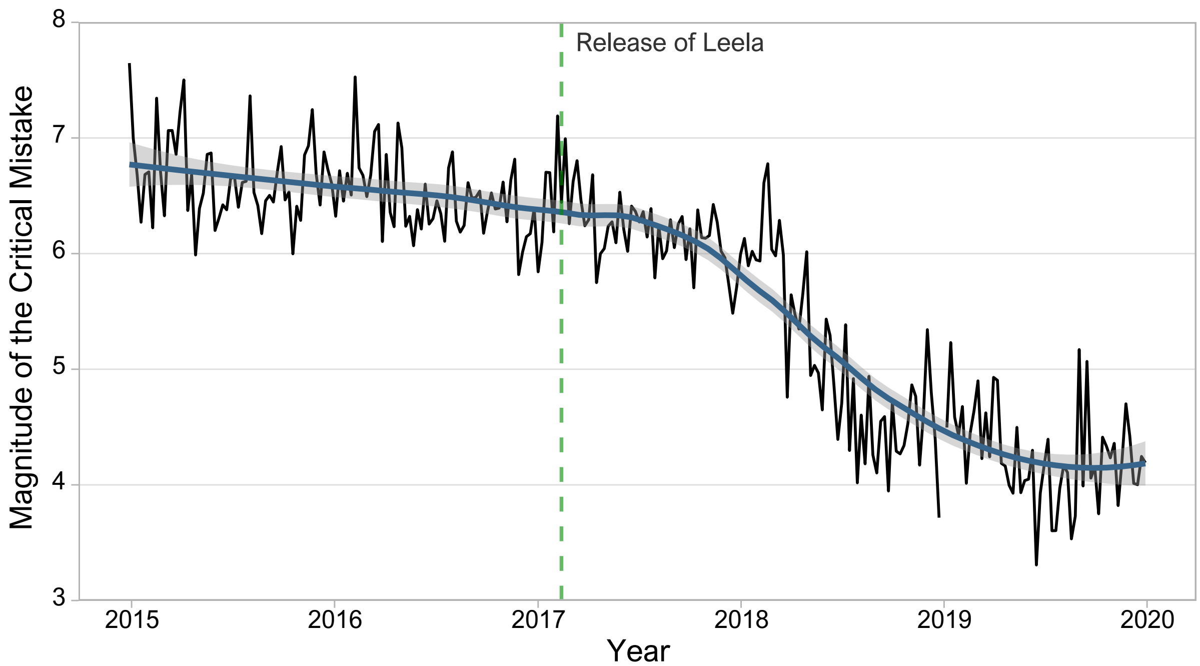


Note: This figure illustrates the effects of APG on $Move\,Quality$, based on the difference-in-differences estimation results reported in Table 5, column 6. The treatment group consisted of players in China and South Korea, countries that held major APG events and exhibited great interest in these games. The control group is comprised of players in Japan, which did not hold a major APG event. Although Japan is, to some extent, treated by the occurrence of APG, the strength of treatment is far weaker for Japanese players. The inclusion of Japan in the control group would bias our estimates towards zero (i.e., *against* our findings), leading to an underestimation. In other words, the resulting estimates provide a lower bound of the effect.

Figure 7: Errors and a critical mistake as mechanisms: Model-free evidence



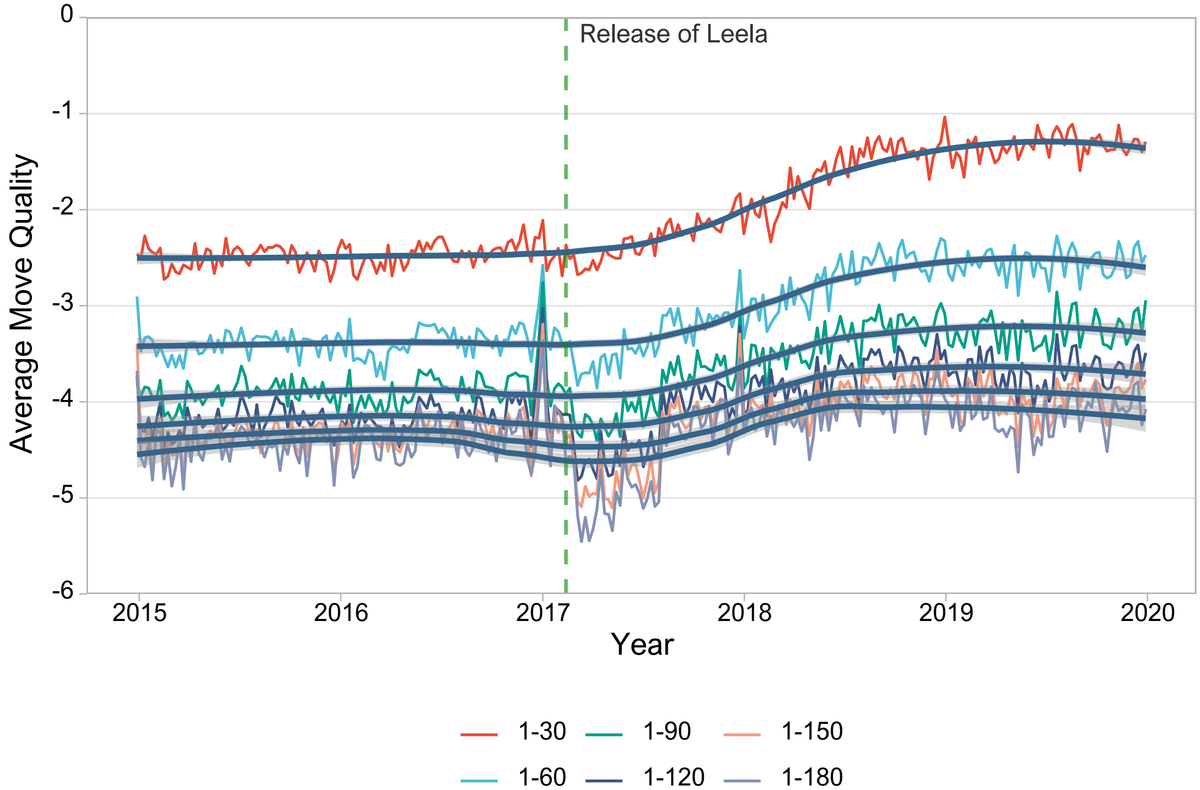
(a) Number of errors



(b) Magnitude of the most critical mistake

Note: This figure illustrates the weekly average of $Number\,of\,Errors$ (Panel a) and the $Magnitude\,of\,the\,Most\,Critical\,Mistake$ (Panel b) from 2015 through 2019. The solid black line represents the raw (unprocessed) weekly average value. The solid blue lines and the gray areas around them show locally smoothed trends and the 95% confidence intervals, respectively. The vertical line on February 2017 represents the first public release of an APG, *Leela*.

Figure 8: Effects of APG on move quality: Heterogeneity by the number of moves



Note: This figure illustrates how the changes in average $Move\,Quality$ differs by the number of moves. Beginning with the opening strategy of the first thirty moves, we incrementally add thirty additional moves (up to 180 moves) and compare the trends; the six colored lines show the raw (unprocessed) weekly average of $Move\,Quality$. The solid blue lines and the gray areas around them show locally smoothed trends and the 95% confidence intervals, respectively. The vertical line on February 2017 represents the first public release of an APG, *Leela*.

Table 1: Descriptive statistics

	N	Mean	Median	SD	P25	P75
Move Quality	50066	-2.01	-1.92	1.07	-2.66	-1.22
Number of Errors	50066	0.13	0.00	0.37	0.00	0.00
Magnitude of the Critical Mistake	50065	5.65	4.80	3.96	2.94	7.36
Age	49214	28.18	24.52	12.57	19.61	31.60
Young	49214	0.62	1.00	0.49	0.00	1.00
Rank	48808	-0.28	-0.18	0.27	-0.43	-0.05
Rank Diff	47712	0.00	0.00	0.22	-0.09	0.09
White	50066	0.50	0.50	0.50	0.00	1.00
7.5 Komi	50066	0.38	0.00	0.49	0.00	1.00

(a) Player-game level

	N	Mean	Median	SD	P25	P75
Move Quality	1242	-2.20	-2.18	0.67	-2.52	-1.79
Number of Errors	1242	0.17	0.12	0.22	0.00	0.20
Magnitude of the Critical Mistake	1242	-6.20	-5.98	2.10	-6.96	-5.04
Age	1149	32.73	27.32	16.15	20.25	42.77
Young	1149	0.49	0.00	0.50	0.00	1.00
Rank	1088	-0.54	-0.54	0.31	-0.81	-0.27
Rank Diff	1080	0.15	0.11	0.19	0.00	0.27

(a) Player level

Note: This table provides the descriptive statistics of the variables at the player-game level (Panel a) and player level (Panel b). Note that, to ease the interpretation of the result, we multiply negative one to the rank of a player and divide it by 1,000 ($Rank$). That is, the higher the value of $Rank$ is, the better the player is. We also divide the rank difference between the focal player and the opponent by 1,000 ($Rank\,Difference$). A negative value for $Rank\,Difference$ indicates that the focal player is a better player.

Table 2: Effects of APG on average move quality of professional players: Event study approach

Dependent Variable:	Move Quality	
Model:	(1)	(2)
Variables		
Post	0.756***	-1.932***
	(0.017)	(0.065)
Trend		0.007**
		(0.003)
Post $\times$ Trend		0.116***
		(0.004)
Fixed-effects		
Player	Yes	Yes
Opponent Player	Yes	Yes
Fit statistics		
Observations	50,066	50,066
R^2	0.264	0.330
Within R^2	0.116	0.195

Note: This table shows the regression estimates on the effects of APG on the $Move\,Quality$ of professional Go players, before and after the first public release of APG, *Leela*. $Post$ takes unity for the games played in the quarters after February 2017. $Trend$ refers to the number of quarters that had passed since the first quarter in our sample. Clustered standard errors at a focal-player level are in parentheses.

$^{*}\ p < 0.1$;$^{**}\ p < 0.05$;$^{***}\ p < 0.01$

Table 3: Differential effects of APG by player age:
Estimates on move quality of young players compared to that of old players

Dependent Variable:	Move Quality			
Model:	(1)	(2)	(3)	(4)
Variables				
Young	0.103***	-0.049**		
	(0.020)	(0.021)		
Rank	0.817***	0.799***	1.626***	2.501***
	(0.036)	(0.037)	(0.240)	(0.292)
Rank Diff	0.137***	0.128***	0.061**	1.023***
	(0.027)	(0.027)	(0.024)	(0.162)
White	-0.134***	-0.133***	-0.131***	-0.131***
	(0.010)	(0.010)	(0.009)	(0.010)
7.5 Komi	0.025	0.023	0.024	0.041**
	(0.017)	(0.017)	(0.016)	(0.019)
Post $\times$ Young		0.273***	0.227***	0.209***
		(0.028)	(0.031)	(0.031)
Fixed-effects				
Quarter	Yes	Yes	Yes	Yes
Player			Yes	Yes
Opponent Player				Yes
Fit statistics				
Observations	46,454	46,454	46,454	46,454
R^2	0.276	0.279	0.325	0.349
Within R^2	0.064	0.069	0.013	0.014

Note: This table shows the regression estimates on the heterogeneous effects of APG by player age; the $Move\,Quality$ of young players compared to that of old players is estimated. $Post$ refers to games played in the quarters after the first public release of APG in February 2017, and $Young$ refers to young professional Go players. Clustered standard errors at a focal player level are in parentheses.
$^{*}\,p < 0.1$; $^{**}\,p < 0.05$; $^{***}\,p < 0.01$

Table 4: Differential effects of APG by player age:
Estimates on move quality of young players compared to that of old players

Dependent Variable:	Move Quality			
Model:	(1)	(2)	(3)	(4)
Variables				
Young	0.103***			
	(0.020)			
Rank	0.817***	0.794***	1.427***	2.291***
	(0.036)	(0.037)	(0.250)	(0.307)
Rank Diff	0.137***	0.128***	0.059**	0.974***
	(0.027)	(0.027)	(0.024)	(0.163)
White	-0.134***	-0.134***	-0.131***	-0.131***
	(0.010)	(0.010)	(0.009)	(0.010)
7.5 Komi	0.025	0.020	0.025	0.043**
	(0.017)	(0.016)	(0.016)	(0.019)
-5 qr prior		-0.082**	-0.013	-0.014
		(0.042)	(0.045)	(0.045)
-4 qr prior		-0.029	0.039	0.019
		(0.054)	(0.054)	(0.056)
-3 qr prior		-0.030	0.011	0.002
		(0.037)	(0.038)	(0.039)
-2 qr prior		-0.049	0.028	0.023
		(0.042)	(0.044)	(0.045)
-1 qr prior		0.002	0.050	0.041
		(0.042)	(0.048)	(0.048)
1 qr after		0.169***	0.207***	0.195***
		(0.047)	(0.049)	(0.050)
2 qr after		0.043	0.085*	0.067
		(0.044)	(0.044)	(0.043)
3 qr after		0.066	0.107**	0.105**
		(0.047)	(0.050)	(0.052)
4 qr after		0.280***	0.301***	0.288***
		(0.052)	(0.056)	(0.055)
5 qr after		0.233***	0.244***	0.208***
		(0.054)	(0.054)	(0.053)
6 qr after		0.325***	0.374***	0.352***
		(0.049)	(0.053)	(0.054)
7 qr after		0.319***	0.332***	0.304***
		(0.048)	(0.053)	(0.055)
8+ qr after		0.267***	0.282***	0.252***
		(0.033)	(0.043)	(0.043)
Fixed-effects				
Quarter	Yes	Yes	Yes	Yes
Player			Yes	Yes
Opponent Player				Yes
Fit statistics				
Observations	46,454	46,454	46,454	46,454
R^2	0.276	0.280	0.325	0.350
Within R^2	0.064	0.071	0.015	0.015

Note: These regressions show the time-varying estimates using specifications described in Section 4.3. The reference quarter is the first quarter of 2017 when APG became publicly available. Clustered standard errors at a focal-player level are in parentheses. $^{*}\ p < 0.1;^{**}\ p < 0.05;^{***}\ p < 0.01$

Table 5: Effects of APG on move quality:
Difference-in-differences estimation using cross-country variation in exposure to APG

Dependent Variable:	Move Quality					
Model:	(1)	(2)	(3)	(4)	(5)	(6)
Variables						
Treated	0.110^{***}	-0.069^{***}				
	(0.024)	(0.026)				
Rank	0.809^{***}	0.810^{***}	2.222^{***}	3.146^{***}	2.174^{***}	3.084^{***}
	(0.040)	(0.040)	(0.221)	(0.275)	(0.221)	(0.278)
Rank Diff	0.133^{***}	0.131^{***}	0.059^{**}	1.115^{***}	0.058^{**}	1.086^{***}
	(0.032)	(0.032)	(0.027)	(0.178)	(0.027)	(0.180)
White	-0.129^{***}	-0.129^{***}	-0.128^{***}	-0.126^{***}	-0.128^{***}	-0.126^{***}
	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)
7.5 Komi	0.011	0.010	0.020	0.030	0.022	0.032^{*}
	(0.018)	(0.017)	(0.017)	(0.020)	(0.016)	(0.019)
Post × Treated		0.315^{***}	0.266^{***}	0.230^{***}		
		(0.031)	(0.031)	(0.031)		
-5 qr prior					-0.037	-0.014
					(0.052)	(0.052)
-4 qr prior					-0.071	-0.084
					(0.059)	(0.060)
-3 qr prior					-0.037	-0.052
					(0.046)	(0.046)
-2 qr prior					-0.013	-0.020
					(0.052)	(0.053)
-1 qr prior					0.039	0.034
					(0.054)	(0.054)
1 qr after					0.165^{***}	0.142^{***}
					(0.054)	(0.055)
2 qr after					0.151^{***}	0.115^{**}
					(0.052)	(0.052)
3 qr after					0.040	0.016
					(0.056)	(0.057)
4 qr after					0.284^{***}	0.251^{***}
					(0.058)	(0.057)
5 qr after					0.353^{***}	0.314^{***}
					(0.053)	(0.052)
6 qr after					0.347^{***}	0.309^{***}
					(0.065)	(0.065)
7 qr after					0.380^{***}	0.329^{***}
					(0.053)	(0.055)
8+ qr after					0.271^{***}	0.230^{***}
					(0.042)	(0.043)
Fixed-effects						
Quarter	Yes	Yes	Yes	Yes	Yes	Yes
Player			Yes	Yes	Yes	Yes
Opponent Player				Yes		Yes
Fit statistics						
Observations	42,783	42,783	42,783	42,783	42,783	42,783
R^2	0.277	0.281	0.327	0.352	0.328	0.353
Within R^2	0.061	0.067	0.014	0.014	0.016	0.015

Note: This table shows the effects of APG on the $Move\,Quality$ by the player's nationality. We consider players in mainland China and South Korea as a treated group, while Japanese players constitute a control group. Models 1 to 4 estimate the differences in the $Move\,Quality$ among country groups before and after the release of APG, while Models 5 and 6 show their time-varying effects. Clustered standard-errors at a focal player level are in parentheses.
$^{*}\ p < 0.1;^{**}\ p < 0.05;^{***}\ p < 0.01$

Table 6: Effects of APG on move quality: Errors and a critical mistake as mechanisms

Dependent Variable:	Number of Errors		Magnitude of the Critical Mistake	
Model:	(1)	(2)	(3)	(4)
Variables				
Post	-0.055***	0.155***	-1.431***	4.124***
	(0.004)	(0.025)	(0.053)	(0.255)
Trend		-0.000		-0.028**
		(0.001)		(0.012)
Post × Trend		-0.009***		-0.233***
		(0.001)		(0.015)
Fixed-effects				
Player	Yes	Yes	Yes	Yes
Opponent Player	Yes	Yes	Yes	Yes
Fit statistics				
Observations	50,066	50,066	50,065	50,065
R^2	0.077	0.081	0.123	0.145
Within R^2	0.005	0.008	0.028	0.052

Note: This table shows the impact of APG on errors and a critical mistake by professional Go players before and after the release of Leela. A dependent variable for Models 1 and 2 is $Number\,of\,Errors$ and for Models 3 and 4 is $Magnitude\,of\,the\,Critical\,Mistake$. $Post$ refers to games played in the quarters after the first public introduction of APG in February 2017, and $Trend$ refers to the number of quarters passed since the first quarter in our sample. Clustered standard errors at a focal-player level are in parentheses. $^{*}\,p < 0.1$;$^{**}\,p < 0.05$;$^{***}\,p < 0.01$

Table 7: Mediation analysis on game winning: Move quality, errors, and a critical mistake

Dependent Variables:	Win	Move Quality	Number of Errors	Magnitude of the Critical Mistake				Win			
Model:	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
	Logit	OLS	OLS	OLS	Logit	Logit	Logit	Logit	Logit	Logit	Logit
Variables											
Rank	-0.461	2.501***	-0.369***	-7.452***	-0.548	-0.220	-0.427	-0.790	-0.546	-0.712	-0.838
	(0.573)	(0.292)	(0.101)	(1.155)	(0.532)	(0.537)	(0.535)	(0.567)	(0.572)	(0.570)	(0.568)
Rank Diff	-10.178***	1.023***	-0.021	-2.209***	-10.333***	-10.174***	-10.276***	-10.334***	-10.195***	-10.286***	-10.355***
	(0.490)	(0.162)	(0.065)	(0.683)	(0.483)	(0.484)	(0.486)	(0.490)	(0.491)	(0.494)	(0.493)
White	0.125***	-0.131***	0.047***	0.622***	0.146***	0.138***	0.148***	0.143***	0.135***	0.145***	0.149***
	(0.025)	(0.010)	(0.004)	(0.037)	(0.025)	(0.025)	(0.025)	(0.025)	(0.025)	(0.025)	(0.025)
7.5 Komi	-0.001	0.041**	-0.011	-0.102	0.001	0.004	0.004	-0.005	-0.002	-0.002	-0.004
	(0.053)	(0.019)	(0.008)	(0.080)	(0.053)	(0.053)	(0.053)	(0.053)	(0.053)	(0.053)	(0.053)
Post $\times$ Young	0.120***	0.209***	-0.025***	-0.508***				0.094**	0.115***	0.106**	0.094**
	(0.044)	(0.031)	(0.008)	(0.101)				(0.045)	(0.045)	(0.045)	(0.045)
Move Quality					0.132***			0.130***			0.084***
					(0.012)			(0.012)			(0.015)
Number of Errors						-0.206***			-0.211***		0.025
						(0.032)			(0.032)		(0.045)
Magnitude of the Critical Mistake							-0.031***			-0.031***	-0.021***
							(0.003)			(0.003)	(0.005)
Fixed-effects											
Player	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Quarter	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Opponent Player	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Fit statistics											
Observations	45,762	46,454	46,454	46,453	46,334	46,334	46,333	45,762	45,762	45,761	45,761
Squared Correlation	0.210	0.349	0.081	0.152	0.214	0.213	0.214	0.212	0.211	0.212	0.213
Pseudo R^2	0.170	0.145	0.102	0.030	0.173	0.172	0.173	0.172	0.171	0.172	0.172
BIC	71,587.716	140,067.468	56,742.687	273,520.002	72,508.324	72,577.867	72,510.675	71,487.035	71,549.677	71,484.283	71,473.918

Note: This table shows how $Move\,Quality$ leads to a winning a game. We test two mechanisms, $Number\,of\,Errors$ and $Magnitude\,of\,the\,Most\,Critical\,Mistake$. Models 1 to 4, respectively, indicate that, after the release of APG, young professional Go players are more likely to win, to improve $Move\,Quality$, to decrease $Number\,of\,Errors$, and to reduce Magnitude of the Critical Mistake. A dependent variable for Models 5 through 11 is whether a player wins a game. Models 5 to 7, respectively, show a player is more likely to win a game if the player's $Move\,Quality$ is greater, if the player's $Number\,of\,Errors$ are fewer, and if the player has a smaller $Magnitude\,of\,the\,Most\,Critical\,Mistake$. The finding is robust when we account for the differences in $Move\,Quality$ by age, as shown in Models 8 through 10. Model 11 presents the full specification that includes all relevant variables. Taken together, young players improve $Move\,Quality$, decrease $Number\,of\,Errors$, and reduce $Magnitude\,of\,the\,Most\,Critical\,Mistake$ after APG; these changes lead to eventual winning of a game. Clustered standard errors at a focal-player level are in parentheses. $^{*}p < 0.1;^{**}\ p < 0.05;^{***}\ p < 0.01$

Table 8: Effects of APG on move quality (by age): Heterogeneity by the number of moves

Dependent Variable:	Move Quality					
Moves:	1-30	1-60	1-90	1-120	1-150	1-180
Model:	(1)	(2)	(3)	(4)	(5)	(6)
Variables						
Rank	2.501***	2.203***	1.491***	1.214***	0.955***	0.944**
	(0.292)	(0.282)	(0.300)	(0.306)	(0.347)	(0.382)
Rank Diff	1.023***	0.647***	0.175	-0.034	-0.168	-0.238
	(0.162)	(0.177)	(0.195)	(0.209)	(0.242)	(0.261)
White	-0.131***	-0.124***	-0.107***	-0.092***	-0.069***	-0.048***
	(0.010)	(0.010)	(0.011)	(0.012)	(0.013)	(0.014)
7.5 Komi	0.041**	0.003	-0.004	-0.026	-0.035	-0.037
	(0.019)	(0.019)	(0.020)	(0.023)	(0.026)	(0.032)
Post $\times$ Young	0.209***	0.183***	0.169***	0.156***	0.131***	0.053
	(0.031)	(0.027)	(0.028)	(0.029)	(0.032)	(0.035)
Fixed-effects						
Quarter	Yes	Yes	Yes	Yes	Yes	Yes
Player	Yes	Yes	Yes	Yes	Yes	Yes
Opponent Player	Yes	Yes	Yes	Yes	Yes	Yes
Fit statistics						
Observations	46,454	46,438	46,284	45,310	41,492	33,510
R^2	0.349	0.267	0.210	0.168	0.141	0.139
Within R^2	0.014	0.010	0.006	0.004	0.003	0.001

Note: This table presents how the APG's influence on $Move\,Quality$ changes depending on the different ranges of moves. Models 1 to 6 increase the range of moves considered by 30 moves: hence, for instance, Model 1 considers moves 1 to 30, while Model 2 presents results from moves 1 to 60. Clustered standard errors at a focal player level are in parentheses.

$^{*}\ p < 0.1$; $^{**}\ p < 0.05$; $^{***}\ p < 0.01$

Appendix

A.1. Leela and Leela Zero

Leela, an AI-powered Go (APG) program, was released in February 2017; it included a stable, deep learning version using Monte Carlo Tree Search (MCTS). It was developed based on the AlphaGo algorithm of Google DeepMind (Silver et al., 2016).[i] Leela was the first APG to surpass human professional players and to be made publicly available on personal computers. A successor with a stronger open-source Go engine, Leela Zero, was released in October 2017 following the publication of an AlphaGo Zero research article by Google DeepMind (Silver et al., 2017). Unlike the original Leela, which uses human knowledge and heuristics in learning, Leela Zero uses only basic rules during self-training. Leela Zero is a faithful reimplementation of the famous Go engine, AlphaGo Zero, and has been made publicly available. We used Leela Zero to evaluate the moves of professional players.

Instead of training a Go engine using expensive Google tensor processing units (TPUs), Leela Zero adopts crowdsourcing infrastructure using graphics processing units (GPUs) via the open computing language (OpenCL) library.[ii] Leela Zero users can contribute their GPU resources to strengthen Leela Zero. Because of this crowdsourcing training, Leela Zero has rapidly improved over time and continues to improve. Leela Zero provides various Go analysis functionalities but these are not meant to be used directly. Several graphical user interface software programs support Leela Zero so that end users may utilize various functionalities without hassle. Examples of these interfaces include Lizzie, Sabaki, and GoReviewPartner.[iii]

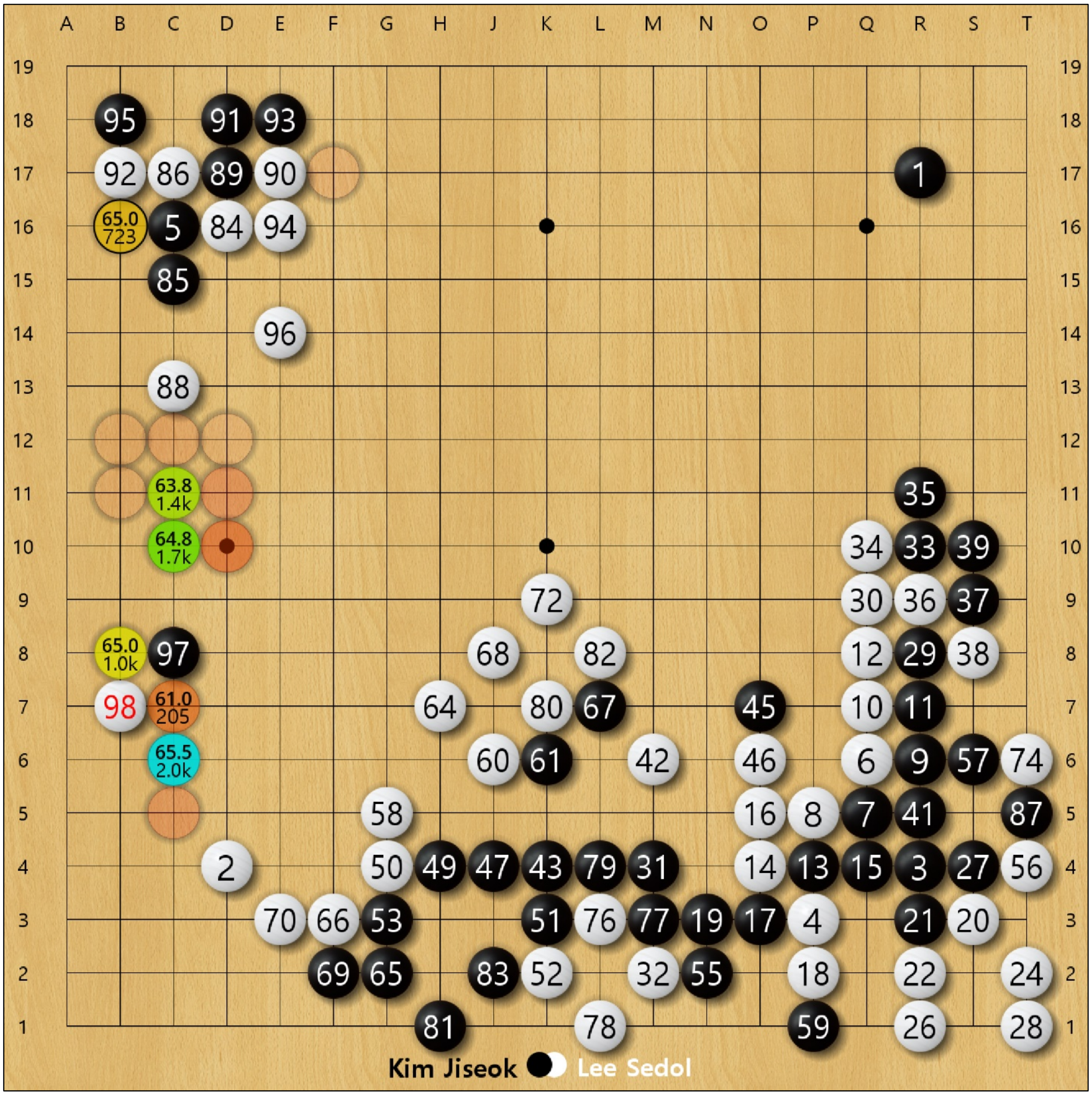
Leela Zero provides an in-depth analysis of the game, including recommendations for next moves. We visualize what Leela Zero provides for Go analysis using the Lizzie graphical user interface (GUI). Figure A.1 shows a recent Go match between the world champion, Lee Sedol, and Jiseok Kim. On the main board, the number on each stone shows the order in which that stone was placed on the board. After the opponent player made the 98th move (the white

---

[i] While AlphaGo and AlphaGo Zero proved the power of AI in Go games, they are not open-source software, nor are user-friendly interfaces provided.
[ii] For more information on OpenCL, please refer to https://www.khronos.org/opencl/.
[iii] These interfaces are available at: https://github.com/featurecat/lizzie/releases/ (Lizzie), https://sabaki.yichuanshen.de/ (Sabaki), and https://github.com/pnprog/goreviewpartner/ (GoReviewPartner).

stone on B7), Leela Zero recommended multiple moves for the focal player based on MCTS simulations. The cyan-colored point represents the recommended next move (i.e., AI's solution), which has a winning probability of 65.5 percent. The number below 65.5 shows that this probability is evaluated with 2,000 ("2.0k") simulations using MCTS.

Figure A.1: Leela Zero and Its Graphical User Interface (GUI)



Notes. This is a game between Sedol Lee (white stones) and Jiseok Kim (black stones) on July 26, 2019.

In addition, although not shown in Figure A.1, users can open additional windows that show future simulations of the game (the next sixteen predicted moves), the current winning probability, and how this has changed from the beginning of the game to the current point. These graphs help the user evaluate the status of the game and analyze how each move changes the winning probability.

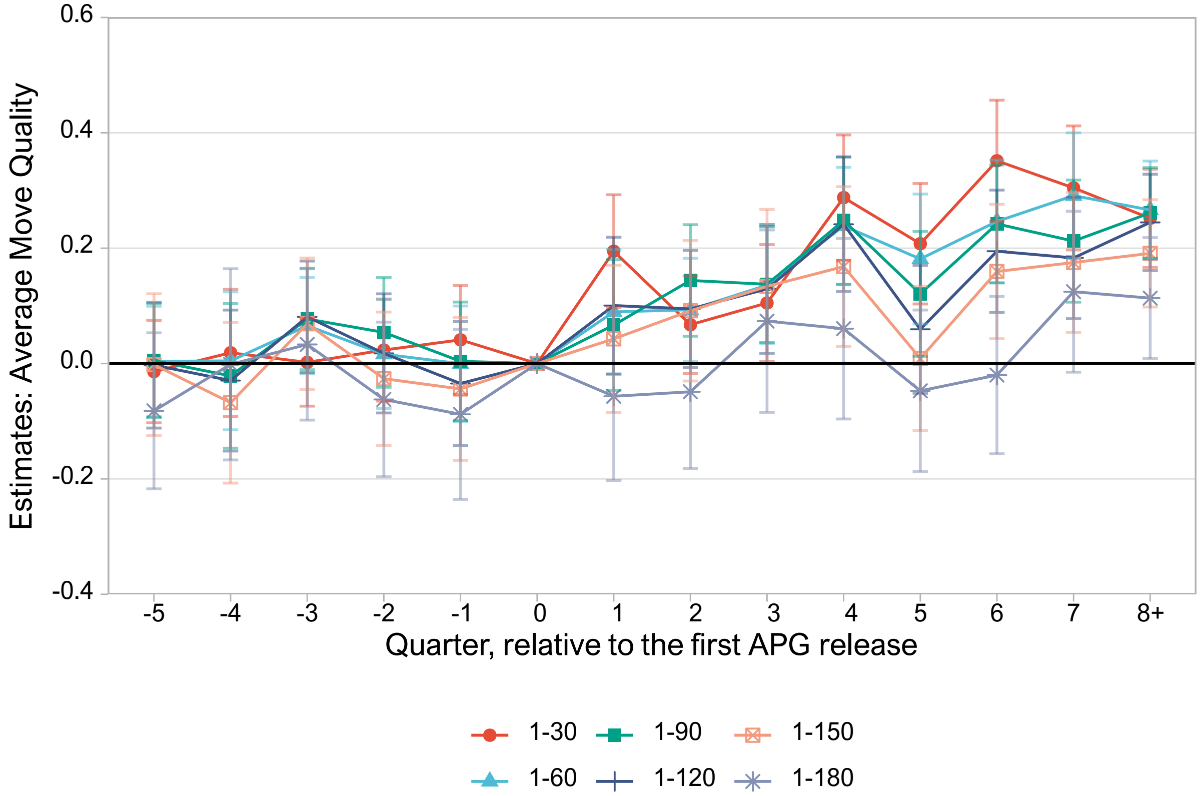
A.2. Implementation Details

We used the official version of Leela Zero to analyze the collected Go games. Since Leela Zero improves over time, for analysis we fixed the Leela Zero model trained on May 23, 2020. We worked with the GoReviewPartner program to analyze a batch of games. We first analyzed each SGF-formatted file using Leela Zero and saved it into an RSGF-formatted file with Leela_zero_analysis.py code. Then each RSGF file was converted to a CSV format file using r2csv.py code for analysis.

We set five seconds as the time budget for Leela Zero to analyze the winning probability of each move. The five-second time budget is the same setting used in the AlphaGo Zero paper (Silver et al. 2017) to analyze the relative performance among AI Go engines. We ran Leela Zero on a Linux system with four Nvidia Titan-X GPUs and an Inter Core i7-6800K CPU. Each game analysis took approximately twenty minutes; with a single GPU, it would have taken 345 full days to analyze all 25,033 games. We finished our game analysis in about three months by running two to eight GPUs in parallel.

A.3. Additional Figures and Tables

Figure A.2: Effects of APG on move quality by age: Heterogeneity by the number of moves



Note: This figure illustrates how the changes in average $Move Quality$ differ by the number of moves. Beginning with the opening strategy of the first thirty moves, we incrementally add thirty additional moves (up to 180 moves). This is a regression version of Figure 8.

Table A.1: Robustness check: Average age as a cutoff for young and old players

Dependent Variable:	Move Quality			
Model:	(1)	(2)	(3)	(4)
Variables				
Young	0.142***	-0.057**		
	(0.023)	(0.024)		
Rank	0.783***	0.750***	1.608***	2.396***
	(0.036)	(0.036)	(0.227)	(0.284)
Rank Diff	0.124***	0.110***	0.058**	0.908***
	(0.027)	(0.027)	(0.024)	(0.163)
white	-0.133***	-0.133***	-0.131***	-0.131***
	(0.010)	(0.010)	(0.009)	(0.010)
7.5 Komi	0.027*	0.026*	0.026	0.043**
	(0.016)	(0.016)	(0.016)	(0.019)
Post × Young		0.370***	0.316***	0.289***
		(0.030)	(0.032)	(0.032)
Fixed-effects				
Quarter	Yes	Yes	Yes	Yes
Player			Yes	Yes
Opponent Player				Yes
Fit statistics				
Observations	46,454	46,454	46,454	46,454
R^2	0.276	0.281	0.325	0.350
Within R^2	0.065	0.072	0.015	0.015

Note: This table re-estimates Table 3 using the average age (instead of median age) as the cutoff separating young versus old players. Clustered standard errors at a focal-player level are in parentheses.
$^{*}\ p < 0.1$; $^{**}\ p < 0.05$; $^{***}\ p < 0.01$

Table A.2: Robustness check: Three age categories (Young, Mid, and Old)

Dependent Variable:	Move Quality			
Model:	(1)	(2)	(3)	(4)
Variables				
Mid	0.095***			
	(0.022)			
Young	0.142***			
	(0.025)			
Rank	0.804***	0.773***	1.503***	2.340***
	(0.036)	(0.036)	(0.243)	(0.296)
Rank Diff	0.132***	0.119***	0.059**	0.962***
	(0.027)	(0.027)	(0.024)	(0.162)
White	-0.133***	-0.134***	-0.131***	-0.131***
	(0.010)	(0.010)	(0.009)	(0.009)
7.5 Komi	0.026	0.020	0.025	0.043**
	(0.016)	(0.016)	(0.016)	(0.019)
Post × Mid		0.224***	0.221***	0.198***
		(0.031)	(0.033)	(0.033)
Post × Young		0.295***	0.328***	0.302***
		(0.032)	(0.040)	(0.040)
Fixed-effects				
Quarter	Yes	Yes	Yes	Yes
Player			Yes	Yes
Opponent Player				Yes
Fit statistics				
Observations	46,454	46,454	46,454	46,454
R^2	0.276	0.280	0.325	0.350
Within R^2	0.065	0.070	0.014	0.014

Note: This table re-estimates Table 3 using three age groups instead of two: those younger than twenty (“Young”), those in their twenties (“Middle”), and those thirty or older (“Old”). Clustered standard errors at a focal-player level are in parentheses.

$^{*}\ p < 0.1;^{**}\ p < 0.05;^{***}\ p < 0.01$

Table A.3: Placebo test: Random reassignment of the age group

Dependent Variable:	Move Quality			
Model:	(1)	(2)	(3)	(4)
Variables				
Young	0.003	-0.008		
	(0.018)	(0.020)		
Rank	0.857***	0.856***	2.366***	3.259***
	(0.036)	(0.036)	(0.228)	(0.274)
Rank Diff	0.142***	0.142***	0.069***	1.169***
	(0.027)	(0.027)	(0.024)	(0.162)
White	-0.133***	-0.133***	-0.131***	-0.131***
	(0.010)	(0.010)	(0.009)	(0.010)
7.5 Komi	0.049***	0.049***	0.024	0.042**
	(0.017)	(0.017)	(0.016)	(0.019)
Post × Young		0.038	0.010	-0.043
		(0.029)	(0.028)	(0.027)
Fixed-effects				
Quarter	Yes	Yes	Yes	Yes
Player			Yes	Yes
Opponent Player				Yes
Fit statistics				
Observations	46,454	46,454	46,454	46,454
R^2	0.274	0.274	0.323	0.348
Within R^2	0.062	0.062	0.010	0.011

Note: This table shows the estimates after players are randomly reassigned to age groups. Clustered standard errors at a focal-player level in are parentheses.
$^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$

Table A.4: Robustness check: Monthly fixed effect

Dependent Variable:	Move Quality			
Model:	(1)	(2)	(3)	(4)
Variables				
Young	0.104***	-0.044**		
	(0.020)	(0.021)		
Rank	0.819***	0.803***	1.683***	2.544***
	(0.037)	(0.037)	(0.240)	(0.295)
Rank Diff	0.138***	0.131***	0.066***	1.021***
	(0.027)	(0.027)	(0.024)	(0.162)
White	-0.134***	-0.133***	-0.131***	-0.131***
	(0.010)	(0.010)	(0.009)	(0.010)
7.5 Komi	0.022	0.019	0.013	0.027
	(0.017)	(0.016)	(0.016)	(0.019)
Post × Young		0.259***	0.208***	0.192***
		(0.028)	(0.031)	(0.031)
Fixed-effects				
Quarter	Yes	Yes	Yes	Yes
Player			Yes	Yes
Opponent Player				Yes
Fit statistics				
Observations	46,454	46,454	46,454	46,454
R^2	0.278	0.282	0.327	0.351
Within R^2	0.064	0.068	0.013	0.013

Note: This table re-estimates Table 3 using month-fixed effects instead of quarter-fixed effects. Clustered standard errors at a focal-player level are in parentheses.

$^{*}\ p < 0.1$; $^{**}\ p < 0.05$; $^{***}\ p < 0.01$

Table A.5: Robustness check: Alternative definitions of early moves

Dependent Variable:	Move Quality					
	First 40 moves (1-40)		First 50 moves (1-50)		First 60 moves (1-60)	
Model:	(1)	(2)	(3)	(4)	(5)	(6)
Variables						
Rank	1.823***	2.755***	1.703***	2.514***	1.588***	2.203***
	(0.224)	(0.277)	(0.239)	(0.296)	(0.232)	(0.282)
Rank Diff	0.041	1.048***	0.035	0.894***	0.013	0.647***
	(0.027)	(0.172)	(0.028)	(0.181)	(0.029)	(0.177)
White	-0.128***	-0.128***	-0.124***	-0.125***	-0.123***	-0.124***
	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)
7.5 Komi	0.023	0.033*	0.013	0.017	0.000	0.003
	(0.016)	(0.019)	(0.015)	(0.019)	(0.016)	(0.019)
Post $\times$ Young	0.220***	0.204***	0.214***	0.202***	0.195***	0.183***
	(0.029)	(0.029)	(0.028)	(0.028)	(0.028)	(0.027)
Fixed-effects						
Quarter	Yes	Yes	Yes	Yes	Yes	Yes
Player	Yes	Yes	Yes	Yes	Yes	Yes
Opponent Player		Yes		Yes		Yes
Fit statistics						
Observations	46,452	46,452	46,448	46,448	46,438	46,438
R^2	0.295	0.320	0.269	0.294	0.242	0.267
Within R^2	0.012	0.013	0.011	0.012	0.010	0.010

Note: This table re-estimates Table 3 with different definitions of early opening moves: the first 40, 50, and 60 moves. Clustered standard errors at a focal-player level are in parentheses.
* $p < 0.1$;** $p < 0.05$;*** $p < 0.01$

Table A.6: Placebo test: Random reassignment of the nationality

Dependent Variable:	Quality of Move			
Model:	(1)	(2)	(3)	(4)
Variables				
Treat	-0.003	-0.001		
	(0.019)	(0.021)		
Rank	0.861***	0.861***	2.507***	3.561***
	(0.038)	(0.038)	(0.240)	(0.281)
Rank Diff	0.154***	0.154***	0.061**	1.325***
	(0.031)	(0.031)	(0.027)	(0.175)
White	-0.129***	-0.129***	-0.129***	-0.126***
	(0.010)	(0.010)	(0.010)	(0.010)
7.5 Komi	0.055***	0.055***	0.020	0.030
	(0.017)	(0.017)	(0.017)	(0.020)
Post $\times$ Treat		-0.003	-0.005	-0.004
		(0.031)	(0.031)	(0.030)
Fixed-effects				
Quarter	Yes	Yes	Yes	Yes
Player			Yes	Yes
Opponent Player				Yes
Fit statistics				
Observations	42,783	42,783	42,783	42,783
R^2	0.275	0.275	0.324	0.350
Within R^2	0.059	0.059	0.010	0.011

Note: This regression shows the re-estimated results of Models 1 to 4 as reported in Table 5. We randomly reassign players' nationality. Clustered standard errors at a focal-player level are in parentheses.
$^{*}\ p < 0.1;^{**}\ p < 0.05;^{***}\ p < 0.01$

Appendix References

Silver D, Huang A, Maddison CJ, Guez A, Sifre L, van den Driessche G, Schrittwieser J, et al. (2016) Mastering the game of Go with deep neural networks and tree search. *Nature* 529(7587): 484–489.

Silver D, Schrittwieser J, Simonyan K, Antonoglou I, Huang A, Guez A, Hubert T, et al. (2017) Mastering the game of Go without human knowledge. *Nature* 550(7676): 354–359.