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

Sukwoong Choi[†]
Job Market Paper

Namil Kim[‡]

Junsik Kim[§]

Hyo Kang[**]

Please click here for the latest version

July 14, 2022

Abstract

Firms increasingly utilize AI to assist or replace human tasks. However, AI can also train humans and make them better. We study how the AI's instructional role improves human decision-making in the professional Go games where an AI-powered Go program (APG) unexpectedly surpassed the best human player, surpassing the best human knowledge and skill accumulated over thousands of years. To isolate the learning-from-AI effect, we compare the quality of human moves to that of AI's superior solutions, before and after the initial public release of an APG. Our analysis of 750,990 moves in 25,033 games suggests that APG's training significantly improved the players' move quality—reducing the number of errors and the magnitude of the most critical mistake. The improvement is most prominent in the early stage of a game when uncertainty is higher. Further, younger players benefit more than older players, suggesting generational inequality in learning from AI.

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

---

[*] We gratefully acknowledge valuable comments from Nur Ahmed, Sinan Aral, Nan Jia, Thorbjørn Knudsen, Amalia R. Miller, Abhishek Nagaraj, Ananya Sen, and Neil Thompson. We also thank organizers and audiences of the 2021 NBER Economics of AI Conference, the Fall 2021 MIT IDE Lunch Seminar, the 2021 Academy of Management Meeting, the 2021 Strategic Management Society (SMS) Annual Conference, the 2021 Conference on Artificial Intelligence, Machine Learning, and Business Analytics, the 2022 Yonsei-Barun ICT Research Colloquium, the 2022 IP and Innovation (IPI) Seminar, the 3rd AI and Strategy Consortium, and the 2022 Wharton Technology and Innovation Conference, the 2022 Wharton/Columbia Management, Analytics and Data Conference, Samsung Advanced Institute of Technology (Targeted to Vice Presidents of Technology), and seminar participants at MIT and Rotterdam School of Management. This paper was previously circulated under the title, "Strategic Choices with Artificial Intelligence." All errors are our own.

[†] MIT Sloan School of Management, MIT Initiative on the Digital Economy. sukwoong@mit.edu.
[‡] Corresponding author. School of Management, Harbin Institute of Technology. namil.kim@hit.edu.cn.
[§] School of Engineering and Applied Sciences, Harvard University. mibastro@gmail.com.
[**] Marshall School of Business, University of Southern California. hyokang@marshall.usc.edu.

1. INTRODUCTION

Artificial intelligence (AI) has developed substantially to date, and its capabilities have reached or even surpassed those of humans (Rai et al., 2019). For instance, AI has outperformed human experts in strategic gameplay (Silver et al., 2017), medical diagnosis (Kim et al., 2021), bioinformatics (Senior et al., 2020), and drug discovery and development (Savage, 2021; Smalley, 2017). The rapid advancement of AI is transforming the future of professional work (De Cremer, 2020; Krakowski et al., 2022). AI, in particular, helps workers yield better performance by providing real-time assistance to their tasks (e.g., Allen & Choudhury, 2022; Choudhury et al., 2020; Lebovitz et al., 2022; Tong et al., 2021). Financial analysts who use AI-based-assistant software make more accurate stock price forecasts than those who do not (Cao et al., 2021). Medical professionals reduce the uncertainty of their diagnoses and improve diagnostic quality by comparing their own judgments to those provided by AI-based solutions (Lebovitz et al., 2022). AI also helps medical coders record patient conditions by suggesting standardized codes for filling in medical charts. Consequently, the quality of charts has improved substantially as has the productivity of medical coders (Wang et al., 2019).

Although extant studies on AI have taken an important step from focusing on the substitute roles of AI toward considering its complementary roles, these studies have largely examined how AI provides real-time assistance to human tasks. The impact of AI could extend far beyond the assistant role, but limited attention has been paid to AI's more fundamental effect, namely, how AI trains human professionals.[1] The performance gap between AI and humans suggests that humans can now learn from AI and catch up with it in areas where they are currently outpaced. AI provides a new metric by quantifying the expected outcomes of each alternative that historically had been guessed, heuristically and tacitly, by long-standing customs or learning-by-doings. In this way, AI could improve humans' intrinsic decision-making abilities—even when AI assistance is not readily available in real time. Put differently, when AI is better than humans, it can make humans better. The distinction between AI's roles in assisting versus instructing is crucial, as the former does not necessarily imply the latter. The assistant role

[1] One study exception is that Gaessler and Piezunka (2021), which focused on how an individual's performance varies by differential learning opportunities with AI.

focuses on how AI could take charge of some tasks—often as a form of division of labor—but the instructional role emphasizes the improvement of fundamental human capabilities.

Filling the gap in the literature, we study how AI trains human professionals by improving their heuristics and decision-making practices. We examine (1) whether AI-based training improves the quality of human decisions and (2) the mechanisms by which performance is improved. Also, to shed light on AI's differential effects (Allen & Choudhury, 2022; Choudhury et al., 2020), we consider openness to new technologies and the ability to utilize them (Barth et al., 2020; M.G. Morris et al., 2005; Tams et al., 2014) to examine (3) age and exposure to AI as key drivers that could affect AI's instructional effects.

Empirical studies of this topic are challenging due to 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, the limited availability of data has constrained researchers from examining its impact (Seamans & Raj, 2018).

We study professional players of Go, a strategy board game that provides a unique opportunity to overcome these challenges. Over thousands of years professional Go players have accumulated knowledge, experience, and skills in this game. Yet the introduction of an 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. In the historic Go match (AlphaGo vs. Sedol Lee) held in 2016, AI beat the best human professional player for the first time and by a large margin. Shortly after this event, the first open-access 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 Leela's release.

The great advantage of the Go context is that it allows us to observe every single decision of professional Go players before and after the public release of the APG; the entire move history is well archived and maintained for all major games. Furthermore, the APG can calculate the probability of winning for every move and can even perform these calculations for earlier games that were played before the APG was released. These calculations were made of 750,990 decisions by 1,242 professional Go players in 25,033 major games held from 2015 through 2019. We assessed the quality of each player's moves by comparing their move-level probability of winning to that of the APG's best solution. It is important to note that professional Go players are not allowed to use APGs (or any type of assistant tools) in a professional match. Thus, any changes in move quality (or winning probability) after training with the APG are attributable to changes in human capabilities (i.e., learning) but not to real-time APG assistance.

The results show that the quality of moves by professional Go players improved substantially following the release of the APG. Before the release, the winning probability of each move by professional Go players averaged 2.47 percentage points lower than the best solutions calculated by the APG. This gap decreased by about 0.756 percentage points (or 30.5 percent) after the public release of the APG.

We also explore the mechanisms through which professional players achieve a higher probability of winning. Our mediation analysis reveals that a focal player's improvement in the quality of moves is achieved 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) 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 (or 31.8 percent) and the magnitude of the most critical mistake decreased by 4–7 percentage points (or 20.9 percent). Additional analyses indicate that the improvement in move quality eventually leads to the final win of the game by decreasing the number of errors and reducing the magnitude of the most critical mistake. This effect is most prominent in the early stage of a game where uncertainty is higher 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 study is one of the first 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 has meaningful implications for AI's instructional role, notably how it could educate and nurture professional decision-making capabilities in a fast-paced, uncertain environment; this aspect is distinct from AI's real-time assistant roles. Further, the fact that the young benefit more from APGs has important implications for digital literacy and for potential inequality in accessing, adopting, and utilizing AI. Finally, our contexts and findings provide meaningful managerial implications. Playing Go is similar to decision-making by executives and managers because it requires players to analyze the environment, make judgments and decisions, and reflect on the results within a limited time (Miric et al., 2020; Reeves & Wittenburg, 2015). Moreover, Go players need to consider the perspective of competitors and to possess intuitive techniques under the uncertainty and time constraints similar to those that arise in real business environments (Anderson, 2004; Roman & Vyas, 2021; Wiseman, 2016). Consequently, the findings of this study relate to how business managers can make better decisions by adopting and learning from AI.

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 & March 1963; Galbraith 1974; Simon 1955, 1958). This 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 & March, 1993).

Acknowledging these limitations, researchers have studied incompleteness in managerial decision-making (Dane & 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 significant decisions for firms, they often fail to follow the procedures for rational choices (Mousavi & 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 & 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 & March 1963; Gavetti & Levinthal 2000). However, benefits of these choice models are marginal in alleviating the aforementioned limitations to optimal decision-making.

Literature on information technology (IT) 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 (Bapna et al., 2013; Bresnahan et al., 2002; Brynjolfsson & 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 & 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 (L. Wu et al., 2020). Overall, the adoption and utilization 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 traditional data-driven approaches for several reasons (Agrawal et al., 2018; Smith, 2019). First, AI can make inferences by self-learning. AI, therefore, is better suited for discovering hidden patterns and can conduct insightful tasks that need human-like "intuition." Second, AI performs predictions and judgments with high accuracy, and the accuracy increases exponentially with the number of training sessions and the quantity of data. With AI, therefore, humans can revisit their decision-making practices which may have yielded inferior decisions otherwise. Thanks to its superior predictive capability, compared to classical statistics and econometric techniques, AI algorithms have been applied to a variety of different decision-making problems (Athey & Imbens, 2019; Blei & Smyth, 2017).

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 court (Kleinberg et al., 2017), discovering protein structure in biology (Senior et al., 2020), and playing strategic games (Schrittwieser et al., 2020), among others. 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. In other words, AI helps 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. For example, when a radiologist uses AI to read a chest X-ray, AI can show the probability of an individual having some predefined disease within in a few seconds. Similarly, when a professional Go players use AI, they can immediately learn the winning probability associated with each possible move and distinguish better moves.

Based on AI's superior predictive power, managers have several incentives to learn from AI. First, classical decision-making theory proposes three conditions that face humans making decisions: certainty, risk, and uncertainty (Langholtz et al., 1993). Without knowing values associated with each choices individuals make decisions under uncertainty, which may lead to unfavorable outcomes. AI, in contrast, provides accurate, predicted values and thereby reduces the uncertainty associated with choices. Managers who learn from AI therefore can make decisions under less uncertainty. Second, the unified theory of acceptance and use of technology (UTAUT) (Venkatesh et al., 2003) emphasizes that managers actively accept and utilize IT when they expect superior performance from its use. Informed managers should thus actively adopt AI in decision-making processes and consequently will achieve superior performance.

Lastly, when managers begin to utilize AI, they learn to improve their decision-making ability. As they choose repeatedly, managers update their choice evaluation rules from feedback (Yechiam & Busemeyer, 2005). Currently AI does not yet explain why a particular choice has better outcomes (Hagendorff & Wezel, 2020), but it can provide feedback on whether an individual choice is good or poor. Repeatedly comparing their choices with those of AI, managers can update or revise their evaluation criteria when comparing alternatives. Therefore, being equipped with the ability to make better evaluative choices, they can make better decision even without real-time AI assistance. In sum, AI can train human professionals and improve the quality of their decisions—especially when the tasks are complex and uncertain.

2.2. Differential adoption and utilization of AI by age

AI has strong potential to train employees and improve their decision-making, but not all professionals benefit from AI to the same extent. Despite its superior prediction performance, AI and its related products and services are relatively new and do not have a proven record in terms of credibility and stability. Professionals thus perceive AI-powered products as generally riskier to adopt and utilize when making important decisions, with a tradeoff between performance and risk (Cadario et al., 2021; Lebovitz et al., 2021). The literature on the differential effects highlights the role of age in digitization and technology (e.g., Barth et al., 2020; Ghasemaghaei et al., 2019; Tams, 2022). Prior studies suggest that age is an important factor in adopting and utilizing new technology (Weinberg, 2004). In the context of AI, extant studies find mixed results. Wang et al. (2019) studied medical coders in hospitals who use AI suggestions for chart coding and found that the productivity of younger employees improved more than that of older employees. In contrast, Choudhury et al. (2020) found that senior employees, who possess greater domain expertise than younger workers, tend to gain more complementary benefits from AI. Allen and Choudhury (2022) then suggest an inverted U-shaped relationship wherein employees with moderate experience are better able to utilize the algorithm tool.

Extending existing theories and arguments on age and digitization/AI, we argue that age could affect the adoption and utilization of AI through three major channels: risk-aversion, search behavior, and absorptive capacity. First, young professionals tend to be more open to new technologies and less risk-averse than established professionals. Hambrick and Mason (1984) argue that managers' ages are closely associated with their strategic decisions. Tyler and Steensma (1998) further find that young managers pursue riskier strategies in terms of adopting and utilizing new technology. These arguments apply to the emerging AI-powered products, and young professionals should be more inclined than senior professionals to utilize AI in their decision-making.

Second, senior professionals tend to rely more on the knowledge and experience they have accumulated than do juniors (Cohen & Levinthal, 1994; Nerkar, 2003). In other words, senior professionals conduct exploitative searches and make choices that are path-dependent on past records. The cumulated experience and record, however, may serve as a liability of success and may delay or deter seniors' adoption of AI. In contrast, junior professionals have less experience and fewer established routines. They are more likely to make decisions that are not dependent on past experience and to have greater incentives to explore new ways of working, leading to active adoption and utilization of AI in training and decision-making.

Third, in addition to the higher incentives, young professionals have better absorptive capacity (Cohen & 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. AI is not a stand-alone technology but is part of a broader digital computing system; it requires a certain understanding of computer algorithms, hardware, operating systems, etc., which young professionals are more likely to possess. Empirical studies support the argument that younger employees are more qualified and more likely to adopt new information and communications technologies (de Koning & Gelderblom, 2006; Meyer, 2011; Michael G. Morris & Venkatesh, 2000; Schleife, 2006).

Taken together, the impact of AI on human decision-making should have differential effects by age, and junior professionals should have greater incentives and ability to benefit from AI.

3. EMPIRICAL STRATEGY

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 a grid 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. More than ten major professional tournaments, sponsored by large corporations, are held throughout the year in each country. 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.[2]

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).[3] 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). The APG context, therefore, can be distinct from the general development of IT; it is about high-dimensional calculations and predictions that only become possible with AI and deep-learning algorithms.

3.1.2 AI's entrance into Go

Even if the latest supercomputers cannot calculate all possible moves in Go, recent advancement in deep-learning algorithms have improved AI remarkably. 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

[2] 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.

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

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).

AlphaGo's success 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 could surpass the best human Go player. The match completely changed how professional Go players learned and practiced Go; since the public release of the APG in 2017, all professional players have learned from APGs —such as Leela Zero, KataGo, and Handol (Somers, 2018).[4] Figure A.1 in Appendix A shows a snapshot of a Go game between two professional players and how an Leela Zero analyzes the game; it illustrates (on the upper left corner) how the winning probability changed with each move made 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 using the Elo rating system.[5] 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, which are based on deep-reinforcement learning, far exceed 4,000. Given this gap in Elo ratings, even top professional Go players have no chance of winning against APGs. Ke Jie, who ranked second in the 2020 world Go rankings, admitted that "AlphaGo is more like the god of Go" (Mozur, 2017). Put differently, the moves selected by APGs yield the highest probability of winning and even the best professional Go player can learn a lot from APGs.

"Insert Figure 1 here"

3.1.4 How does Go resemble managerial decision-making?

The decisions made in each move in Go share many aspects of managerial decision-making in a

[4] Before AI, professional Go players learned from books and past games. They also held group discussions (e.g., post-match game reviews), but it was generally impossible to quantitatively analyze the moves and games.

[5] 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.

complex, competitive environment. Studies have indicated the link between decision-making in games and real businesses because both players and managers should analyze the environment, make judgments and decisions, and reflect on the results (Anderson, 2004; Miric et al., 2020; Reeves & Wittenburg, 2015). For example, Simon (1987) argues that the intuitive skills required for managers are similar to the intuitive skills of chess masters. Mintzberg (1971, 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, 1971). Similarly, professional Go players face a complex, competitive environment and are forced to make a series of important decisions within strict time restrictions (Roman & Vyas, 2021; Wiseman, 2016); when players' allotted time ends, they must make their move within thirty seconds. To win, a player must think from their competitor's point of view. Go strategy involves securing unique and superior positioning by making moves in consideration of the opponent's point of view. Strategic, differentiated positioning gives a player greater advantage, increasing the chances of defeating the opponent. In businesses, likewise, it is necessary for firms to be unique and to offer products and services that are differentiated from those of competitors, as confirmed by executives and practitioners. 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 very beginning. Business management is the same" (Kwon et al., 2014). Also, LG Economic Research Institute (2004) published a report, *Learning Business Strategy from the Principles of Go*, highlighting the observation 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 an 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; its February 2017 update utilized the deep-learning algorithm used in AlphaGo. A few months later, a new version, Leela Zero, was developed after AlphaGo Zero. Upon their release, Leela and Leela Zero gained wide attention from media as well as professional players; the two programs have been viewed as the world's most successful open-source Go engines based on AI (Somers 2018). Leela—which provided a set of best possible moves with the winning probability of each alternative—had substantial impact on Go players. We describe how Leela and Leela Zero work in Appendix A.

Importantly, the development of APGs did not arise from demands of Go players. Before AlphaGo, Go programs could only play at the level of human amateurs, and professional players did not believe that they could ever be beaten by computer programs. 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, who had no interest in playing Go himself, also made it clear that he wanted to understand how deep learning worked. 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 APGs on the quality of moves by professional Go players. In February 2017 a major update of Leela adopted the AlphaGo-based deep-learning algorithm and outperformed the best human player; this is our event of interest. 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 thirty 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 to the way players lead the game (Noh, 2016). 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 & 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., Chinese, Japanese, South Korean, Taiwanese, and others), genders, and annual rankings of professional players. We multiply 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 use it to compare the difference in winning probability between a move made by a professional player and Leela Zero's suggested move, a move that would increase the winning probability more than any alternative move.

Our main dependent variable is $Move\ Quality_{ig}$, which represents the average difference in winning probability of the focal player i's move compared to the APG's corresponding solution for the first thirty moves of a game g (i.e., the game's 1, 3, 5, …, 29th move if the focal player moves first or 2, 4, 6, …, 30th moves otherwise). For each game, we separately calculate the value of the move qualities for each player i (black stone holder and white stone holder):

$$Move\ Quality_{ig} = \frac{\sum_{n=1}^{15} \left(\begin{matrix} The\ winning\ prob.\ of\ the\ focal \\ player\ i's\ n^{th}\ move\ in\ a\ game\ g \end{matrix} - \begin{matrix} The\ winning\ prob.\ of\ the\ APG's \\ solution\ to\ the\ move\ in\ a\ game\ g \end{matrix} \right)}{15}$$

where n represents the order of the focal player's move. 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 the APG indicates higher-quality moves by the player. For instance, if a player places stones as suggested by the APG for all moves, the average difference in winning probability between the player and the APG 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 the APG.

In Appendix A.1, we describe how Leela Zero provides in-depth analysis of a game between two professional players. 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. The implementation and calculation details are provided in Appendix A.2.

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 shows that the professional player's winning probability for the first thirty moves in a game averages 2.01 percentage points less than that of the APG's best move. This is a substantial difference because the difference of two percentage points for each move accumulates throughout the game—from the first move to the last. 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 identify 1,242 professional players from Go matches held 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 Leela, the first APG that surpassed human performance. This model-free illustration shows that, before APGS, $Move\ Quality_{ig}$ was relatively low and stable over time. 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 opposing player. Y_{ig} is $Move\ Quality_{ig}$. $Post_g$ is equal to 1 if a Go game is played in a quarter 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 APGs 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 the APG's public release.

"Insert Table 2 here"

It is possible that professional players' performance had been improving over time and drove 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 improved slowly over time. Importantly, the coefficient of the interaction term ($\beta = 0.116, p < 0.01$) shows that there were much larger—i.e., about seventeen times greater—improvements following the public release of the APG, even after performance trends are taken into account. The effect in the 10th quarter (i.e., the first quarter after the APG release) is 0.222 ($-1.008 + 0.007 \times 10 + 0.116 \times 10$).

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, Panel (a) the model-free illustration of two different age groups: young and old. This figure shows that $Move\ Quality_{ig}$ was relatively stable and similar among the two groups before the APG while the increase in $Move\ Quality_{ig}$ is notably greater for the young group after the release.

"Insert Figure 3 here"

We thus test whether the 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 control variables at the player or game levels, such as *Komi*, *White*, *Rank*, and *Rank differences between players*. $Young_i$ is an indicator variable equal to 1 if the player's age is below the median age of all players (i.e., less than twenty-eight years) as of Leela's public release on February 2017, 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 younger players is 0.273 percentage points (or 11 percent) greater than that for older players.

To check whether our results are robust when players' inborn characteristics are considered, 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 younger group, whose quality of moves improved by 0.209–0.273 percentage points (or 8.5–11 percent) over that of the older group, even after including the players' fixed effects. 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 numbers of different moves for opening strategies (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. The symbol z represents the indicators for time leads and lags—that is, the number of quarters before or after the public release of the APG.

Table B.1 of Appendix B, columns 1 and 2, show the detailed regression results, and Figure 3(b) 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 the APG, the estimates are large and statistically significant. The improved quality by younger players is large and persistent.

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 younger versus the older group; this increases the cutoff age from twenty-eight years to thirty-three years. The results, provided in Table B.2 of Appendix B, 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 age thirty or older ("Old"). The results are provided in Table B.3 of Appendix B. The estimates for $Post_g \times Young_i$ ($\beta = 0.302, p < 0.01$ in column 4) and $Post_g \times Middle_i$ ($\beta = 0.198, p < 0.01$ in column 4) 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 have captured spurious variations when testing 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 B.4 of Appendix B, 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 B.5 of Appendix B shows that the results are consistent with this alternative. Figure B.1 of Appendix B graphically illustrates the results obtained from the models with the distributed time leads and lags at the month level; these results are similar to those shown in Figure 3(b). We confirm the parallel time trend before the release of the APG and the significant effect of APG on move quality after the event.

4.3.5 Opening strategy with different numbers 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 of the game. The results, shown in Table B.6 of Appendix B, are robust to these alternative definitions, confirming that the operationalization of early moves does not drive the findings.

5. FURTHER ANALYSES

5.1. Exposure to and interest in AI by country

It is possible that there was a general improvement in the performance of professional Go players around the APG release 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 with 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 (AlphaGo vs. Sedol Lee, February 2016 held in South Korea and AlphaGo vs. Ke Jie, May 2017 held in China) were held between each country's best player and an APG, but none was held in Japan. This reflects the countries' interest levels in APG, which further affected their post-event adoption and utilization of it. A Google Trends search 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 4 (compared to China's score of 100 and South Korea's score of 92, as reference points; for details, see Appendix C).

"Insert Figure 4 here"

This motivated us to estimate a difference-in-difference model that compares players in countries that are significantly affected by APGs (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 APGs. Figure 4(a) illustrates the similarity among the three countries in the model-free average $Move\ Quality_{ig}$ before the APG. However, the average $Move\ Quality_{ig}$ increases more rapidly for Chinese and South Korean players, while the improvement is less for Japanese players. The fact that our control group is 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. The symbol 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 4 shows the results from the difference-in-differences estimation. In the previous main analysis, the quality increased by 0.756 percentage points (Table 2, column 1). In the difference-in-differences model, the magnitude is smaller: 0.315 to 0.230 percentage points (Table 4, 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 the 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 4 here"

Columns 3 and 4 in Table B.1 of Appendix further show the time-varying effects of the release of the APG, *Leela*, on move quality. The results in column 4 are illustrated graphically in Figure 4(b). We find no evidence for the increase in move quality for Chinese and South Korean players compared to Japanese players 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 responsible for improvement in the quality of moves by professional players.

We also conducted a placebo test to check whether our findings are driven by spurious variations in players' nationalities. We *randomly* reassign players to one of three nationalities—Chinese, Japanese, and South Korean—and estimate the same models. Under this placebo test, we expect that the nationality effect is minimal and statistically not distinguishable from zero. As shown in Table B.7 of Appendix B, the estimate for $Post_g \times Treat_{ic}$ is close to zero and not statistically significant.

5.2. Did professional players really learn from an APG?

We argue that improvement in move quality is achieved by players' learning from APGs. A stable trend in move quality before the availability of APGs and a gradual yet substantial increase after their availability supports this idea. Yet several alternative explanations can also be made. For instance, professional players may have changed their playing styles (without using the APG) after realizing that the APG beats established routines that players have developed. To validate learning from APGs as the key driver of the effect, we test the match between players' moves and the APG's top choices. If players indeed learn from APGs, the likelihood of a player making the exact same moves as the APG's top suggested moves should increase. Given that APGs are not available while a game is played, a player's move that exactly matches those of the APG should provide strong evidence that, prior to the game, the player learned from the APG. If players have not learned from the APG and instead try unprecedented moves, we do not expect an exact match.

We create an indicator variable, $Move\ Match_{ig}^{k}$, that captures, on average, how many moves of the focal player i are the same as the APG's top k suggestions among the first thirty moves in a game g. If $k = 1$, we consider there to be an exact match between the player's move and the APG's top suggestion. If $k = 3$, we check whether the player's move is among the APG's top three suggestions:

$$Move\ Match_{ig}^{k} = \frac{\sum_{n=1}^{15} \mathbf{1}\left(Player\ move_{ing} \in \left\{APG\ move_{ng}^{1}, APG\ move_{ng}^{2}, \ldots, APG\ move_{ng}^{k}\right\}\right)}{15}$$

Table B.8 of Appendix B shows the results from estimations with $Move\ Match_{ig}^{k}$ as a dependent variable. Columns 1–3 test the age effects. After the public release of the APG, younger players $(Post_g \times Young_i)$ were more likely than older players to make moves that match the APG's top 1, 3, and 5 recommendations.

What is more interesting and convincing is that the estimates for $Post_g \times Young_i$ in Table B.8 shrinks gradually as we broaden the set: 0.032 (top 1, column 1), 0.025 (top 3, column 2), and 0.017 (top 5, column 3). When players learn from an APG, they should be more inclined to learn the best move (i.e., the APG's top 1 suggestion). The fact that the estimate is largest in magnitude for top 1 moves therefore supports our argument that players did learn from an APG.

Columns 4–6 in Table B.8 show the country effects. The match between players' moves and the APG's top 1, 3, and 5 recommendations is greater for Chinese and Korean players than for Japanese players, after the introduction of APGs. Yet again, the estimate for $Post_g \times Treat_{ic}$ is largest for the top 1 recommendations; this supports the argument that results are driven by AI learning and not by other idiosyncratic channels.

5.3. Mechanisms for quality improvement: Errors and 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 mistakes*. 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, Jin-seo Shin (who was ranked first in the world in 2020) stated:

> *Before APG, players and their peers replayed games and discussed which move was an error and which was a critical mistake. After the public release of APGs, this replay and discussion by players became almost meaningless. APGs teach 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 move is an error. If the probability drops from 80 percent to 20 percent, that*

is a critical mistake. ... I have to admit that APG-based training provides limitless help in developing my Go skills (Sohn 2021).

To test these mechanisms, we measure the error in a game as the number of "bad" moves, those 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 B.2 of Appendix B shows the model-free trend of errors (in Panel A) and the critical mistake (in Panel B). Both the errors and the critical mistake show a substantial decrease after the release of the APG.

We then conduct regression analyses on errors and the critical mistake. Table 5, columns 1 and 3, shows that the number of errors and the magnitude of the 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 ($\beta = -0.009$, $p < 0.01$) is discontinuously accelerated after the introduction of APG($\beta = -0.233$, $p < 0.01$). These results confirm that AI improved the quality of moves of professional players by reducing both the number of errors (31.8%) and the magnitude of the critical mistake (20.9%).

"Insert Table 5 here"

5.4. Do AI-driven improvements of move quality lead to winning?

Building upon our finding that younger players improve more than older players after APG training, we further investigate whether this improvement leads to a higher probability of winning a game. We conduct the three-step mediation analysis suggested by (Baron & Kenny, 1986). As a baseline model, we run a logit regression of winning a game on an interaction between an indicator for a younger player and an indicator for a post-APG period. We find from Table 6, 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 6, columns 2–4, shows that *Move Quality* is positively associated with the younger group after APG, while *Errors* and *Critical mistake* are negatively associated.

The second step is to check whether move quality is positively associated with the probability of winning, while errors and the magnitude of the 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 6, 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 6, columns 8–11, the estimates for the explanatory variable ($Post_g \times Young_i$) are shown to decrease for all cases after adding the mediator variables, compared to those in the baseline model (column 1).

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

"Insert Table 6 here"

5.5. 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 B.3 of Appendix B. The AI effect is most prominent in early opening moves (for moves 1–30) and gradually decreases as we include later moves in the analysis.

Formal analyses confirm these observations. Table 7 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 B.4 of Appendix B. This clearly shows that younger 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 7 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 APG training, 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 an 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 setting in which to observe every single decision made by human players and to evaluate the quality of these decisions. Furthermore, provided that Go players did not use a non-AI Go program that lagged far behind the performance of professional players, our context provides a unique opportunity to measure the effect of AI but not the effect of broader IT. We find that by learning from AI algorithms, professional players improve the quality of their moves (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 during early stages of the game where higher uncertainty and unexpectedness are exhibited and players have had more training opportunity with AI. Younger players benefited more from AI than older players, suggesting a potential inequality in AI utilization by age and generation.

The findings from AI in professional Go games provide 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 (i.e., an educational role beyond a mere substitution for or assistance to human tasks) on the practices and performance of individuals and organizations; it can pave the way for new paradigms.

Importantly, there exists a widespread concern that AI will replace human jobs. Contrary to this prediction, depending on the context, AI may be utilized as an instructional tool to improve the skills and performance of humans. In the professional Go industry, this happened as AI provided quantified evaluations for decision-making by human professionals. Our findings can also be generalized to some domains where AI has already outperformed or will outperform human activities For example, in radiology AIs perform as well as trained radiologists in triaging chest and breast x-rays and in detecting lung cancers; doctors learn from and rely on AI's analysis as it provides better diagnoses and predictions (Grady, 2019; Lebovitz et al., 2021, 2022; Reardon, 2019). Yet, it is important to carefully assess different contexts and to 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 et al., 2020). We add to this stream of literature by showing that openness to new technologies and the ability to utilize it could also contribute to reaping gains from AI; the young benefited more from APGs than did the old. Further investigation is needed 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 a game. This boundary condition of the AI's effect is consistent with the findings in drug discovery and development (Lou & 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 provide new theoretical and empirical accounts of how AI transforms human decision-making (Brynjolfsson et al., 2021). As AI technology advances, researchers have recently expanded their interest in the role of AI in supporting human judgment (Choudhury et al., 2020; Kleinberg et al., 2017; Wang et al., 2019). However, existing studies have focused on AI's real-time assistant role in boosting task-related performance but paid little attention to the direct impact on human decision-making abilities. Our study provides new insights by showing that AI can help reevaluate knowledge and train professionals, reduce human errors and mistakes (that stem from heuristics and routines), and thereby improve the quality of decisions.

Second, our findings contribute to the literature on age and AI (Ghasemaghaei et al., 2019; Tams, 2022; Tams et al., 2014). Although researchers have begun to pay attention to AI literacy (e.g., Allen & Choudhury, 2022; Barth et al., 2020; Choudhury et al., 2020, Wang et al., 2019) there is little study of AI's impact—in particular, its instructional roles—on human decision-making by those of different ages. Our finding—that learning through AI is more significant for young, emerging professionals than for established veterans—provides new, important empirical evidence and calls for future studies.

Third, we contribute to the discussion on the boundary conditions of AI's instructional roles in human capital and strategic decision-making. While we show that AI can train professionals and nurture their capabilities, AI's effects vary by the uncertainty and unexpectedness of the task. Our study thus suggests that managers need to consider these factors when they seek to learn from the AI to improve their performance.

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. Similarly, the codification of routines is indeed important in a firm (e.g., Bingham et al. 1997; Foss 2003). Still, while managers and employees also face a set of rules and restrictions, they 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

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

Ahmed, N., & Wahed, M. (2020). The De-democratization of AI: Deep Learning and the Compute Divide in Artificial Intelligence Research. *ArXiv*. http://arxiv.org/abs/2010.15581

Allen, R., & Choudhury, P. (Raj). (2022). Algorithm-Augmented Work and Domain Experience: The Countervailing Forces of Ability and Aversion. *Organization Science*, *33*(1), 149–169. https://doi.org/10.1287/orsc.2021.1554

Anderson, T. (2004). *The way of Go: 8 ancient strategy secrets for Success in Business and Life*. Simon and Schuster.

Athey, S., & Imbens, G. W. (2019). Machine Learning Methods That Economists Should Know About. *Annual Review of Economics*, *11*(1), 685–725. https://doi.org/10.1146/annurev-economics-080217-053433

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. *Management Science*, *59*(3), 641–658. https://doi.org/10.1287/mnsc.1120.1586

Baron, R. M., & Kenny, D. A. (1986). The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology*, *51*(6), 1173–1182. https://doi.org/10.1037/0022-3514.51.6.1173

Barth, E., Davis, J. C., Freeman, R. B., & McElheran, K. (2020). Twisting the Demand Curve: Digitalization and the Older Workforce. *NBER Working Paper*.

Bingham, C. B., Eisenhardt, K. M., & Davis, J. P. (1997). Opening the Black Box of Organizational Expertise: Understanding What Firms Learn from their Process Experience and How that learning Unfolds over Time. *Working Paper, University of Maryland*.

Bingham, C. B., Eisenhardt, K. M., & Furr, N. R. (2007). What makes a process a capability? Heuristics, strategy, and effective capture of opportunities. *Strategic Entrepreneurship Journal*, *1*(1–2), 27–47. https://doi.org/10.1002/sej.1

Bingham, C. B., & Haleblian, J. J. (2012). How firms learn heuristics: Uncovering missing components of organizational learning. *Strategic Entrepreneurship Journal*, *6*(2), 152–177. https://doi.org/10.1002/sej.1132

Blackwell, D. (1953). Equivalent Comparisons of Experiments. *The Annals of Mathematical Statistics*, *24*(2), 265–272. https://doi.org/10.1214/aoms/1177729032

Blei, D. M., & Smyth, P. (2017). Science and data science. *Proceedings of the National Academy of Sciences*, *114*(33), 8689–8692. https://doi.org/10.1073/pnas.1702076114

Bresnahan, T. F., Brynjolfsson, E., & Hitt, L. M. (2002). Information Technology, Workplace Organization, and the Demand for Skilled Labor: Firm-Level Evidence. *The Quarterly Journal of Economics*, *117*(1), 339–376. https://doi.org/10.1162/003355302753399526

Brynjolfsson, E., & Hitt, L. M. (2000). Beyond Computation: Information Technology. *Journal of Economic Perspectives*, *14*(4), 23–48.

Brynjolfsson, E., Hitt, L. M., & Kim, H. H. (2011). Strength in Numbers: How Does Data-Driven Decision-making Affect Firm Performance? *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.1819486

Brynjolfsson, E., Wang, C. (Alex), & Zhang, X. (Michael). (2021). The Economics of IT and

Digitization: Eight Questions for Research. *MIS Quarterly*, *45*(1), 473–477. https://doi.org/10.25300/MISQ/2021/15434.1.4

Burton-Hill, C. (2016, February 16). The superhero of artificial intelligence: can this genius keep it in check? *The Guardian*. http://www.theguardian.com/technology/2016/feb/16/demis-hassabis-artificial-intelligence-deepmind-alphago

Cadario, R., Longoni, C., & Morewedge, C. K. (2021). Understanding, explaining, and utilizing medical artificial intelligence. *Nature Human Behaviour*. https://doi.org/10.1038/s41562-021-01146-0

Cao, S., Jiang, W., Wang, J., & Yang, B. (2021). From Man vs. Machine to Man + Machine: The Art and AI of Stock Analyses. In *NBER Working Paper*. https://doi.org/10.3386/w28800

Chao, X., Kou, G., Li, T., & Peng, Y. (2018). Jie Ke versus AlphaGo: A ranking approach using decision making method for large-scale data with incomplete information. *European Journal of Operational Research*, *265*(1), 239–247. https://doi.org/10.1016/j.ejor.2017.07.030

Chen, W. R. (2008). Determinants of firms' backward- and forward-looking R&D search behavior. *Organization Science*, *19*(4), 609–622. https://doi.org/10.1287/orsc.1070.0320

Choudhury, P., Starr, E., & Agarwal, R. (2020). Machine learning and human capital complementarities: Experimental evidence on bias mitigation. *Strategic Management Journal*, *41*(8), 1381–1411. https://doi.org/10.1002/smj.3152

Cohen, W. M., & Levinthal, D. A. (1990). Absorptive Capacity: A New Perspective on Learning and Innovation. *Administrative Science Quarterly*, *35*(1), 128. https://doi.org/10.2307/2393553

Cohen, W. M., & Levinthal, D. a. (1994). *Absorptive Capacity : A New Perspective on and Innovation Learning*. *35*(1), 128–152.

Cyert, R. M., & March, J. G. (1963). A behavioral theory of the firm. In *Englewood Cliffs, NJ* (Vol. 2, Issue 4).

Dane, E., & Pratt, M. G. (2007). Exploring Intuition and its Role in Managerial Decision Making. *Academy of Management Review*, *32*(1), 33–54. https://doi.org/10.5465/amr.2007.23463682

Danziger, S., Levav, J., & Avnaim-Pesso, L. (2011). Extraneous factors in judicial decisions. *Proceedings of the National Academy of Sciences*, *108*(17), 6889–6892. https://doi.org/10.1073/pnas.1018033108

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

De Cremer, D. (2020). What Does Building a Fair AI Really Entail? *Harvard Business Review Digital Articles*, *September*. https://hbr.org/2020/09/what-does-building-a-fair-ai-really-entail

de Koning, J., & Gelderblom, A. (2006). ICT and older workers: No unwrinkled relationship. *International Journal of Manpower*, *27*(5), 467–490. https://doi.org/10.1108/01437720610683967

Eisenhardt, K. M. (1990). Speed and Strategic Choice: How Managers Accelerate Decision Making. *California Management Review*, *32*(3), 39–54. https://doi.org/10.2307/41166616

Foss, N. J. (2003). Bounded rationality and tacit knowledge in the organizational capabilities approach: an assessment and a re-evaluation. *Industrial and Corporate Change*, *12*(2), 185–201. https://doi.org/10.1093/icc/12.2.185

Gaessler, F., & Piezunka, H. (2021). Where AI is a Game Changer: Evidence from Chess Computers. *Working Paper*.

Galbraith, J. R. (1974). Organization Design: An Information Processing View. *Interfaces*, *4*(3), 28–36. https://doi.org/10.1287/inte.4.3.28

Gavetti, G., & Levinthal, D. (2000). Looking forward and looking backward: Cognitive and experiential search. *Administrative Science Quarterly*, *45*(1), 113–137. https://doi.org/10.2307/2666981

Ghasemaghaei, M., Hassanein, K., & Benbasat, I. (2019). Assessing the Design Choices for Online Recommendation Agents for Older Adults: Older Does Not Always Mean Simpler. *MIS Quarterly*, *43*(1), 329–346. https://doi.org/10.25300/MISQ/2019/13947

Grady, D. (2019, May 20). A.I. Took a Test to Detect Lung Cancer. It Got an A. *The New York Times*. https://www.nytimes.com/2019/05/20/health/cancer-artificial-intelligence-ct-scans.html

Guest, R. H. (1956). Of time and the foreman. *Personnel*, *32*, 478–486.

Hagendorff, T., & Wezel, K. (2020). 15 challenges for AI: or what AI (currently) can't do. *AI & SOCIETY*, *35*(2), 355–365. https://doi.org/10.1007/s00146-019-00886-y

Hambrick, D. C., & Mason, P. A. (1984). Upper Echelons: The Organization as a Reflection of Its Top Managers. *The Academy of Management Review*, *9*(2), 193. https://doi.org/10.2307/258434

He, K., Zhang, X., Ren, S., & Sun, J. (2015). Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification. *2015 IEEE International Conference on Computer Vision (ICCV)*, *6*(4), 1026–1034. https://doi.org/10.1109/ICCV.2015.123

Kahn, J. (2020, December 23). DeepMind unveils new details of game-mastering A.I. it says may help in complex real-world environments. *Fortune*. https://fortune.com/2020/12/23/deepmind-ai-artificial-intelligence-game-mastering-muzero/

Kahneman, D. (2003). Maps of Bounded Rationality: Psychology for Behavioral Economics. *American Economic Review*, *93*(5), 1449–1475. https://doi.org/10.1257/000282803322655392

Kalberg, S. (1980). Max Weber's Types of Rationality: Cornerstones for the Analysis of Rationalization Processes in History. *American Journal of Sociology*, *85*(5), 1145–1179. https://doi.org/10.1086/227128

Kim, J. H., Han, S. G., Cho, A., Shin, H. J., & Baek, S. E. (2021). Effect of deep learning-based assistive technology use on chest radiograph interpretation by emergency department physicians: a prospective interventional simulation-based study. *BMC Medical Informatics and Decision Making*, *21*(1), 1–9. https://doi.org/10.1186/s12911-021-01679-4

Kleinberg, J., Lakkaraju, H., Leskovec, J., Ludwig, J., & Mullainathan, S. (2017). Human Decisions and Machine Predictions. *The Quarterly Journal of Economics*, *January*, 237–293. https://doi.org/10.1093/qje/qjx032

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

Krakowski, S., Luger, J., & Raisch, S. (2022). Artificial intelligence and the changing sources of competitive advantage. *Strategic Management Journal*, *1*(2), 57–72. https://doi.org/10.1002/smj.3387

Kwon, N.-K., Hong, S., Sung, Y.-J., Bae, J., Yoon, H., Min, S., Kim, H., & Lee, H. (2014, December 8). [SUPER RICH] Game of go makes great businesspeople. *Korea Herald*. http://www.koreaherald.com/view.php?ud=20141208000697

Langholtz, H., Gettys, C., & Foote, B. (1993). Resource-Allocation Behavior under Certainty, Risk, and Uncertainty. *Organizational Behavior and Human Decision Processes*, *54*(2), 203–224. https://doi.org/10.1006/obhd.1993.1010

Lebovitz, S., Levina, N., & Lifshitz-Assa, H. (2021). Is AI Ground Truth Really True? The Dangers of Training and Evaluating AI Tools Based on Experts' Know-What. *MIS Quarterly*, *45*(3), 1501–1526. https://doi.org/10.25300/MISQ/2021/16564

Lebovitz, S., Lifshitz-Assaf, H., & Levina, N. (2022). To Engage or Not to Engage with AI for Critical Judgments: How Professionals Deal with Opacity When Using AI for Medical Diagnosis. *Organization Science*, *January*, 1–45. https://doi.org/10.1287/orsc.2021.1549

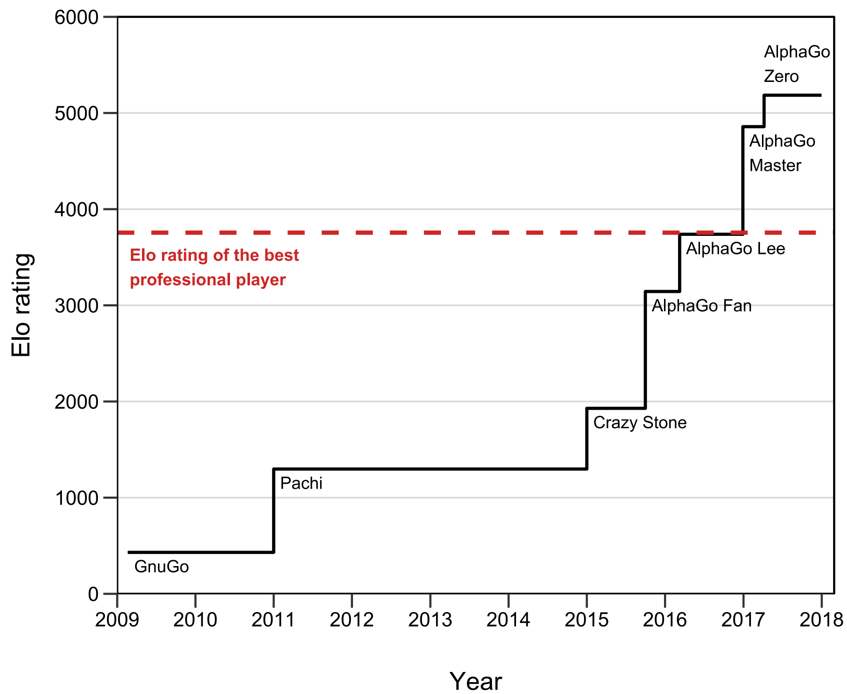
Levinthal, D. A., & March, J. G. (1993). The myopia of learning. *Strategic Management Journal*,

14(S2), 95–112. https://doi.org/10.1002/smj.4250141009
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. *Academy of Management Review*, *45*(1), 247–263. https://doi.org/10.5465/amr.2018.0181
Lou, B., & Wu, L. (2021). Artificial Intelligence on Drugs: A Large-Scale Examination of AI and Drug Innovation. *MIS Quarterly*.
Loveman, G. (2003). Diamonds in the Data Mine. *Harvard Buiness Review*.
Meyer, J. (2011). Workforce age and technology adoption in small and medium-sized service firms. *Small Business Economics*, *37*(3), 305–324. https://doi.org/10.1007/s11187-009-9246-y
Mintzberg, H. (1987). Crafting strategy: Harvard Business School Press Boston. In *MA, USA*.
Mintzberg, Henry. (1971). Managerial Work: Analysis from Observation. *Management Science*, *18*(2), B-97-B-110. https://doi.org/10.1287/mnsc.18.2.b97
Mintzberg, Henry. (1990). The design school: Reconsidering the basic premises of strategic management. *Strategic Management Journal*, *11*(3), 171–195. https://doi.org/10.1002/smj.4250110302
Mintzberg, Henry. (1994). Rethinking strategic planning part I: Pitfalls and fallacies. *Long Range Planning*, *27*(3), 12–21. https://doi.org/10.1016/0024-6301(94)90185-6
Miric, M., Lu, J., & Teodoridis, F. (2020). Decision-Making Skills in an AI World: Lessons from Online Chess. *SSRN Electronic Journal*, 1–43. https://doi.org/10.2139/ssrn.3538840
Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A. a, Veness, J., Bellemare, M. G., Graves, A., Riedmiller, M., Fidjeland, A. K., Ostrovski, G., Petersen, S., Beattie, C., Sadik, A., Antonoglou, I., King, H., Kumaran, D., Wierstra, D., Legg, S., & Hassabis, D. (2015). Human-level control through deep reinforcement learning. *Nature*, *518*(7540), 529–533. https://doi.org/10.1038/nature14236
Morris, M.G., Venkatesh, V., & Ackerman, P. L. (2005). Gender and Age Differences in Employee Decisions About New Technology: An Extension to the Theory of Planned Behavior. *IEEE Transactions on Engineering Management*, *52*(1), 69–84. https://doi.org/10.1109/TEM.2004.839967
Morris, Michael G., & Venkatesh, V. (2000). Age Differences in Technology Adoption Decisions: Implications for a Changing Work Force. *Personnel Psychology*, *53*(2), 375–403. https://doi.org/10.1111/j.1744-6570.2000.tb00206.x
Mousavi, S., & Gigerenzer, G. (2014). Risk, uncertainty, and heuristics. *Journal of Business Research*, *67*(8), 1671–1678. https://doi.org/10.1016/j.jbusres.2014.02.013
Mozur, P. (2017, May 23). Google's AlphaGo Defeats Chinese Go Master in Win for A.I. *The New York Times*, 2–3. https://www.nytimes.com/2017/05/23/business/google-deepmind-alphago-go-champion-defeat.html
Muoio, D. (2016, March 10). Why Go is so much harder for AI to beat than chess. *Business Insider*. https://www.businessinsider.com/why-google-ai-game-go-is-harder-than-chess-2016-3
Nerkar, A. (2003). Old Is Gold? The Value of Temporal Exploration in the Creation of New Knowledge. *Management Science*, *49*(2), 211–229. https://doi.org/10.1287/mnsc.49.2.211.12747
Noh, S. (2016, February 22). 5 Reasons AlphaGo Can't "Never" Beat Lee Sedol - Weak Paving Stones, Narrow Vision... still one number down. *Maeil Business Newspaper*. https://www.mk.co.kr/news/economy/view/2016/02/140725/ (Written in Korean)
Press, G. (2021, May 19). 114 Milestones In The History Of Artificial Intelligence (AI). *Forbes*, 1–20.

Rai, A., Constantinides, P., & Sarker, S. (2019). Next-Generation Digital Platforms: Toward Human–AI Hybrids. *MIS Quarterly*, *43*(February), 1–9.
Ramon, J., & Struyf, J. (2003). Computer science issues in baduk. *Proceedings of the 2nd International Conference on Baduk*, 163–182. http://www.cs.kuleuven.ac.be/~dtai/publications/files/40886.pdf
Reardon, S. (2019). Rise of Robot Radiologists. *Nature*, *576*(7787), S54–S58. https://doi.org/10.1038/d41586-019-03847-z
Reeves, M., & Wittenburg, G. (2015). Games Can Make You a Better Strategist. *Harvard Business Review*, 3. https://hbr.org/2015/09/games-can-make-you-a-better-strategist
Roman, B., & Vyas, K. (2021, November 5). This 5,000-Year-Old Chinese Board Game Can Help Your Business Make Millions. *Entrepreneur*. https://www.entrepreneur.com/article/393599
Saunders, A., & Tambe, P. (2013). A Measure of Firms' Information Practices Based on Textual Analysis of 10-K Filings. *Working Paper*.
Savage, N. (2021). Tapping into the drug discovery potential of AI. *Biopharma Dealmakers*, *May*, 1–12. https://doi.org/10.1038/d43747-021-00045-7
Schleife, K. (2006). Computer Use and Employment Status of Older Workers - An Analysis Based on Individual Data. *Labour*, *20*(2), 325–348. https://doi.org/10.1111/j.1467-9914.2006.00341.x
Schrittwieser, J., Antonoglou, I., Hubert, T., Simonyan, K., Sifre, L., Schmitt, S., Guez, A., Lockhart, E., Hassabis, D., Graepel, T., Lillicrap, T., & Silver, D. (2020). Mastering Atari, Go, chess and shogi by planning with a learned model. *Nature*, *588*(7839), 604–609. https://doi.org/10.1038/s41586-020-03051-4
Seamans, R., & Raj, M. (2018). AI, Labor, Productivity and The Need for Firm-Level Data. *NBER Working Paper*.
Senior, A. W., Evans, R., Jumper, J., Kirkpatrick, J., Sifre, L., Green, T., Qin, C., Žídek, A., Nelson, A. W. R., Bridgland, A., Penedones, H., Petersen, S., Simonyan, K., Crossan, S., Kohli, P., Jones, D. T., Silver, D., Kavukcuoglu, K., & Hassabis, D. (2020). Improved protein structure prediction using potentials from deep learning. *Nature*, *577*(7792), 706–710. https://doi.org/10.1038/s41586-019-1923-7
Shepherd, D. A., Williams, T. A., & Patzelt, H. (2015). Thinking About Entrepreneurial Decision Making. *Journal of Management*, *41*(1), 11–46. https://doi.org/10.1177/0149206314541153
Silver, D., Huang, A., Maddison, C. J., Guez, A., Sifre, L., van den Driessche, G., Schrittwieser, J., Antonoglou, I., Panneershelvam, V., Lanctot, M., Dieleman, S., Grewe, D., Nham, J., Kalchbrenner, N., Sutskever, I., Lillicrap, T., Leach, M., Kavukcuoglu, K., Graepel, T., & Hassabis, D. (2016). Mastering the game of Go with deep neural networks and tree search. *Nature*, *529*(7587), 484–489. https://doi.org/10.1038/nature16961
Silver, D., Schrittwieser, J., Simonyan, K., Antonoglou, I., Huang, A., Guez, A., Hubert, T., Baker, L., Lai, M., Bolton, A., Chen, Y., Lillicrap, T., Hui, F., Sifre, L., van den Driessche, G., Graepel, T., & Hassabis, D. (2017). Mastering the game of Go without human knowledge. *Nature*, *550*(7676), 354–359. https://doi.org/10.1038/nature24270
Simon, H. A. (1955). A Behavioral Model of Rational Choice. *The Quarterly Journal of Economics*, *69*(1), 99. https://doi.org/10.2307/1884852
Simon, H. A. (1991). Bounded Rationality and Organizational Learning. *Organization Science*, *2*(1), 125–134. https://doi.org/10.1287/orsc.2.1.125
Simon, H. A., & March, J. G. (1958). *Organizations*. Wiley-Blackwell.
Smalley, E. (2017). AI-powered drug discovery captures pharma interest. *Nature Biotechnology*, *35*(7), 604–605. https://doi.org/10.1038/nbt0717-604
Smith, B. C. (2019). *The promise of artificial intelligence: reckoning and judgment*. MIT Press.

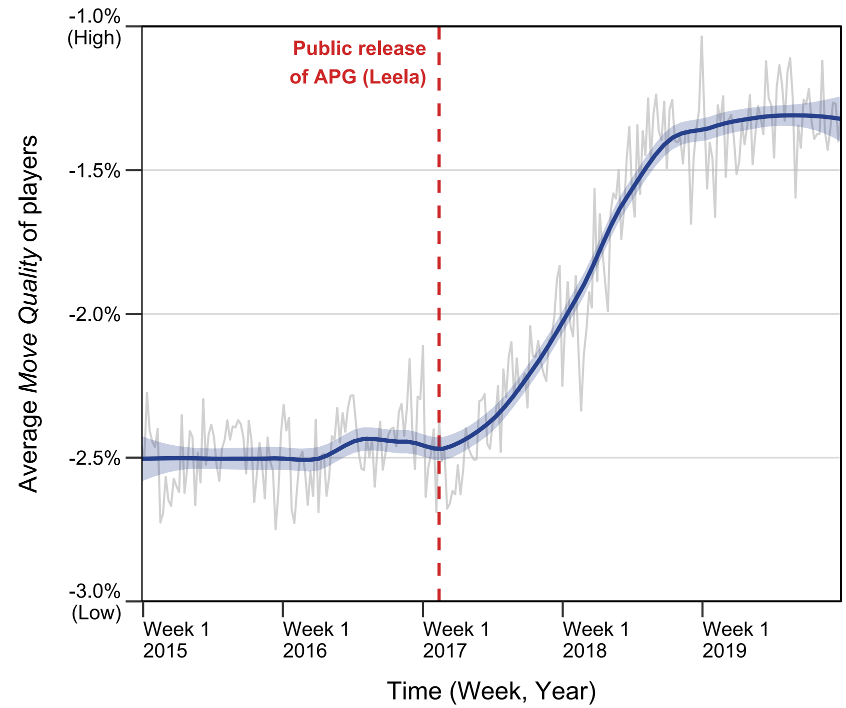
Somers, J. (2018, December 28). How the Artificial-Intelligence Program AlphaZero Mastered Its Games. *The New Yorker*, 1–6.
Tams, S. (2022). Helping Older Workers Realize Their Full Organizational Potential: a Moderated Mediation Model of Age and It-Enabled Task Performance. *MIS Quarterly*. https://doi.org/10.25300/MISQ/2022/16359
Tams, S., Grover, V., & Thatcher, J. (2014). Modern information technology in an old workforce: Toward a strategic research agenda. *Journal of Strategic Information Systems*, *23*(4), 284–304. https://doi.org/10.1016/j.jsis.2014.10.001
Thaler, R. (1993). Advances in Behavioral Finance Russell Sage Foundation. In *Russell Sage Foundation*.
Tong, S., Jia, N., Luo, X., & Fang, Z. (2021). The Janus face of artificial intelligence feedback: Deployment versus disclosure effects on employee performance. *Strategic Management Journal*, *42*(9), 1600–1631. https://doi.org/10.1002/smj.3322
Tyler, B. B., & Steensma, H. K. (1998). The effects of executives' experiences and perceptions on their assessment of potential technological alliances. *Strategic Management Journal*, *19*(10), 939–965. https://doi.org/10.1002/(SICI)1097-0266(199810)19:10<939::AID-SMJ978>3.0.CO;2-Z
Venkatesh, Morris, Davis, & Davis. (2003). User Acceptance of Information Technology: Toward a Unified View. *MIS Quarterly*, *27*(3), 425. https://doi.org/10.2307/30036540
Wang, W., Gao, G. (Gordon), & Agarwal, R. (2019). Friend or Foe? The Influence of Artificial Intelligence on Human Performance in Medical Chart Coding. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3405759
Weinberg, B. (2004). Experience and Technology Adoption. *IZA Discussion Paper*, *1051*.
Wiseman, B. O. (2016, March 25). The AI with an Intuition. *PrimeMind*. https://primemind.com/the-ai-with-an-intuition-a7bc0317ba9a
Wu, L., Hitt, L., & Lou, B. (2020). Data Analytics, Innovation, and Firm Productivity. *Management Science*, *66*(5), 2017–2039. https://doi.org/10.1287/mnsc.2018.3281
Wu, T.-R., Wu, I.-C., Chen, G.-W., Wei, T.-H., Wu, H.-C., Lai, T.-Y., & Lan, L.-C. (2018). Multilabeled Value Networks for Computer Go. *IEEE Transactions on Games*, *10*(4), 378–389. https://doi.org/10.1109/TG.2018.2852806
Yechiam, E., & Busemeyer, J. R. (2005). Comparison of basic assumptions embedded in learning models for experience-based decision making. *Psychonomic Bulletin & Review*, *12*(3), 387–402. https://doi.org/10.3758/BF03193783
Zarley, B. D. (2021, May 27). Liverpool and DeepMind are bringing AI to the beautiful game. *Freethink*. https://www.freethink.com/culture/ai-in-sports

Figure 1. Elo rating comparing the best professional player to Go programs



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 and Go4Go. The performance of APGs has surpassed that of the best human player since March 2016 when *AlphaGo Lee* was introduced.
GoRatings: https://www.goratings.org/en/
Go4Go: https://www.go4go.net/go/players/rank/

Figure 2. Effects of APG on average *Move Quality* of professional players: Model-free evidence

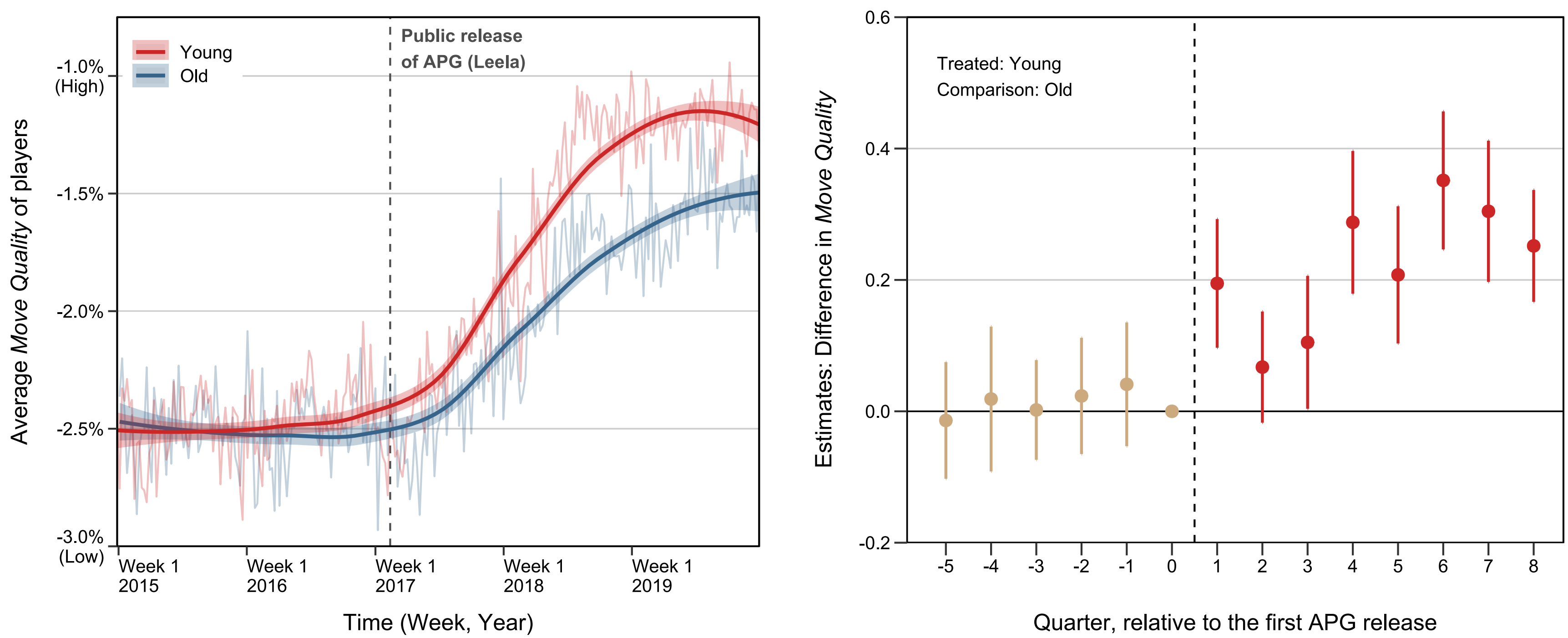


Note. This figure illustrates the weekly average *Move Quality* of players from 2015 through 2019. The gray solid line represents the raw (unprocessed) weekly average value. The blue solid line and the blue area around it show the smoothed trend (loess; span=0.4) and the 95% confidence interval, respectively. The vertical line on February 2017 represents the first public release of an APG, Leela.

Figure 3. Differential effects of APG on move quality by player age

(a) Model-free evidence

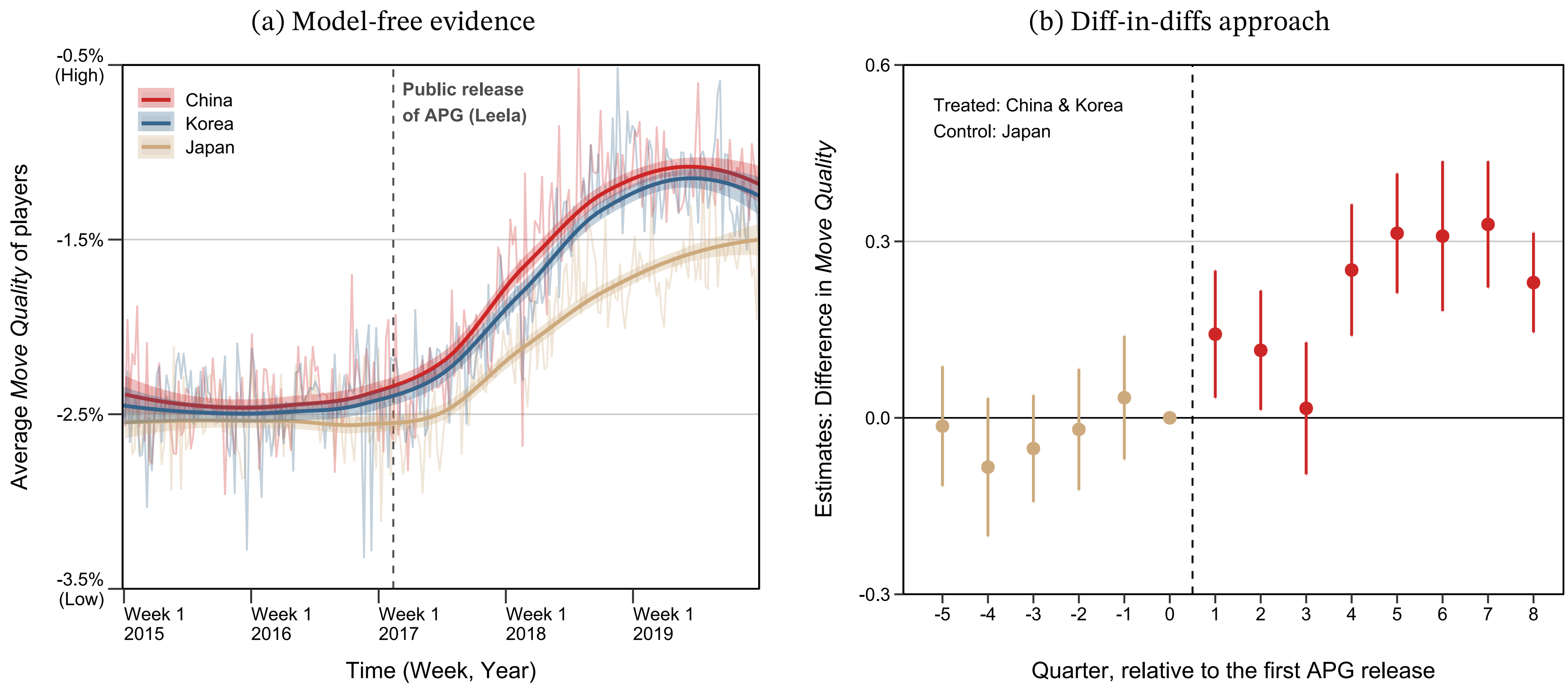
(b) Diff-in-diffs approach



Note. This figure illustrates the average *Move Quality* of professional players by player age. The red and blue fluctuating lines show the raw (unprocessed) weekly average values for younger players (below median age) and older players (above median age), respectively. The red and blue smooth lines and the shaded areas around them show the locally smoothed trends (loess; span=0.7) and the 95% confidence intervals. The vertical line on February 2017 represents the first public release of an APG, Leela.

Note. This figure illustrates the differential effects of APG on *Move Quality* by player age. The points graphically present the *Move Quality* of younger players (below median age) compared to that of older players (above median age), based on the regression estimates in Table B.1 of Appendix B, column 2. The vertical error bars show the 95% confidence intervals. Before the APG, we do not find a difference in *Move Quality* by age. After the APG, the increase in *Move Quality* is greater for young players than for old players.

Figure 4. Differential effects of APG on move quality by player country



Note. This figure illustrates the average *Move Quality* of professional players by their nationality. The red, blue, and brown lines show the raw (unprocessed) weekly average of *Move Quality* for Chinese, Korean, and Japanese players, respectively. The red, blue, and brown smooth lines and the shaded areas around them show the locally smoothed trend (loess, span=0.7) and the 95% confidence interval. The vertical line on February 2017 represents the first public release of an APG, Leela.

Note. This figure illustrates the effects of APG on *Move Quality*, based on the difference-in-differences estimation results reported in Table B.1 of Appendix B, column 4. 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 the APG, the strength of treatment is far weaker for Japanese players. The inclusion of Japan in the control group would bias our estimates toward zero (i.e., *against* our findings), leading to an underestimation. In other words, the resulting estimates provide a lower bound of the effect.

Table 1. Descriptive statistics

(a). Player-game level

	N	Mean	Median	SD	P25	P75
Move Quality	50,066	–2.01	–1.92	1.07	–2.66	–1.22
Number of Errors	50,066	0.13	0.00	0.37	0.00	0.00
Magnitude of the Critical Mistake	50,065	5.65	4.80	3.96	2.94	7.36
Age	49,214	28.18	24.52	12.57	19.61	31.6
Young	49,214	0.62	1.00	0.49	0.00	1.00
Rank	48,808	–0.28	–0.18	0.27	–0.43	–0.05
Rank Diff	47,712	0.00	0.00	0.22	–0.09	0.09
White	50,066	0.50	0.50	0.50	0.00	1.00
7.5 Komi	50,066	0.38	0.00	0.49	0.00	1.00

(b). Player level

	N	Mean	Median	SD	P25	P75
Move Quality	1,242	–2.20	–2.18	0.67	–2.52	–1.79
Number of Errors	1,242	0.17	0.12	0.22	0.00	0.20
Magnitude of the Critical Mistake	1,242	–6.20	–5.98	2.10	–6.96	–5.04
Age	1,149	32.73	27.32	16.15	20.25	42.77
Young	1,149	0.49	0.00	0.50	0.00	1.00
Rank	1,088	–0.54	–0.54	0.31	–0.81	–0.27
Rank Diff	1,080	0.15	0.11	0.19	0.00	0.27

Notes. 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 results, we multiply negative one by the rank of a player and divide it by 1,000 (*Rank*). That is, the higher the value of *Rank*, 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	−1.008
	(0.017)	(0.038)
	[p=0.000]	[p=0.000]
Trend		0.007
		(0.003)
		[p=0.023]
Post × Trend		0.116
		(0.004)
		[p=0.000]
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

Notes. 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 an 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 beginning of 2015; *Trend* takes the value of 10 in the first quarter after Leela's release (Q2 2017). Clustered standard errors at a focal-player level are in parentheses and p-values are in squared brackets.

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)		
	[p=0.000]	[p=0.019]		
Rank	0.817	0.799	1.626	2.501
	(0.036)	(0.037)	(0.240)	(0.292)
	[p=0.000]	[p=0.000]	[p=0.000]	[p=0.000]
Rank Diff	0.137	0.128	0.061	1.023
	(0.027)	(0.027)	(0.024)	(0.162)
	[p=0.000]	[p=0.000]	[p=0.011]	[p=0.000]
White	−0.134	−0.133	−0.131	−0.131
	(0.010)	(0.010)	(0.009)	(0.010)
	[p=0.000]	[p=0.000]	[p=0.000]	[p=0.000]
7.5 Komi	0.025	0.023	0.024	0.041
	(0.017)	(0.017)	(0.016)	(0.019)
	[p=0.126]	[p=0.170]	[p=0.142]	[p=0.031]
Post × Young		0.273	0.227	0.209
		(0.028)	(0.031)	(0.031)
		[p=0.000]	[p=0.000]	[p=0.000]
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

Notes. 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 an APG in February 2017, and *Young* refers to young professional Go players. Clustered standard errors at a focal player level are in parentheses and p-values are in squared brackets.

Table 4. Effects of APG on move quality: DID estimation using cross-country variation in exposure to APG

Dependent Variable:	Move Quality			
Model:	(1)	(2)	(3)	(4)
Variables				
Treat	0.110	−0.069		
	(0.024)	(0.026)		
	[*p*=0.000]	[*p*=0.008]		
Rank	0.809	0.810	2.222	3.146
	(0.040)	(0.040)	(0.221)	(0.275)
	[*p*=0.000]	[*p*=0.000]	[*p*=0.000]	[*p*=0.000]
Rank Diff	0.133	0.131	0.059	1.115
	(0.032)	(0.032)	(0.027)	(0.178)
	[*p*=0.000]	[*p*=0.000]	[*p*=0.029]	[*p*=0.000]
White	−0.129	−0.129	−0.128	−0.126
	(0.010)	(0.010)	(0.010)	(0.010)
	[*p*=0.000]	[*p*=0.000]	[*p*=0.000]	[*p*=0.000]
7.5 Komi	0.011	0.010	0.020	0.030
	(0.018)	(0.017)	(0.017)	(0.020)
	[*p*=0.515]	[*p*=0.557]	[*p*=0.229]	[*p*=0.130]
Post × Treat		0.315	0.266	0.230
		(0.031)	(0.031)	(0.031)
		[*p*=0.000]	[*p*=0.000]	[*p*=0.000]
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	0.277	0.281	0.327	0.352
Within R^2	0.061	0.067	0.014	0.014

Notes. This table shows the effects of APGs 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 estimate the differences in *Move Quality* among country groups before and after the release of the APG. Clustered standard errors at a focal-player level are in parentheses and p-values are in squared brackets.

Table 5. 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	0.083	−1.431	2.262
	(0.004)	(0.013)	(0.053)	(0.143)
	[*p*=0.000]	[*p*=0.000]	[*p*=0.000]	[*p*=0.000]
Trend		−0.000		−0.028
		(0.001)		(0.012)
		[*p*=0.739]		[*p*=0.021]
Post × Trend		−0.009		−0.233
		(0.001)		(0.015)
		[*p*=0.000]		[*p*=0.000]
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

Notes. This table shows the impact of APGs on errors and the 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 the 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 and p-values are in squared brackets.

Table 6. Mediation analysis on game winning: Move quality, errors, and a critical mistake

Dependent Variables:	Win	Move Quality	Number of Errors	Critical Mistake	Win						
Model:	(1) Logit	(2) OLS	(3) OLS	(4) OLS	(5) Logit	(6) Logit	(7) Logit	(8) Logit	(9) Logit	(10) Logit	(11) Logit
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)
	[*p*=0.421]	[*p*=0.000]	[*p*=0.000]	[*p*=0.000]	[*p*=0.303]	[*p*=0.682]	[*p*=0.424]	[*p*=0.163]	[*p*=0.339]	[*p*=0.212]	[*p*=0.140]
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)
	[*p*=0.000]	[*p*=0.000]	[*p*=0.749]	[*p*=0.001]	[*p*=0.000]	[*p*=0.000]	[*p*=0.000]	[*p*=0.000]	[*p*=0.000]	[*p*=0.000]	[*p*=0.000]
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)
	[*p*=0.000]	[*p*=0.000]	[*p*=0.000]	[*p*=0.000]	[*p*=0.000]	[*p*=0.000]	[*p*=0.000]	[*p*=0.000]	[*p*=0.000]	[*p*=0.000]	[*p*=0.000]
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)
	[*p*=0.989]	[*p*=0.031]	[*p*=0.173]	[*p*=0.205]	[*p*=0.979]	[*p*=0.935]	[*p*=0.943]	[*p*=0.923]	[*p*=0.969]	[*p*=0.965]	[*p*=0.933]
Post × 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)
	[*p*=0.007]	[*p*=0.000]	[*p*=0.003]	[*p*=0.000]				[*p*=0.036]	[*p*=0.010]	[*p*=0.018]	[*p*=0.036]
Move Quality					0.132			0.130			0.084
					(0.012)			(0.012)			(0.015)
					[*p*=0.000]			[*p*=0.000]			[*p*=0.000]
Number of Errors						−0.206			−0.211		0.025
						(0.032)			(0.032)		(0.045)
						[*p*=0.000]			[*p*=0.000]		[*p*=0.582]
Magnitude of the Critical Mistake							−0.031			−0.031	−0.021
							(0.003)			(0.003)	(0.005)
							[*p*=0.000]			[*p*=0.000]	[*p*=0.000]
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	46,337	46,334	46,333	45,762	45,762	45,761	45,761
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	71,587.71	140,067.46	56,742.68	273,520.00	72,540.67	72,577.86	72,510.67	71,487.03	71,549.67	71,484.28	71,473.91

Notes. 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 the APG, young professional Go players were 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 the introduction of the APG; these changes lead to eventually winning a game. Clustered standard errors at a focal-player level are in parentheses and p-values are in squared brackets.

Table 7. 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)
	[p=0.000]	[p=0.000]	[p=0.000]	[p=0.000]	[p=0.006]	[p=0.014]
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)
	[p=0.000]	[p=0.000]	[p=0.370]	[p=0.869]	[p=0.487]	[p=0.361]
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)
	[p=0.000]	[p=0.000]	[p=0.000]	[p=0.000]	[p=0.000]	[p=0.001]
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)
	[p=0.031]	[p=0.856]	[p=0.852]	[p=0.266]	[p=0.185]	[p=0.257]
Post × 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)
	[p=0.000]	[p=0.000]	[p=0.000]	[p=0.000]	[p=0.000]	[p=0.136]
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

Notes. This table presents how an 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 presents 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 and p-values are in squared brackets.

Online Appendix

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

A. AI-powered Go Program (APG)

A.1 Leela and Leela Zero

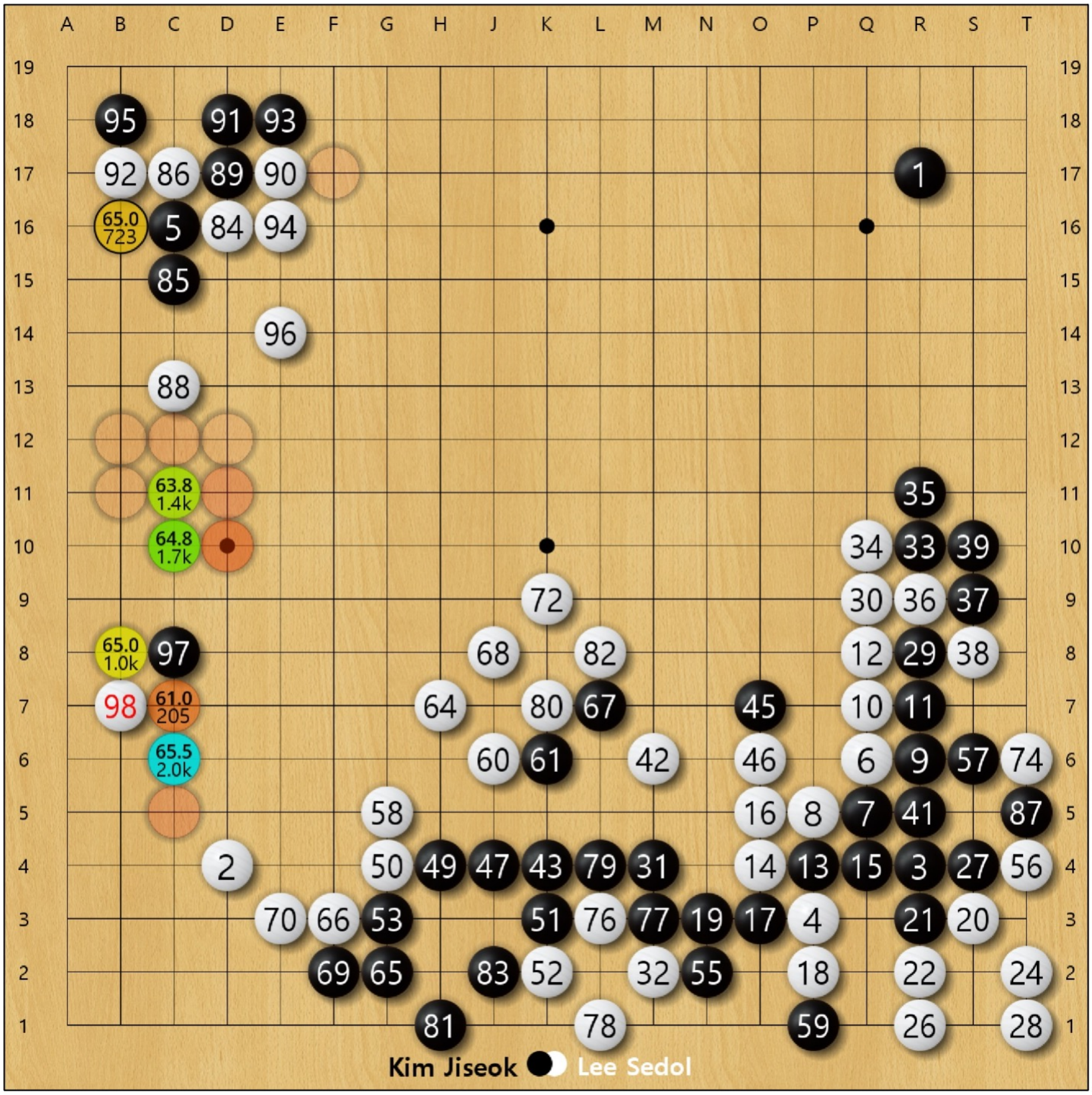
Leela, an AI-powered Go (APG) program, was released in February 2017; it included a stable, deep learning version using Monte Carlo Tree Search (MCTS). It was developed based on the AlphaGo algorithm of Google DeepMind (Silver et al., 2016).[1] Leela was the first APG to surpass human professional players 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.[2] 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.[3]

[1] 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.
[2] For more information on OpenCL, please refer to https://www.khronos.org/opencl/.
[3] These interfaces are available at: https://github.com/featurecat/lizzie/releases/ (Lizzie), https://sabaki.yichuanshen.de/ (Sabaki), and https://github.com/pnprog/goreviewpartner/ (GoReviewPartner).

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 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. 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.

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



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

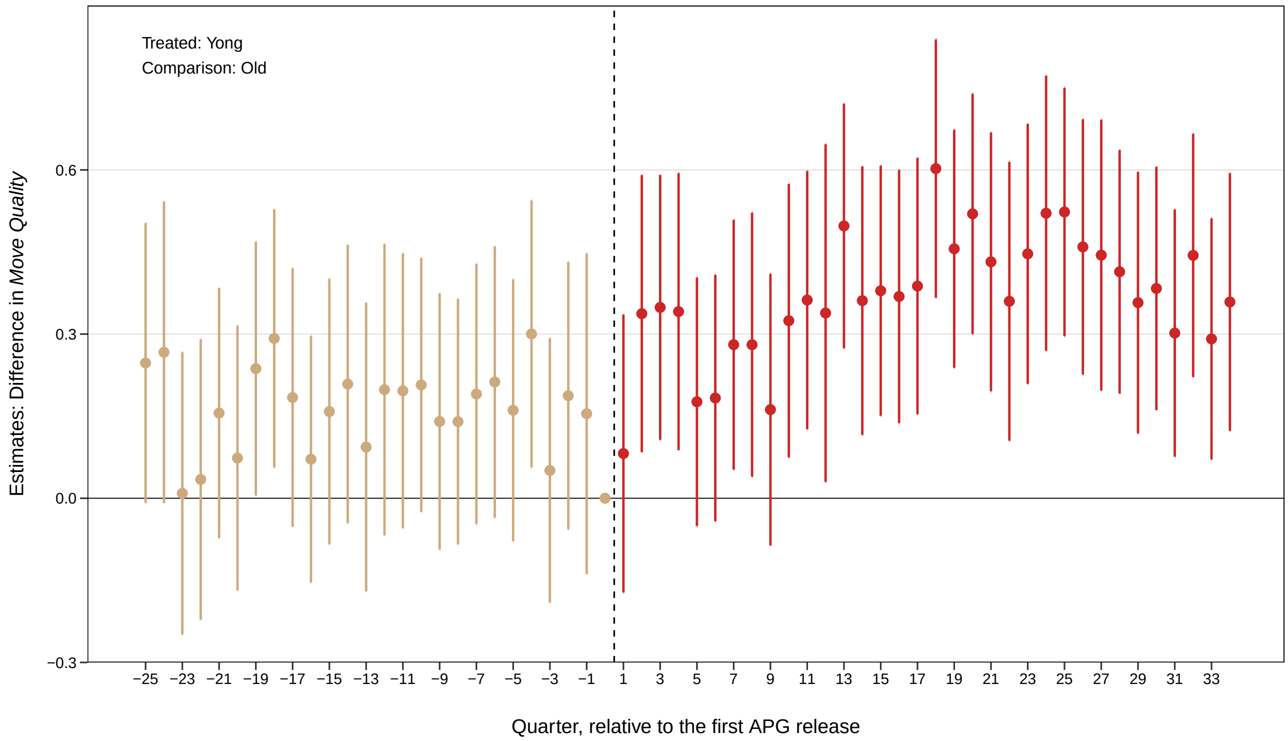
A.2 Implementation Details

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.

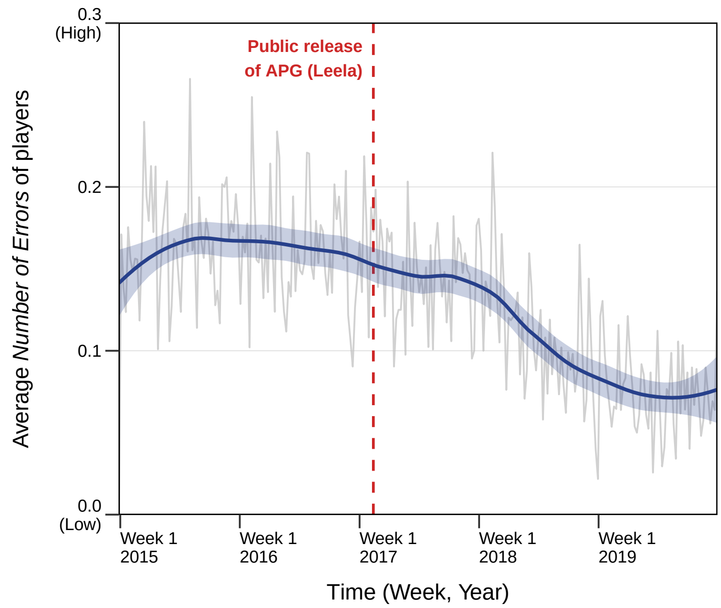
B. Additional Figures and Tables

Figure B.1: Robustness check: Differential effects using a month fixed effect

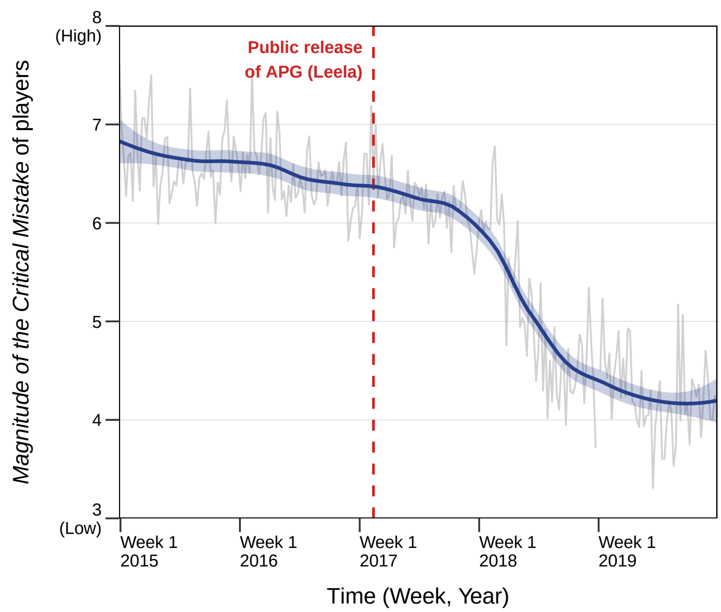


Note: This figure illustrates the differential effects of APG's using a month fixed effect instead of a quarter fixed effect.

Figure B.2: Errors and a critical mistake as mechanisms: Model-free evidence



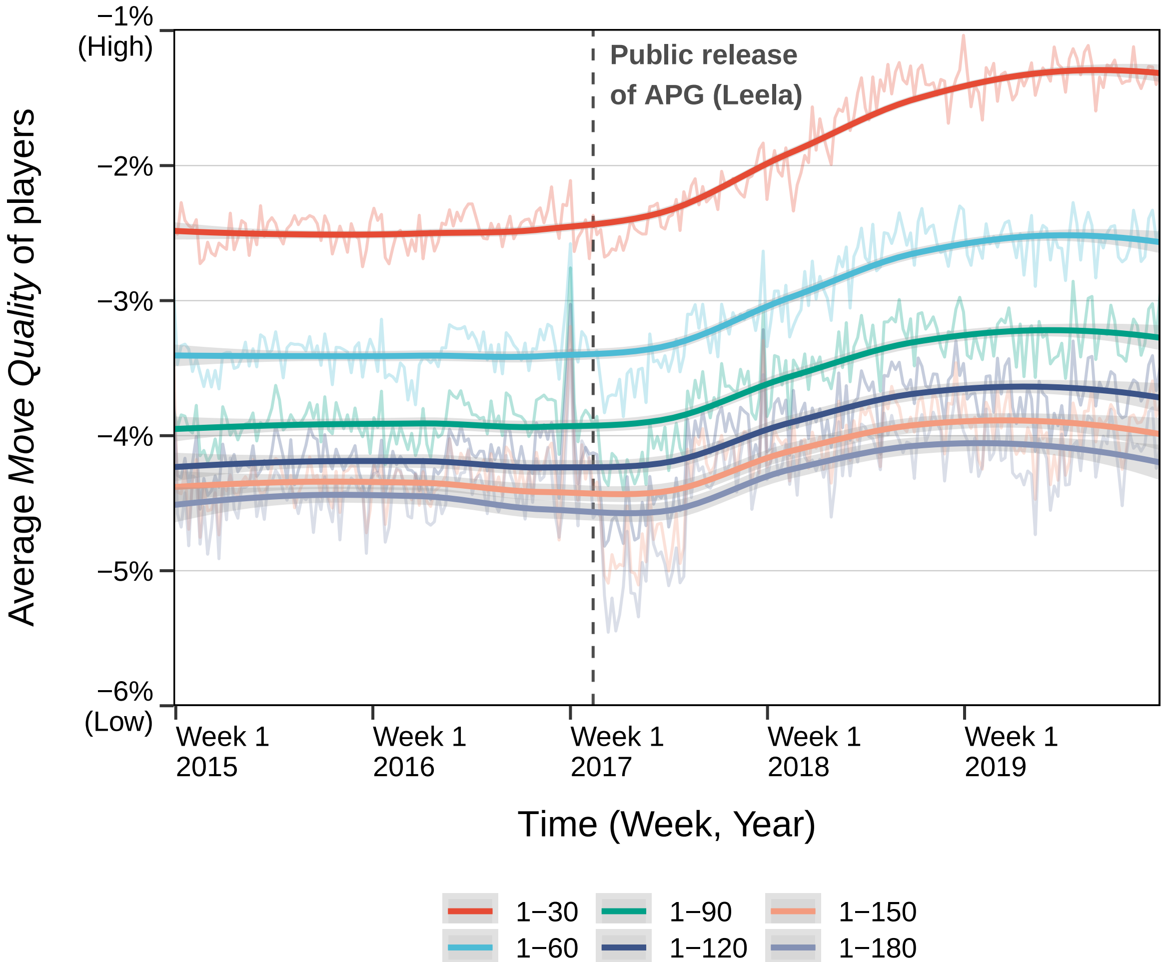
(a) Number of errors



(b) Magnitude of the most critical mistake

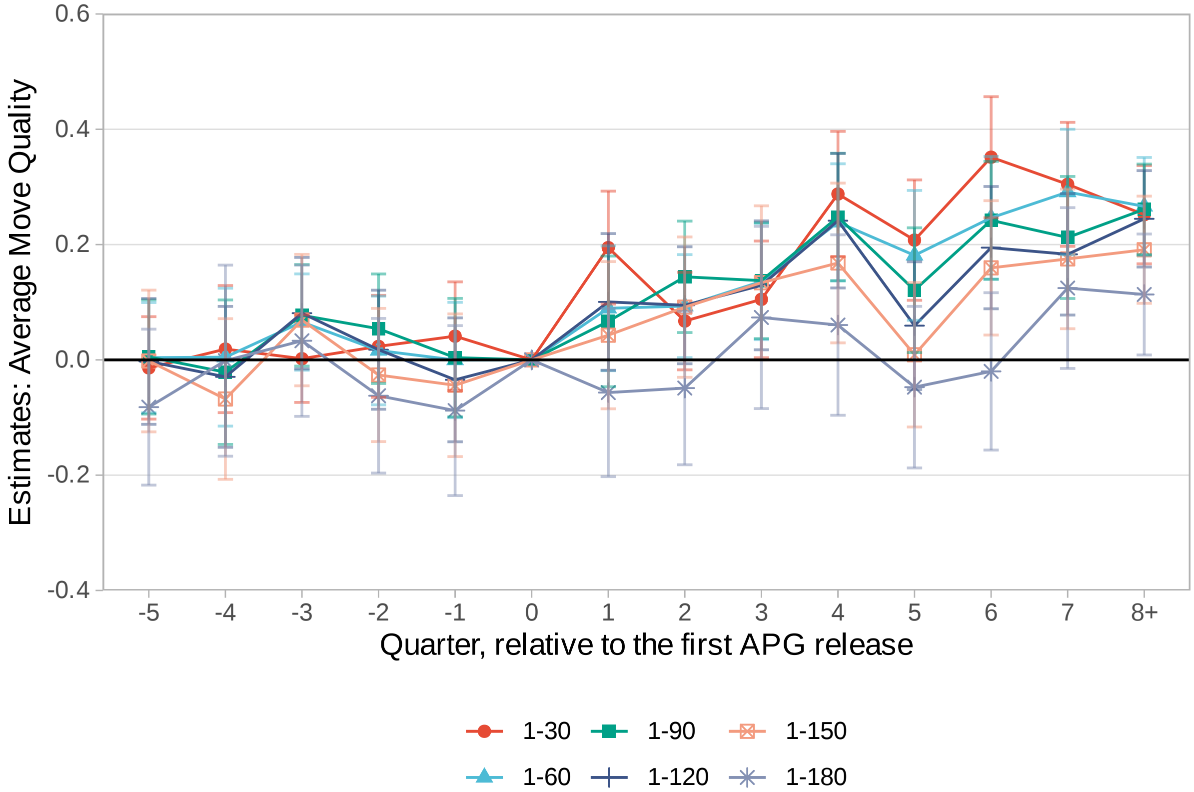
Note: This figure illustrates the weekly average of *Number of Errors* (Panel a) and the *Magnitude of the Most Critical Mistake* (Panel b) from 2015 through 2019. The solid black line represents the raw (unprocessed) weekly average value. The solid blue lines and the gray areas around them show locally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

Figure B.3: 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*.

Figure B.4: 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 B.1: Distributed leads and lags

Dependent Variable:	Move Quality			
	Age		Country	
Model:	(1)	(2)	(3)	(4)
Variables				
Rank	1.427***	2.291***	2.174***	3.084***
	(0.250)	(0.307)	(0.221)	(0.278)
	[p=0.000]	[p=0.000]	[p=0.000]	[p=0.000]
Rank Diff	0.059**	0.974***	0.058**	1.086***
	(0.024)	(0.163)	(0.027)	(0.180)
	[p=0.014]	[p=0.000]	[p=0.032]	[p=0.000]
White	-0.131***	-0.131***	-0.128***	-0.126***
	(0.009)	(0.010)	(0.010)	(0.010)
	[p=0.000]	[p=0.000]	[p=0.000]	[p=0.000]
7.5 Komi	0.025	0.043**	0.022	0.032*
	(0.016)	(0.019)	(0.016)	(0.019)
	[p=0.112]	[p=0.022]	[p=0.189]	[p=0.092]
-5 qr prior	-0.013	-0.014	-0.037	-0.014
	(0.045)	(0.045)	(0.052)	(0.052)
	[p=0.774]	[p=0.759]	[p=0.484]	[p=0.787]
-4 qr prior	0.039	0.019	-0.071	-0.084
	(0.054)	(0.056)	(0.059)	(0.060)
	[p=0.474]	[p=0.739]	[p=0.229]	[p=0.163]
-3 qr prior	0.011	0.002	-0.037	-0.052
	(0.038)	(0.039)	(0.046)	(0.046)
	[p=0.761]	[p=0.957]	[p=0.429]	[p=0.261]
-2 qr prior	0.028	0.023	-0.013	-0.020
	(0.044)	(0.045)	(0.052)	(0.053)
	[p=0.531]	[p=0.604]	[p=0.803]	[p=0.707]
-1 qr prior	0.050	0.041	0.039	0.034
	(0.048)	(0.048)	(0.054)	(0.054)
	[p=0.296]	[p=0.393]	[p=0.476]	[p=0.523]
1 qr after	0.207***	0.195***	0.165***	0.142***
	(0.049)	(0.050)	(0.054)	(0.055)
	[p=0.000]	[p=0.000]	[p=0.002]	[p=0.010]
2 qr after	0.085*	0.067	0.151***	0.115**
	(0.044)	(0.043)	(0.052)	(0.052)
	[p=0.051]	[p=0.119]	[p=0.003]	[p=0.027]
3 qr after	0.107**	0.105**	0.040	0.016
	(0.050)	(0.052)	(0.056)	(0.057)
	[p=0.035]	[p=0.042]	[p=0.469]	[p=0.777]
4 qr after	0.301***	0.288***	0.284***	0.251***
	(0.056)	(0.055)	(0.058)	(0.057)
	[p=0.000]	[p=0.000]	[p=0.000]	[p=0.000]
5 qr after	0.244***	0.208***	0.353***	0.314***
	(0.054)	(0.053)	(0.053)	(0.052)
	[p=0.000]	[p=0.000]	[p=0.000]	[p=0.000]
6 qr after	0.374***	0.352***	0.347***	0.309***
	(0.053)	(0.054)	(0.065)	(0.065)
	[p=0.000]	[p=0.000]	[p=0.000]	[p=0.000]
7 qr after	0.332***	0.304***	0.380***	0.329***
	(0.053)	(0.055)	(0.053)	(0.055)
	[p=0.000]	[p=0.000]	[p=0.000]	[p=0.000]
8+ qr after	0.282***	0.252***	0.271***	0.230***
	(0.043)	(0.043)	(0.042)	(0.043)
	[p=0.000]	[p=0.000]	[p=0.000]	[p=0.000]
Fixed-effects				
Quarter	Yes	Yes	Yes	Yes
Player	Yes	Yes	Yes	Yes
Opponent Player		Yes		Yes
Fit statistics				
Observations	46,454	46,454	42,783	42,783
R^2	0.325	0.350	0.328	0.353
Within R^2	0.015	0.015	0.016	0.015

Note: These regressions show the time-varying estimates using specifications described in Section 4.3 (age effect) for columns 1 and 2, and Section 5.1 (country effect) for columns 3 and 4. 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 and p-values are in squared brackets.

Table B.2: 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)		
	[p=0.000]	[p=0.018]		
Rank	0.783	0.750	1.608	2.396
	(0.036)	(0.036)	(0.227)	(0.284)
	[p=0.000]	[p=0.000]	[p=0.000]	[p=0.000]
Rank Diff	0.124	0.110	0.058	0.908
	(0.027)	(0.027)	(0.024)	(0.163)
	[p=0.000]	[p=0.000]	[p=0.015]	[p=0.000]
White	-0.133	-0.133	-0.131	-0.131
	(0.010)	(0.010)	(0.009)	(0.010)
	[p=0.000]	[p=0.000]	[p=0.000]	[p=0.000]
7.5 Komi	0.027	0.026	0.026	0.043
	(0.016)	(0.016)	(0.016)	(0.019)
	[p=0.096]	[p=0.097]	[p=0.104]	[p=0.022]
Post × Young		0.370	0.316	0.289
		(0.030)	(0.032)	(0.032)
		[p=0.000]	[p=0.000]	[p=0.000]
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 and p-values are in squared brackets.

Table B.3: Robustness check: Three age categories (Young, Mid, and Old)

Dependent Variable:	Move Quality			
Model:	(1)	(2)	(3)	(4)
Variables				
Mid	0.095			
	(0.022)			
	[p=0.000]			
Young	0.142			
	(0.025)			
	[p=0.000]			
Rank	0.804	0.773	1.503	2.340
	(0.036)	(0.036)	(0.243)	(0.296)
	[p=0.000]	[p=0.000]	[p=0.000]	[p=0.000]
Rank Diff	0.132	0.119	0.059	0.962
	(0.027)	(0.027)	(0.024)	(0.162)
	[p=0.000]	[p=0.000]	[p=0.014]	[p=0.000]
White	-0.133	-0.134	-0.131	-0.131
	(0.010)	(0.010)	(0.009)	(0.009)
	[p=0.000]	[p=0.000]	[p=0.000]	[p=0.000]
7.5 Komi	0.026	0.020	0.025	0.043
	(0.016)	(0.016)	(0.016)	(0.019)
	[p=0.109]	[p=0.201]	[p=0.109]	[p=0.022]
Post × Mid		0.224	0.221	0.198
		(0.031)	(0.033)	(0.033)
		[p=0.000]	[p=0.000]	[p=0.000]
Post × Young		0.295	0.328	0.302
		(0.032)	(0.040)	(0.040)
		[p=0.000]	[p=0.000]	[p=0.000]
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 and p-values are in squared brackets.

Table B.4: 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)		
	[p=0.880]	[p=0.675]		
Rank	0.857	0.856	2.366	3.259
	(0.036)	(0.036)	(0.228)	(0.274)
	[p=0.000]	[p=0.000]	[p=0.000]	[p=0.000]
Rank Diff	0.142	0.142	0.069	1.169
	(0.027)	(0.027)	(0.024)	(0.162)
	[p=0.000]	[p=0.000]	[p=0.005]	[p=0.000]
White	-0.133	-0.133	-0.131	-0.131
	(0.010)	(0.010)	(0.009)	(0.010)
	[p=0.000]	[p=0.000]	[p=0.000]	[p=0.000]
7.5 Komi	0.049	0.049	0.024	0.042
	(0.017)	(0.017)	(0.016)	(0.019)
	[p=0.004]	[p=0.004]	[p=0.135]	[p=0.028]
Post × Young		0.038	0.010	-0.043
		(0.029)	(0.028)	(0.027)
		[p=0.186]	[p=0.735]	[p=0.107]
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 and p-values are in squared brackets.

Table B.5: Robustness check: Month fixed effect

Dependent Variable:	Move Quality			
Model:	(1)	(2)	(3)	(4)
Variables				
Young	0.104	-0.044		
	(0.020)	(0.021)		
	[p=0.000]	[p=0.036]		
Rank	0.819	0.803	1.683	2.544
	(0.037)	(0.037)	(0.240)	(0.295)
	[p=0.000]	[p=0.000]	[p=0.000]	[p=0.000]
Rank Diff	0.138	0.131	0.066	1.021
	(0.027)	(0.027)	(0.024)	(0.162)
	[p=0.000]	[p=0.000]	[p=0.006]	[p=0.000]
White	-0.134	-0.133	-0.131	-0.131
	(0.010)	(0.010)	(0.009)	(0.010)
	[p=0.000]	[p=0.000]	[p=0.000]	[p=0.000]
7.5 Komi	0.022	0.019	0.013	0.027
	(0.017)	(0.016)	(0.016)	(0.019)
	[p=0.192]	[p=0.244]	[p=0.414]	[p=0.147]
Post × Young		0.259	0.208	0.192
		(0.028)	(0.031)	(0.031)
		[p=0.000]	[p=0.000]	[p=0.000]
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 and p-values are in squared brackets.

Table B.6: 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	1.946	2.888	1.812	2.636	1.688	2.318
	(0.226)	(0.280)	(0.241)	(0.300)	(0.233)	(0.285)
	[p=0.000]	[p=0.000]	[p=0.000]	[p=0.000]	[p=0.000]	[p=0.000]
Rank Diff	0.043	1.073	0.038	0.917	0.015	0.668
	(0.027)	(0.172)	(0.028)	(0.181)	(0.029)	(0.177)
	[p=0.113]	[p=0.000]	[p=0.185]	[p=0.000]	[p=0.615]	[p=0.000]
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)
	[p=0.000]	[p=0.000]	[p=0.000]	[p=0.000]	[p=0.000]	[p=0.000]
7.5 Komi	0.022	0.032	0.013	0.017	-0.000	0.003
	(0.016)	(0.019)	(0.015)	(0.019)	(0.016)	(0.019)
	[p=0.158]	[p=0.089]	[p=0.411]	[p=0.383]	[p=0.982]	[p=0.875]
Post × Young	0.184	0.170	0.183	0.171	0.167	0.154
	(0.030)	(0.029)	(0.029)	(0.029)	(0.028)	(0.027)
	[p=0.000]	[p=0.000]	[p=0.000]	[p=0.000]	[p=0.000]	[p=0.000]
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	0.294	0.319	0.268	0.293	0.242	0.267
Within R^2	0.012	0.012	0.010	0.011	0.009	0.009

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 and p-values are in squared brackets.

Table B.7: Placebo test: Random reassignment of the nationality

Dependent Variable:	Move Quality			
Model:	(1)	(2)	(3)	(4)
Variables				
Treated	-0.003	-0.001		
	(0.019)	(0.021)		
	[p=0.856]	[p=0.946]		
Rank	0.861	0.861	2.507	3.561
	(0.038)	(0.038)	(0.240)	(0.281)
	[p=0.000]	[p=0.000]	[p=0.000]	[p=0.000]
Rank Diff	0.154	0.154	0.061	1.325
	(0.031)	(0.031)	(0.027)	(0.175)
	[p=0.000]	[p=0.000]	[p=0.024]	[p=0.000]
White	-0.129	-0.129	-0.129	-0.126
	(0.010)	(0.010)	(0.010)	(0.010)
	[p=0.000]	[p=0.000]	[p=0.000]	[p=0.000]
7.5 Komi	0.055	0.055	0.020	0.030
	(0.017)	(0.017)	(0.017)	(0.020)
	[p=0.002]	[p=0.002]	[p=0.232]	[p=0.126]
Post × Treated		-0.003	-0.005	-0.004
		(0.031)	(0.031)	(0.030)
		[p=0.913]	[p=0.866]	[p=0.879]
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 and p-values are in squared brackets.

Table B.8: Learning from APG by comparing the match between human players' moves and APG's top 1, 3, and 5 suggestions

	Age			Country		
Dependent Variable:	Top 1	Top 3	Top 5	Top 1	Top 3	Top 5
Model:	(1)	(2)	(3)	(4)	(5)	(6)
Variables						
Rank	0.343	0.314	0.204	0.451	0.422	0.280
	(0.049)	(0.046)	(0.036)	(0.050)	(0.043)	(0.035)
	[p=0.000]	[p=0.000]	[p=0.000]	[p=0.000]	[p=0.000]	[p=0.000]
Rank Diff	0.088	0.081	0.065	0.102	0.096	0.080
	(0.027)	(0.024)	(0.020)	(0.030)	(0.026)	(0.022)
	[p=0.001]	[p=0.001]	[p=0.001]	[p=0.001]	[p=0.000]	[p=0.000]
White	0.045	0.091	0.041	0.046	0.092	0.042
	(0.002)	(0.002)	(0.001)	(0.002)	(0.002)	(0.001)
	[p=0.000]	[p=0.000]	[p=0.000]	[p=0.000]	[p=0.000]	[p=0.000]
7.5 Komi	0.002	0.002	-0.000	0.002	0.002	-0.001
	(0.003)	(0.003)	(0.002)	(0.003)	(0.003)	(0.002)
	[p=0.490]	[p=0.387]	[p=0.921]	[p=0.625]	[p=0.485]	[p=0.537]
Post × Young	0.032	0.025	0.017			
	(0.005)	(0.005)	(0.004)			
	[p=0.000]	[p=0.000]	[p=0.000]			
Post × Treated				0.038	0.031	0.022
				(0.005)	(0.005)	(0.004)
				[p=0.000]	[p=0.000]	[p=0.000]
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	46,454	46,454	42,783	42,783	42,783
R^2	0.293	0.388	0.307	0.296	0.391	0.311
Within R^2	0.031	0.120	0.039	0.033	0.124	0.042

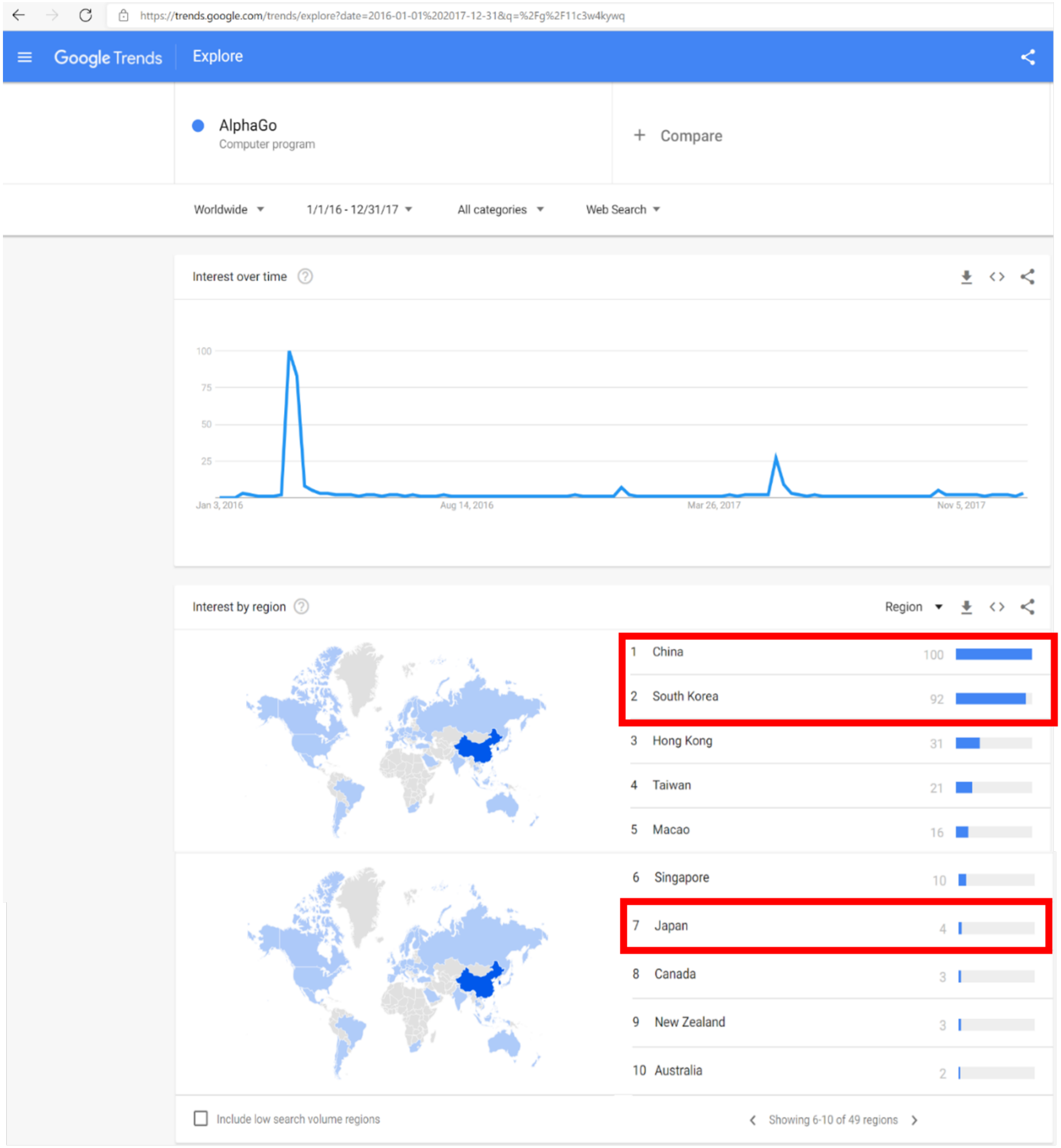
Note. This table shows the regression estimates the learning from APG by comparing the match between human players' moves and APG's top 1, 3, and 5 suggestions. Columns 1–3 test the age effects (equivalent to Table 3), while Columns 4–6 then show the country effects (equivalent to a simplified version of Table 5). Clustered standard errors at a focal player level are in parentheses and p-values are in squared brackets.

C. Interest in AI and APG by country

We consider different exposure to AI between major Go-playing countries: China, Japan, and South Korea. Since the AlphaGo events took place in China (date) and South Korea (date) but not in Japan, Japan should have relatively low awareness of APG. To verify, we did a Google Trend search with the keyword "AlphaGo" around the AlphaGo events—i.e., from January 2016 through December 2017.[4] Figure C.1 shows the screenshot of the search result. As expected, China and South Korea are the top two countries in terms of their interests in AlphaGo. When we set the interest level in China as 100 percent (benchmark), South Korea's interest was 92 percent. In contrast, Japan's interest was only four percent of China's. This verifies our argument that Japan exhibited little interest in APG (compared to China and South Korea) and supports our approach in Section 5.1. to use Japanese players and their moves as a comparison group.

[4] https://trends.google.com/trends/explore?date=2016-01-01%202017-12-31q=%2Fg%2F11c3w4kywq. Accessed November 16, 2021.

Figure C.1: Google Trend Search on AlphaGo by Country



Appendix References

Silver, D., Huang, A., Maddison, C. J., Guez, A., Sifre, L., Van Den Driessche, G., Schrittwieser, J., Antonoglou, I., Panneershelvam, V., Lanctot, M., 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., Baker, L., Lai, M., Bolton, A., et al. (2017). Mastering the game of go without human knowledge. *nature*, *550*(7676), 354–359.