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

Sukwoong Choi† Hyo Kang‡ Namil Kim§ Junsik Kim**

October 2023

Abstract

We study how humans learn from AI, exploiting an introduction of an AI-powered Go program (APG) that unexpectedly outperformed the best professional player. We compare the move quality of professional players to that of APG's superior solutions around its public release. Our analysis of 749,190 moves demonstrates significant improvements in players' move quality, accompanied by decreased number and magnitude of errors. The effect is pronounced in the early stages of the game where uncertainty is highest. In addition, younger players and those in AI-exposed countries experience greater improvement, suggesting potential inequality in learning from AI. Further, while players of all levels learn, less skilled players derive higher marginal benefits. These findings have implications for managers seeking to adopt and utilize AI effectively within their organizations.

Keywords: Artificial Intelligence, Learning from AI, Decision-making, Professional Go players, AI and Inequality

---

* We gratefully acknowledge valuable comments from Sinan Aral, Paola Criscuolo, Fabian Gaessler, Alfonso Gambardella, Nan Jia, Thorbjørn Knudsen, Myriam Mariani, Amalia R. Miller, Abhishek Nagaraj, Henning Piezunka, Ananya Sen, Neil Thompson, Christopher Tucci, Dashun Wang, and Yian Yin. 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 IP and Innovation (IPI) Seminar, the 3rd AI and Strategy Consortium, the 2022 Wharton Technology and Innovation Conference, the 2022 Columbia/Wharton Management, Analytics and Data Conference, and seminar participants at Bocconi University, Imperial College London, MIT, Northwestern University, Rotterdam School of Management, Stevens of Institute Technology, Stony Brook University, and University at Albany. We extend our gratitude to Hong-Ryeol Lee and Minho Kwak for providing proprietary data on Go playing time and their participation in interviews. This paper was previously circulated under the title, "Strategic Choices with Artificial Intelligence." All errors are our own.
† School of Business, University at Albany, State University of New York. schoi27@albany.edu.
‡ Marshall School of Business, University of Southern California. hyokang@marshall.usc.edu.
§ Corresponding author. Konkuk University. namilkim@konkuk.ac.kr.
** School of Engineering and Applied Sciences, Harvard University. mibastro@gmail.com.

1. INTRODUCTION

Artificial intelligence (AI) has developed substantially to date, and its capabilities have reached or even surpassed those of humans in numerous domains (Rai et al., 2019).[1] 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). In particular, AI helps workers perform better because it provides real-time assistance with 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). By comparing their own judgments to those provided by AI-based solutions, medical professionals reduce the uncertainty of their diagnoses and improve diagnostic quality (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).

Extant studies, which have taken an important step from focusing on AI's substitute roles toward considering its complementary roles, 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.[2] The performance gap between AI and humans suggests that humans can now learn from AI and can catch up with it in areas where they are currently outpaced. AI provides a new

[1] We focus on deep reinforcement learning algorithms as a technical definition of AI, as detailed in Plaat (2022).
[2] A concurrent study, Gaessler and Piezunka (2023), examines how chess players' performance varies by differential training opportunities with AI. We discuss how our study complements and expands theirs in Discussion, Section 6.1.

metric as it quantifies the expected outcome of alternatives that historically have been guessed—heuristically and tacitly—by long-standing custom or through learning by doing. AI's new metric 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 focuses on how AI can 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 heuristic 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) ages of individuals and their exposure to AI as key drivers that could affect AI's instructional effects.

Studying this topic empirically can be 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's dramatic progress is only recent, 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 skill 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, in 2017, the first open-access APG, Leela, became available to players. 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. We calculated the winning probability of 749,190 moves by 1,241 professional Go players in 24,973 major games held from 2015 through 2019. We then 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 in the probability of winning) after 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 solution suggested by the APG. This gap decreased by about 0.756 percentage points on average (or 30.5%) and *up to* 1.3 percentage points (or 47.6%) after the public release of the APG. Importantly, the improvement was accompanied by an increased *match* between players' moves and AI's top suggestions, confirming the effects were indeed driven by learning from AI.

We also explore the mechanisms through which professional players achieve a higher winning probability. Our mediation analysis reveals that a focal-player's improvement in the move quality is achieved mainly by reducing *Errors* (the number of moves where the winning probability drops by 10 or more percentage points compared to the wining probability of the immediately preceding move) and by reducing *Critical mistake* (the magnitude of the biggest drop in winning probability during the game). Additional analyses indicate that the improvement in move quality eventually leads to the final win of the game. 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, the improvement is more prominent among younger 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 749,190 moves in Go games has meaningful implications for AI's instructional role, notably for 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 context and findings provide meaningful managerial implications. Playing Go is similar to the 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 must consider the perspective of competitors and must possess intuitive techniques under the constraints of uncertainty and time similar to the constraints that arise in real business environments (Anderson, 2004; Roman & Vyas, 2021; Wiseman, 2016). Our findings also offer boundary conditions and heterogeneity of AI's effectiveness (e.g., by age, exposure to AI by country, and skill level of workers) of which managers should be aware for successful adoption and utilization of AI in organizations.

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 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). 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, the 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 decision-making (Brynjolfsson et al., 2011; Wu et al., 2019). For example, Brynjolfsson et al. (2011) find that the adoption of data-driven decision-making practices increases financial returns. Saunders and Tambe (2013) reveal that firms with data-driven decision-making at an executive level have higher productivity and market valuations. Data analytics also support decision-making for R&D search and incremental process improvements (Wu et al., 2020). Overall, the adoption 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 that may otherwise have yielded inferior decisions. Thanks to superior predictive capability, compared to that of 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 other settings. 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, within a few seconds AI can show the probability of the patient having some predefined disease. Similarly, when professional Go players use AI, they can immediately learn the winning probability associated with each possible move and can 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 choice, 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 (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, managers who utilize AI learn to improve their decision-making ability. 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. By repeatedly comparing their choices with those of AI, managers can update or revise their evaluation criteria based on AI's feedback (Yechiam & Busemeyer, 2005). For instance, AI-powered simulations present managers with opportunities for experiential learning, enabling them to understand superior choices through direct experience (Gaessler & Piezunka, 2023). Therefore, being equipped with the ability to make better evaluative choices, managers can make better decisions even without real-time AI assistance.

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 tools 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 the ages of individuals 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). Notably, the learning-by-doing literature indicates that the marginal effect of learning from new technology varies with age (or tenure) (Allen & Choudhury, 2022; Foster & Rosenzweig, 1995).

In the context of AI, extant studies find mixed results. Wang et al. (2019) studied medical coders in hospitals who used 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 possessed greater domain expertise than younger workers, tended 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. These studies tend to assume that seniority is associated with the workers who have accumulated experience and breadth of knowledge. To better understand the differential effects of age-related learning on AI adoption and utilization, we draw on the literature on algorithmic aversion and vintage-specific human capital.

Algorithmic aversion. Algorithmic aversion is the tendency of individuals to distrust or avoid algorithms in decision-making (Dietvorst et al., 2015). Individuals tend to undervalue the performance of algorithms, even when they are presented with evidence of the algorithm's superiority (Logg et al., 2019). Prior studies suggest that the aversion is exacerbated when individuals exhibit a higher level of risk aversion (Kahneman et al., 2016; Kahneman & Tversky, 1979), find a lack of transparency in the algorithm's workings (Shin & Park, 2019), and demonstrate low familiarity with technology (Dietvorst et al., 2015).

The degree of algorithmic aversion can also vary depending on the age of individuals (Mahmud et al., 2022). Young professionals tend to exhibit lower levels of risk aversion than the old (Tyler & Steensma, 1998). Furthermore, older professionals tend to view algorithmic decisions as less beneficial (Araujo et al., 2020) and exhibit lower trust in them (Lourenço et al., 2020). Allen and Choudhury (2022) show that senior professionals are more reluctant to accept algorithmic advice than are the young because seniors have greater confidence in their expertise and a greater sense of accountability for their actions. Building on these insights, we argue that younger professionals are less prone to algorithmic aversion.

Vintage-specific human capital. Vintage-specific human capital refers to the unique set of skills and knowledge that are specific to a certain time period or technology (Chari & Hopenhayn, 1991). As technology evolves and tasks change, individuals with vintage-specific human capital are better equipped to adapt to and utilize new technologies effectively (Autor et al., 2003; Gibbons &

Waldman, 2004). Younger professionals, having grown up in a digital environment from an early age, typically possess a better understanding of new technologies than their elders. The learning by doing literature suggests that these experiences improve their knowledge and skills (Arrow, 1962; Foster & Rosenzweig, 1995) and equip them with a rich vintage-specific human capital on emerging technologies (Michael G. Morris & Venkatesh, 2000; Schleife, 2006). Hence, younger professionals have greater absorptive capacity for AI intricacies (Choudhury et al., 2022) and more likely to learn from AI and do so more effectively.

2.3 Research questions

Based on the arguments above, we ask two primary questions about the relationship between AI and human decision-making. First, does AI improve human experts' decision-making and, if so, how? Second, how does the influence of AI vary across human professionals by their age and other characteristics? We argue that young professionals have greater incentives and ability to utilize AI-powered tools and benefit from them. In what follows, we empirically examine these research questions and conduct a series of post-hoc analyses to unpack what drives the observed patterns.

3. EMPIRICAL STRATEGY

3.1. Setting

3.1.1. The game of Go and professional tournaments

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 ($413,000) to the first-place winner in addition to per-game compensation.[3]

3.1.2. AI's entrance into Go

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).[4] 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).

Even if the latest supercomputers cannot calculate all possible moves in Go, recent advancement in deep reinforcement learning algorithms have improved AI remarkably. Instead of evaluating all possible solutions, AI uses these algorithms to reduce the potential moves to be considered and predicts sequential outcomes and winning probabilities.[5] AlphaGo, the initial APG with these algorithms, was invented by Google DeepMind. After several quality tests, Google held a historic Go match in 2016 between AlphaGo and the human Go master, Sedol Lee. Prior to this match, Lee and other Go experts expected that Lee would sweep all five games. Yet AlphaGo beat Lee 4–1, "a feat previously thought to be at least a decade away" (Silver et al. 2016). This event

[3] 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 Jin-seo 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.
[4] For comparison, chess has about 35^{80} possible moves. After the first two moves, chess has 400 possible next moves, while Go has 130,000 possible next moves (Muoio, 2016).
[5] The APG context, therefore, is distinct from the general development of IT; it is about high-dimensional calculations and predictions that only become possible with AI and deep reinforcement learning algorithms.

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 inferior 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).[6]

3.1.3. How much better at Go is AI compared to humans?

Go players are ranked and evaluated using the Elo rating system.[7] 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. 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. Learning from APGs

In the game of Go, AI technology is employed as a learning tool. APGs are not designed to provide real-time predictions to players on the spot during professional matches (which is strictly prohibited) but rather are used as a superior training tool to enhance players' decision-making capabilities. This distinction is important, as it highlights the potential of AI not only as a

[6] 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.
[7] 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.

productivity-enhancing tool but also as a means of experiential learning.

The testimonials of professional Go players highlight the learning effect it brings, largely attributable to its superior performance. Jin-seo Shin (who was ranked first in the world in 2020) provides further insights (Noh, 2019):

> "I have been using an APG since 2017. … I look at the APG's suggested move(s) and review other positions. … An APG is also used to predict the moves of opponents in the early stages. … Comprehending the move-level winning probability offered by APG is the new way I learn."

Figure A.1 in the Appendix A provides a practical example of the information that professional Go players obtain from an APG. At any point in the game, the APG displays several optimal moves with the winning probability associated with each suggested move; the different color schemes of the suggested moves make it easy to distinguish the very best move from others. Further, as the player chooses a move, the APG displays the optimal responses to that move, helping the player predict the opponent's responses in the following move. Repeating this training process enables professional Go players to substantially improve their understanding of strategic interactions in the game as well as their decision-making skills.

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 outperformed the best human player was Leela with its February 2017 update that utilized the deep-reinforcement-learning algorithm used in AlphaGo. A few months later, a new version, Leela Zero, was developed based on the algorithm of AlphaGo Zero; it had substantial impact on professional players. For example, the Korea Baduk (Go) Association and the South Korean National Go Team use Leela Zero for learning and training.

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 decided to develop the AlphaGo program solely because of Go's profound complexity (Burton-Hill, 2016). The developer of Leela, Gian-Carlo Pascutto, also made it clear that, although he had no interest in playing Go himself, 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. The event of interest is a major update of Leela in February 2017 that adopted the AlphaGo-based deep-learning algorithm. We conduct the analyses at the player-game level. Our sample consists of major professional Go games held from 2015 through 2019.

We then conduct a version of difference-in-differences estimation to understand the differential effects of APG. In an ideal world, we want to observe individual-level APG usage over time; unfortunately, such data is not available. Alternatively, we identify different age groups and compare the effects for young players ("treated") as opposed to old players ("comparison"). Although we do not have a clean control group—some players in the old group may have also adopted APG—we expect younger players to have adopted APG to a greater extent. Comparing the relative effect size for the young and old groups will produce a smaller estimate (i.e., biased toward zero) than an ideal estimation that uses a clean control group with no APG usage (see, for example, Agrawal et al., 2016; Kang & Lee, 2022; Lipsitz & Starr, 2022). We run a set of robustness checks and adopt a similar approach for a country comparison.

We focus primarily on early moves—the first thirty moves for each game—because, like many other games, 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 (Seungwook Noh, 2016).

We also analyze later stages and compare the results.

3.3 Data

Go games and professional 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), 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.

Measuring the quality of moves. Since Leela Zero provides the probability of winning for any possible move made at any particular point of the game, we use it to calculate the difference in winning probability between a move made by a professional player and Leela Zero's suggested move, a move that would achieve the highest winning probability 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 1st, 3rd, 5th, …, 29th moves if the focal player moves first or the 2nd, 4th, 6th, …, 30th moves otherwise). For each game, we separately calculate the value of the move qualities for each player i (the black stone holder and the 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. *Move Quality*$_{ig}$ takes a *non-positive* value (since APG is superior) and ranges from –100 (lowest quality) to 0 (highest quality). A smaller absolute number indicates higher-quality moves by the player. 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*=0). This variable becomes larger in absolute value as a player's moves deviate (worsen) from the best moves suggested by the APG.

We used Leela Zero (May 23, 2020 version) along with the GoReviewPartner program to analyze all 749,190 moves in 24,973 games played from 2015 through 2019. The computation took about three months; Appendix A.2 provides the calculation and implementation details.

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. That is, the players' 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 as the game progresses. 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 each other).

Table 1(b) shows the descriptive statistics at the player level. We identify 1,241 players from 2015 through 2019. The average age of players is 32.41, and the median age is 26.98.

"Insert Table 1 here"

4. RESULTS

4.1. Does APG improve the quality of moves by professional players?

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, but it increased immediately after Leela's public release.

"Insert Figure 2 here"

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. Focal-player fixed effects are represented by γ_i 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 (β=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.[8]

"Insert Table 2 here"

It is possible that the performance of professional players had been improving over time and drove the results, although Figure 2 does not indicate evidence of this. To control for this trend,

[8] All percentage changes are calculated as relative changes from the average move quality of games played by the players of interest in the preceding quarters throughout the sample period leading up to the release of the APG.

we add a $Trend_g$ variable (i.e., the number of quarters elapsed 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 (β=0.007, p<0.05), suggesting that the performance of professional players improved slowly over time. Importantly, the coefficient of the interaction term (β=0.116, p<0.01) shows that there are much larger—that is, 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 (–1.007+0.007×10+0.116×10).

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 then formally test whether the APG indeed has differential effects on the move quality of professional players of different ages. We estimate the following 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 such as *Komi*, *White*, *Rank*, and *Rank differences between players* at the player or game levels. $Young_i$ 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-seven years) as of Leela's public release in 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 (β=0.268, p<0.01) is positive and significant; the quality improvement for younger players is 0.268 percentage points (or 10.9 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 0.203–0.268 percentage points (or 8.2%–10.9%) over that of the older group, even after including the players' fixed effects.[9]

"Insert Table 3 here"

4.3. Robustness checks

We further check the robustness of the results in six 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, 5) the different numbers of moves for opening strategies (the first 15, 20, 40, 50, or 60 moves), and 6) a test for earlier Go programs.

Estimation 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, shows the detailed regression results, and

[9] Note that the magnitude of the effect is smaller than that in the main analysis (0.756 percentage points in Table 2, column 2). As discussed in Section 3.2, this is because our empirical design uses the old as the comparison group, and this group was also affected by APG in the same way (although to a lesser extent) as the young group.

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 APG's release, the improved quality by younger players is large and persistent.

Sensitivity test for age groups. We test whether the results are sensitive to our operationalization of age groups. First, we use the average age (instead of the median age) as the cutoff for the young and old groups; this increases the cutoff age from twenty-eight years to thirty-three years. The results in Table B.2 of Appendix B are robust to this alternative classification (β=0.268, p<0.01 in column 4). Second, we investigate the same model with three age groups based on the age tertile: "Young" (bottom tertile); "Middle" (middle tertile); and "Old" (top tertile). The results are provided in Table B.3 of Appendix B. The estimates for $Post_g \times Young_i$ (β=0.338, p<0.01 in column 4) and $Post_g \times Middle_i$ (β=0.248, p<0.01 in column 4) are large and statistically significant. Importantly, the effect is most pronounced among Young players compared to Old players, and the magnitude is smaller among Middle players. We obtain similar results when classifying players' ages into three categories: under age twenty; twenties (ages 20–29); and thirties or older.

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 no effects and thus cannot reject the null hypothesis that the age effect is zero. The estimate for $Post_g \times Young_i$ is close to zero and not statistically significant for the randomly assigned age group (see Table B.4 of Appendix B).

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 once again confirm the parallel time trend before the release of the APG and the substantial effect post-APG.

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 15, 20, 40, 50, and 60 moves of the game. The results, shown in Table B.6 of Appendix B, are robust to these alternative definitions.

The effect of earlier Go programs. Although Go programs prior to AlphaGo or Leela did not perform at the level of top human players, these programs may have offered training opportunities for professional players similar to how training sessions with early chess computers have been shown to improve the skills of chess players (Gaessler and Piezunka, 2023). The introduction of earlier Go programs therefore provides a valuable opportunity to check whether the effects are driven by learning from APG's superior performance or by having more frequent training opportunities (albeit with an inferior performance). We examined the impact of the earlier Go program, Crazy Stone, released in 2015; its Elo rating was just below 2,000 and inferior to the best human level (around 3,800; see Figure 1). Figure B.5 in Appendix B illustrates the results. We do not find any improvement in move quality after the release of Crazy Stone. This result rules out the possibility that more frequent training sessions (with inferior programs) are driving the findings and supports the proposition that learning from AI (i.e., from superior APGs) is the key channel through which players have improved their move quality.

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 2017 for reasons not directly related to APGs. To further address this concern and to explore the heterogeneous effects of exposure to AI, we exploit country-level variations in APG utilization. Among the three major countries with the largest professional Go leagues—(mainland)

China, Japan, and South Korea—exposure to or interest in APGs has been relatively lower in Japan. For example, two historical matches were held between each country's best player and an APG (AlphaGo vs. Sedol Lee, held in March 2016 in South Korea and AlphaGo vs. Ke Jie, held in May 2017 in China), but no match was held in Japan. A Google Trends search also reveals that, from 2016 through 2017, the term *AlphaGo* was searched most by China (interest score 100; a reference point) and South Korea (interest score 92); in contrast, Japan was ranked seventh with an interest score of 4 (see Appendix C for details). In an in-person interview, an expert shared that Japan has been reluctant to utilize APGs due partly to the deeply rooted mindset of Japanese players who consider Go to be an art form rather than a mere competition; these players were not comfortable with the APG's invasion of a game they consider to embody unique Japanese craftsmanship.

This motivates us to estimate a difference-in-difference model that compares players in countries that are significantly affected by APGs (i.e., China and Korea) with those in a country that is less affected (i.e., Japan). Note that, as discussed in Section 3.2, we do not have a clean control group. This will bias our estimates toward zero (or *against* our findings). Figure 4(a) shows 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. We then formally 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.

"Insert Figure 4 here"

Table 4 shows the results. The estimates for $Post_g \times Treat_{ic}$ are 0.315 to 0.227 percentage points depending on player- and opponent-fixed effects (Table 4, columns 2–4), and these are statistically significant at the 0.01 level. This suggests that the impact of AI is notably more substantial, by approximately 9.2 to 12.8 percent, for players who are more exposed to or who have a greater interest in AI, as compared with those who were less exposed or reluctant to adopt AI.

"Insert Table 4 here"

Figure 4(b) further show the time-varying effects of the APG on move quality (see also columns 3 and 4 in Table B.1 of the Appendix). We find no evidence of an increase in move quality for Chinese and South Korean players compared with 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 for both Chinese and South Korean players. From this stringent model, we once again find evidence that AI is responsible for improvement in the quality of moves. (For results up to 2022, see Appendix B, Figure B.7)

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

5.2. Move Match: 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 APGs became available supports this idea. Yet several alternative explanations can also be made. For instance, professional players may have changed their playing styles (without using an APG) after realizing that the APG beats the 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 have learned from APGs, the likelihood of a player making the exact same moves as the APG's top suggestions should increase. Given that APGs are not available while a game is played, a player's moves that exactly match those of the APG should provide strong evidence that, prior to the game, the player learned from the APG.

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. We consider k=1 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:

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

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 one, three, and five recommendations.

What is more interesting and convincing is that the estimates for $Post_g \times Young_i$ in Table B.8 shrink as we broaden the set: 0.031 (top one, column 1), 0.025 (top three, column 2), and 0.018 (top five, column 3). When players learn from an APG, they should be more inclined to learn the best move (i.e., the APG's top suggestion). The fact that the estimate is largest in magnitude for the top one move therefore supports our argument that players did learn from an APG.

Columns 4–6 in Table B.8 show the country effects. After the introduction of APGs, the matches between players' moves and the APG's top recommendations are greater for Chinese and Korean players than for Japanese players. Yet again, the estimate for $Post_g \times Treat_{ic}$ is largest for the top one recommendation, supporting the argument that results are driven by learning from AI.

5.3. How did players improve when their opponents used AI?

So far, we have focused on the focal player. An important question is how the quality of players' moves varies by the extent to which opponent players learn from AI. If the improvement in move quality is indeed driven by learning from AI, the effect should be greater for a pair of players where both have heavily utilized AI. This is because play between such players most resembles situations where a player has learned from AI. To check this, we split the sample by player age and by country and conduct a series of event-study analysis examining how the effect varies across different dyadic pairs. The results are illustrated as a heatmap in Figure 5. The move quality has improved across all pairs, but the effect is particularly marked among Young versus Young pairs (Panel a) and Chinese versus Chinese (or Chinese versus Korean) pairs (Panel b). In contrast, the improvement is relatively smaller for Old versus Old and Japanese versus Japanese pairs. This finding—that the effect is magnified when a player's counterpart has similarly learned from AI—once again bolsters our argument that players indeed learn from AI.

"Insert Figure 5 here"

5.4. Did better players improve more?

The increase in *average* move quality does not necessarily mean that players of different skill levels improved to the same extent. For example, it may be the case that positive average effects are driven by high (or low) performers. To this end, we first examine the effects across the *distribution* of players' skill levels (à la Athey & Imbens, 2006; Lipsitz & Starr, 2022). The effects at deciles of the skill level distribution are illustrated in Figure 6. The effects are positive across the whole range of the distribution, and the effect size is greater for the low-skilled.

"Insert Figure 6 here"

Further, we compare the improvement over time for players in the top decile (10^{th} decile) to those in the bottom (1^{st} and 2^{nd}) deciles. The model-free evidence is illustrated in Figure B.6 of Appendix B. Panel (a) shows there was improvement in move quality even among the top decile.

Further, in Panel (b), although the top performers experienced a notable improvement (up to 44.8%), the magnitude of improvement was more prominent for players in bottom deciles (up to 49.3%). Players with lower skills improved more than top players, reducing the performance gap.

5.5. 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 focal player's immediately preceding move. 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 (β=–0.009, p<0.01) is discontinuously accelerated after the introduction of APG (β=–0.233, p<0.01). These results confirm that learning from AI improved the quality of moves of professional players by reducing both the number of errors (33.7%) and the magnitude of the critical mistake (21.9%).

"Insert Table 5 here"

5.6. Did AI-driven improvements in 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 and Kenny (1986). As a baseline model, we run an OLS regression, or a linear probability model of winning a game on an interaction between an indicator for a younger player and an indicator for a post-APG period. In Table 6, column 1, the estimate for $Post_g \times Young_i$ (β=0.024, p<0.01) is positive and statistically significant. The improvements in move quality indeed led to a higher chance of winning; the changes of young players winning are on average 2.4 percentage points (4.6%) 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 with those in the baseline model (column 1). In the separately estimated mediation models for the three moderators—*Move quality*, *Errors*, and *Critical mistake*—the indirect effects through the explanatory variable account for 18.9 percent, 4.2 percent, and 13.1 percent of the total separate mediation effects, respectively, and these mediation effects are statistically significant under the Sobel test. In sum, younger players are more likely to win after the introduction of APG through their improvements in three dimensions: *Move quality*, *Errors*, and *Critical mistake*.

"Insert Table 6 here"

5.7. How did 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. Here we extend the analysis to include later stages of the game, incrementally adding thirty moves (up to 180 moves) to our analysis. We 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 0.203 (for moves 1–30) to 0.050 (for moves 1–180). The estimates with distributed leads and lags are graphically illustrated in Figure B.4 of Appendix B; younger players' improvement by using APG is highest for the opening strategy and becomes weaker as moves from later stages of the game are included.

"Insert Table 7 here"

To further investigate, we used data on a time stamp of each move for a subset of games and measured the time elapsed before each move (Choi et al., 2023). To the extent that players have learned from an APG, they should spend less time during the game. We compare changes in time spent by the stages of the game. As reported in Appendix D, players spent less time early in the game (moves 1–30) and more time in the later stages (moves 60–120), consistent with our argument that learning is more pronounced in the early stages of the game.

One explanation for this can be uncertainty. At the early stage of a game, when only a few stones have been 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 learning from AI can most help players to improve the quality of moves. 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 potential moves. The results suggest that the learning effect from AI can vary depending on the uncertainty of the environment and the opportunity to learn from AI.

6. DISCUSSION

6.1. APG and chess computers

Both this study and that of Gaessler and Piezunka (2023), who consider chess computers as an early form of AI and study their role as a training partner, explore the effect of AI on human decision-making; both studies share a similar empirical approach that uses granular data in the context of complicated board games.

The differences in the present study bring additional values to the understanding of the relationship between AI and human decision-making. First, the APG is powered by advanced (deep) reinforcement learning. Furthermore, unlike the chess computers of the 1970s–1990s, APG performance has surpassed that of the best human players.[10] Second, APGs were released *free of charge* and *around the same time* (similarly to the release of large language models such as

[10] Coincidently, the Elo rating of a pre-APG software, Crazy Stone, is similar to that of an advanced (later-stage) chess computer in Gaessler and Piezunka (2023). Unlike the chess setting, as examined in section 4.3.6, Crazy Stone did not improve the move quality of professional Go players.

ChatGPT), providing a unique opportunity to assess the impact of AI penetration.[11] Third, APG provides players with more information. APG makes three to five suggestions, gives the winning probabilities associated with those moves, and further suggests the likely next five to ten moves; chess computers, in contrast, provided only one deterministic move and no further information.

Fourth, taking advantage of APGs' superiority, we study how humans "learn" from AI; this takes a step forward from extant studies that view AI as a training (sparring) partner that is not necessarily better than the trainee. This feature also highlights the democratization of high-quality learning opportunities, as players are able to learn from the strategies and decisions of the very best player, APG. Interestingly, unlike what chess computers did earlier, training with pre-APG Go computers could not improve move quality (see Section 4.3.6). The superior performance of an APG program was paramount in enhancing the skills of professional Go players, perhaps because Go is much more complex than chess. Fifth, we expand our discussion on the boundary conditions—namely, age, exposure to AI by country, and the stage of the game—of AI's instructional roles. We also conduct dyadic analyses by players' age and country to study interactions between players and explore how the effect size varies across the *distribution* of players' ability. We hope this scholarly dialogue deepens our understanding of how humans learn from AI.

6.2. Generalizability and limitations

Our findings may not be representative of all kinds of human-AI interactions and the consequences; for instance, AI may replace humans in certain tasks or domains rather than help improve human skills. Still, cases where humans learn from AI and continue performing tasks is not rare or special. AI is not (yet) universally superior to humans in all domains, and it is not clear whether and when AI can achieve above-human level capabilities in all its bearings and completely replace humans.

[11] In contrast, chess computers came at a relatively high cost and were gradually diffused over several decades. For instance, the first commercial chess computer in 1977 costed $200 (approximately $1,000 in today's value).

AI is also costly in terms of algorithm development and computational power (Thompson et al., 2020), which may not justify its use in every situation. The continuation of human professionals performing tasks is necessary and so is learning from AI. The findings of this study offer the potential for AI tools to educate human professionals and offer insights that can be indicative of their impact in other settings. For example, workers can benefit from AI-powered training programs, which can enhance their decision-making abilities, leadership skills, and strategic thinking in complicated and uncertain business environments. In Appendix E we discuss several examples where the findings of this study may be applicable.

There are several limitations to this study. First, since we do not have individual-level data on APG usage, we use player characteristics as a proxy for APG adoption and utilization. While our auxiliary empirical analyses on move match rule out potential alternative explanations, future research could benefit from direct measure of AI usage. Second, the application of our findings to different contexts requires careful consideration. Although the domains where AI outperforms humans have broadened to include various organizations such as hospitals (Cadario et al., 2021), law firms (Kahn, 2020), and sports teams (Zarley, 2021), the Go context presents unique characteristics and strategic dynamics that may differ from other domains when it comes to AI.

6.3. Contributions and implications

The findings from AI in professional Go games provide timely implications for the ever-expanding role of AI and its relationship with humans. First, AI could reveal that what humans have believed to be a solution may not be the best approach; AI could bring breakthroughs in human knowledge, heuristics, or routines and pave the way for new paradigms. This study underscores the promise of AI algorithms in workforce training and AI-driven human resource management and contributes to the discussions on AI within the strategic human capital domain such as algorithmic aversion and vintage-specific skills (e.g., Choudhury et al. 2020; Gaessler & Piezunka 2023; Krakowski et al., 2022; Tong et al., 2021). Second, AI has broader application than merely substituting for or assisting with human tasks; we provide new theoretical and empirical accounts of how AI transforms human decision-making (Brynjolfsson et al., 2021). Although 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), existing studies have focused on AI's real-time assistant role in boosting task-related performance. AI's influence on altering human capabilities, such as decision-making skills, has received limited attention. We highlight the instructional role of AI as it may be utilized to improve the skills and performance of humans (contrary to a widespread concern that AI will replace human jobs). Our findings may be applied to certain domains where AI has already outperformed or will outperform human activities. For example, AI's performance in radiology is as good as that of 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).

Third, not everyone may enjoy the benefits of AI at the same level. We show that openness to new technologies and the ability to utilize it (characterized by individuals age, experience, or cultural background) could contribute to reaping benefits from AI. These findings add to a growing stream of literature on differential effects or the potential inequality implications of AI (e.g., Beane & Anthony, 2023; Choudhury et al., 2020; Miric et al., 2020).

Fourth, the impact of AI also depends on the complexity and uncertainty of a situation. In Go, AI-driven improvement is most prominent in the early stages of a game. This boundary condition of 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 optimal outcome and could lead to inefficient allocation of AI and human resources. A careful consideration of where to adopt and utilize AI and to what extent is therefore required.

This study also offers managerial implications for firms. In today's rapidly evolving AI era, a central question facing firms is how to utilize AI to achieve competitive advantage and enhance performance. Our study suggests that organizations can leverage AI tools in line with firm strategy to boost managers' skills. By integrating AI tools, such as AI-powered simulations, firms can offer valuable learning opportunities to help employees improve their decision-making capabilities (and mitigate their errors and mistakes). In so doing, it is critical for firms to be aware of the boundary conditions and heterogeneity and to tailor these programs. For example, the gains from using AI tools are not uniform across workers, and relatively low-skilled workers could gain more. When facing budget constraints, managers may prioritize the use of AI tools for lower performers to maximize the marginal returns from using AI tools. Understanding how AI's effectiveness is influenced by age and prior exposure to AI could also lead to successful adoption and utilization of AI, thereby fostering a firm's innovation and growth.

REFERENCES

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

Agrawal, A., Goldfarb, A., & Teodoridis, F. (2016). Understanding the Changing Structure of Scientific Inquiry. *American Economic Journal: Applied Economics*, *8*(1), 100–128.

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.

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

Araujo, T., Helberger, N., Kruikemeier, S., & de Vreese, C. H. (2020). In AI we trust? Perceptions about automated decision-making by artificial intelligence. *AI and Society*, *35*(3), 611–623.

Arrow, K. J. (1962). The Economic Learning Implications of by Doing. *The Review of Economic Studies*, *29*(3), 155–173.

Athey, S., & Imbens, G. W. (2006). Identification and inference in nonlinear difference-in-differences models. *Econometrica*, *74*(2), 431–497.

Athey, S., & Imbens, G. W. (2019). Machine Learning Methods That Economists Should Know About. *Annual Review of Economics*, *11*(1), 685–725.

Autor, D. H., Levy, F., & Murnane, R. J. (2003). The Skill Content of Recent Technological Change: An Empirical Exploration. *The Quarterly Journal of Economics*, *118*(4), 1279–1333.

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.

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.

Barth, E., Davis, J. C., Freeman, R. B., & McElheran, K. (2020). Twisting the Demand Curve: Digitalization

and the Older Workforce. *NBER Working Paper*.
Beane, M., & Anthony, C. (2023). Inverted Apprenticeship: How Senior Occupational Members Develop Practical Expertise and Preserve Their Position When New Technologies Arrive. *Organization Science*.
Blackwell, D. (1953). Equivalent Comparisons of Experiments. *The Annals of Mathematical Statistics*, *24*(2), 265–272.
Blei, D. M., & Smyth, P. (2017). Science and data science. *Proceedings of the National Academy of Sciences*, *114*(33), 8689–8692.
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.
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*.
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.
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*.
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*.
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.
Chari, V. V, & Hopenhayn, H. (1991). Vintage Human Capital, Growth, and the Diffusion of New Technology. *The Journal of Political Economy*, *99*(6), 1142–1165.
Chen, W. R. (2008). Determinants of firms' backward- and forward-looking R&D search behavior. *Organization Science*, *19*(4), 609–622.
Choi, S., Kang, H., Kim, N., & Kim, J. (2023). Standing on the Shoulders of AI: Pushing the Knowledge Frontier by Learning from AI. *Academy of Management Proceedings*, 15478.
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.
Cyert, R. M., & March, J. G. (1963). A behavioral theory of the firm. In *Englewood Cliffs, NJ* (Vol. 2, Issue 4).
Danziger, S., Levav, J., & Avnaim-Pesso, L. (2011). Extraneous factors in judicial decisions. *Proceedings of the National Academy of Sciences*, *108*(17), 6889–6892.
Davenport, T., & Harris, J. (2017). *Competing on analytics: Updated, with a new introduction: The new science of winning*. Harvard Business Press.
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
Dietvorst, B. J., Simmons, J. P., & Massey, C. (2015). Algorithm aversion: People erroneously avoid algorithms after seeing them err. *Journal of Experimental Psychology: General*, *144*(1), 114–126.
Foster, A. D., & Rosenzweig, M. R. (1995). Learning by Doing and Learning from Others: Human Capital and Technical Change in Agriculture. *Journal of Political Economy*, *103*(6), 1176–1209.
Gaessler, F., & Piezunka, H. (2023). Training with AI: Evidence from chess computers. *Strategic Management Journal*, *July 2020*, 1–27.
Galbraith, J. R. (1974). Organization Design: An Information Processing View. *Interfaces*, *4*(3), 28–36.
Gavetti, G., & Levinthal, D. (2000). Looking forward and looking backward: Cognitive and experiential search. *Administrative Science Quarterly*, *45*(1), 113–137.
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.
Gibbons, R., & Waldman, M. (2004). Task-Specific Human Capital. *American Economic Review*, *94*(2), 203–207.
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

Hagendorff, T., & Wezel, K. (2020). 15 challenges for AI: or what AI (currently) can't do. *AI & SOCIETY*, *35*(2), 355–365.

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.

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.

Kahneman, D., Rosenfield, A. M., Gandhi, L., & Blaser, T. (2016). Noise: How to Overcome the High, Hidden Cost of Inconsistent Decision Making. *Harvard Business Review*, *October*.

Kahneman, D., & Tversky, A. (1979). On the interpretation of intuitive probability: A reply to Jonathan Cohen. *Cognition*, *7*(4), 409–411.

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.

Kang, H., & Lee, W. (2022). How innovating firms manage knowledge leakage: A natural experiment on the threat of worker departure. *Strategic Management Journal*, *March*, 1–22.

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.

Kleinberg, J., Lakkaraju, H., Leskovec, J., Ludwig, J., & Mullainathan, S. (2017). Human Decisions and Machine Predictions. *The Quarterly Journal of Economics*, *January*, 237–293.

Knight, W. (2016, January 27). Google's AI Masters the Game of Go a Decade Earlier Than Expected. *MIT Technology Review*.

Krakowski, S., Luger, J., & Raisch, S. (2022). Artificial intelligence and the changing sources of competitive advantage. *Strategic Management Journal*, *1*(2), 57–72.

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.

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.

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.

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.

Lipsitz, M., & Starr, E. (2022). Low-Wage Workers and the Enforceability of Noncompete Agreements. *Management Science*, *68*(1), 143–170.

Logg, J. M., Minson, J. A., & Moore, D. A. (2019). Algorithm appreciation: People prefer algorithmic to human judgment. *Organizational Behavior and Human Decision Processes*, *151*(December 2018), 90–103.

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

Lourenço, C. J. S., Dellaert, B. G. C., & Donkers, B. (2020). Whose Algorithm Says So: The Relationships between Type of Firm, Perceptions of Trust and Expertise, and the Acceptance of Financial Robo-Advice. *Journal of Interactive Marketing*, *49*(1), 107–124.

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

Mahmud, H., Islam, A. K. M. N., Ahmed, S. I., & Smolander, K. (2022). What influences algorithmic decision-making? A systematic literature review on algorithm aversion. *Technological Forecasting and Social Change*, *175*(August 2021), 121390.

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

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

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

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.

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.

Morris, Michael G., & Venkatesh, V. (2000). Age Differences in Technology Adoption Decisions: Implications for a Changing Work Force. *Personnel Psychology*, *53*(2), 375–403.

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

Noh, Seongyeol. (2019, March 5). AI Shakes Up Millennia-Old Opening Strategy, Unveils New Strategy to Players. *The Munhwa Il-Bo*. http://www.munhwa.com/news/view.html?no=2019030501032003009001.

Noh, Seungwook. (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/.

Plaat, A. (2022). *Deep reinforcement learning* (Vol. 10). Springer.

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.

Reeves, M., & Wittenburg, G. (2015). Games Can Make You a Better Strategist. *Harvard Business Review*, 3.

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.

Schleife, K. (2006). Computer Use and Employment Status of Older Workers - An Analysis Based on Individual Data. *Labour*, *20*(2), 325–348.

Schrittwieser, J., Antonoglou, I., Hubert, T., Simonyan, K., Sifre, L., Schmitt, S., Guez, A., 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.

Seamans, R., & Raj, M. (2018). AI, Labor, Productivity and The Need for Firm-Level Data. *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.

Shin, D., & Park, Y. J. (2019). Role of fairness, accountability, and transparency in algorithmic affordance. *Computers in Human Behavior*, *98*, 277–284.

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.

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.

Simon, H. A. (1955). A Behavioral Model of Rational Choice. *The Quarterly Journal of Economics*, *69*(1), 99.

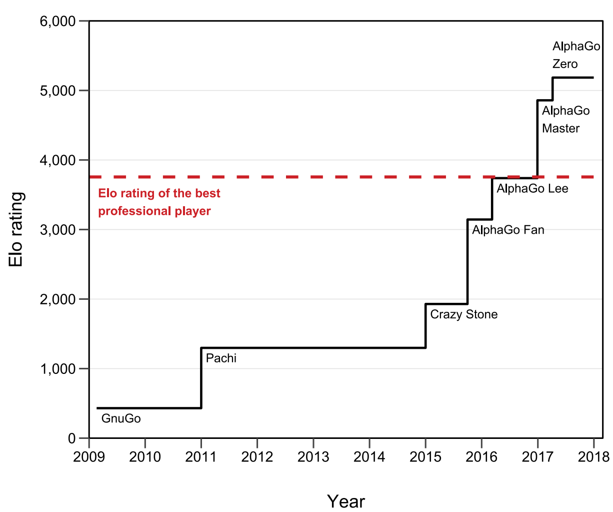
Simon, H. A., & March, J. G. (1958). *Organizations*. Wiley-Blackwell.

Smalley, E. (2017). AI-powered drug discovery captures pharma interest. *Nature Biotechnology*, *35*(7), 604–605.

Smith, B. C. (2019). *The promise of artificial intelligence: reckoning and judgment*. MIT Press.

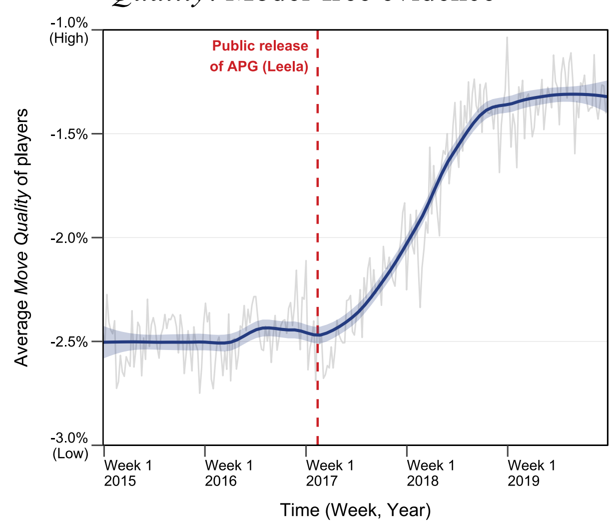
Somers, J. (2018, December 28). How the Artificial-Intelligence Program AlphaZero Mastered Its Games. *The New Yorker*, 1–6.
Tams, S. (2022). Helping Older Workers Realize Their Full Organizational Potential: a Moderated Mediation Model of Age and It-Enabled Task Performance. *MIS Quarterly*.
Tams, S., 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.
Thompson, N. C., Greenewald, K., Lee, K., & Manso, G. F. (2020). The Computational Limits of Deep Learning. *ArXiv*. http://arxiv.org/abs/2007.05558
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.
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.
Venkatesh, Morris, Davis, & Davis. (2003). User Acceptance of Information Technology: Toward a Unified View. *MIS Quarterly*, *27*(3), 425.
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*.
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.
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.
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.
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. *Data*. GoRatings: https://www.goratings.org/en/. Go4Go: https://www.go4go.net/go/players/rank/ .

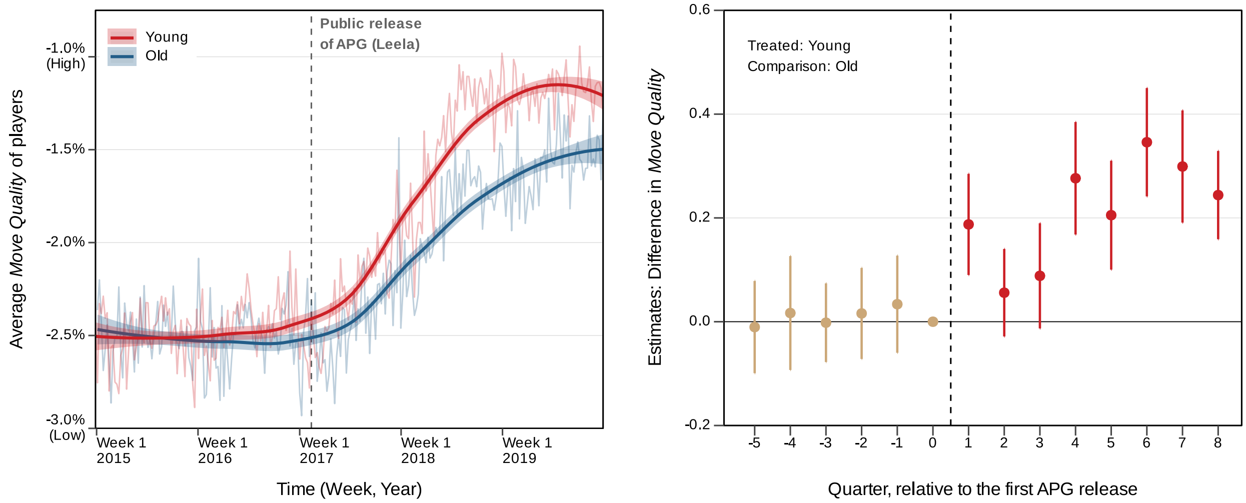
Figure 2. Effects of APG on average *Move Quality*: Model-free evidence



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=0.7) 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

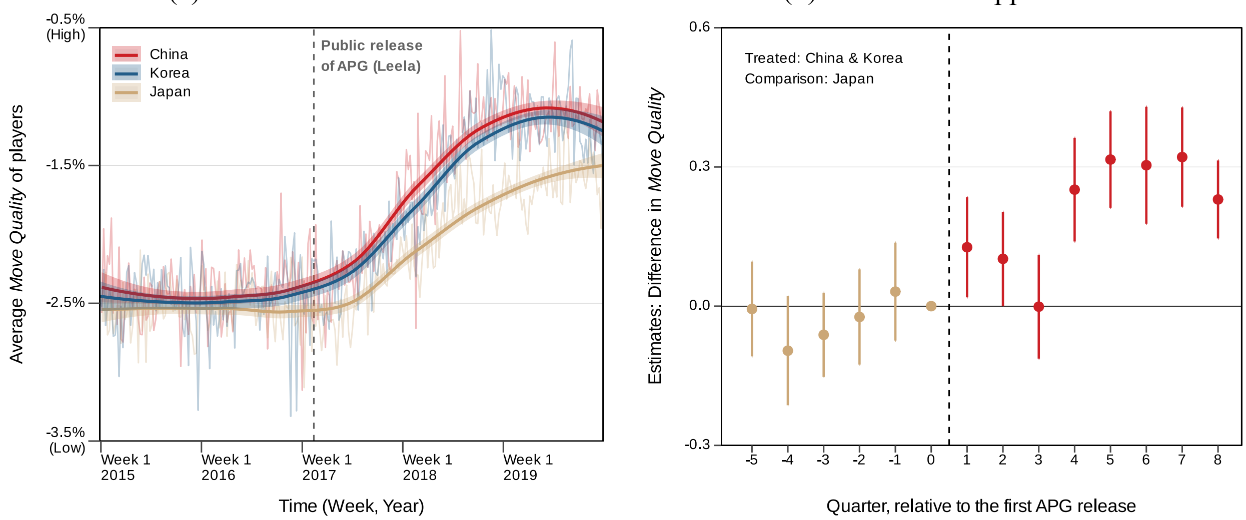


Note. This figure illustrates the average *Move Quality* of professional players by player age. The red and blue fluctuating lines show the raw (unprocessed) weekly average values for younger players (below median age) and older players (above median age), respectively. The red and blue smooth lines 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 (those below the 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 exposure to AI by players' country

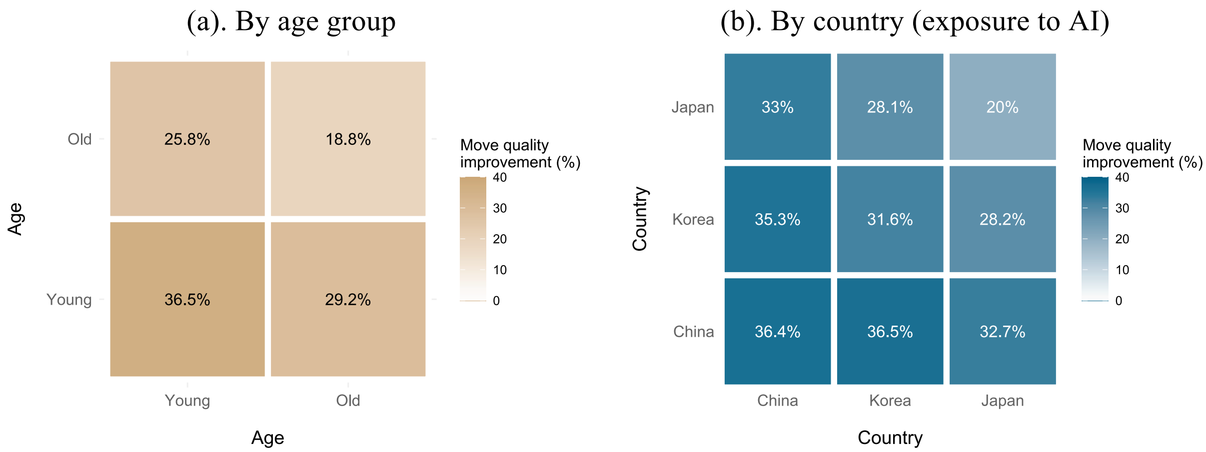
(a) Model-free evidence (b) Diff-in-diffs approach



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.

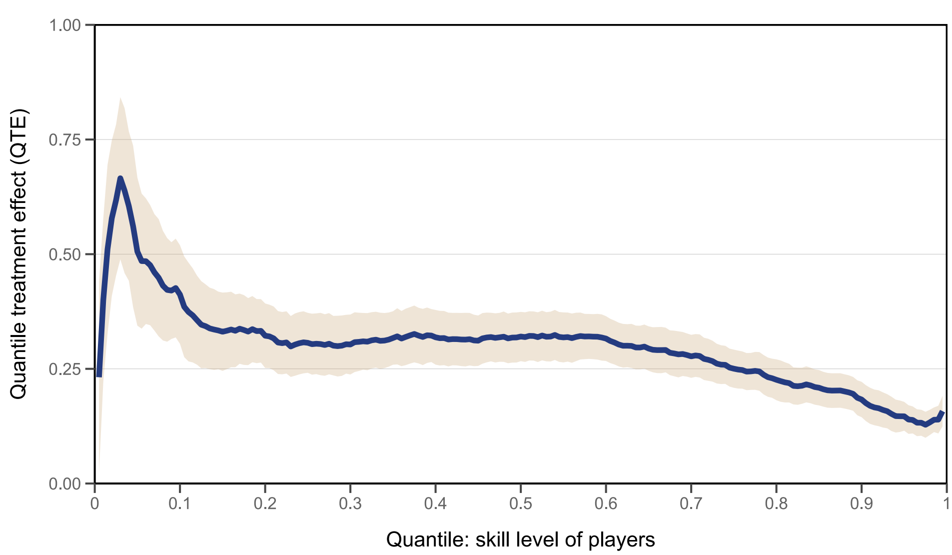
Note. This figure illustrates the effects of APG on *Move Quality*, based on the difference-in-differences estimation results reported in Table B.1 of Appendix B, column 4. The inclusion of Japan in the control group would bias our estimates toward zero (i.e., *against* our findings), leading to an underestimation.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

Figure 5. The effects of APG on *Move Quality* across dyadic relationships of age and country groups



Note. This figure illustrates the improvement in *Move Quality* across various dyadic relationships, categorized by pairs of age groups as shown in Panel (a) and by exposure to AI by county as shown in Panel (b), around the introduction of APG.

Figure 6. Quantile treatment effects on *Move Quality*



Note. This figure illustrates treatment effects at quantiles of *Move Quality* before and after the first public release of an APG, Leela, following the method suggested by Athey and Imbens (2006). The shaded areas show bootstrapped 95% confidence intervals.

Table 1. Descriptive statistics

(a). Player-game level

	N	Mean	Median	SD	P25	P75
Move Quality	49,946	–2.01	–1.92	1.07	–2.66	–1.22
Number of Errors	49,946	0.13	0.00	0.37	0.00	0.00
Magnitude of the Critical Mistake	49,945	5.66	4.80	3.96	2.94	7.36
Age	49,613	28.07	24.31	12.58	19.55	31.52
Young	49,613	0.62	1.00	0.48	0.00	1.00
Rank	48,813	–0.27	–0.17	0.27	–0.42	–0.05
Rank Diff	47,826	0.00	0.00	0.21	–0.09	0.09
White	49,946	0.50	0.50	0.50	0.00	1.00
7.5 Komi	49,946	0.38	0.00	0.49	0.00	1.00

(b). Player level

	N	Mean	Median	SD	P25	P75
Move Quality	1,241	–2.20	–2.18	0.67	–2.52	–1.79
Number of Errors	1,241	0.17	0.12	0.22	0.00	0.20
Magnitude of the Critical Mistake	1,241	–6.20	–5.98	2.09	–6.96	–5.05
Age	1,188	32.41	26.98	16.11	20.12	42.55
Young	1,188	0.50	1.00	0.50	0.00	1.00
Rank	1,104	–0.52	–0.52	0.31	–0.79	–0.26
Rank Diff	1,097	0.14	0.11	0.18	0.00	0.25

Note. This table provides descriptive statistics of variables at the player-game level in Panel (a) and at the player level in 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	–1.007
	(0.017)	(0.038)
	[p<0.001]	[p<0.001]
Trend		0.007
		(0.003)
		[p=0.023]
Post × Trend		0.116
		(0.004)
		[p<0.001]
Fixed effects		
Player	Yes	Yes
Opponent Player	Yes	Yes
Fit statistics		
Observations	49,946	49,946
R^2	0.264	0.330
Within R^2	0.116	0.195

Note. This table shows the regression estimates on the effects of APG on the *Move Quality* of professional Go players, before and after the first public release of 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 elapsed 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	0.096	–0.053		
	(0.020)	(0.021)		
	[p<0.001]	[p=0.010]		
Rank	0.846	0.828	1.723	2.582
	(0.036)	(0.037)	(0.246)	(0.292)
	[p<0.001]	[p<0.001]	[p<0.001]	[p<0.001]
Rank Diff	0.128	0.120	0.067	1.040
	(0.028)	(0.028)	(0.025)	(0.164)
	[p<0.001]	[p<0.001]	[p=0.008]	[p<0.001]
White	–0.133	–0.133	–0.130	–0.131
	(0.010)	(0.010)	(0.009)	(0.009)
	[p<0.001]	[p<0.001]	[p<0.001]	[p<0.001]
7.5 Komi	0.021	0.019	0.021	0.037
	(0.016)	(0.016)	(0.016)	(0.018)
	[p=0.191]	[p=0.248]	[p=0.190]	[p=0.047]
Post × Young		0.268	0.220	0.203
		(0.028)	(0.031)	(0.031)
		[p<0.001]	[p<0.001]	[p<0.001]
Fixed effects				
Quarter	Yes	Yes	Yes	Yes
Player			Yes	Yes
Opponent Player				Yes
Fit statistics				
Observations	47,292	47,292	47,292	47,292
R^2	0.277	0.281	0.325	0.350
Within R^2	0.065	0.070	0.013	0.014

Note. This table shows the regression estimates on the heterogeneous effects of APG by player age; the *Move Quality* of young players compared to that of old players is estimated. *Post* refers to games played in the quarters after the first public release of 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: Difference-in-differences estimation using cross-country variation in exposure to APG

Dependent Variable:	Move Quality			
Model:	(1)	(2)	(3)	(4)
Variables				
Treated	0.106	–0.074		
	(0.024)	(0.026)		
	[p<0.001]	[p=0.005]		
Rank	0.844	0.845	2.375	3.276
	(0.040)	(0.040)	(0.219)	(0.275)
	[p<0.001]	[p<0.001]	[p<0.001]	[p<0.001]
Rank Diff	0.123	0.120	0.064	1.115
	(0.033)	(0.033)	(0.028)	(0.183)
	[p<0.001]	[p<0.001]	[p=0.024]	[p<0.001]
White	–0.130	–0.130	–0.129	–0.126
	(0.010)	(0.010)	(0.010)	(0.010)
	[p<0.001]	[p<0.001]	[p<0.001]	[p<0.001]
7.5 Komi	0.006	0.005	0.019	0.027
	(0.017)	(0.017)	(0.017)	(0.019)
	[p=0.708]	[p=0.757]	[p=0.240]	[p=0.159]
Post × Treated		0.315	0.263	0.227
		(0.031)	(0.030)	(0.031)
		[p<0.001]	[p<0.001]	[p<0.001]
Fixed effects				
Quarter	Yes	Yes	Yes	Yes
Player			Yes	Yes
Opponent Player				Yes
Fit statistics				
Observations	43,151	43,151	43,151	43,151
R^2	0.278	0.283	0.327	0.352
Within R^2	0.063	0.069	0.015	0.014

Note. This table shows the effects of APGs on *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	0.082	–1.430	2.261
	(0.004)	(0.013)	(0.053)	(0.143)
	[p<0.001]	[p<0.001]	[p<0.001]	[p<0.001]
Trend		0.000		–0.028
		(0.001)		(0.012)
		[p=0.730]		[p=0.021]
Post × Trend		–0.009		–0.233
		(0.002)		(0.015)
		[p<0.001]		[p<0.001]
Fixed effects				
Player	Yes	Yes	Yes	Yes
Opponent Player	Yes	Yes	Yes	Yes
Fit statistics				
Observations	49,946	49,946	49,945	49,945
R^2	0.077	0.081	0.123	0.145
Within R^2	0.005	0.008	0.028	0.052

Note. This table shows the impact of 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	Magnitude of the Critical Mistake	Win						
Model:	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Variables											
Rank	–0.095	2.582	–0.350	–7.373	–0.096	–0.027	–0.066	–0.161	–0.110	–0.140	–0.167
	(0.106)	(0.292)	(0.101)	(1.139)	(0.098)	(0.099)	(0.098)	(0.105)	(0.106)	(0.105)	(0.105)
	[*p*=0.372]	[*p*<0.001]	[*p*<0.001]	[*p*<0.001]	[*p*=0.327]	[*p*=0.782]	[*p*=0.499]	[*p*=0.125]	[*p*=0.300]	[*p*=0.183]	[*p*=0.111]
Rank Diff	–1.802	1.040	0.016	–1.897	–1.826	–1.796	–1.811	–1.829	–1.801	–1.815	–1.829
	(0.087)	(0.164)	(0.065)	(0.692)	(0.086)	(0.086)	(0.086)	(0.087)	(0.087)	(0.087)	(0.087)
	[*p*<0.001]	[*p*<0.001]	[*p*=0.804]	[*p*=0.006]	[*p*<0.001]	[*p*<0.001]	[*p*<0.001]	[*p*<0.001]	[*p*<0.001]	[*p*<0.001]	[*p*<0.001]
White	0.026	–0.131	0.047	0.628	0.029	0.028	0.030	0.029	0.027	0.029	0.030
	(0.005)	(0.009)	(0.004)	(0.037)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)
	[*p*<0.001]	[*p*<0.001]	[*p*<0.001]	[*p*<0.001]	[*p*<0.001]	[*p*<0.001]	[*p*<0.001]	[*p*<0.001]	[*p*<0.001]	[*p*<0.001]	[*p*<0.001]
7.5 Komi	0.000	0.037	–0.013	–0.129	–0.001	0.000	–0.001	–0.001	–0.001	–0.001	–0.001
	(0.010)	(0.018)	(0.008)	(0.080)	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)
	[*p*=0.990]	[*p*=0.047]	[*p*=0.093]	[*p*=0.107]	[*p*=0.945]	[*p*=0.978]	[*p*=0.958]	[*p*=0.918]	[*p*=0.949]	[*p*=0.930]	[*p*=0.906]
Post × Young	0.024	0.203	–0.024	–0.487				0.018	0.023	0.021	0.018
	(0.009)	(0.031)	(0.009)	(0.103)				(0.009)	(0.009)	(0.009)	(0.009)
	[*p*=0.008]	[*p*<0.001]	[*p*=0.006]	[*p*<0.001]				[*p*=0.041]	[*p*=0.012]	[*p*=0.021]	[*p*=0.041]
Move Quality					0.026			0.026			0.017
					(0.002)			(0.002)			(0.003)
					[*p*<0.001]			[*p*<0.001]			[*p*<0.001]
Number of Errors						–0.042			–0.042		0.003
						(0.006)			(0.006)		(0.009)
						[*p*<0.001]			[*p*<0.001]		[*p*=0.695]
Magnitude of the Critical Mistake							–0.006			–0.006	–0.004
							(0.001)			(0.001)	(0.001)
							[*p*<0.001]			[*p*<0.001]	[*p*<0.001]
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	47,280	47,292	47,292	47,291	47,538	47,538	47,537	47,280	47,280	47,279	47,279
R^2	0.214	0.350	0.082	0.153	0.217	0.215	0.217	0.216	0.215	0.216	0.216
Within R^2	0.018	0.014	0.005	0.011	0.020	0.019	0.020	0.020	0.019	0.020	0.021

Note. This table shows how *Move Quality* leads to 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	2.582	2.209	1.623	1.351	1.131	0.964
	(0.292)	(0.285)	(0.303)	(0.307)	(0.341)	(0.384)
	[*p*<0.001]	[*p*<0.001]	[*p*<0.001]	[*p*<0.001]	[*p*<0.001]	[*p*=0.012]
Rank Diff	1.040	0.647	0.214	0.010	–0.113	–0.254
	(0.164)	(0.178)	(0.196)	(0.211)	(0.245)	(0.272)
	[*p*<0.001]	[*p*<0.001]	[*p*=0.276]	[*p*=0.962]	[*p*=0.644]	[*p*=0.351]
White	–0.131	–0.123	–0.107	–0.092	–0.069	–0.048
	(0.009)	(0.010)	(0.011)	(0.012)	(0.013)	(0.014)
	[*p*<0.001]	[*p*<0.001]	[*p*<0.001]	[*p*<0.001]	[*p*<0.001]	[*p*<0.001]
7.5 Komi	0.037	0.008	–0.004	–0.029	–0.039	–0.043
	(0.018)	(0.019)	(0.020)	(0.022)	(0.026)	(0.032)
	[*p*=0.047]	[*p*=0.678]	[*p*=0.844]	[*p*=0.201]	[*p*=0.130]	[*p*=0.174]
Post × Young	0.203	0.179	0.159	0.148	0.120	0.050
	(0.031)	(0.028)	(0.029)	(0.029)	(0.032)	(0.035)
	[*p*<0.001]	[*p*<0.001]	[*p*<0.001]	[*p*<0.001]	[*p*<0.001]	[*p*=0.153]
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	47,292	47,276	47,120	46,132	42,238	34,122
R^2	0.350	0.266	0.210	0.168	0.141	0.139
Within R^2	0.014	0.010	0.006	0.004	0.003	0.001

Note. This table presents how 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.

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

Sukwoong Choi Hyo Kang Namil Kim Junsik Kim

October 2023

Table of Contents

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]

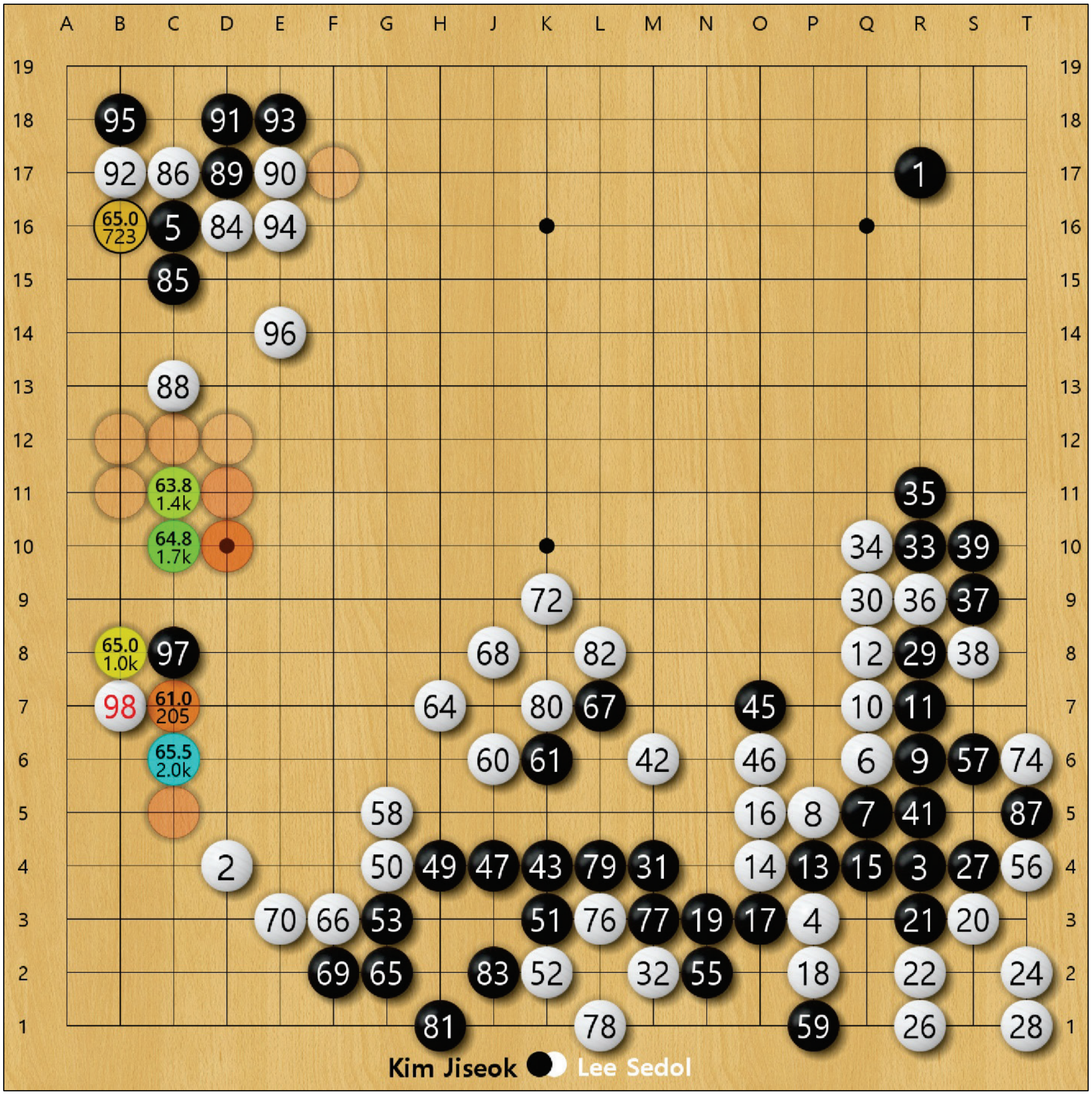
Leela Zero provides an in-depth analysis of the game, including recommendations for next moves. We visualize what Leela Zero provides for Go analysis using the Lizzie graphical user interface (GUI). Figure A.1 shows a recent Go match between the world champion, Lee Sedol, and Jiseok Kim. On the main 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

[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).

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.

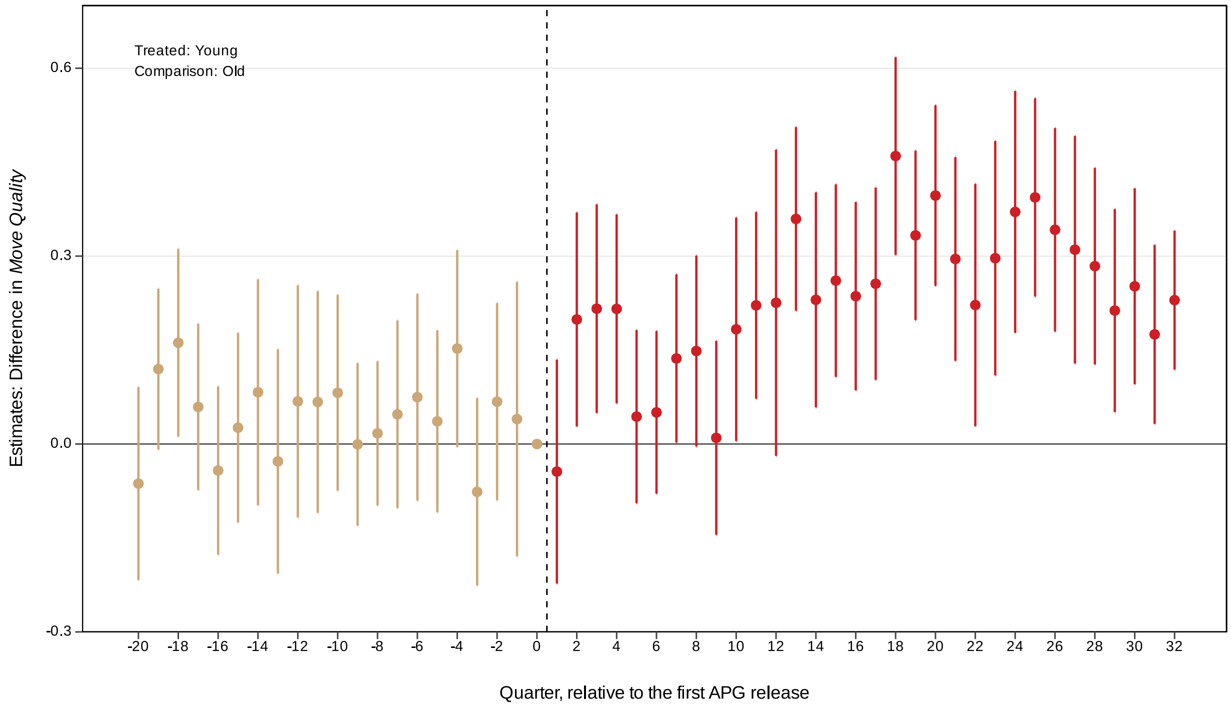
A.2. Implementation Details

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

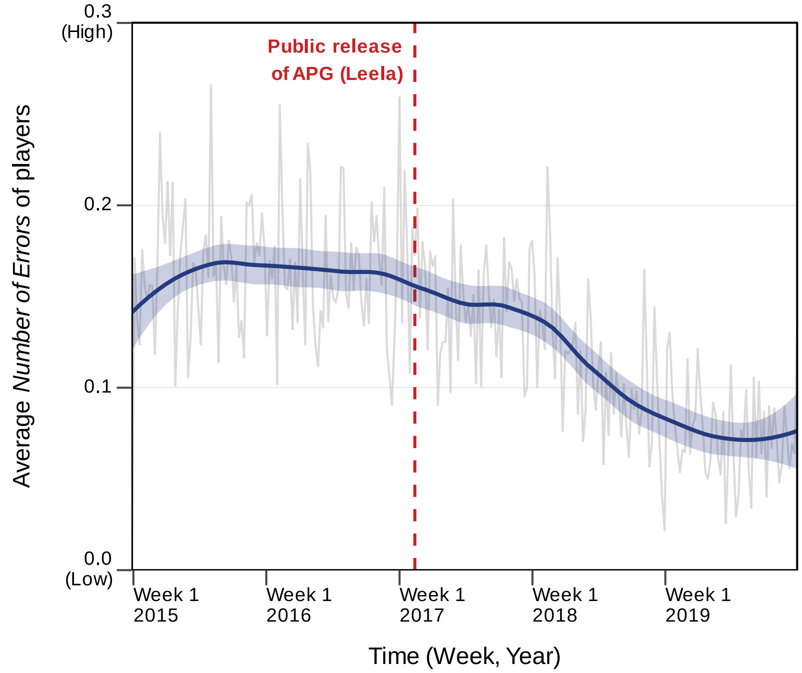
Figure B.1: Robustness check: Differential effects using a month fixed effect



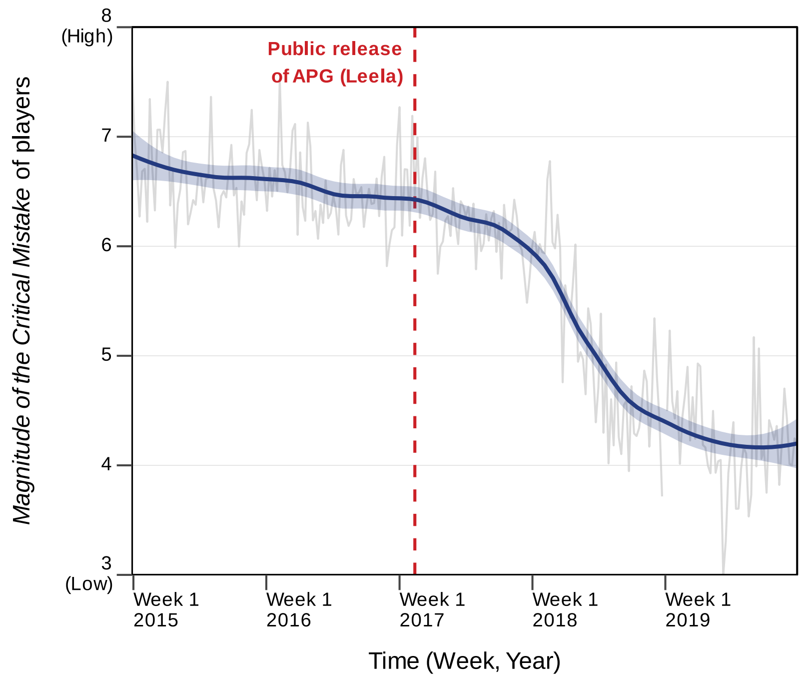
Note. This figure illustrates the differential effects of APG *Move Quality* by player age using a month fixed effect instead of a quarter fixed effect, as shown in Figure 3(b). The points graphically present the *Move Quality* of younger players (those below the median age) compared to that of older players (those above the median age). The vertical error bars show the 95% confidence intervals.

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

(a). Number of errors

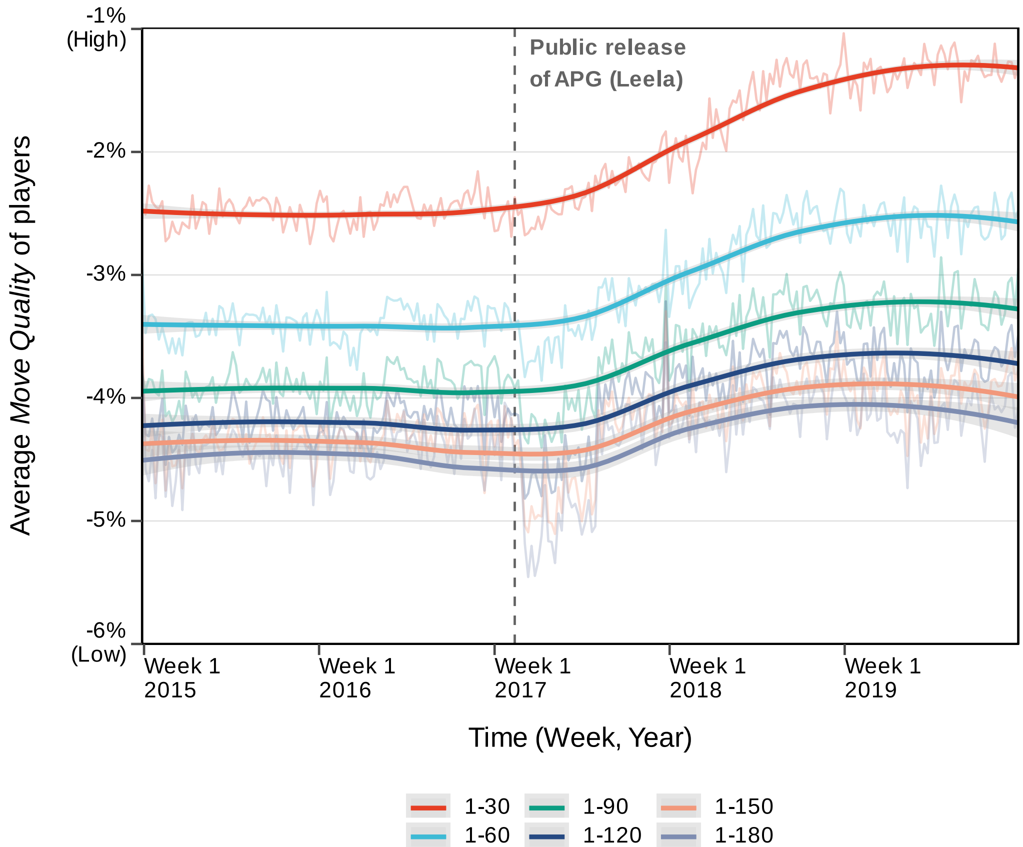


(b). Magnitude of the most critical mistake



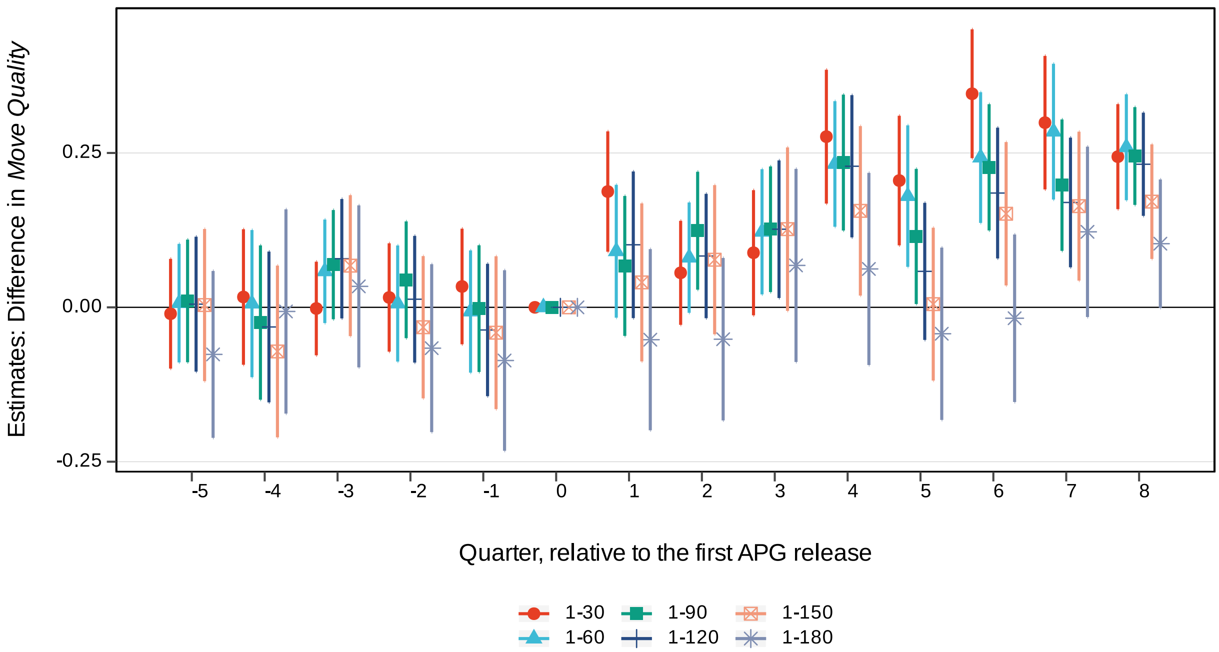
Note. This figure illustrates the weekly average of *Number of Errors* (Panel a) and the *Magnitude of the Most Critical Mistake* (Panel b) from 2015 through 2019. The gray solid lines represent the raw (unprocessed) weekly average value. The blue solid lines and the blue areas around them show locally 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



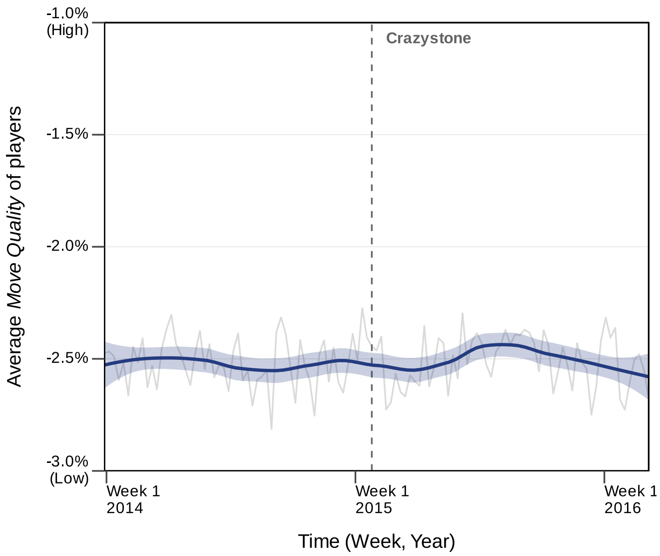
Note. This figure illustrates how the changes in average *Move Quality* differ by the number of moves. Beginning with the opening strategy of the first thirty moves, we incrementally add thirty additional moves (up to 180 moves) and compare the trends; the six colored lines show the raw (unprocessed) 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 we consider. Beginning with the opening strategy of the first thirty moves, we incrementally add thirty additional moves (up to 180 moves). The estimates for this figure are derived from the similar regressions of Table 7 with distributed leads and lags.

Figure B.5: The effect of earlier Go programs: Crazy Stone

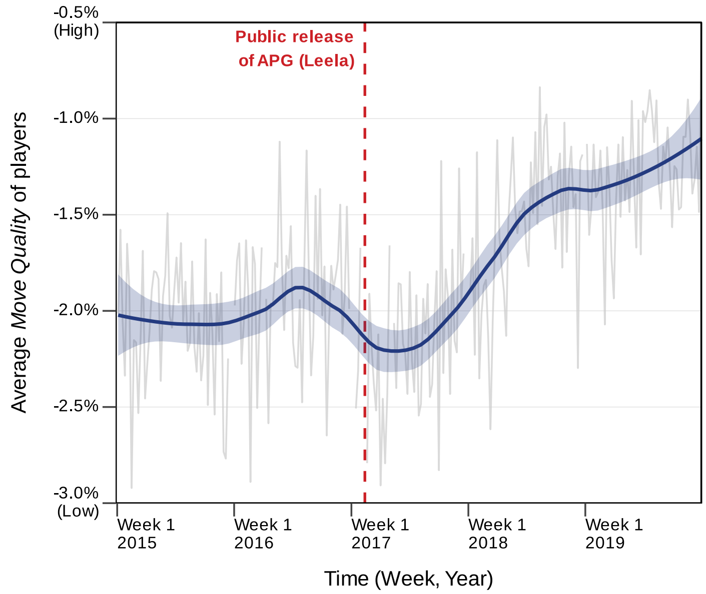


Dependent Variable:	Move quality
Model:	(1)
Variables	
Crazy stone	0.114
	(0.082)
	[*p*=0.162]
Trend	–0.001
	(0.007)
	[*p*=0.923]
Crazy Stone × Trend	–0.011
	(0.013)
	[*p*=0.386]
Fixed effects	
Player	Yes
Opponent Player	Yes
Fit statistics	
Observations	18,686
R^2	0.144
Within R^2	0.000

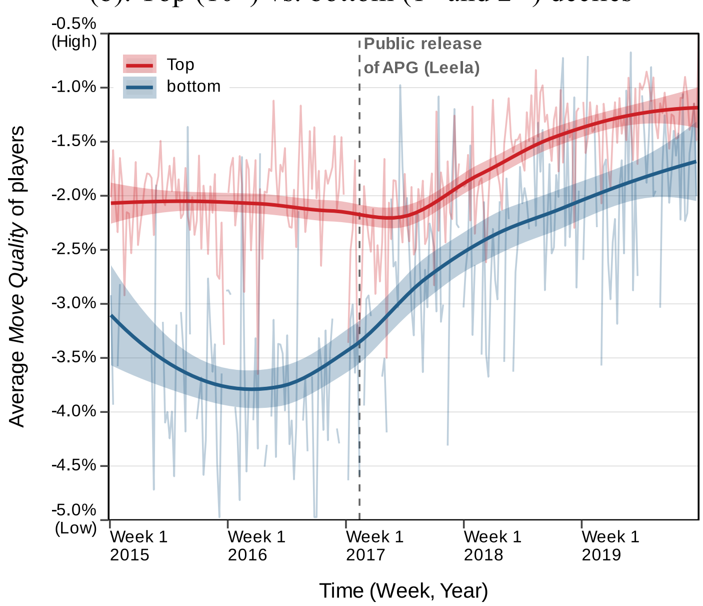
Note. This left figure illustrates the weekly average *Move Quality* before and after the release of Crazy Stone in 2015. For this figure, we used data collected from the period spanning January 1, 2014, to March 8, 2016, one day before the AlphaGo match. The gray solid line represents the raw (unprocessed) weekly average value. The blue solid line and the blue area around it show the smoothed trend (loess; span=0.7) and the 95% confidence interval, respectively. The vertical line at the end of January 2015 represents the release of the earlier Go program called Crazy Stone that year. The right table provides a re-estimation of Model 2 of Table 2 by considering Crazy Stone. *Crazy stone* takes unity for the games played in the quarters after January 2015. Trend refers to the number of quarters that had elapsed since the beginning of 2014. Clustered standard errors at a focal-player level are in parentheses and p-values are in squared brackets.

Figure B.6: Auxiliary to quantile treatment effects analysis: Top vs. bottom deciles

(a). Top (10th) decile

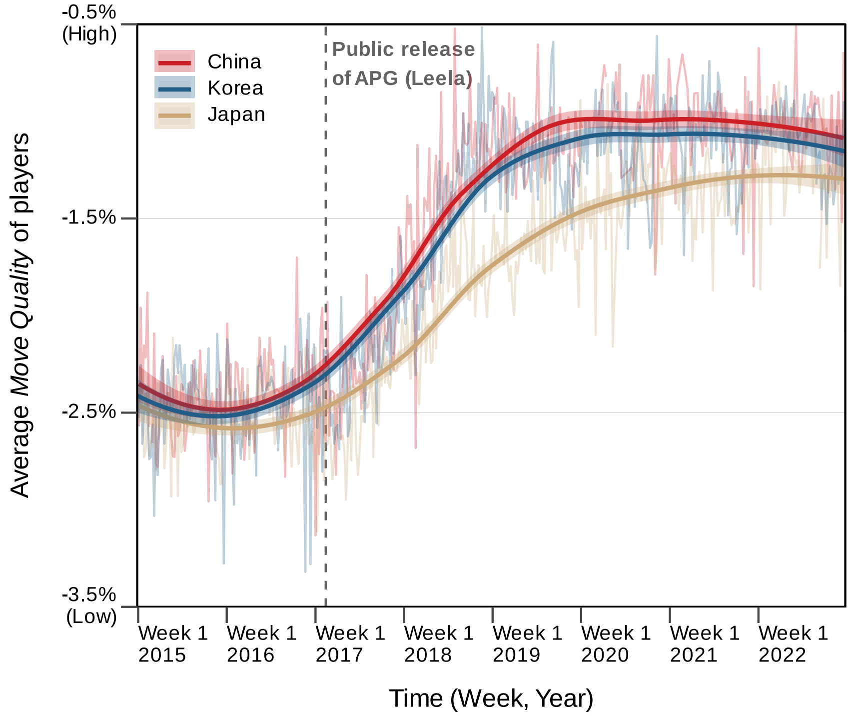


(b). Top (10th) vs. bottom (1st and 2nd) deciles



Note. This figure illustrates the weekly average *Move Quality* for players in the top decile (Panel a) and for both the top and bottom deciles (Panel b) from 2015 through 2019, based on *Move Quality*. For both panels, the gray solid lines represent the raw (unprocessed) weekly average value. For Panel a, blue solid line and blue areas around it show locally smoothed trends and the 95% confidence intervals, respectively. In contrast, the red represents the top decile, and the blue represents the bottom deciles in Panel b. In both figures, the vertical lines on February 2017 represents the first public release of an APG, *Leela*.

Figure B.7: Effects of APG on move quality by players' nationality:
Model-free evidence over an extended time period



Note. This figure extends the time period of Figure 4(a) to illustrate the average *Move Quality* of professional players by their nationality, now including data up to the end of 2022. 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*.

Table B.1: Distributed leads and lags

Dependent Variable:	*Move Quality*			
	Age		Country	
Model:	(1)	(2)	(3)	(4)
Variables				
Rank	1.523	2.374	2.347	3.250
	(0.256)	(0.306)	(0.220)	(0.276)
	[p<0.001]	[p<0.001]	[p<0.001]	[p<0.001]
Rank Diff	0.065	1.001	0.063	1.104
	(0.025)	(0.165)	(0.028)	(0.185)
	[p=0.010]	[p<0.001]	[p=0.025]	[p<0.001]
White	–0.130	–0.131	–0.129	–0.126
	(0.009)	(0.009)	(0.010)	(0.010)
	[p<0.001]	[p<0.001]	[p<0.001]	[p<0.001]
7.5 Komi	0.023	0.039	0.021	0.030
	(0.016)	(0.018)	(0.016)	(0.019)
	[p=0.151]	[p=0.034]	[p=0.197]	[p=0.113]
5 quarter before	–0.009	–0.010	–0.028	–0.006
	(0.045)	(0.045)	(0.052)	(0.052)
	[p=0.838]	[p=0.820]	[p=0.590]	[p=0.908]
4 quarter before	0.036	0.017	–0.081	–0.096
	(0.054)	(0.056)	(0.058)	(0.060)
	[p=0.505]	[p=0.765]	[p=0.167]	[p=0.110]
3 quarter before	0.006	–0.002	–0.044	–0.062
	(0.037)	(0.039)	(0.046)	(0.046)
	[p=0.870]	[p=0.961]	[p=0.340]	[p=0.183]
2 quarter before	0.020	0.016	–0.016	–0.023
	(0.044)	(0.045)	(0.052)	(0.052)
	[p=0.656]	[p=0.721]	[p=0.762]	[p=0.656]
1 quarter after	0.040	0.034	0.037	0.031
	(0.047)	(0.048)	(0.054)	(0.054)
	[p=0.394]	[p=0.480]	[p=0.491]	[p=0.560]
1 quarter after	0.200	0.188	0.154	0.127
	(0.049)	(0.050)	(0.054)	(0.055)
	[p<0.001]	[p<0.001]	[p=0.005]	[p=0.021]
2 quarter after	0.074	0.056	0.142	0.102
	(0.044)	(0.043)	(0.051)	(0.052)
	[p=0.087]	[p=0.194]	[p=0.006]	[p=0.049]
3 quarter after	0.091	0.088	0.026	–0.001
	(0.051)	(0.052)	(0.056)	(0.057)
	[p=0.072]	[p=0.088]	[p=0.634]	[p=0.985]
4 quarter after	0.288	0.276	0.283	0.251
	(0.055)	(0.055)	(0.057)	(0.057)
	[p<0.001]	[p<0.001]	[p<0.001]	[p<0.001]
5 quarter after	0.239	0.205	0.357	0.316
	(0.054)	(0.054)	(0.054)	(0.053)
	[p<0.001]	[p<0.001]	[p<0.001]	[p<0.001]
6 quarter after	0.367	0.346	0.342	0.304
	(0.053)	(0.053)	(0.064)	(0.064)
	[p<0.001]	[p<0.001]	[p<0.001]	[p<0.001]
7 quarter after	0.328	0.299	0.374	0.321
	(0.054)	(0.055)	(0.053)	(0.055)
	[p<0.001]	[p<0.001]	[p<0.001]	[p<0.001]
8+ quarter after	0.272	0.244	0.270	0.230
	(0.043)	(0.043)	(0.041)	(0.043)
	[p<0.001]	[p<0.001]	[p<0.001]	[p<0.001]
Fixed effects				
Quarter	Yes	Yes	Yes	Yes
Player	Yes	Yes	Yes	Yes
Opponent Player		Yes		Yes
Fit statistics				
Observations	47,292	47,292	43,151	43,151
R^2	0.326	0.351	0.328	0.353
Within R^2	0.015	0.015	0.016	0.016

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	0.128	–0.060		
	(0.023)	(0.023)		
	[p<0.001]	[p=0.010]		
Rank	0.816	0.782	1.705	2.481
	(0.037)	(0.036)	(0.231)	(0.281)
	[p<0.001]	[p<0.001]	[p<0.001]	[p<0.001]
Rank Diff	0.116	0.104	0.065	0.919
	(0.028)	(0.028)	(0.025)	(0.165)
	[p<0.001]	[p<0.001]	[p=0.010]	[p<0.001]
White	–0.133	–0.133	–0.130	–0.131
	(0.010)	(0.010)	(0.009)	(0.009)
	[p<0.001]	[p<0.001]	[p<0.001]	[p<0.001]
7.5 Komi	0.022	0.022	0.023	0.038
	(0.016)	(0.016)	(0.016)	(0.018)
	[p=0.157]	[p=0.160]	[p=0.147]	[p=0.036]
Post × Young		0.350	0.295	0.268
		(0.029)	(0.031)	(0.032)
		[p<0.001]	[p<0.001]	[p<0.001]
Fixed effects				
Quarter	Yes	Yes	Yes	Yes
Player			Yes	Yes
Opponent Player				Yes
Fit statistics				
Observations	47,292	47,292	47,292	47,292
R^2	0.278	0.282	0.326	0.350
Within R^2	0.066	0.072	0.015	0.015

Note. This table presents re-estimated results from Table 3, which shows the differential effects of APG by age, 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				
Middle	0.108			
	(0.027)			
	[p<0.001]			
Young	0.158			
	(0.029)			
	[p<0.001]			
Rank	0.813	0.772	1.483	2.248
	(0.038)	(0.036)	(0.250)	(0.301)
	[p<0.001]	[p<0.001]	[p<0.001]	[p<0.001]
Rank Diff	0.119	0.104	0.063	0.904
	(0.028)	(0.028)	(0.025)	(0.164)
	[p<0.001]	[p<0.001]	[p=0.012]	[p<0.001]
White	–0.132	–0.133	–0.130	–0.131
	(0.010)	(0.010)	(0.009)	(0.009)
	[p<0.001]	[p<0.001]	[p<0.001]	[p<0.001]
7.5 Komi	0.019	0.015	0.023	0.039
	(0.016)	(0.016)	(0.016)	(0.018)
	[p=0.228]	[p=0.348]	[p=0.144]	[p=0.035]
Post × Middle		0.254	0.276	0.248
		(0.035)	(0.036)	(0.036)
		[p<0.001]	[p<0.001]	[p<0.001]
Post × Young		0.340	0.367	0.338
		(0.034)	(0.040)	(0.041)
		[p<0.001]	[p<0.001]	[p<0.001]
Fixed effects				
Quarter	Yes	Yes	Yes	Yes
Player			Yes	Yes
Opponent Player				Yes
Fit statistics				
Observations	47,292	47,292	47,292	47,292
R^2	0.278	0.282	0.326	0.350
Within R^2	0.066	0.072	0.015	0.015

Note. This table re-estimates Table 3, illustrating the differential effects of APG by age, using three age groups instead of the previous two-group classification: 'Young' (bottom tertile), 'Middle' (middle tertile), and 'Old' (top tertile). 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	0.008	0.024		
	(0.018)	(0.020)		
	[*p*=0.633]	[*p*=0.236]		
Rank	0.884	0.879	2.491	3.393
	(0.036)	(0.036)	(0.232)	(0.272)
	[*p*<0.001]	[*p*<0.001]	[*p*<0.001]	[*p*<0.001]
Rank Diff	0.130	0.133	0.074	1.182
	(0.028)	(0.028)	(0.025)	(0.164)
	[*p*<0.001]	[*p*<0.001]	[*p*=0.004]	[*p*<0.001]
White	–0.132	–0.133	–0.131	–0.131
	(0.010)	(0.010)	(0.009)	(0.009)
	[*p*<0.001]	[*p*<0.001]	[*p*<0.001]	[*p*<0.001]
7.5 Komi	0.043	0.042	0.021	0.038
	(0.016)	(0.016)	(0.016)	(0.018)
	[*p*=0.008]	[*p*=0.010]	[*p*=0.184]	[*p*=0.039]
Post × Young		0.022	0.006	0.029
		(0.029)	(0.028)	(0.027)
		[*p*=0.437]	[*p*=0.836]	[*p*=0.271]
Fixed effects				
Quarter	Yes	Yes	Yes	Yes
Player			Yes	Yes
Opponent Player				Yes
Fit statistics				
Observations	47,292	47,292	47,292	47,292
R^2	0.275	0.276	0.323	0.348
Within R^2	0.063	0.064	0.011	0.012

Note. This table presents the re-estimated results from Table 3, displaying the regression estimates of the heterogeneous effects of APG on player age. The re-estimation occurs after players have been 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	0.097	–0.049		
	(0.020)	(0.021)		
	[p<0.001]	[p=0.020]		
Rank	0.847	0.831	1.781	2.626
	(0.037)	(0.037)	(0.248)	(0.295)
	[p<0.001]	[p<0.001]	[p<0.001]	[p<0.001]
Rank Diff	0.129	0.122	0.072	1.039
	(0.028)	(0.028)	(0.025)	(0.164)
	[p<0.001]	[p<0.001]	[p=0.004]	[p<0.001]
White	–0.133	–0.133	–0.130	–0.130
	(0.010)	(0.010)	(0.009)	(0.009)
	[p<0.001]	[p<0.001]	[p<0.001]	[p<0.001]
7.5 Komi	0.017	0.015	0.010	0.023
	(0.016)	(0.016)	(0.016)	(0.018)
	[p=0.291]	[p=0.357]	[p=0.536]	[p=0.217]
Post × Young		0.255	0.201	0.186
		(0.028)	(0.031)	(0.031)
		[p<0.001]	[p<0.001]	[p<0.001]
Fixed effects				
Quarter	Yes	Yes	Yes	Yes
Player			Yes	Yes
Opponent Player				Yes
Fit statistics				
Observations	47,292	47,292	47,292	47,292
R^2	0.280	0.283	0.327	0.352
Within R^2	0.065	0.069	0.013	0.013

Note. This table re-estimates Table 3, which shows the regression estimates on the heterogeneous effects of APG by player age, 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 15 moves (1-15)		First 20 moves (1-20)		First 40 moves (1-40)		First 50 moves (1-50)		First 60 moves (1-60)	
Model:	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Variables										
Rank	1.522	2.125	1.652	2.427	2.037	2.949	1.869	2.683	1.709	2.323
	(0.243)	(0.280)	(0.251)	(0.286)	(0.233)	(0.281)	(0.247)	(0.300)	(0.241)	(0.289)
	[p<0.001]	[p<0.001]	[p<0.001]	[p<0.001]	[p<0.001]	[p<0.001]	[p<0.001]	[p<0.001]	[p<0.001]	[p<0.001]
Rank Diff	0.090	0.793	0.075	0.953	0.048	1.081	0.036	0.933	0.009	0.667
	(0.022)	(0.152)	(0.023)	(0.162)	(0.028)	(0.174)	(0.029)	(0.179)	(0.030)	(0.178)
	[p<0.001]	[p<0.001]	[p=0.001]	[p<0.001]	[p=0.086]	[p<0.001]	[p=0.210]	[p<0.001]	[p=0.753]	[p<0.001]
White	–0.053	–0.052	–0.130	–0.129	–0.128	–0.128	–0.124	–0.125	–0.122	–0.123
	(0.008)	(0.008)	(0.009)	(0.009)	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)
	[p<0.001]	[p<0.001]	[p<0.001]	[p<0.001]	[p<0.001]	[p<0.001]	[p<0.001]	[p<0.001]	[p<0.001]	[p<0.001]
7.5 Komi	–0.005	0.005	0.005	0.019	0.023	0.034	0.016	0.020	0.003	0.007
	(0.013)	(0.016)	(0.014)	(0.017)	(0.016)	(0.019)	(0.015)	(0.019)	(0.015)	(0.019)
	[p=0.695]	[p=0.767]	[p=0.730]	[p=0.260]	[p=0.137]	[p=0.072]	[p=0.307]	[p=0.280]	[p=0.841]	[p=0.695]
Post × Young	0.091	0.072	0.146	0.128	0.178	0.165	0.179	0.168	0.164	0.152
	(0.029)	(0.029)	(0.031)	(0.030)	(0.030)	(0.029)	(0.029)	(0.029)	(0.028)	(0.028)
	[p=0.002]	[p=0.013]	[p<0.001]	[p<0.001]	[p<0.001]	[p<0.001]	[p<0.001]	[p<0.001]	[p<0.001]	[p<0.001]
Fixed effects										
Quarter	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Player	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Opponent Player		Yes		Yes		Yes		Yes		Yes
Fit statistics										
Observations	47,292	47,292	47,292	47,292	47,290	47,290	47,286	47,286	47,276	47,276
R^2	0.355	0.380	0.350	0.375	0.294	0.319	0.267	0.293	0.241	0.266
Within R^2	0.007	0.006	0.013	0.013	0.012	0.012	0.010	0.011	0.009	0.009

Note. This table provides a re-estimation of Models 3 and 4 from Table 3, which shows the effects of APGs on Move Quality by the player's age. The re-estimation is conducted with varying definitions of early opening moves, segmented into five categories: the first 15 moves (corresponding to Models 1–2), the first 20 moves (Models 3–4), the first 40 moves (Models 5–6), the first 50 moves (Models 7–8), and the first 60 moves (Models 9–10). Clustered standard errors at a focal-player level in are 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	–0.014	–0.038		
	(0.019)	(0.022)		
	[p=0.447]	[p=0.091]		
Rank	0.895	0.895	2.705	3.731
	(0.038)	(0.038)	(0.237)	(0.278)
	[p<0.001]	[p<0.001]	[p<0.001]	[p<0.001]
Rank Diff	0.144	0.144	0.066	1.331
	(0.032)	(0.032)	(0.028)	(0.180)
	[p<0.001]	[p<0.001]	[p=0.019]	[p<0.001]
White	–0.130	–0.130	–0.129	–0.127
	(0.010)	(0.010)	(0.010)	(0.010)
	[p<0.001]	[p<0.001]	[p<0.001]	[p<0.001]
7.5 Komi	0.049	0.050	0.020	0.027
	(0.017)	(0.017)	(0.017)	(0.019)
	[p=0.004]	[p=0.004]	[p=0.239]	[p=0.155]
Post × Treated		0.041	0.035	0.037
		(0.031)	(0.029)	(0.028)
		[p=0.191]	[p=0.231]	[p=0.187]
Fixed effects				
Quarter	Yes	Yes	Yes	Yes
Player			Yes	Yes
Opponent Player				Yes
Fit statistics				
Observations	43,151	43,151	43,151	43,151
R^2	0.277	0.277	0.324	0.350
Within R^2	0.061	0.061	0.011	0.012

Note. This regression shows the re-estimated results of Models 1 to 4 as reported in Table 4, which shows the effects of APGs on *Move Quality* by the player's nationality. In contrast to Table 4, this table includes a random reassignment of players' nationalities instead of their true 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

Dependent Variable:		*Age*			*Country*	
Model:	(1)	(2)	(3)	(4)	(5)	(6)
Variables						
Rank	0.354	0.318	0.218	0.465	0.420	0.300
	(0.048)	(0.046)	(0.037)	(0.048)	(0.042)	(0.036)
	[p<0.001]	[p<0.001]	[p<0.001]	[p<0.001]	[p<0.001]	[p<0.001]
Rank Diff	0.096	0.075	0.049	0.102	0.082	0.060
	(0.027)	(0.024)	(0.022)	(0.030)	(0.026)	(0.024)
	[p<0.001]	[p=0.002]	[p=0.025]	[p<0.001]	[p=0.002]	[p=0.012]
White	0.045	0.091	0.092	0.046	0.092	0.092
	(0.002)	(0.002)	(0.001)	(0.002)	(0.002)	(0.001)
	[p<0.001]	[p<0.001]	[p<0.001]	[p<0.001]	[p<0.001]	[p<0.001]
7.5 Komi	0.001	0.001	0.000	0.001	0.002	0.000
	(0.003)	(0.003)	(0.002)	(0.004)	(0.003)	(0.002)
	[p=0.680]	[p=0.679]	[p=0.940]	[p=0.714]	[p=0.570]	[p=0.884]
Post × Young	0.031	0.025	0.018			
	(0.005)	(0.005)	(0.004)			
	[p<0.001]	[p<0.001]	[p<0.001]			
Post × Treated				0.037	0.030	0.022
				(0.005)	(0.005)	(0.004)
				[p<0.001]	[p<0.001]	[p<0.001]
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	47,292	47,292	47,292	43,151	43,151	43,151
R^2	0.294	0.388	0.407	0.297	0.391	0.410
Within R^2	0.031	0.120	0.157	0.033	0.123	0.160

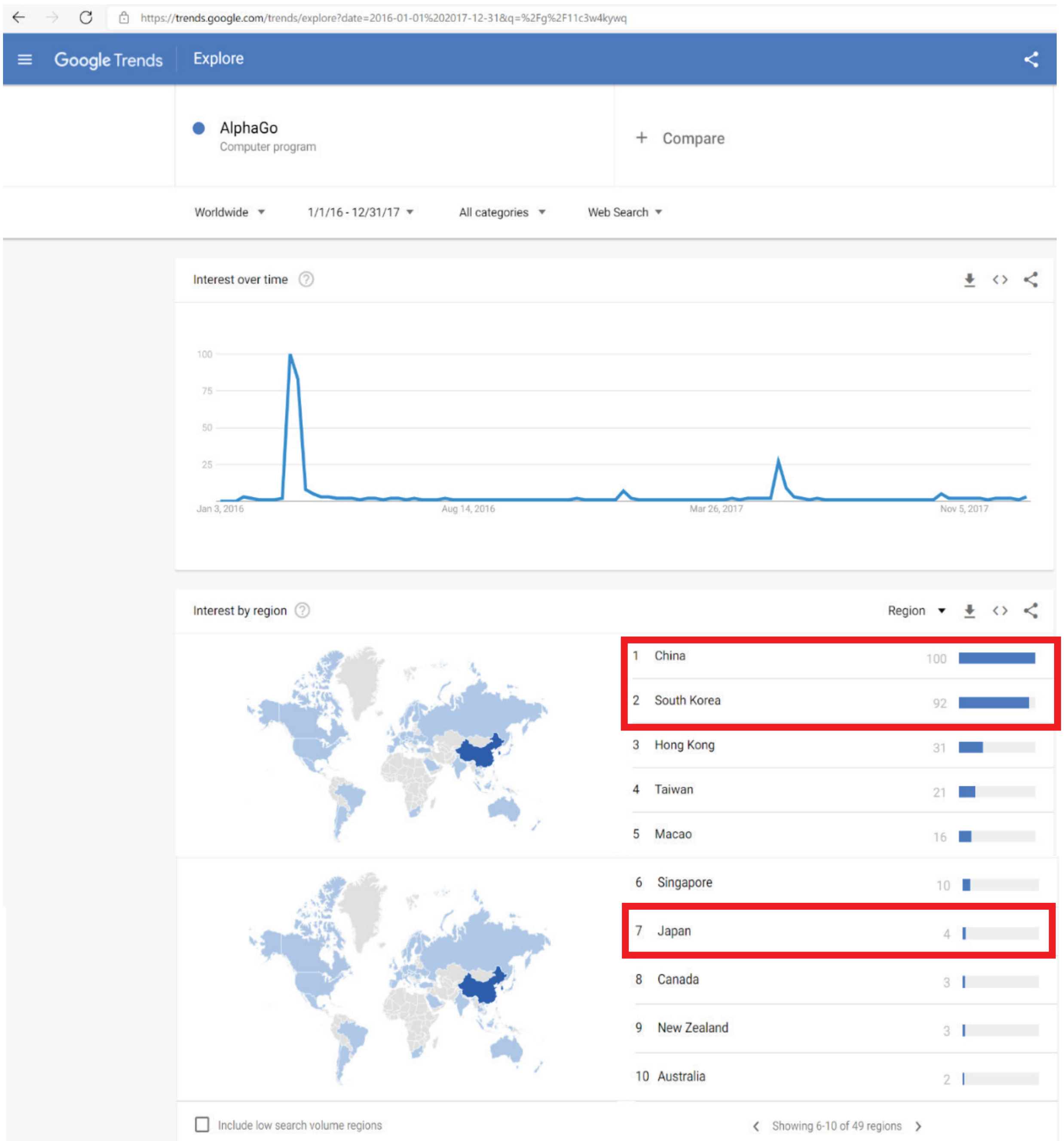
Note. This table shows the regression estimates for the learning from APG by comparing the match between players' moves and APG's top 1, 3, and 5 suggestions. Models 1–3 test the age effects (equivalent to Models 2–4 of Table 3), while Models 4–6 then show the country effects (equivalent to Models 2–4 of Table 4). 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 (May 2017) and South Korea (March 2016) 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.2 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%202017-12-31&q=%2Fg%2F11c3w4kywq. Accessed November 16, 2021.

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



D. Time Reallocation: Time Spent by the Stages of the Game

Table D.1: Proportion of time spent by the stages of the game

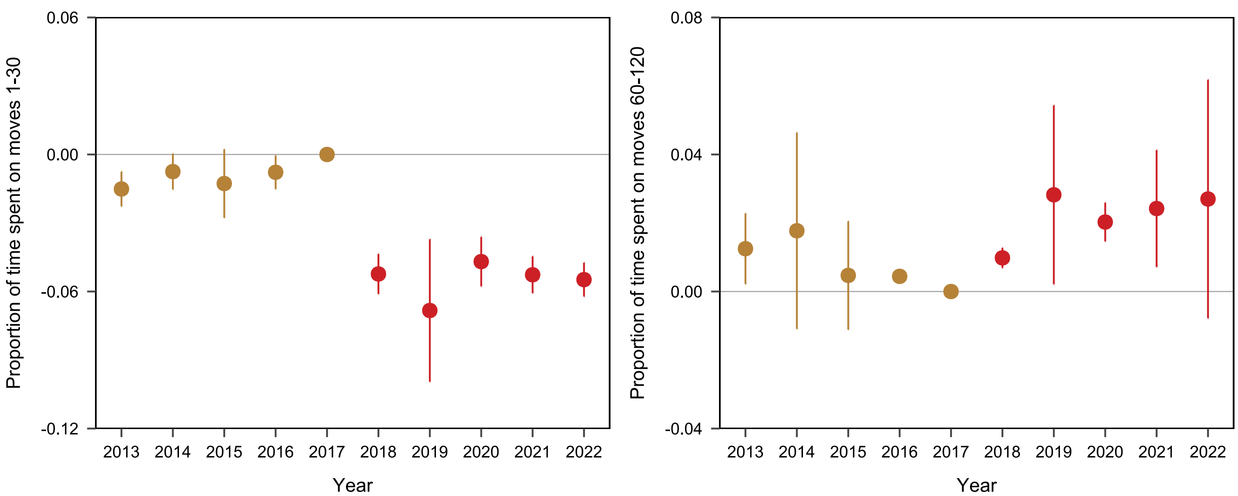
	Dependent variable: Proportion of time spent							
	Stage of the game (moves)							
	[1,30]	(30,60]	(60,90]	(90,120]	(120,150]	(150,180]	(180,210]	210+
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Post	−0.046	0.016	0.020	0.008	0.001	−0.001	−0.002	−0.001
	p = 0.000	p = 0.00000	p = 0.000	p = 0.00001	p = 0.631	p = 0.333	p = 0.244	p = 0.714
Constant	0.171	0.204	0.178	0.143	0.118	0.097	0.077	0.092
	p = 0.000	p = 0.000	p = 0.000	p = 0.000	p = 0.000	p = 0.000	p = 0.000	p = 0.000
Observations	2,626	2,626	2,626	2,626	2,623	2,390	1,902	1,277
R^2	0.106	0.011	0.026	0.008	0.0001	0.0004	0.001	0.0001
Adjusted R^2	0.105	0.010	0.026	0.007	−0.0003	−0.00003	0.0002	−0.001

Note. This table shows the regression estimates on the effects of APG on the proportion of time spent, before and after the first public release of an APG, Leela. Models 1 through 8 sequentially increase the range of moves considered by 30 moves; for instance, Model 1 presents moves 1 to 30, while Model 2 presents results from moves 31 to 60. *Post* takes unity for the games played after 2017. Standard errors are in parentheses, and p-values are in squared brackets.

Figure D.1: Proportion of time spent by the stages of the game: Event study estimates

(a). Moves 1–30

(b). Moves 60–120



Note. This figure illustrates the annual changes in the proportion of time spent in different states of the game, compared to the reference year 2017 when an APG became publicly available. Panel (a) focuses on the first 30 moves (moves 1-30), while panel (b) examines moves 61–120. The vertical error bars show the 90% confidence intervals.

E. Examples of learning-from-AI

In this section, we discuss several examples where human professionals learn from AI but perform the tasks by themselves.

1 | ***Education***. Educational systems are increasingly employing AI tools to provide customized learning experiences and improve the overall quality of education. Students are guided and tutored by AI-powered educational programs, leading to enhanced learning and skill development (Qadir, 2023). However, during examinations or assessments meant to evaluate their learning, AI tools aren't usually permitted (Mearian, 2023). That is, ChatGPT can teach students how to write a good essay but is not permitted in exams.

2 | ***Training Fighter Pilots***. There is an emerging trend of AI-embedded pilot training, evident in both academic research and real-world practice (Guevarra et al., 2023; Halpern, 2022). This is different from the traditional flight *simulation* and mirrors our context of APG. In AI-embedded pilot training, human pilots learn by engaging with AI counterparts (enemies) that had been proven to outperform human pilots. Learning occurs as human pilots repeatedly engage in a combat with AI pilots, gaining insights on winning strategies in dogfights. It is worth noting that AI-operated flight is not yet available in the real world, and AI-pilots would not completely replace human pilots in the near future. Therefore, learning from AI-embedded pilots provides vital opportunities for human pilots to learn and prepare the real combat.

3 | ***Corporate Training***. In the corporate world, managers can benefit from AI-powered training programs. Such programs can help managers refine their decision-making abilities, leadership skills, and strategic thinking. But when it comes to applying these skills in real-world scenarios, such as crisis management, field operations in locations with limited connectivity, unexpected yet time-pressing decision-making, or leading team meetings outside their workplaces, AI tools may not be readily available. It thus is important to nurture the intrinsic human capability to perform tasks even if AI tools could perform them better or faster.

Appendix References

Geis, J. R., Brady, A. P., Wu, C. C., Spencer, J., Ranschaert, E., Jaremko, J. L., ... & Kohli, M. (2019). Ethics of artificial intelligence in radiology: summary of the joint European and North American multisociety statement. *Radiology*, *293*(2), 436-440.

Guevarra, M., Das, S., Wayllace, C., Epp, C. D., Taylor, M., & Tay, A. (2023). Augmenting Flight Training with AI to Efficiently Train Pilots. In *Proceedings of the AAAI Conference on Artificial Intelligence* (Vol. 37, No. 13, pp. 16437-16439).

Halpern, S. (2022, January 17). The Rise of A.I. Fighter Pilots. *The New Yorker.* https://www.newyorker.com/magazine/2022/01/24/the-rise-of-ai-fighter-pilots

Mearian, L. (2023, April 25). Schools look to ban ChatGPT, students use it anyway. *Computerworld.* https://www.computerworld.com/article/3694195/schools-look-to-ban-chatgpt-students-use-it-anyway.html

Qadir, J. (2023). Engineering education in the era of ChatGPT: Promise and pitfalls of generative AI for education. In *2023 IEEE Global Engineering Education Conference (EDUCON)* (pp. 1-9). IEEE.

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

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