Internet Matching Technologies and the Geographic Distribution of Chess Skills[1]

Sara Fisher Ellison, *MIT and CESifo*; and Maren Young, *UC Berkeley*

December, 2022

[1]We thank Molly Dalzell, Will de Rubertis, J. William Hegelmeyer, Jessica Li, Eileen Liu, Fiona Paine, James Simon, and Wentao Zhou, for their valuable data gathering efforts and other excellent research assistance. José de Sousa, Glenn Ellison, and Pai-ling Yin, as well as a number of conference participants, provided useful suggestions. Sara thanks Eric Meyer and Chris Avery for a dinnertime conversation about how their sons' experiences learning chess had differed from theirs a generation ago, which inspired this project. Finally, Sara thanks the Paris School of Economics for their hospitality while much of this research was completed. E-mail: sellison@mit.edu and mareny@berkeley.edu.

Abstract

There is broad consensus that the internet has the potential to become an important tool for learning. There is less consensus on its likely distributional effects. On the one hand, it could provide previously isolated regions, countries lacking high-quality educational systems, or under-resourced individuals and communities access to skills and knowledge. On the other hand, its offerings might be relatively more accessible to well-resourced communities for many reasons, widening gaps. The COVID pandemic has put these questions into sharp relief, with millions worldwide being moved to online or distance-learning. This paper leverages a historic example of distance-acquisition of one particular skill, chess-playing. We look for evidence of either this democratizing effect or a widening divide on chess skills from internet penetration. We also look broadly at how the geographic distribution of chess skills has changed since the introduction of the internet. Finally, we speculate on what, if anything, can be said about the relevant mechanism given the particular characteristics of how one typically learns to play chess.

Keywords: Internet, skills acquisition, distance learning, economic geography, digitization

JEL Classification Codes: L86, D83, O15, R12, J24, N30

1 Introduction

There is a broadly-held and hopeful belief that the internet holds promise as an important tool for learning, providing previously isolated regions, countries lacking high-quality educational systems, or under-resourced individuals and communities access to skills and knowledge. Still, evidence of this democratization of knowledge and skills has been elusive. Lack of hard evidence to date suggests an alternative, and less optimistic, narrative: rich and well-resourced areas are the ones poised to benefit from the internet, both because they can afford the infrastructure and because they have the baseline skills on which to build. (A third, and even more discouraging, possibility is that the internet is not a valuable tool for learning at all but rather just a very expensive gaming platform.)

These questions became immediately resonant to millions of students, teachers, and parents worldwide when the COVID-19 pandemic precipitated an immediate shift from in-person to distance learning in much of the developed world. While students have, largely, moved back to in-person classrooms, we are left with an important set of questions about both the efficacy of this COVID-era distance learning and whether its effects were uniform or unevenly distributed. We have no doubt that this unfortunate natural experiment will serve as useful fodder for research for years. Our focus is both broader than this question—we are interested in potential effects of learning and acquiring skills online over the entire existence of the internet—and more focused—we seek to explore these issues in the context of one particular and easy-to-measure skill, chess playing.

We perform a number of analyses using comprehensive data on the world's top chess players going back five decades to try to address these questions. First, we estimate descriptive regressions at the individual level to undercover potential important factors in determining high-level chess ability. Second, we perform a different regression analysis, at the country level, to help explain the concentration of high-level chess talent in particular countries and years. In this analysis, we use a number of demographic variables as reported by the World Bank, including internet penetration. Finally, we would like to document the extent to which the geographic concentration of chess skill has changed over the last 50 years in a single index. Creating such an index is not entirely straightforward.

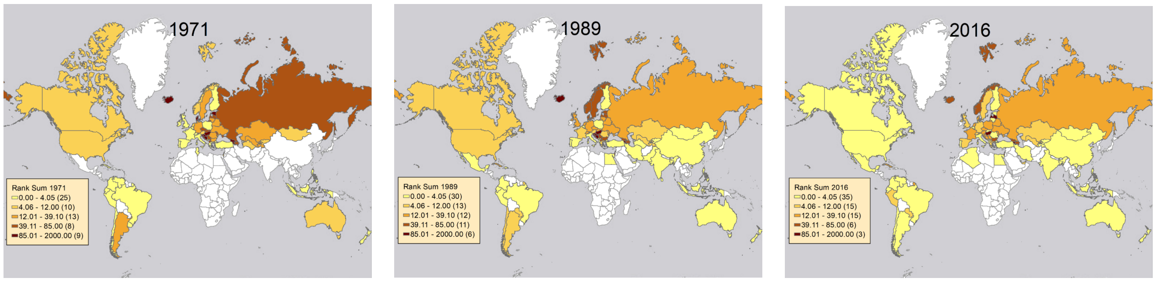


Figure 1: Heat Maps of Worldwide Distribution of Chess Skill from 1971-2016

Many of the complications, however, can be addressed by adapting the Ellison and Glaeser (1997) index of industrial geographic concentration, which controls for phenomena such as randomly-occurring clustering and predicted concentration based on population and other demographic factors. Using the fact that we can choose which demographic factors to condition on, we can compute different versions of the index, one conditioning on internet penetration and one not, for instance, to identify the role that one particular factor such as internet penetration has played in the evolving geographic concentration of chess skill. We find evidence that higher levels of internet penetration are associated with higher levels of chess-playing skill, and that the internet appears to have differential effects depending on the characteristics of the player. We also find evidence that the internet has had a significant effect on the geographic concentration of chess-playing skill, resulting in less concentration over time as internet penetration rates have increased.

Figure 1, a series of heat maps of worldwide distribution of chess skill since 1971, suggests some "flattening" of the distribution, but effects are subtle. Concentrations of skill in some powerhouse countries, such as USSR/Russia and USA, have declined somewhat, whereas a number of countries with negligible concentrations in 1971 have moved up the ranks. We can, though, perform a more systemic calculation. Figure 2 shows the Ellison-Glaeser Index (EGI) of geographic concentration of chess-playing skill and how it has evolved over the past fifty years. This index measures how far the distribution of chess skills is from the counterfactual distribution if it were randomly assigned over the current geographical distribution of people on earth. In other words, the index controls for population distribution linearly but for no other demographics or covariates that might affect the distribution of chess-playing skill.

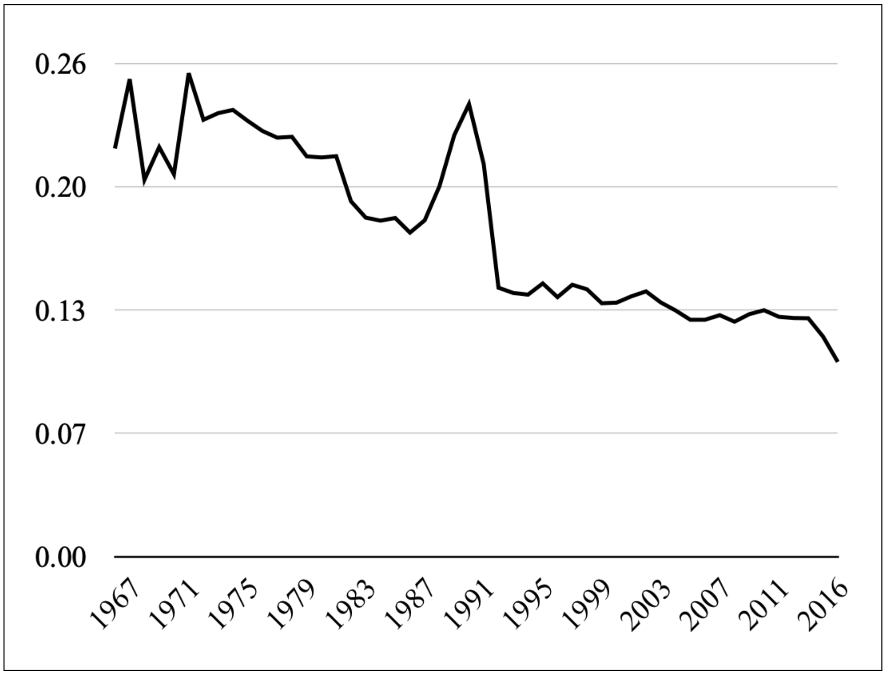


Figure 2: Index of Geographic Concentration of Chess Skill

The interpretation of the scale of the graph is the excess probability of the co-location of pairs of chess players above the counterfactual probability. So, in other words, a value of 0.1 means that players are 0.1 more likely to locate in the same country than they would be at random, a substantial magnitude.

Perhaps the most salient feature of the graph is a steady decline over the past fifty years, albeit not a monotonic one. We know that these changes are not driven by underlying (linear) changes in population distribution, since we have controlled for that, but we want to answer the question as to whether the diffusion of internet technologies has driven some or all of these changes. This question is central to the exercise here.

We will discuss the calculation of this index as well as compare it with another in the results section of the paper. For now, simply note the decline.[1]

[1] Another salient feature is the large spike in concentration—a jump up and subsequent drop—around 1990. One might worry that the index is exhibiting an artifact of an important redefinition of countries that occurred around this time, namely the break-up of the Soviet Union, Yugoslavia, and Czechoslovakia. As we explain later, the index does accommodate these changes in geographic boundaries. Furthermore, there was, in fact, a sharp increase and then decrease in the fraction of top players from the Soviet Union, the world's preeminent producer of chess skill, around that time. In any case, internet penetration was at very low levels in all countries at that time, so our analysis of the effect of the internet will focus mostly on the time period after the spike.

Chess has an unusual feature that makes it particularly interesting to study: it is a skill largely learned in competition, either formal or informal, with an opponent. In other words, being matched repeatedly to opponents who are approximately at the same level as you is a crucial element to learning and progressing in chess. This feature stands in contrast to characteristics of many other skills, such as solving mathematical problems. There, skill acquisition comes through watching lectures and demonstrations as well as individual study, and has little or nothing to do with being matched to someone of a similar level. So chess provides a valuable setting in which to study the effects, not just of the internet more generally, or of the broadcasting technologies inherent in the internet, but specifically of the internet's matching technologies. Put another way, one can think of at least two internet mechanisms relevant for learning: 1) the internet can serve as a conduit for written knowledge, essentially an enormous text available on electronic platforms or a forum for lectures with arbitrarily large audiences, or 2) it can serve as a matchmaker, where person-to-person interaction is still important but can be facilitated with internet matching technologies. The relative importance of these two mechanisms could have implications on the future transmission and distribution of skills and knowledge. The particular skill we study here, chess-playing, is one where the matching function should be quite important.

As ideas of social learning and cultural evolution are gaining increasing traction across many fields in social science, understanding the potential role of the internet in these mechanisms is central. In *The Secret of Our Success,* Joseph Henrich (2016) directly addresses the immense potential of the internet to "dramatically expand" our collective brains and accelerate cultural evolution, but adds the following caveats.[2] First, he notes that language differences may slow the process. Second, he cautions that prosocial norms for information sharing over the internet might be difficult to sustain. Chess could be an interesting exception to both of those caveats. To the extent that learning is happening in game play, language barriers should be minimal. Furthermore, both parties would have similar incentives to participate, reducing reliance on prosocial norms. Note that we will not address Heinrich's broader thesis about the increase in knowledge and skill over time since we will focus on a potential side-effect: the change in the global distribution of knowledge and skill

[2] See page 326.

over time.

There are many strands of economics literature that are woven into this study. We hope it can contribute, in a small way at least, to fields such as urban economics and digitization and, within those, topics as diverse as distance learning, agglomeration economies, the digital divide, the economics of chess, and geography and the internet. Below, we try to flesh out some of these connections and relationships with the literature.

Urban economics has long been interested in agglomeration economies and how geographic co-location can have important economic effects. Classic works highlighting the importance of agglomeration include Thünen (1826), Marshall (1920), and Krugman (1991). Among those positive effects are, of course, many types of learning, such as skills and knowledge transmission and spillovers. See, for instance, Jaffe et al. (1993) on the localization of knowledge spillovers. Studying the geographic concentration of a skill which so directly benefits from agglomeration, chess, could be interesting in its own right, but examining how it is affected by the internet is an especially useful addition to this literature on geography. If internet effects are large, there could be implications for the future distribution of economic activities for which matching is important.

We rely on results from a paper in this literature that makes methodological contributions in the computation of geographic concentration indexes. Ellison and Glaeser (1997) were interested in measuring the geographic concentration of various types of economic activity. They realized, however, that existing measures made two important mistakes in calibrating observed concentration to a no-agglomeration-economies baseline: 1) clustering will exist absent any agglomeration economies just due to simple random variation, and 2) the existence of a size distribution of firms will also mimic clustering in the absence of agglomeration economies. They also wanted a measure that could compare concentration to a baseline that conditioned on underlying demographic or geographic variation which could exogenously give rise to clustering. We use the index they developed and adapt it to concentration of a skill as opposed to concentration of an economic activity. Details are provided in the Methods and Results section.

There is a growing literature studying the matching functions of the internet. An obvious setting is online job search. Faberman and Kudlyak (2019) offer a survey of the economics literature on the effect of internet job sites on the functioning of labor markets. Academics have also studied online dating and whether the phenomenon results in better romantic matches. Finkel et al. (2012) offer an interesting look at internet dating from a psychology perspective and study changes in quality of match and the overall functioning of the dating market. Finally, internet matching technologies have been studied in the context of consumer goods. I mention two empirical papers while acknowledging the existence of a theory literature on matching as well: Ellison and Ellison (2018) note the importance of internet matching technologies in the transformation of the market for used books. Adams and Williams (2019) study the limitations of them in consumer goods markets when the geographic distribution of preferences is far from uniform.

We are not aware of studies looking specifically at internet matching technologies and learning or skills acquisition, though. One should not be so surprised at this absence, perhaps. Most examples of distance learning—Kahn Academy, MOOCs, YouTube videos about making your own cat food or tying a bow-tie—would exploit the internet's broadcasting, not matching, technologies. So this paper will expand the literature on internet matching technologies to the area of learning and skills acquisition. There is, however, an extensive literature on the effects of internet technologies, broadly construed, in a variety of settings, such as labor productivity, job growth, crime, political engagement, and academic research. Without claiming to offer an exhaustive list, we refer to Agrawal and Goldfarb (2008), Forman et al. (2012), Bhuller et al. (2013), Hjort and Poulsen (2019), and Gavazza et al. (2018). Some papers in this literature treat internet penetration as exogenous, while others consider the potential endogeneity of internet spread, availability, and use. In particular, some exploit fortunate "natural experiments" arising from geological characteristics or physical constraints of infrastructure. Such natural experiments provide clean identification of the internet's effects in various settings, but they are neither available for a study of global effects of a technology nor, we would argue, as important when studying the effects on a skill as specialized as chess-playing. Put differently, if we were interested in studying the effect of internet penetration on, say, job growth, reverse causality would be an undeniable issue. It is harder to think that reverse causality could be at play in a study of internet penetration and chess-playing. (The related challenge of omitted variable bias is still present. Additional discussion can be found in the results section.)

A small literature has already emerged studying the effects of the abrupt shift to distance learning during the pandemic. A leading and interesting example is Carlana and Ferrara (2021). The authors analyze an online tutoring program offered to a fraction of low-income middle school students in addition to their online pandemic-era schooling and find positive and significant impacts. (In fact, they were instrumental in establishing the program, not just analyzing it.) Related is an interesting study on distance learning in higher education, which also exploits a randomized experiment (Cacault et al. (2021)). Education scholars and others outside of economics have been keenly interested in the question of the effectiveness of online learning. Some have taken advantage of the "natural experiment" that COVID provided as well. This literature is large and will be further afield methodologically and substantively from what we study here, but we mention a few contributions: Ikeda and Yamaguchi (2021), Champeaux et al. (2021), Agostinelli et al. (2020), Bacher-Hicks et al. (2021),Doyle (2020), Hardt et al. (2020), Escueta et al. (2017), and Dahiya et al. (2021).

Finally, we contribute to a small but interesting literature on the economics of chess. Some examples include de Sousa and Hollard (2015), de Sousa and Niederle (2018), González-Díaz et al. (2021), and Anderson and Green (2018). The first three of these papers examine issues of gender in the world of competitive chess. As we will see in the data we present, the gender mix in elite chess is extraordinarily lopsided. We will have a few comments about gender issues in this paper and will discuss de Sousa and Niederle, in particular, later when we do. The third and fourth papers use detailed data on individual chess games played to comment on the setting of reference points and players' behavioral biases. Levitt et al. (2011) is not a paper about chess per se, but uses world-class chess players in a lab experiment to test agents' abilities to backward induct.

2 A Short History of Chess-Playing over Distance

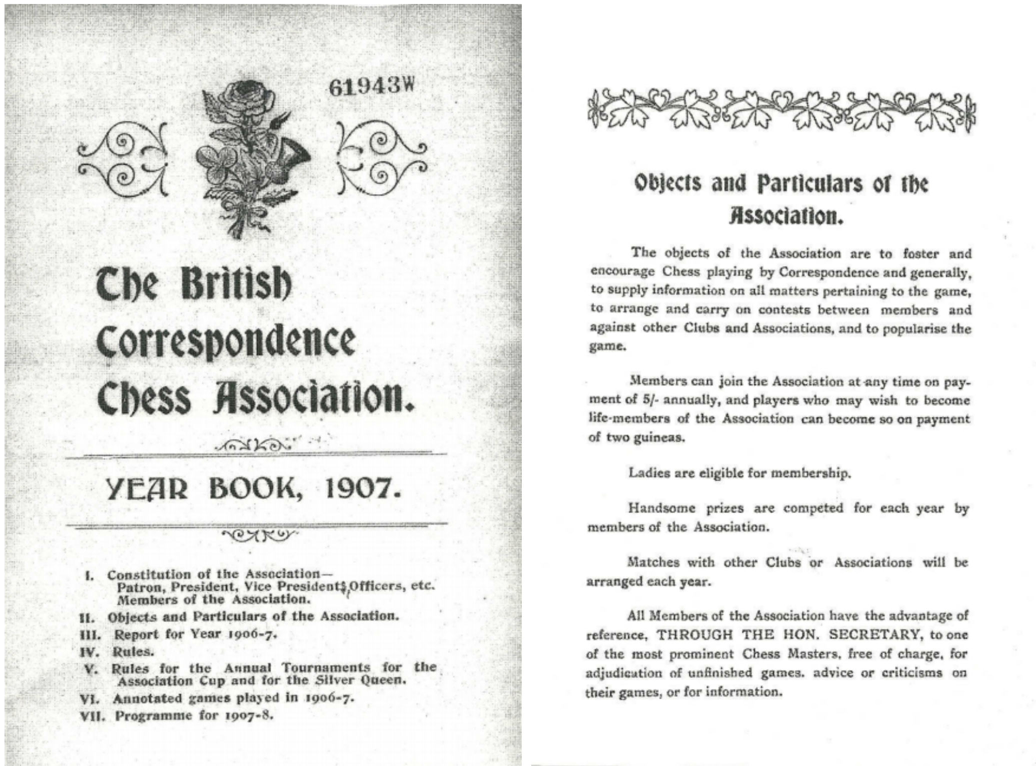
The problem of finding opponents, matching to appropriate-level players, and playing games remotely, is, of course, as old as the game itself. A game of correspondence chess was thought to be played between Emperor Nicephorus and the Caliph of Baghdad, Harun al-Rashid in the 9th Century.[3] Other participants in correspondence games might have included British and French royalty, Venetian and Croatian merchants, and Dutch army officers. A famous four-year match between the London Chess Club and the Edinburgh Chess Club starting in 1824 necessitated letters being carried 400 miles between the two clubs, and they were delivered within three days each round. Games of correspondence chess using homing pigeons have also been documented. The introduction of the postage stamp in London in 1840 resulted in a large increase in the number of correspondence games, as delivery of letters became inexpensive enough for ordinary people to afford.

Sending letters through the post to play a game of chess was a satisfactory, if extremely slow, technology for the game itself, but it did not solve the problem of finding the opponent in the first place. Chess clubs did, of course, mitigate this problem for their members, but one needed to be in proximity to a chess club of the appropriate level or be willing to travel to one in order to take advantage of the benefits of membership. There was still an unmet demand for players seeking appropriate games, and, by the middle of the 19th Century, chess magazines were established as an alternative technology for matching to potential opponents, and to promote correspondence chess more generally. See Figure 3 for an example of such a magazine.

In 1895, the first chess moves were transmitted over cable by telegraph between the British Chess Club and the Manhattan Chess Club. And in 1902, a chess game was played via radio transmissions between two ships in the Atlantic Ocean. The 20th Century saw the development of chess leagues devoted entirely to correspondence chess. It also saw growing suspicion of correspondence chess as a way to communicate in secret code, especially during WWII. In 1943, the FBI prevented Humphrey Bogart from playing correspondence chess with American GIs overseas for this reason. After the war, though, correspondence chess grew, and the International Correspondence Chess Federation was formed in 1951 with 65 member states and over 100,000 individual members.

By the early 1990's, email started to take over as the preferred medium for correspondence chess, due to decreased latency, but that technology would not dominate for long. In

[3] This and some additional facts here are taken from chess.com in its article section. Additional information was found on the Wikipedia entries for "Internet Chess Server" and "Correspondence Chess."



61943W

The British Correspondence Chess Association.

YEAR BOOK, 1907.

I. Constitution of the Association—
Patron, President, Vice Presidents, Officers, etc.
Members of the Association.
II. Objects and Particulars of the Association.
III. Report for Year 1906-7.
IV. Rules.
V. Rules for the Annual Tournaments for the Association Cup and for the Silver Queen.
VI. Annotated games played in 1906-7.
VII. Programme for 1907-8.

Objects and Particulars of the Association.

The objects of the Association are to foster and encourage Chess playing by Correspondence and generally, to supply information on all matters pertaining to the game, to arrange and carry on contests between members and against other Clubs and Associations, and to popularise the game.

Members can join the Association at any time on payment of 5/- annually, and players who may wish to become life-members of the Association can become so on payment of two guineas.

Ladies are eligible for membership.

Handsome prizes are competed for each year by members of the Association.

Matches with other Clubs or Associations will be arranged each year.

All Members of the Association have the advantage of reference, THROUGH THE HON. SECRETARY, to one of the most prominent Chess Masters, free of charge, for adjudication of unfinished games, advice or criticisms on their games, or for information.

Figure 3: The British Correspondence Chess Association

parallel, the technology for playing online chess was developing. Even before the internet, players could trade moves over networks like ARPANET and see the game evolving using graphical interfaces on their own computers. A dedicated chess server was first hosted at the University of Utah in 1992, and it soon moved to Carnegie Mellon. The capabilities of this and other chess servers evolved over time, but, at their core, they provided two benefits: a means for finding good matches between players and a means for actually playing the games. Note that these technologies emerged very soon after the commercial internet—there was not a substantial lag between the availability of internet connections and the accessibility of online matching and chess-playing technologies.

Today, the largest online venue for chess games is chess.com. It has grown from a membership of 250,000 in 2008 100-fold to 25 million in 2018. (The most recent figure reported, post-pandemic, is 93 million users.) At any given time, there are tens of thousands of players online either actively participating in a game or looking for a game. See Figure 4 for a screenshot of the chess.com website and Figure 5 for an example of their graphical

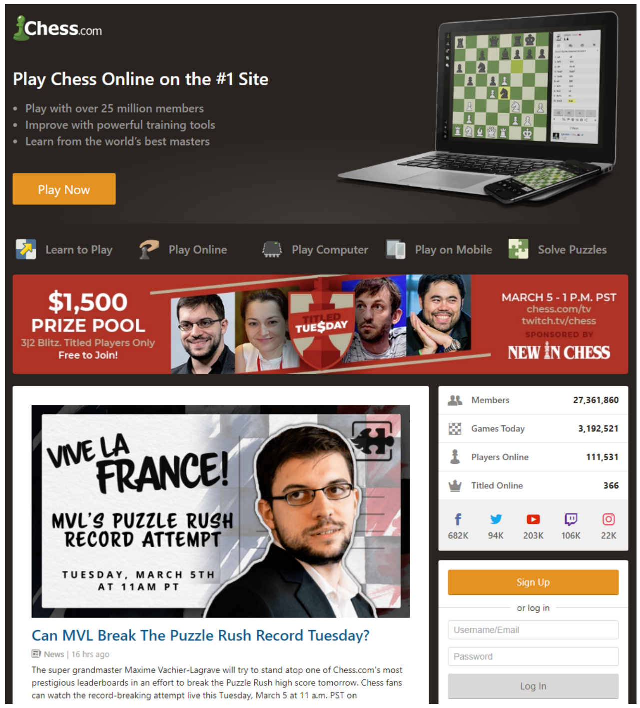


Figure 4: Chess.com Website

Figure 5: Graphical Interface for Chess Games on Chess.com

interface for games. Although one can access many learning resources on websites like chess.com, they are not visited only by beginners or hobbyists. Elite, or "titled," players also use chess.com—Figure 4 notes that, at the moment of the screenshot, there were 366 titled players online.

As one can see in the timeline in Figure 6[4] below, technological innovation over the last several decades has impacted chess in other ways as well, most notably in the development

[4] Wall, Bill (2017) and Mullen et al. (2009) were consulted to construct this timeline.

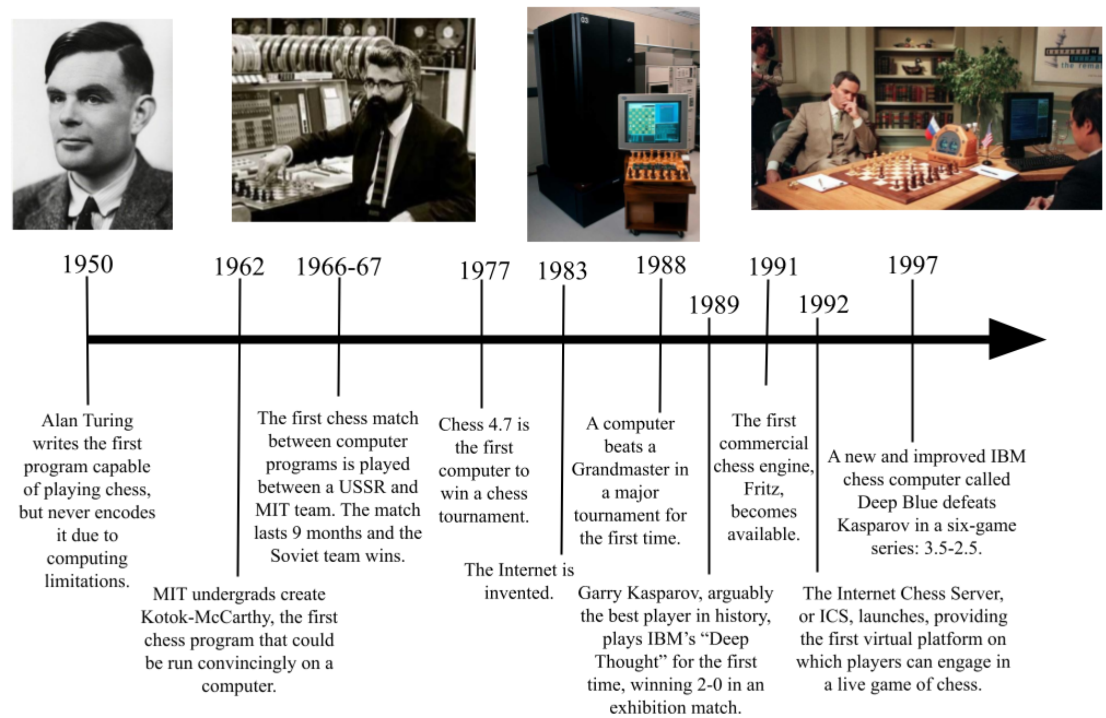


Figure 6: Timeline of Chess Bot Technology

and improvement of computer programs capable of playing chess, or "chess bots." As early as 1950, Alan Turing wrote a program capable of playing chess, but it was never encoded. Several attempts followed Turing, with the first chess program able to play chess convincingly being created by MIT undergraduate students in 1962. Named Kostok-McCarthy, this program played a match against a newly created Soviet program in 1966. After nine months, the USSR program emerged victorious.

The first chess program to be consistently successful against competitive human opponents was Chess 4.5, created by graduate students at Northwestern University. In 1977, Chess 4.5 earned a record of five wins and one loss against human opponents to win the Minnesota Open. This program was still not capable of defeating the Grandmaster-level human chess players, however. A computer would not be able to beat a chess Grandmaster (GM) in a major tournament until 1988, when Carnegie Mellon University's HITECH chess machine defeated GM Bent Larsen in the American Open. The decade finished out with a 1989 exhibition match between top-ranked Garry Kasparov and IBM's Deep Thought chess computer, where Kasporov won both games played. Computer chess superiority over humans was finally achieved nearly a decade later in 1997 when a new and improved version of Deep Thought called Deep Blue narrowly won a six game series against Kasparov: 3.5-2.5.

As our focus here will be on the effect of the internet technologies that allow distance matching and game-playing, one might be concerned that any such effect could be confounded with the potential effects of these chess bots on teaching chess skills. Note, though, that access to high-quality chess programs that could help train the world's elite chess players was extremely limited until quite recently. IBM, for instance, did not make Deep Blue widely available to all aspiring young chess players for training purposes in the 1990s and 2000s. We, therefore, believe that any effects that we detect are more likely due to internet matching of players as opposed to players training with bots. Still it is useful to keep this timeline and caveat in mind.

We close this section with a quote we received from Henrik Carlsen, father and manager of the current number one player in the world, Magnus Carlsen, when we told him that we were writing this paper:

> As Magnus's father I can tell you that online chess was an absolutely crucial factor allowing the development of his world class talent coming from Norway without that much of a chess "culture." However, computer literacy and access as well as internet access in Norway have generally been very good, something Magnus took advantage of being from the middle class in prosperous Norway.

3 A Data Set of Chess Players

We collected a comprehensive data set of the top chess players in the world since 1967 from various internet chess sites. Hundreds of thousands of players were represented at some point over the 50-year period, with more recent years listing as many as 200,000 but earlier years having as few as the top 400 or so players. For each year, "top" players are defined as the players with the highest current Elo rating.[5] The data include name, rating, birthdate, sex, and country for each player each year he or she is listed. A unique player ID also appears in the data, although it proved fairly unreliable, so extensive cleaning was required to match up players across years. Based on rating, we also create a ranking variable, where the player with the highest rating in each year was ranked first, second highest was ranked second, and so forth. Players with the same rating were considered tied and given the same rank.

For two different reasons, we will want a weight associated with each rank, essentially a non-linear transformation of ranks. First, when creating a geographic index of high-level chess ability, we will need a way to measure "embodied chess skill" in a country in a year. We will not want the 1000th best player to "embody" the same amount of chess skill that the top-ranked player in the world does. That argument suggests weights decreasing in rank. Furthermore, we would think that a reasonable and robust weighting scheme would allow weights decreasing at a decreasing rate, so that the difference in "embodied chess skill" between the first and 10th ranked players was greater than that between the 990th and 1000th best players. To that end, we created a non-linear weighting scheme based on the rank r_i of each player i in year t,

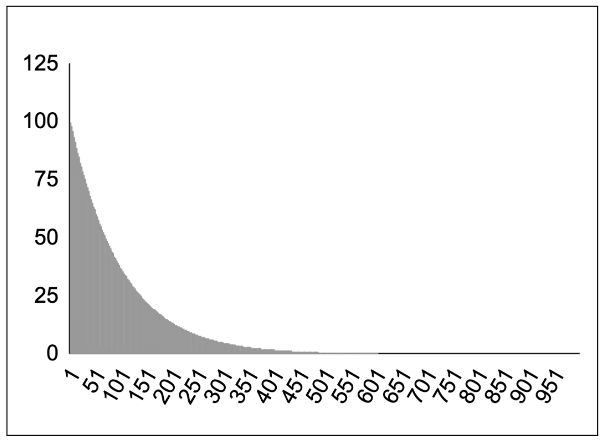


Figure 7: Weighting Function

$$w_i = \frac{100}{0.99} 0.99^{r_i}.$$

See Figure 7. As we said, we can think of w_i as embodied chess skill. This function assigns a weight of 100 to the top player in the world and a weight 0.004 to the the 1000th best

[5]The Elo Rating was first adopted in 1960 by the US Chess Federation and more widely after that. It differed from the previous rating system in that it took account of the strength of one's opponents, not simply wins and losses. It is a self-referential rating system in the sense that how much your rating changes upon a win or loss is a function of the difference between the before-match ratings of you and your opponent. It also has a zero-sum property, where rating points gained by the winner of a match will equal rating points lost by the loser of the match.

player. (The scale is, of course, irrelevant. Only the relative weights matter.) All players would, in theory, have positive weights, but many are rounded to zero given computing constraints. And ones with very small values would contribute essentially nothing to the index. Therefore, we retain only approximately the top 1000 chess players every year for ease of manual data cleaning. (The raw data sets since the early 70s have included at least the top 1000.) This weighting scheme is arbitrary but not unreasonable. It places more weight on top players but positive weight on many players every year. Furthermore, the difference of one rank is treated as more important for the top players than for much lower-ranked players. A second benefit of this scheme is that, given that there seem to be different inclusion criteria in the data set every year that would be difficult to model, it is helpful to have a weighting scheme that renders that issue largely unimportant.

Computing a ranking (and then a weighted ranking) for each player each year also allows us to sidestep the issue of "rating inflation," a phenomenon well-recognized by chess players whereby the numerical value of a rating has failed to retain the same meaning over time.[6] In other words, the interpretation of rank in a particular year is clear, so we prefer to base our analysis on that metric. Another feature of the data set is that we are able to follow the players across years.[7] We have a small minority of players who switch countries mid-career. Almost all of those players did not actually move from one country to another but rather had their country change under them. For instance, the Soviet Union had a large number of players early in the data set, which were subsequently assigned to their particular member state after the break-up of the Soviet Union. We will discuss how changes in the boundaries of countries affect the computing of the index of geographic concentration in the results section.

Finally, using each player's country, we merge country-level covariates from the World

[6] Basing our analysis on rank in a particular year does have the disadvantage that we cannot say anything about changes in the overall level of quality of chess play. Some chess players have created a ratings inflation index, meant to correct ratings to make them comparable over time. Since those indices are somewhat controversial, we make the conservative decision to look only at changes in distribution over time.

[7] Much of the matching was easily accomplished with consistent player IDs across years, but a number of the players needed to be matched manually, due to missing or inconsistent IDs. Manual matching was performed initially with exact matches of name and then checked based on the other individual characteristics for mistakes in both directions—failing to match people with small differences in their names across years or matching people with identical names but different characteristics.

Bank, such as population, land size, and, importantly, internet penetration.[8] The units on each should be fairly self-explanatory. Note that the variables on population density, percent urban, and percent in the largest city were collected to help capture any potential effects of traditional geographic agglomeration on chess skill. Life expectancy was collected to proxy for overall wealth of a country.

Table presents summary statistics from this data set. Panel A includes variables at the player-year level. We use data from 1967 to 2016. Panel B includes variables at the country-year level from 1980 to 2016, corresponding to the observations we use in our country-level analyses.[9]

Note that a very large majority of our players are male, 99%. Sources of such dramatic discrepancies would be interesting to investigate, in particular whether the necessity to match to opponents is an important factor. As mentioned before, both de Sousa and Hollard (2015) and de Sousa and Niederle (2018) make interesting observations on gender differences in chess.

We have a lot of variation, both over time and across countries, in internet penetration, the range being 0 to over 98%. The World Bank statistics on internet penetration start in 1990. At that time, even the most advanced countries had low penetration, so we impute 0 to all countries before 1990. Note also that the best chess players in the world originate from a very heterogeneous group of countries, some very rural, some entirely urban, and with almost a 40-year difference between the highest and lowest life expectancies at the country-year level.

Our chess players range in age from 12 to 89, suggesting that age should be considered in the analysis due to age-based differential exposure to the internet. In other words, a player who is 70 years old in the year 2000 would not have had access to the internet during what were likely her formative chess years, regardless of internet penetration in her

[8] The country codes for chess differ from those used by the World Bank, an easy enough fix, but sometimes the actual country definitions differ as well. For instance, the World Bank provides historical statistics *for the countries that currently exist*, not for the countries as they existed in the relevant year. We had to essentially reconstruct country covariates for the Soviet Union, Czechoslovakia, and Yugoslavia, based on the statistics reported by the World Bank for the constituent countries. In addition, the year chess players started reporting their residence in a new country did not necessarily correspond perfectly with the year that that country first existed. We had to create counterfactual covariate series in those circumstances.

[9] We choose to use a shorter time period for the country-level analyses due to sparser individual-level data in the beginning as well as the fact that the internet could not have any impact until 1990 at the earliest.

Table 1: Summary Statistics—Top Chess Players and Their Countries

Variable	Variable Description	Mean	StDev	Min	Max	Obs
Panel A: Player Level						
$BirthYear$	Birth year	1958	17.0	1896	2001	46,315
Age	Age	35.1	11.0	12	89	46,315
$Male$	Male indicator	0.99	0.12	0	1	47,994
$Rank$	Rank	507.6	299.6	1	1030.5	48,017
$WtdRank$	Weighted rank	10.2	20.4	0.003	100	48,017
$IntExp$	Internet exposure	1.2	6.4	0	97.3	46,315
$CumIntExp$	Cumulative internet exposure	11.4	49.6	0	615.1	48,017
Panel B: Country Level						
Variable	Variable Description	Mean	StDev	Min	Max	Obs
$Internet$	Percent internet penetration	20.4	28.2	0	98.2	3,539
$Population$	Population	53.1	164.9	0.03	1371.2	3,539
$Area$	Area	1.0	2.6	0	21.5	3,539
$PopDensity$	Population density	415.5	1832.9	1.1	18,865.5	3,539
$Urban$	Percent urban	64.7	20.6	14.3	100	3,539
$LargestCity$	Percent largest city	21.9	20.1	0.6	100	3,539
$LifeExpect$	Life expectancy	72.7	6.1	50.0	89.5	3,539
$AvgAge$	Average age	34.2	5.7	15	64	2,445
$SumWtdRank$	Sum of weighted ranks	104.5	314.5	0	4286.7	3,539
$IntIndPool$	Internet-induced pool	1.1	2.0	0	13.5	3,539

Notes: $Male$ is coded 1 for male and 0 for female. *Population* is in millions. *Area* is in millions of square kilometers. *AvgAge* is the average age only among the chess players in our data set.

home country in 2000. A 20-year-old player in the same year, however, grew up and learned to play chess during a time when the internet was available, widely in some countries and not in others. In fact, even the average age by country-year of chess players in our data set ranges from 15 to 73. The variable $IntExp$ is meant to capture this notion. It is the weighted sum of internet penetration rates in a particular player's country during his or her formative years. We constructed the variable with piecewise linear weights equaling 0 from birth until age 6, jumping up to 0.25 and then linearly increasing until reaching 1 at age 12, maintaining that level until age 20, and then decreasing back to 0 at age 25. In other words, a 16-year-old chess player receives the full benefit of internet penetration in her country in her 16th year but a 6-year-old player would only receive a fraction of the benefit in his 6th year. In addition, we create $CumIntExp$, which is the lifetime cumulative value of that variable for each player discounted annually by 0.1. (These variables can be computed over years when the player is not included in the data set because we know how old each player was in each year and also what the internet penetration in her country was.)

The shape of our age-weighting function is reflected in Figure 8. We based the specifics of this function on player stories and anecdotes. Note, though, that small changes in these specifics will not have large effects on the values of the variables.

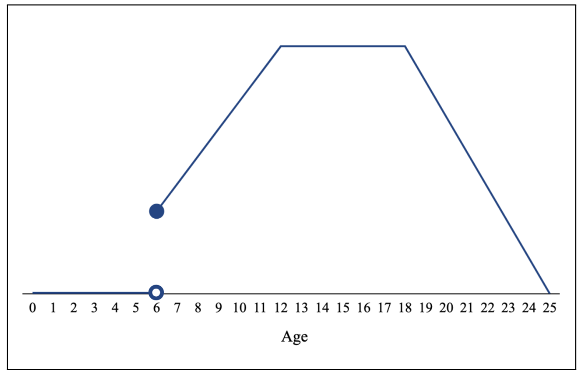
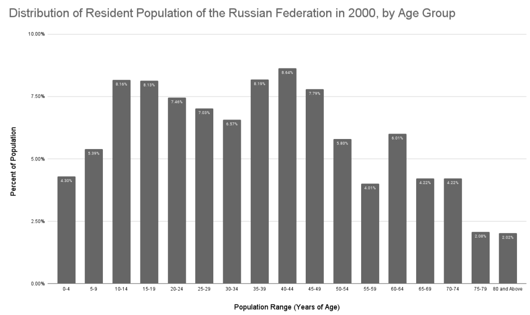


Figure 8: Age Profile for Chess Learning

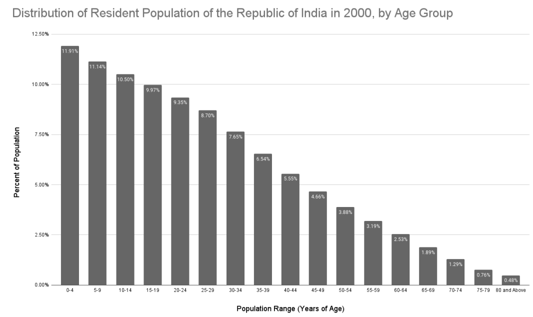
IntExp and *CumIntExp* are both defined at the individual level. Aggregating this measure up to the country-year level is not entirely straightforward. We do not, in fact, want a measure of the average internet exposure of everyone in a country that *ended up* becoming a top chess player. Rather, we want a measure of the average internet exposure of the part of the population that potentially *could have* become top chess players. We, therefore, created the variable *IntIndPool*, our abbreviation for Internet-induced pool of chess talent. It is a function which looks back retrospectively at internet penetration and age distribution and creates a sort of cumulative interaction of them. We feel that including data on retrospective age distributions is the only natural way to aggregate *CumIntExp* up to country-year level. Just to emphasize the need for including a measure of age distribution in *IntIndPool*, we include Figure 9 with the graphs of the very different age profiles of two countries which are quite important in our data set, Russia and India.[10] Russia has relatively small fractions of their population in the youngest age groups in contrast to India. Any measure attempting to reflect a pool of potential chess stars created by internet exposure in a country would need to accommodate those differences.

The schematic below in Figure 10 illustrates how we construct the internet-induced pool of chess talent using the cohort born in 1990 in a hypothetical country C as an example. In 1990, none of the cohort would have been using the internet for two reasons: they were

[10] These graphs were constructed using subnational population data from the US Census Department at www.census.gov.



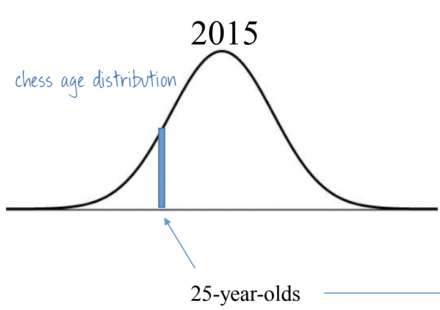
(a) Russia



(b) India

Figure 9: Population Distribution by Age Group in 2000

newborns and the internet had not yet reached their country. So, in the year 1990, at least, the internet was not contributing to anyone in that cohort in that country learning how to play chess. By the time they were ten years old, in 2000, internet penetration had reached 20% in country C, and ten-year-olds would be approaching prime age for honing their chess skills, but not be at the maximum quite yet. (They would, according to our age profile function, be at 78% of their maximum.) Furthermore, that cohort represents 2% of the population of this hypothetical country. We multiply these three numbers, 0.20, 0.78, and 0.02, to obtain the contribution of this cohort to the internet-induced pool in the year 2000. We simply do this calculation for every cohort in every year before year t and weight the total contribution by cohort by its prevalence in the actual age distribution of elite chess players to obtain our measure for country C in year t.



Sample Years	Age	Fraction of Pop	Internet Penetration	Age-Based Weight
1990	0	0.02	0	0
1995	5	0.02	0.10	0
2000	10	0.02	0.20	0.78
2005	15	0.02	0.40	1.0
2010	25	0.02	0.45	0.71

Multiply across and sum down

Figure 10: Constructing Internet-Induced Pool of Chess Talent

See Figure 11 below for a graph of $IntIndPool$ for a few selected countries. You can see that all countries have an upward trajectory, as one would expect, but that the slopes vary. Russia had early high levels of internet penetration, but its aging population ensures that its slope is very flat relative to other countries. India has a young population, but its internet penetration, even as recently as 2016, dampens the level and growth rate of its internet-induced pool.

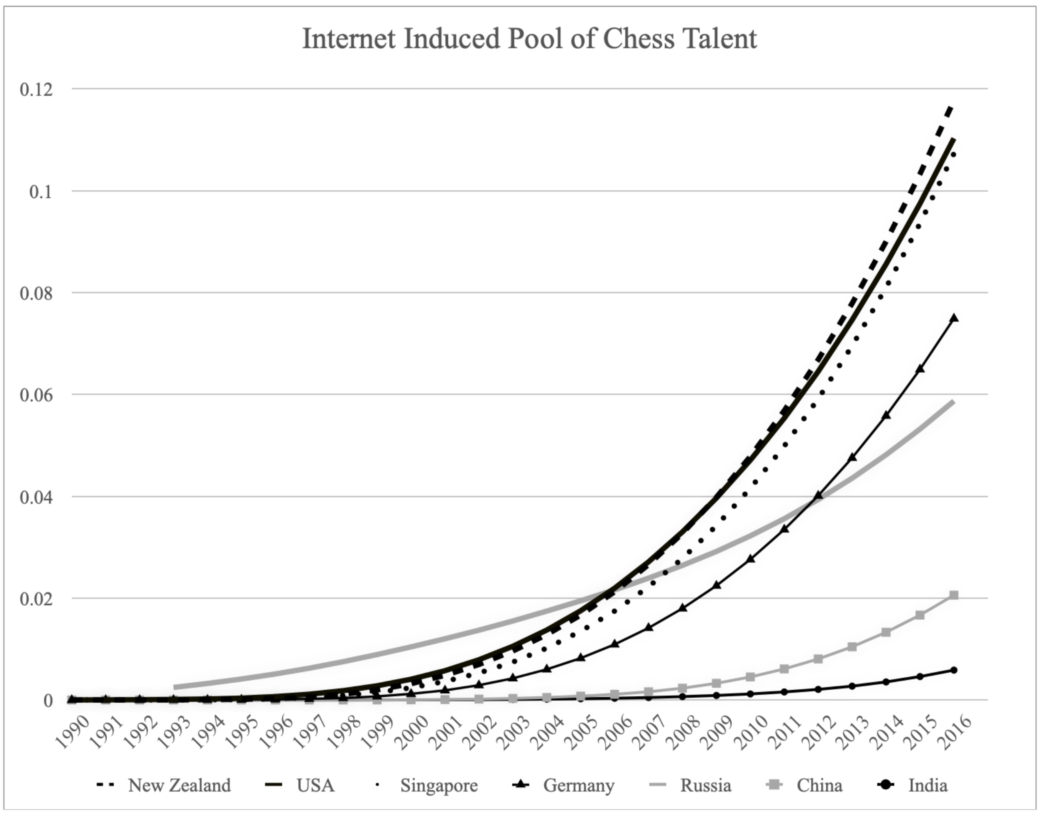


Figure 11: Internet-Induced Pool of Chess Talent from 1990 to 2016

4 Empirical Strategy and Results

With these data in hand, we turn back to the core analyses. We rely on two methodologies, standard panel data methods for the player-level and country-level regression analysis, and construction of an index of geographic concentration based on the Ellison-Glaeser Index (EGI). Let us first discuss the calculation and use of the EGI. Note that it was developed as a measure of industrial geographic concentration, and we need to adapt it for measuring the geographic concentration of a skill. As we explained in the previous section, we created a non-linear transformation of rank to capture the notion that higher-ranked chess players embody more chess skill than lower-ranked ones, and that skill decreases in rank at a decreasing rate. We denote this measure of embodied chess skill w_i. When constructing an index of geographic concentration, we can think of chess skill as being analogous to firm size—a country with the single best player has a large concentration of chess skill, just like the country with the world's largest airplane manufacturer has a large concentration of airplane manufacturing activity. The first step to construct the index is to create a standardized version of this weighted rank measure,

$$s_i = w_i / \sum_j w_j,$$

where the denominator is the sum over all weighted ranks, or players, for a particular year. Then, the index for year t is

$$I_t = \{\sum_c (\sum_{i \in c} s_i - p_c)^2 - \sum_k s_k^2\} / \{1 - \sum_k s_k^2\},$$

where the sum indexed by c is taken over countries, the sums indexed by i and k are taken over players, and the subscript t on all variables has been suppressed. p_c is the predicted share of chess players in country c (in year t) given whatever demographics one wants to condition on. Typically, one would simply use population share—the EGI of chess skill presented in the Introduction used population share—but we will also compute an EGI that controls for internet-related variables. Put differently, one can produce "predicted share" in the index by using a single variable like population share, using a sophisticated non-linear model incorporating many variables to produce a predicted share, or something in between. The index reflects the *excess* probability of two random players coming from the same country, where "excess" is defined relative to whatever model gives us predicted share. Furthermore, we can use a richer share model to essentially decompose the effects of additional variables on changes in the geographic distribution of chess skill.

We will begin with some regression results using the player-year level data. We offer them more in the spirit of descriptive statistics due to the potential issues arising from the selection of the sample. (Recall that we argued that sample selection should not be an issue in the construction of the index since the marginal players are contributing so little to the index. Sample selection should also not be an issue with the country-year level regressions we present further below since weighted averages and sums over players are computed and, again, the marginal players will contribute very little to those values. For the individual regressions, however, each player, no matter how small his or her weighted rank, contributes equally to the regression results.) See Table 2 for results of the player-year level regressions. We have included three separate specifications, all with weighted rank as the dependent variable and including age-related and internet-related explanatory variables. The third specification also includes gender and the second includes country fixed effects.

Table 2: Player-Year Level Results

	Dep Variable: $WeightedRank$		
Exp. Variables	(1)	(2)	(3)
Age	**0.512**	**0.600**	**0.497**
	(0.040)	(0.041)	(0.041)
Age^2	**-0.007**	**-0.008**	**-0.007**
	(0.001)	(0.001)	(0.001)
$Internet$	**-3.556**	**-1.846**	1.634
	(0.380)	(0.399)	(2.314)
$CumIntExp$	**1.235**	**1.170**	**1.202**
	(0.247)	(0.248)	(0.247)
$Male$			**6.335**
			(0.651)
$Male * Internet$			**-5.189**
			(2.321)
Country FE		✓	
Observations	46,315	46,315	46,294
R^2	0.005	0.055	0.006

Notes: Standard errors are robust. Coefficient estimates in boldface are significant at 5% or better.

In all cases, Age and Age^2 are important variables, suggesting that an additional year of age improves one's ranking on average initially but at a decreasing rate that eventually turns negative. The internet-related variables should be considered together. First note that $IntExp$ was not significant in any specification and is not included. This result is not surprising. The variable measures age-weighted internet exposure of that player *in the current year*—it would be somewhat surprising if current year exposure had a large effect. $CumIntExp$ is positive and significant in all specifications. That variable measures the cumulative age-weighted exposure over the player's life and is meant to proxy for the opportunities the player has had to use the internet for learning chess and improving his or her rank. Finally, the variable $Internet$ has an inconsistent effect across specifications. Keep in mind that, in the presence of $CumIntExp$, $Internet$ may be picking up general differences in wealth or socioeconomic factors in the player's country. In that light, it is not surprising its estimated effect changes somewhat with the inclusion of country fixed effects.

The estimated coefficient on $Male$ is large, positive, and significant, whereas the estimated coefficient on the interaction term is large, negative, and significant. The caveat about interpreting these individual results notwithstanding, these results are suggestive of women benefiting disproportionately from increased internet penetration. For example, high-level women chess players might have a more difficult time matching with opponents (who know they are female) than their male counterparts and therefore benefit more by the introduction of (potentially anonymous) internet matching technologies. Top chess player Susan Polgar famously said "I have never beaten a healthy man," suggesting that men are particularly averse to losing to women and when they do, they feel compelled to invent excuses for their defeat. It would stand to reason that they might try to avoid playing chess with a woman who may beat them, but the internet could offer a veil of anonymity to women looking for a chess opponent. In addition, this suggestion is related to the results in de Sousa and Niederle (2018) which finds that an affirmative action program for women chess players had positive and lasting effects on the quality of their play, even after the program ended, perhaps because the women were matched with male players they would not have played against absent the program, and what they learned from those experiences had lasting effects.

The fact that the inclusion of country fixed effects had a modest impact on the main age- and internet-related variables of interest suggests that these effects may not vary too much from country to country. We could not include a fixed-effects specification with the gender-related variables due to very small numbers of women in the sample.

We do want to emphasize again that this first set of results needs to be interpreted with the caveat on sample selection mentioned above, and we would, therefore, not place causal interpretations on any of these estimates.

Our second set of regression results is using data at the country-year level. We will further investigate some of the issues suggested by the first set of regressions, without the econometric concerns raised by sample selection. Our model is an exponential one, where the sum of the weighted ranks in a particular country and year—the total embodied chess skill—is explained by e raised to an index of the explanatory variables. In other words,

$$\mathrm{E}(SumWtdRank_{ct}) = \exp(\beta_0 + \beta_1 X_1 + \cdots + \beta_k X_k + \epsilon_{ct}).$$

This model can be estimated as a standard Poisson regression even though the dependent variable is not a count variable.[11]

At this point, anticipating that we will attempt to explain some of the variation in chess skill over countries and time by measures of internet penetration, one might be concerned about either endogeneity, omitted variable bias, or both. On the former, we feel relatively comfortable treating internet penetration as exogenous to the use of the internet for playing chess—there are myriad factors such as type of government, power of political groups, state of technology, terrain, market structure of ISPs, and so forth, that would play much more central roles in decisions about whether and when to extend the physical network and enable connectivity in a country. On the latter, we acknowledge that correlation between internet penetration and other variables could be more problematic. As you will see below, we deal with this issue in two ways. First, we include demographic measures, such as life expectancy, and rich fixed effects in various specifications, which should mitigate much of this concern. Second, we also will rely primarily on a variable constructed with both country demographics and the path of internet penetration over time to identify the internet-related effects we are most interested in. Finally, as explained in the introduction, the type of small-scale, local natural experiments which others have used to address these econometric concerns in previous studies of the effect of the internet would not be useful or feasible in a study whose focus is on changes in global distribution of skills.

[11] Of course one would also want to loosen the Poisson restriction of variance equaling mean, which we do.

See Table 3 for the output of several such models we investigate. We include additional variables in the specifications as we move from left to right in the table. The right-most three specifications also add fixed effects, at the year-, subregion-, and country-level.

Since one of the characteristics of chess skill acquisition that we have emphasized throughout this paper is its dependence on matching to an appropriate opponent, we would want our specification to allow for the possibility of agglomeration economies at various levels. Therefore, we do not impose that the total embodied chess skill in a country will increase one-for-one with its population and instead allow estimation of the coefficient on $log(population)$. Note, however, that its estimated coefficient does not suggest agglomeration economies, at least on the scale of country size. It is positive and highly significant, though. $LifeExpectancy$ is included to proxy for overall socioeconomic status of a country, and it is also consistently positive and significant in some specifications. The coefficients on $PopDensity$ and $PercLargestCity$ in the presence of a population control could identify agglomeration economies within countries: for a given population size, countries with high population density or a large fraction living in a dominant city could be ones where matches may be relatively easy to make, producing higher levels of chess skill. We do not find significant coefficient estimates for those variables in any of our specifications, though.

Again looking across specifications, the coefficient on $Internet$ is inconsistent and mostly insignificant. This finding comports with the idea that *current* internet penetration in a country has little effect on the level of chess skill, which seems quite plausible. The estimated coefficient on $IntIndPool$, though, is positive and significant in specifications (3)-(5). So, controlling for current internet penetration, our retrospective measure of the internet's effect on the country's pool of chess talent is explaining a significant amount of variation we see in chess outcomes.

This result holds (and is strengthened) with the inclusion of year fixed effects. When subregion fixed effects are added, the significance is lost but the positive point estimate remains. When we add country fixed effects, the sign on the $IntIndPool$ flips to negative. These results are puzzling, but interpretation is clouded by the fact that country definitions are not stable over time. (For instance, the Soviet Union fixed effect exists until 1980, at which point it disappears and a number of fixed effects for its constituent countries appear.)

Table 3: Country-Year Level Results

Exp. Variables	Dep Variable: $SumWeightedRank$ (1)	(2)	(3)	(4)	(5)	(6)	(7)
$Log(Population)$	**0.361**	**0.363**	**0.396**	**0.409**	**0.439**	**0.846**	-0.657
	(0.117)	(0.117)	(0.121)	(0.124)	(0.122)	(0.100)	(1.347)
$LifeExpectancy$	**0.048**	0.053	0.062	0.063	0.059	0.015	**0.100**
	(0.023)	(0.037)	(0.036)	(0.037)	((0.036)	(0.038)	(0.0416)
$PercLargestCity$	-4.476	-4.459	-3.998	-4.158	-3.835	3.399	3.201
	(2.905)	(2.886)	(2.765)	(2.773)	(2.707)	(1.808)	(4.503)
$PopDensity$	-0.003	-0.003	-0.002	0.001	-0.004	-0.001	0.002
	(0.002)	(0.002)	(0.002)	(0.003)	(0.003)	(0.001)	(0.007)
$Internet$		-0.170	**-1.252**	0.266	-0.825	-1.079	**-0.871**
		(0.631)	(0.618)	(0.633)	(0.598)	(0.598)	(0.381)
$IntIndPool$			**15.754**	**16.241**	**23.836**	9.654	**-10.790**
			(5.723)	(5.792)	(7.112)	(7.255)	(4.482)
$PopDensity * Internet$				0.005	0.004	0.001	0.001
				(0.003)	(0.003)	(0.001)	(0.001)
Constant	-3.965	-4.319	-5.602	-5.715	**-7.503**	**-10.637**	-3.431
	(2.336)	(3.126)	(3.001)	(2.980)	(3.193)	(4.165)	(22.801)
Year FE					✓	✓	✓
Subregion FE						✓	
Country FE							✓
Observations	3,539	3,539	3,539	3,539	3,539	3,539	3,539
Pseudo R^2	0.28	0.28	0.29	0.31	0.33	0.93	0.74

Notes: Standard errors are robust and clustered at the country level. Coefficient estimates in boldface are significant at 5% or better.

Subregion fixed effects are consistently defined throughout our time period, though.[12]

Table reports similar results for a data set where we compute the dependent variable—our measure of the concentration of current chess talent—based only on the women chess players in our data set. Recall that about 1% of the top chess players are women. Therefore, although we have many fewer women, we still have the same number of observations since we estimate these models on country-year-level data. (We have many more zeros in our dependent variable.) Note also that we could not report a specification with country fixed effects due to the large number of zeros in our dependent variable.

We find it interesting that, in contrast to the main results using all players, the *Internet* variable plays a stronger and more consistent role here. It is positive in all specifications

[12]To investigate whether the potential effects of the internet were language-specific, we estimated a model with an interaction between *IntIndPool* and a dummy for English language usage. The interaction had a negative estimated coefficient, which we found difficult to interpret at face value: the internet should not be *easier* for people from non-English-speaking countries to access. The language dummy could have been proxying for something else, like cultural differences.

Table 4: Country-Year Level Results (Women)

Exp. Variables	Dep Variable: $SumWeightedRank$ (1)	(2)	(3)	(4)	(5)	(6)
$Log(Population)$	0.246	0.242	0.202	0.120	0.182	**0.904**
	(0.259)	(0.259)	(0.258)	(0.259)	(0.257)	(0.301)
$LifeExpectancy$	0.019	-0.008	-0.023	-0.023	-0.044	-0.051
	(0.028)	(0.038)	(0.041)	(0.041)	(0.051)	(0.073)
$PercLargestCity$	-2.407	-2.424	-2.881	-2.905	-3.044	14.184
	(1.553)	(1.612)	(1.726)	(1.709)	(1.822)	(9.654)
$PopDensity$	-0.001	-0.001	-0.002	-0.001	-0.001	-0.002
	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)	(0.001)
$Internet$		0.937	**3.055**	**3.220**	**3.590**	5.224
		(0.511)	(1.245)	(1.379)	(1.684)	(2.817)
$IntIndPool$			-45.985	-46.161	-51.213	**-107.565**
			(24.656)	(24.582)	(27.935)	(38.559)
$PopDensity * Internet$				-0.001	-0.001	-0.004
				(0.002)	(0.003)	(0.004)
Constant	-4.954	-3.181	-1.337	-1.345	0.841	**-17.753**
	(6.543)	(7.114)	(7.338)	(7.333)	(7.726)	(9.048)
Year FE					✓	✓
Subregion FE						✓
Observations	3,539	3,539	3,539	3,539	3,539	3,539
Pseudo R^2	0.05	0.06	0.09	0.09	0.13	0.58

Notes: Standard errors are robust and clustered at the country level. Coefficient estimates in boldface are significant at 5% or better.

and significant with similar magnitudes across three of them. Also in contrast to the main results, the $IntIndPool$ variable is mostly insignificant. (It is significant and negative when we add year and subregion fixed effects.) Given the very small number of women, we want to interpret these results with caution. The interpretation they suggest, though, is that the internet has played, perhaps, a less important role in early training of women chess players and a more important role in their ability to maintain their status through continued practice and competition. (Current internet penetration is more important in most specifications than our retrospective measure.) Even if we are hesitant to place any weight on the specifics of these results, the contrasts between them and the overall results (dominated by men) suggest that the role of the internet in women's chess careers could be quite different than that for men's chess careers.

These regression results suggest a role for the internet in the dissemination of chess knowledge and skill. One large question remains unanswered: can we say anything about the internet's role in the changing geographic distribution of chess skill? Recall that, in the Introduction, we saw a graph of the declining index of concentration of chess skill over the last half-century. Recall also that we can calculate that index using different predicted shares (from models with multiple explanatory variables) to decompose the effects of different sets of factors. To that end, we estimate two more pared-down versions of our regression model with per capita sum of weighted rank explained by some of the variables of greatest interest. We feed predicted values from those regressions into predicted shares of chess skill by country and year for a more sophisticated EGI than the one we presented in the Introduction. See Figure 12. Index 1 conditions only on population (linearly), so it corresponds to the EGI in Figure 2, except that we only graph the post-internet period. Index 2 conditions on internet penetration and internet-induced pool in addition to population, so a comparison between indexes 1 and 2 isolates the role of those two internet-related variables in the change of the global distribution of chess skill.

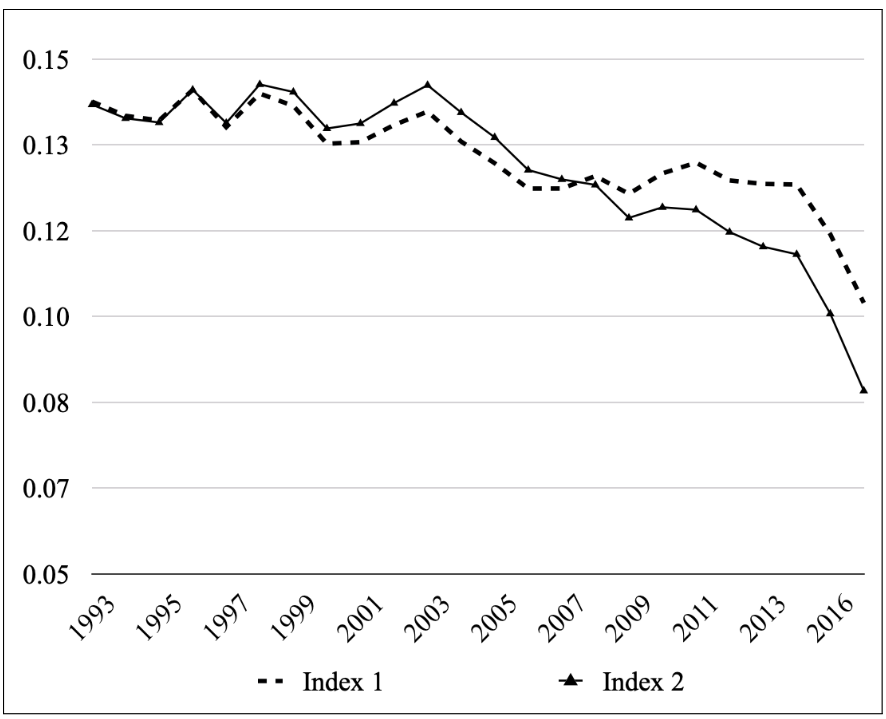


Figure 12: Geographic Concentration of Chess Skill 1993-2016

Recall the interpretation of the index: it is measuring the *excess* probability of co-location relative to what would be predicted from our set of conditioning variables. And when we condition on additional variables, we are controlling for more of what causes co-location. In other words, we could think, roughly speaking, of the difference between the two indexes as representing the effect of internet penetration on geographic concentration, conditioning on population.

At the beginning of the period, the two indexes coincide. This is not surprising given that the internet played a small role in the early 90's. Starting a few years later, we see that the internet-related variables do a worse job of explaining chess skill until about 2008, at which point they start doing a much better job of explaining it. In other words, as the geographic concentration of chess skill in the world has been decreasing, the internet had little or no power to explain that decrease in its early years but seems increasingly important in more recent years.

We find the timing very interesting in light of the classic puzzle of the early 2000s in the productivity literature: computers were widely adopted and used but evidence of their effect on productivity was difficult to come by. In fact, Robert Solow famously said "You can see computers everywhere except in the productivity statistics." Here, we see timing echoing that of computers' measured effect on productivity: no evidence of an effect of the internet in the two or so decades after access started diffusing into various countries, but a noticeable and accelerating effect thereafter. (See Syverson (2018) for discussion of the evolution of the overall productivity statistics.). Our mechanism, however, could be different. Access to the internet does not make playing most chess tournaments more efficient in any meaningful sense. Rather, the internet is a tool allowing young people to learn chess more efficiently, an effect that only becomes apparent years after those young people began learning it.

5 Conclusion

Our goal was to present evidence on the role of the internet in learning and skill acquisition. A second, related, question concerned the geographic distribution of knowledge and skills and whether we could find evidence of it changing as a result of the internet. We chose to study a particular skill that was easy to measure and quantify, chess playing. That skill is special in another way as well: it is one that, arguably, draws more intensively on one particular internet technology, match making, than many other types of knowledge and skills. For that reason, it could provide interesting information about the effects of that specific internet technology on learning and skills acquisition.

We find evidence that higher levels of internet penetration, appropriately lagged and mediated by demographic changes, do seem to be associated with higher levels of chess skill at both the player- and country-level. In addition, we measure a decrease in geographical concentration of chess skills since the advent of the internet and can attribute this decrease in part to internet penetration. These findings complement a growing literature on the effects of internet matching technologies in areas as diverse as job search, dating, and consumer goods.

Although these findings are useful and provocative, they fall short of dispositive evidence that there has been substantial global improvement in chess skill—expansion of Heinrich's "collective brain"—and that the internet has driven that improvement. Still, the results are highly suggestive of an important role of the internet in the acquisition of chess skill and point towards internet matching technologies being potentially useful in the acquisition of other skills as well. These results add to a nascent broader literature on the role of the internet in education, skill acquisition, and, ultimately, the distribution of knowledge and skill.

References

Adams, Brian, and Kevin Williams. 2019. "Zone Pricing in Retail Oligopoly." *American Economics Journal: Microeconomics* 11 (1): 124–56, doi: 10.1257/mic.20170130.

Agostinelli, Francesco, Matthias Doepke, Giuseppe Sorrenti, and Fabrizio Zilibotti. 2020. "When the Great Equalizer Shuts Down: Schools, Peers, and Parents in Pandemic Times." IZA Discussion Paper No. 13965, doi: 10.3386/w28264.

Agrawal, Ajay, and Avi Goldfarb. 2008. "Restructuring Research: Communication Costs and the Democratization of University Innovation." *American Economic Review* 98 (4): 1578–90, doi: 10.1257/aer.98.4.1578.

Anderson, Ashton, and Etan A. Green. 2018. "Personal Bests as Reference Points." *Proceedings of the National Academy of Sciences* 115 (8): 1772–76, doi: 10.1073/pnas.1706530115.

Bacher-Hicks, Andrew, Joshua Goodman, and Christine Mulhern. 2021. "Inequality in Household Adaptation to Schooling Shocks: Covid-Induced Online Learning Engagement in Real Time." *Journal of Public Economics* 193: 104345, doi: 10.1016/j.jpubeco.2020.104345.

Bhuller, Manudeep, Tarjei Havnes, and Edwin Leuven. 2013. "Broadband Internet: An Information Superhighway to Sex Crimes." *Review of Economic Studies* 80 (4): 1237–66, doi: 10.1093/restud/rdt013.

Cacault, Paula, Christian Hildebrand, Jérémy Laurent-Lucchetti, and Michele Pellizzari. 2021. "Distance Learning in Higher Education: Evidence from a Randomized Experiment." *Journal of the European Economic Association*, doi: 10.1093/jeea/jvaa060.

Carlana, Michela, and Eliana La Ferrara. 2021. "Apart but Connected: Online Tutoring and Student Outcomes During the COVID-19 Pandemic." HKS Working Paper RWP21-001, doi: 10.2139/ssrn.3777556.

Champeaux, Hugues, Lucia Mangiavacchi, Francesca Marchetta, and Luca Piccoli. 2021. "Learning at Home: Distance Learning Solutions and Child Development During the COVID-19 Lockdown." *Covid Economics: Vetted and Real-Time Papers* 68 (10.2): 69–99.

Dahiya, Shefali, Lila N. Rokanas, Surabhi Singh, Melissa Yang, and Jon M. Peha. 2021. "Lessons from Internet Use and Performance During COVID-19." *Journal of Information Policy* 11: 202–21, doi: `10.5325/jinfopoli.11.2021.0202`.

de Sousa, José, and Guillaume Hollard. 2015. "Gender Differences: Evidence From Field Tournaments." CEPREMAP Working Paper No. 1506.

de Sousa, José, and Muriel Niederle. 2018. "Trickle Down Affirmative Action: A Case Study in Chess." mimeo.

Doyle, Orla. 2020. "COVID-19: Exacerbating Educational Inequalities." *UCD Geary Institute for Public Policy.*

Ellison, Glenn, and Sara Fisher Ellison. 2018. "Match Quality, Search, and the Internet Market for Used Books." NBER Working Paper No. 24197, doi: `10.3386/w24197`.

Ellison, Glenn, and Edward Glaeser. 1997. "Geographic Concentration in US Manufacturing Industries: A Dartboard Approach." *Journal of Political Economy* 105 (5): 889–927, doi: `10.1086/262098`.

Escueta, Maya, Vincent Quan, Andre Joshua Nickow, and Philip Oreopoulos. 2017. "Education Technology: An Evidence-Based Review." NBER Working Paper No. 23744, doi: `10.3386/w23744`.

Faberman, Jason R., and Mariana Kudlyak. 2019. "The Intensity of Job Search and Search Duration." *American Economic Journal: Macroeconomics* 11 (3): 327–57, doi: `10.1257/mac.20170315`.

Finkel, Eli, Paul Eastwick, Benjamin Karney, Harry Reis, and Susan Sprecher. 2012. "Online Dating: A Critical Analysis From the Perspective of Psychological

Science." *Psychological Science in the Public Interest* 13 (1): 3–66, doi: 10.1177/1529100612436522.

Forman, Chris, Avi Goldfarb, and Shane Greenstein. 2012. "The Internet and Local Wages: A Puzzle." *American Economic Review* 102 (1): 556–75, doi: 10.1257/aer.102.1.556.

Gavazza, Alessandro, Mattia Nardotto, and Tommaso Valletti. 2018. "Internet and Politics: Evidence from U.K. Local Elections and Local Government Policies." *Review of Economic Studies* 86 (5): 2092–135, doi: 10.1093/restud/rdy028.

González-Díaz, Julio, Ignacio Palacios-Huerta, and José M. Abuín. 2021. "Personal Bests and Gender." *The Review of Economics and Statistics*, doi: 10.1162/rest_a_01145.

Hardt, David, Markus Nagler, and Johannes Rincke. 2020. "Can Peer Mentoring Improve Online Teaching Effectiveness? An RCT During the COVID-19 Pandemic." CESifo Working Paper No. 8671.

Henrich, Joseph. 2016. *The Secret of Our Success: How Culture Is Driving Human Evolution, Domesticating Our Species, and Making Us Smarter*. Princeton University Press.

Hjort, Jonas, and Jonas Poulsen. 2019. "The Arrival of Fast Internet and Skilled Job Creation in Africa." *American Economic Review* 109 (3): 1032–79, doi: 10.1257/aer.20161385.

Ikeda, Masato, and Shintaro Yamaguchi. 2021. "Online Learning During School Closure Due to COVID-19." *Japanese Economic Review* 72: 471–507, doi: 10.1007/s42973-021-00079-7.

Jaffe, Adam, Manuel Trajtenbern, and Rebecca Henderson. 1993. "Geographic Localization of Knowledge Spillovers as Evidenced by Patent Citations." *The Quarterly Journal of Economics* 108 (3): 577–98, doi: 10.2307/2118401.

Krugman, Paul. 1991. "Increasing Returns and Economic Geography." *Journal of Political Economy* 99 (3): 483–99, doi: 10.1086/261763.

Levitt, Steven D., John A. List, and Sally E. Sadoff. 2011. "Checkmate: Exploring Backward Induction Among Chess Players." *American Economic Review* 101 (2): 975–90, doi: 10.1257/aer.101.2.975.

Marshall, Alfred. 1920. *Principles of Economics: An Introductory Volume.* London: Macmillan and Co.

Mullen, Matt, Amanda Onion, and Missy Sullivan. 2009. "Chess Champion Garry Kasparov Defeats IBM's Deep Blue." *History.com.* Available: https://www.history.com/this-day-in-history/kasparov-defeats-chess-playing-computer.

Syverson, Chad. 2018. "Why Hasn't Technology Sped Up Productivity?." *Chicago Booth Review.* Available: https://www.chicagobooth.edu/review/why-hasnt-technology-sped-productivity.

Thünen, Johann Heinrich von. 1826. *Der Isolirte Staat in Beziehung auf Landwirtschaft und Nationalökonomie.* Darmstadt: Wirtschaft and Finan.

Wall, Bill. 2017. "Computers and Chess – A History." *Chess.com.* Available: https://www.chess.com/article/view/computers-and-chess---a-history.